NASA
Reference
Publication
1145

March 1986

Introduction to Time Series Analysis

Jay C. Hardin

NASA

NASA
Reference
Publication
1145

1986

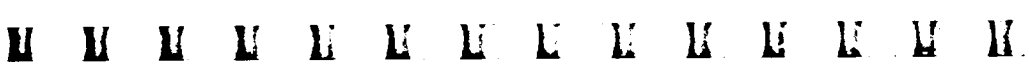
Introduction to Time Series Analysis

Jay C. Hardin

Langley Research Center
Hampton, Virginia

NASA
National Aeronautics
and Space Administration

Scientific and Technical
Information Branch



Preface

This presentation of time series analysis techniques has been developed by the author in the process of teaching (since 1971) a graduate level course on the subject to scientists, engineers, and computer analysts at NASA Langley Research Center. The intent is to develop, from the beginning, the basic understanding necessary to properly apply modern spectral analysis techniques. The subject rests on a firm foundation in the theory of probability, which will be reviewed in this monograph. Thus, the only prerequisites are an ordinary engineering knowledge of calculus and some acquaintance with linear system theory. However, familiarity with random process theory, as provided in *Probability, Random Variables, and Stochastic Processes* by Papoulis, and with Fourier analysis techniques, as provided in *The Fourier Transform and Its Applications* by Bracewell, would be helpful.

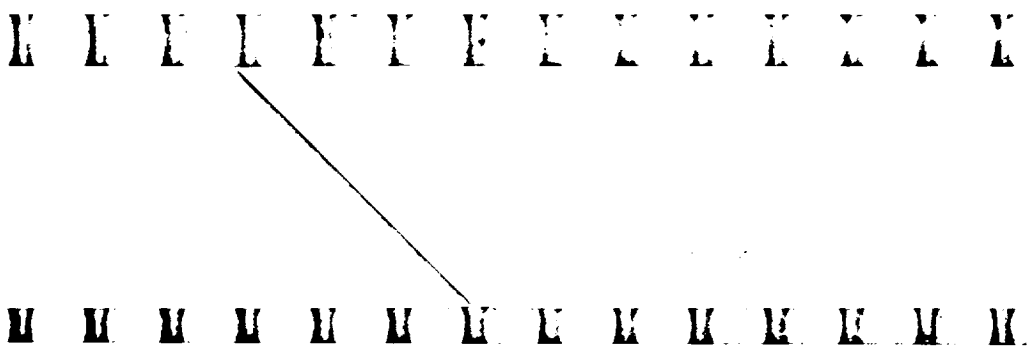
Although there are many textbooks on time series analysis, several of which the author has used in his courses, this monograph takes a different approach from most. First, the theory in this presentation has been developed, insofar as possible, for continuous data. This postpones the inevitable use of discrete mathematics, which the author believes tends to obscure physical understanding, until after the reader has gained some familiarity with the concepts. Only then are the computational details for digital data introduced. Second, the author assumes that most readers will have access to either standard computer software or hard-wired spectral analyzers to do the work of computation. One big danger of such standard analysis techniques, however, is that they will always yield an output, even if the input does not satisfy the assumptions on which the analysis is based. Thus, this monograph seeks to provide the theoretical overview necessary to correctly apply the full range of these powerful techniques. Finally, time series analysis is a vast and rapidly changing field. In an attempt to remain complete and current, the last chapter introduces the reader to many specialized techniques and areas where research is presently in progress.

The author would like to express his appreciation to William E. Zorumski and Stephen K. Park, who worked almost as hard in reviewing this manuscript as the author did in writing it.

Jay C. Hardin
NASA Langley Research Center
Hampton, VA 23665-5225

July 16, 1985

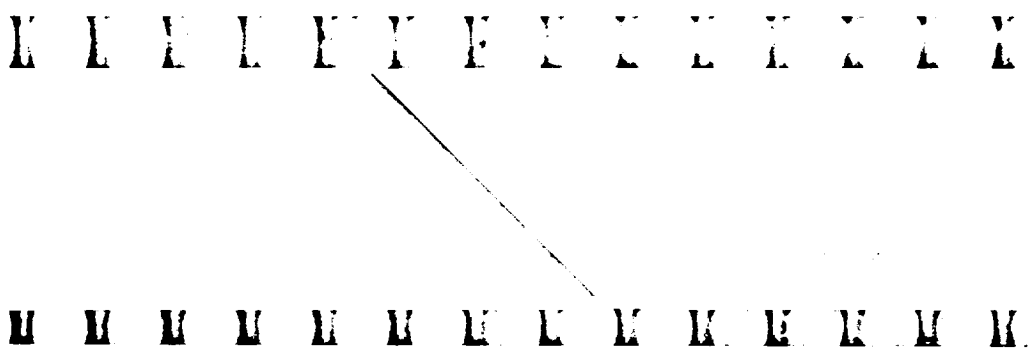
Table of Contents



Symbols

a_n, b_n	series coefficients
$d(t)$	data window function, equal to 0 for $t < 0$ or $t > T$
$E\{\ \}$	expectation operator
$F(\omega)$	Fourier integral transform of $f(t)$ (eq. (2.3))
f	cyclic frequency, Hz
$f(t)$	real function of independent variable t
$f_X(x;t)$	first order density function of random process $X(t)$ (eq. (3.1))
$f_X(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n)$	nth order density function of random process $X(t)$
f_{xy}	joint density function of random processes $X(t)$ and $Y(t)$
$H(\omega)$	frequency response function of linear, shift-invariant system (eq. (5.2))
$h(t)$	impulse response function of linear, shift-invariant system (eq. (5.3))
k	number of degrees of freedom of chi-square random variable
$m_X(t)$	mean value taken by random process $X(t)$ at time t (eq. (3.2))
N	number of samples (or data points) taken of a random process
N_B	number of blocks of data
$P\{\ \}$	probability of event $\{\ \}$
p	period of periodic signal $f(t)$
$R_X(t_1, t_2)$	autocorrelation of random process $X(t)$ at times t_1 and t_2 (eq. (3.5))
$R_{XY}(t_1, t_2)$	cross correlation of random process $X(t)$ at time t_1 and random process $Y(t)$ at time t_2 (eq. (3.11))
S	chi-square random variable (eq. (3.9))

$S_X(\omega)$	power spectral density of random process $X(t)$ (eq. (4.1))
$S_{XY}(\omega)$	cross power spectral density of random processes $X(t)$ and $Y(t)$ (eq. (4.12))
$\mathrm{sinc}(x)$	sinc function, equal to $(\sin x)/x$
T	length of data record
T_B	length of data block
T_m	half-length of lag window
T_r	response time of linear, shift-invariant system
t	independent variable, not necessarily time
$U(\omega)$	Fourier transform of lag window function (eq. (7.8))
$u(\tau)$	lag window function
W_R	window correction factor in autocorrelation estimate
W_S	window correction factor in spectral estimate
$X(t), Y(t), Z(t)$	random processes
$X(\omega), Y(\omega)$	Fourier transforms of random processes $X(t)$ and $Y(t)$ (eq. (4.10))
$X_F(\omega)$	Fourier transform of random process $X(t)$ through data window (eq. (7.16))
$X_T(\omega)$	finite Fourier transform of random process $X(t)$, calculated from sample function of length T (eq. (7.12))
$\Gamma_X(t_1, t_2)$	covariance of random process $X(t)$ at times t_1 and t_2
Δf	bandwidth of spectral estimate
Δt	sampling interval
$\Delta\omega$	bandwidth of spectral estimate in rad/sec
$\delta_{k,n}$	Kronecker delta function



$\delta(\omega - \omega')$	Dirac delta function
$\sigma_X^2(t)$	variance of random process $X(t)$ at time t (eq. (3.4))
φ_n	random phase angles of sinusoidal signals
τ	time lag, equal to $t_2 - t_1$
ω	frequency, units are radians per second if t is time
ω_c	Nyquist frequency, equal to $\pi/\Delta t$
ω_n	frequencies of periodic function, equal to $2n\pi/p$ for $-\infty < n < \infty$; also, set of frequencies not necessarily related; also, set of frequencies at which spectral estimates are calculated
*	complex conjugate
$\wedge$	estimate

Chapter I

Introduction

Consider a record of length T of a real function $f(t)$ as shown in figure 1. By convention, the independent variable is called "time," although it need not actually be time. Instead, the function may depend on distance or angle or any other variable of interest. The data record shown is of finite length, since that is all that is ever available in the real world, and need not be continuous but may, in fact, consist of digital data taken at a set of discrete times. This monograph will be concerned with the development, interpretation, and use of various techniques to extract information from such a record.

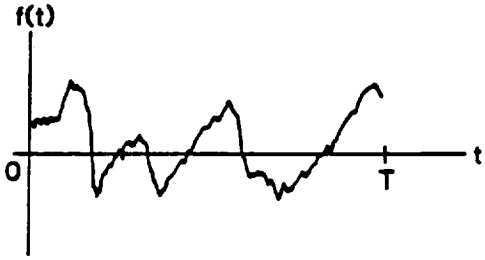


Figure 1. Record of finite length of real function $f(t)$.

1.1 Why Harmonic Analysis?

Many time series analysis techniques involve harmonic analysis, that is, decomposition of the record $f(t)$ into a collection of sines and cosines of various frequencies. Before considering these techniques, it is relevant to ask why it would ever be necessary or advantageous to represent a function by harmonic functions. Certainly many records of interest look very different from the well-behaved periodic sine and cosine functions. There are at least three answers to this question.

Simple input/output relations for linear systems. Consider a signal $x(t)$ that is passed through a linear, shift-invariant physical system to

Figure 2. Schematic of linear system.

produce an output signal $y(t)$ as shown in figure 2. Although the input and output are related by a convolution integral in the time domain, the harmonic representations of the input $X(\omega)$, where ω is the frequency in radians per second, and the output $Y(\omega)$ are related by the simple expression

$$Y(\omega) = H(\omega)X(\omega) \tag{1.1}$$

where $H(\omega)$ is called the frequency response function for the system. This fact, which is known as the convolution theorem, is the basis of many techniques for the solution of differential and integral equations and is an aid to understanding the response of linear systems.

Ease of interpretation (diagnostics). Many time signals are not easily analyzed in the time domain. For example, figure 3 displays the voltage output time history of a microphone recording the noise radiated by a supersonic jet operating in an off-design condition. Such time histories are nearly unintelligible. However, although the time and frequency domain representations contain precisely the same information in the sense that one may be recovered from the other by integration, the generation and potential effects of a signal may often be more easily understood in the frequency domain. For example, figure 4 shows the power spectral density, as a function of frequency, for the time history in figure 3. From the frequency domain perspective in figure 4, it can be seen that most of the power in the signal is concentrated near 5 kHz. In addition, a screech tone, caused by oscillation of the shocks in the jet due to a natural instability of the jet plume, is apparent near 5.2 kHz.

Another dramatic example of such analysis was given by Blackman and Tukey[1] in *Measurement of Power Spectra* when they quoted a letter from Walter E. Munk: "... we were able to discover in the general wave record a very weak low-frequency peak which would surely have escaped our attention without spectral analysis. This peak, it turns out, is almost certainly due to a swell from the Indian Ocean, 10,000 miles distant. Physical dimensions are: 1 mm high, a kilometer long."

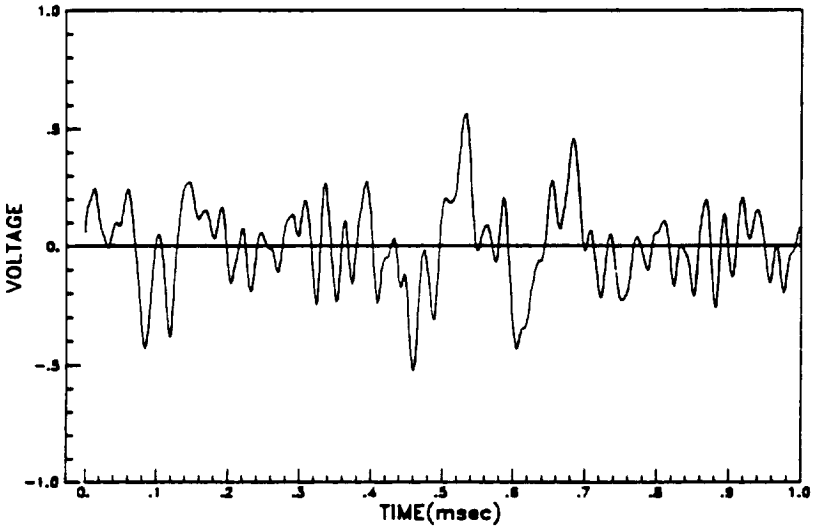


Figure 3. Noise radiation by supersonic jet.

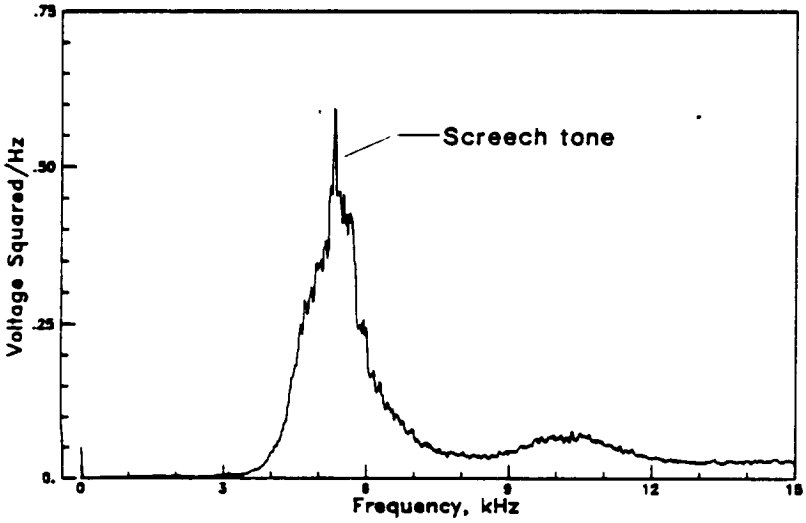


Figure 4. Power spectral density of noise radiation.

Because of their ability to extract information from highly variable records, spectral analysis techniques are widely applied in fluid dynamics, acoustics, and vibration. In addition, such analyses are readily accomplished with either modern digital computers or specialized hardware.

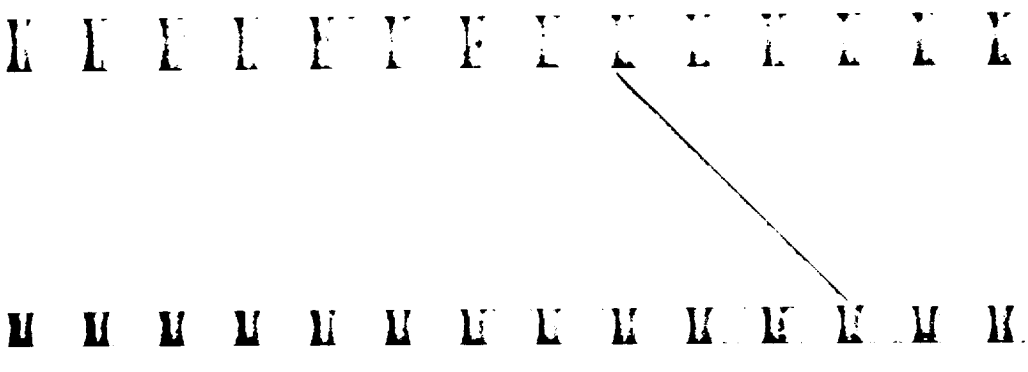
Ease of simulation. Often it is necessary or desirable to excite a system with a particular time history or class of time histories, either in the laboratory or on a computer. However, it is not practical to develop an excitation system for each individual signal. Thus, if the signal can be decomposed into its constituent harmonics and ordinary oscillators (or harmonic functions on a computer) used to produce the excitation, simulation becomes appreciably easier and less expensive. This technique is used by electrodynamic shakers in missile and automobile vibration testing, for example.

1.2 Deterministic or Random?

An important question in the extraction of information from a record like that shown in figure 1 is as follows: Is the record unique or is it merely *representative* of an ensemble of records which might have been obtained? For example, a smoother version of figure 1 might be a record of the elevation as a function of distance along the track of an amusement park ride. If one were designing a cart to traverse that particular track, then this would be the unique (deterministic) record of interest. On the other hand, figure 1 might be a record of the vertical gust velocity as a function of time experienced by an aircraft flying through a thunderstorm. If one were designing an airplane, then the record would be viewed as merely representative of an ensemble of data that might have been obtained in many different thunderstorms. In this random case, the particular properties of the record at hand are not as interesting as the average properties of the whole ensemble of records which might have been obtained. To discuss such data, the concept of a random process must be introduced.

Although many of the techniques developed in this monograph are equally applicable to deterministic data records, the monograph will primarily be concerned with the extraction of information from records that may be considered sample functions of random processes. The analysis of such nondeterministic records is a rapidly changing field with new techniques being devised continually. It is also a field requiring sound engineering judgment in the application of techniques and interpretation of results; many pitfalls await the unwary. It is hoped that this monograph will give the reader the understanding necessary

for the proper application and interpretation of time series analysis techniques.



Chapter II

Harmonic Analysis

Although several seemingly different forms of harmonic analysis are in common use today, they are all special cases of what is now called generalized harmonic analysis. While this monograph will occasionally require use of the full power of this elegant theory, for the most part, the ordinary idea of a Fourier transform as developed in advanced calculus courses will suffice.

2.1 Fourier Transform Pair

If a function $f(t)$ is such that the integral

$$\int_{-\infty}^{\infty} \frac{|f(t)|}{(1+t^2)^{\alpha}}\, dt \tag{2.1}$$

converges for some $\alpha > 0$, then $f(t)$ may be written

$$f(t) = \int_{-\infty}^{\infty} F(\omega)\, e^{i\omega t}\, d\omega \tag{2.2}$$

where

$$F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t)\, e^{-i\omega t}\, dt \tag{2.3}$$

is called the Fourier integral transform[2] of $f(t)$. If the variable t is actually time, then ω is the frequency in radians per second. Since $e^{i\omega t} = \cos\omega t + i \sin\omega t$, equation (2.2) provides a representation of the function $f(t)$ in terms of the periodic sine and cosine functions. *Equations (2.2) and (2.3)* form what is called a Fourier transform pair and *are the fundamental Fourier transform pair that will be used throughout this monograph.*

As will be discussed further in chapter IV, one has considerable freedom in defining a Fourier transform pair. In particular, the

factor $(2\pi)^{-1}$ may be placed before either the time or the frequency integration. The reason for the above choice is that it allows the total power in a random process to be obtained by integrating its power spectral density over all frequencies with no proportionality factor required. The inherent simplicity and elegance of this relationship seems to the author to be worthy of achieving. On the other hand, with the above definition of a transform pair, the $(2\pi)^{-1}$ factor must be reserved for the frequency integration in defining the frequency response function of a linear, shift-invariant system, in order to preserve the simplicity of fundamental relations such as equation (1.1). Thus, this monograph will violate the above convention in the single case of a frequency response function.

The Fourier integral representation for the function $f(t)$ (eq. (2.2)) converges to $f(t)$ at every point where $f(t)$ is continuous. If $f(t)$ is discontinuous at some point $t = t_0$, then the integral in equation (2.2) will converge to

$$\frac{f(t_0^+) + f(t_0^-)}{2}$$

the average of the right- and left-hand limits of $f(t)$ at $t = t_0$, provided that these limits exist.

2.2 Examples

The Fourier integral (eq. (2.2)) has different characteristics dependent on the properties of $f(t)$. Consider the following special cases.

Transient functions. If $f(t)$ is bounded and approaches zero as $|t| \to \infty$, it certainly satisfies the condition that integral (2.1) converge for $\alpha > 0$. In this case, $F(\omega)$ is an ordinary (often complex) continuous function. For example, if

$$f(t) = \begin{cases} e^{-\beta t} & (t \geq 0) \\ 0 & (\text{Otherwise}) \end{cases}$$

where β is real and positive, then

$$F(\omega) = (\beta + i\omega)^{-1}/2\pi$$

which is complex. If the magnitude of $F(\omega)$,

$$|F(\omega)| = (\beta^2 + \omega^2)^{-1/2}/2\pi$$

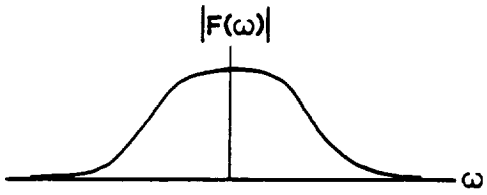


Figure 5. Fourier transform of transient function.

is plotted as shown in figure 5, it can be seen that $f(t)$ is represented by harmonic functions having a continuum of frequencies.

Periodic functions. If $f(t)$ is bounded and periodic with period p, it again satisfies the condition associated with integral (2.1). However, unlike the situation illustrated in figure 5, $F(\omega)$ is nonzero only at a discrete set of frequencies. To see this, note that periodic functions may be expanded in the familiar Fourier series, that is,

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos \omega_n t + b_n \sin \omega_n t)$$

where $\omega_n = 2n\pi/p$. This series also converges to $f(t)$ at all points where $f(t)$ is continuous. Introducing the mathematically convenient concept of negative frequencies allows this series to be written as

$$f(t) = \sum_{n=-\infty}^{\infty} F_n e^{i\omega_n t} \tag{2.4}$$

where

$$F_n = \frac{a_n - ib_n}{2} = \frac{1}{p} \int_0^p f(t)\, e^{-i\omega_n t}\, dt \tag{2.5}$$

and $\omega_{-n} = -\omega_n$.

Now, consider the so-called Dirac delta function $\delta(\omega - \omega')$, which is defined by the relations,

$$\delta(\omega - \omega') = 0 \qquad (\omega \neq \omega')$$

and

$$\int_{-\infty}^{\infty} g(\omega)\, \delta(\omega - \omega')\, d\omega = g(\omega')$$

where $g(\omega)$ is any "fairly good" function,[3] that is, differentiable and well behaved at infinity. Most functions met in the real world are of

this class. The delta function, which is defined by its integral property above, is actually a generalized function; that is, it lies outside the class of ordinary functions. It may be envisioned as the limit as $\Delta\omega \to 0$ of the rectangular function

$$D(\omega - \omega') = \begin{cases} \frac{1}{\Delta\omega} & (|\omega - \omega'| < \frac{\Delta\omega}{2}) \\ 0 & (\text{Otherwise}) \end{cases}$$

whose width is $\Delta\omega$, whose height is $1/\Delta\omega$, and whose integral is unity. As $\Delta\omega \to 0$, the amplitude of this function grows without bound, but its integral is unchanged. The delta function has a long and controversial history, being first introduced without proper mathematical justification by those who found it exceedingly convenient. Only in recent years has it been placed on a rigorous foundation[3] to the satisfaction of the mathematicians.

The delta function arises in the analysis of periodic functions because periodic functions are not square integrable; that is, $\int_{-\infty}^{\infty} f^2(t)\,dt$ is unbounded if $f(t)$ is periodic. As the existence of this integral is a requirement for the nongeneralized Fourier transform to exist.[2] harmonic analysis of such functions must rely on the full power of generalized harmonic analysis. From the definition of the delta function. it can be seen that the function $f(t) = e^{i\omega' t}$ may be expressed as

$$f(t) = e^{i\omega' t} = \int_{-\infty}^{\infty} \delta(\omega - \omega')\, e^{i\omega t}\, d\omega$$

If this relation is compared with equation (2.2), it follows that the Fourier transform of $f(t) = e^{i\omega' t}$ is

$$F(\omega) = \delta(\omega - \omega')$$

Thus, by equation (2.3), the highly useful relation

$$\boxed{\delta(\omega - \omega') = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i(\omega - \omega')t}\, dt} \qquad (2.6)$$

is derived. This equation is of fundamental importance and will be used many times in this monograph.

As an example, only because of equation (2.6) is it possible to develop the Fourier integral transform of a periodic function. If equation (2.4) is used in equation (2.3) and equation (2.6) is applied.

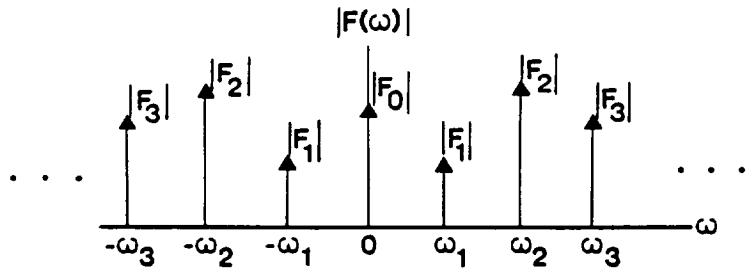


Figure 6. Fourier transform of periodic function.

then

$$F(\omega) = \sum_{n=-\infty}^{\infty} F_n \delta(\omega - \omega_n) \tag{2.7}$$

which can be seen to be nonzero only at a discrete set of equally spaced frequencies. Thus, the Fourier transform of a periodic function has a discrete structure. The magnitude of this relation is shown in figure 6, since the F_n's are generally complex. In this figure, the arrows have been used to represent the delta functions. It should be mentioned that, even though the delta functions are unbounded, discrete transforms are generally plotted as an amplitude spectrum with the height of the arrow indicating the coefficient magnitude, $|F_n|$, which is actually the contribution to the integral of the spectrum at the frequency $\omega = \omega_n$. Such amplitude spectra are even functions of ω since

$$F_{-n} = F_n^*$$

2.3 Convolution Theorems

One property of the Fourier integral transform that will be used repeatedly in this monograph is its ability to transform products. Consider the transform of the product of two time functions:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) g(t)\, e^{-i\omega t}\, dt$$

If the individual time functions $f(t)$ and $g(t)$ have Fourier integral transforms $F(\omega)$ and $G(\omega)$, respectively, then

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

and

$$g(t) = \int_{-\infty}^{\infty} G(\omega)\, e^{i\omega t}\, d\omega$$

These relations allow the transform of the product to be written

$$\begin{aligned}
&\frac{1}{2\pi}\int_{-\infty}^{\infty} f(t)g(t)\, e^{-i\omega t} dt \\
&\quad = \frac{1}{2\pi}\int_{-\infty}^{\infty} dt\, e^{-i\omega t} \int_{-\infty}^{\infty} d\omega'\, F(\omega')\, e^{i\omega' t} \int_{-\infty}^{\infty} d\omega''\, G(\omega'')\, e^{i\omega'' t} \\
&\quad = \int_{-\infty}^{\infty} d\omega' \int_{-\infty}^{\infty} d\omega''\, F(\omega')G(\omega'')\, \frac{1}{2\pi}\int_{-\infty}^{\infty} dt\, e^{-i(\omega-\omega'-\omega'')t}
\end{aligned}$$

where the last integral may be recognized from equation (2.6) as $\delta(\omega - \omega' - \omega'')$. Thus, carrying out the integration over either ω' or ω'' yields the fundamental relation

$$\boxed{\begin{aligned}
&\frac{1}{2\pi}\int_{-\infty}^{\infty} f(t)g(t)\, e^{-i\omega t}\, dt \\
&\quad = \int_{-\infty}^{\infty} d\omega'\, F(\omega')G(\omega - \omega') = \int_{-\infty}^{\infty} d\omega''\, F(\omega - \omega'')G(\omega'')
\end{aligned}} \tag{2.8}$$

called the frequency convolution theorem.[4] Often this theorem is written

$$f(t)g(t) \longleftrightarrow F(\omega) * G(\omega)$$

where the asterisk indicates convolution and the double-headed arrow indicates a Fourier transform pair. In words, this theorem states that the Fourier integral transform of the product of two time functions is equal to the convolution integral of the Fourier integral transforms of the two functions in the frequency domain. This result will be applied extensively throughout this monograph.

A converse of this theorem, called the time convolution theorem.[4] may also be developed. The Fourier transform of a convolution integral in the time domain

$$\int_{-\infty}^{\infty} f(\tau)g(t-\tau)\, d\tau = f(t) * g(t)$$

satisfies the theorem

$$f(t) * g(t) \longleftrightarrow 2\pi F(\omega)G(\omega) \tag{2.9}$$

that its Fourier transform is given by 2π times the product of the Fourier transforms of the individual functions $f(t)$ and $g(t)$. This theorem greatly simplifies the application of Fourier transforms in linear, shift-invariant systems.

With the definition of the fundamental Fourier transform pair and some understanding of its properties, the tools are available to begin consideration of the application of harmonic analysis to time histories that may be considered to result from random phenomena.

Chapter III

Random Process Theory

The theoretical foundation underlying the harmonic analysis of random time histories is random process theory. For much of the actual practice of such analysis, this foundation is buried so deeply that the user may not even be aware of its existence. However, proper understanding and application of time series analysis techniques require its consideration.

3.1 The Concept of a Random Process

The subject of all random analyses is an experiment $\mathcal{E}$, which could be performed repeatedly, at least conceptually. For example, one might test an aircraft part to failure in fatigue. Each performance of such an experiment is called a *trial* and its result is an *outcome* ς. *If it is impossible to predict, independent of the number of times the experiment has been performed previously, what the result of a given trial will be, the experiment is said to be* random. In this case, there are more than one and, perhaps, an infinite number of possible outcomes to the experiment. The set of all possible outcomes is called the *sample space* $S = \{\varsigma_1, \varsigma_2, \ldots\}$ of the experiment. In the real world, the sample space generally contains an uncountable number of possible outcomes. In analyzing the experiment, an attempt is made to statistically describe this whole set of possible outcomes.

Now, consider an operator that yields a function $x(t)$ of the parameter t for each outcome of the experiment, such as a strain gage measuring the strain at some point on the specimen as a function of time in the fatigue test mentioned above. This operation is shown schematically in figure 7. One time history $x_j(t)$ is related to each outcome ς_j of the experiment $\mathcal{E}$. The ensemble of all possible time functions that might be obtained in this way, $X(t;\varsigma)$, is called a *random process*. Three of the possible time histories included in this ensemble are shown in figure 7. The single function obtained on a given trial is called a *sample*

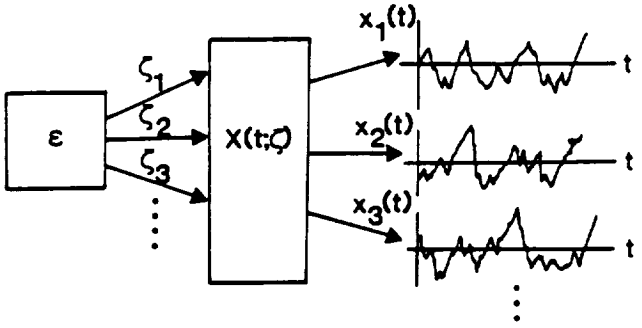


Figure 7. Schematic of a random process.

function, or *realization*, of the random process. For example, the experiment might be launching the Space Shuttle and the operator might be a pressure sensor mounted somewhere on the fuselage. Although the gross characteristics of the pressure signal would be predictable deterministically, there will be random variations due to wind loading, fuel burn rates, etc. Thus, the pressure time history produced by this transducer on a given launch would be a sample function from the random process made up of all possible pressure time histories that might be produced in this way, some of which have occurred on past Shuttle flights or will occur on future Shuttle flights.

Description of random processes. The most complete description possible for a random process is given in terms of its distribution or density functions. Consider the *event* $\{X(t_0) \leq x_0\}$ that the random process $X(t)$ (dependence on the outcome ς is understood) takes a value less than or equal to some chosen number x_0 at time t_0. Imagine that the experiment $\mathcal{E}$ is repeated many times. Each time the experiment is conducted, any one of the possible outcomes ς and, thus, any one of the time histories making up the ensemble representing the random process $X(t)$ could occur. However, on any given trial, one could examine the resulting time history to see whether its value at time t_0 was less than or equal to x_0. If the experiment is repeated N times, then one could form the ratio N_E/N of the number of times N_E that the event occurred to the number of times N that the experiment was repeated. This ratio will be between zero and one. The probability of the event may then be defined as the limit of this ratio as the number of repetitions approaches infinity, that is,

$$P\{X(t_0) \leq x_0\} = \lim_{N \to \infty} \frac{N_E}{N}$$

assuming the limit to exist. Although some mathematicians are not totally comfortable with this intuitive definition,[5] it is the ultimate relationship between the theory and the real world.[6] In words, $P\{X(t_0) \leq x_0\}$ is the likelihood of occurrence of a sample function whose value is less than or equal to x_0 at time t_0.

The value of $P\{X(t_0) < x_0\}$ generally depends on t_0 and x_0. Thus, one may expand the concept by dropping the subscripts and thinking of x and t as variables to define the *first order distribution function* of the random process $X(t)$:

$$F_X(x;t) = P\{X(t) \leq x\}$$

This function generally depends on both x and t and always satisfies

$$0 \leq F_X(x;t) \leq 1$$

An exactly equivalent, but mathematically more convenient, description is given by the *first order density function* of the random process $X(t)$:

$$f_X(x;t) = \frac{\partial F_X(x;t)}{\partial x} \tag{3.1}$$

The density function satisfies

$$\int_{-\infty}^{\infty} f_X(x;t)\,dx = 1$$

and

$$f_X(x;t) \geq 0$$

since the distribution function is a monotonically nondecreasing function of x for each t.

For many applications, the amount of information contained in the distribution or density function is more than is feasible, or even desirable, to know about the random process. In these cases, it is possible to introduce certain averages, or *expected values*, over the whole ensemble of functions comprising the random process. Suppose that $g[X(t)]$ is any function of the values taken by the random process at time t, such as $\sin X(t)$ or $X^3(t)$. Then one may define the expected value of $g[X(t)]$ by

$$E\{g[X(t)]\} = \int_{-\infty}^{\infty} g(x) f_X(x;t)\,dx$$

where E is called the expectation operator. Thus, the operation of taking expectation merely amounts to multiplying the function by the density function for the random process at time t and integrating over all possible values of x. It is equivalent to averaging the values of the function at time t obtained by conducting the experiment repeatedly.

The most useful of these expected values are

Mean

$$m_X(t) = E\{X(t)\} = \int_{-\infty}^{\infty} x f_X(x;t)\,dx \tag{3.2}$$

which is the average value taken by the random process at time t.

Mean square

$$E\{X^2(t)\} = \int_{-\infty}^{\infty} x^2 f_X(x;t)\,dx \tag{3.3}$$

which is often called the "power" in the random process $X(t)$.

Variance

$$\sigma_X^2(t) = E\{[X(t) - m_X(t)]^2\} = \int_{-\infty}^{\infty} [x - m_X(t)]^2 f_X(x;t)\,dx \tag{3.4}$$

which is a measure of the variation of the random process about its mean at time t. The square root of the variance, $\sigma_X(t)$, is called the *standard deviation*. Note that the expectation operation, which is an integration, is linear. Thus, the expected value of a sum is the sum of the expected values. Further, the expected value of a constant or deterministic function is equal to that constant or deterministic function. Therefore,

$$\begin{aligned}\sigma_X^2(t) &= E\{[X(t) - m_X(t)]^2\} = E\{X^2(t) - 2m_X(t)X(t) + m_X^2(t)\} \\ &= E\{X^2(t)\} - 2m_X(t)E\{X(t)\} + m_X^2(t) \\ &= E\{X^2(t)\} - m_X^2(t)\end{aligned}$$

This expression indicates the fundamental relationship between the mean, mean square, and variance and provides a useful alternative way of calculating the variance.

In the same spirit in which the first order distribution function was developed, one may define the *second order distribution function* of the random process $X(t)$:

$$F_X(x_1, x_2; t_1, t_2) = P\{X(t_1) \le x_1 \cap X(t_2) \le x_2\}$$

Here $\cap$ indicates the set operation of intersection and may be read as "and." Thus, the second order distribution is the probability that the random process is less than or equal to x_1 at time t_1 and less than or equal to x_2 at time t_2. Again, it can be interpreted as the likelihood of occurrence of a sample function having these properties.

Likewise, the *second order density function*

$$f_X(x_1, x_2; t_1, t_2) = \frac{\partial^2 F_X(x_1, x_2; t_1, t_2)}{\partial x_1 \, \partial x_2}$$

and an expected value that depends on the values taken by the random process at two times

$$E\{g[X(t_1), X(t_2)]\} = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \, g(x_1, x_2) f_X(x_1, x_2; t_1, t_2)$$

may be defined. Here $g[X(t_1), X(t_2)]$ is again any arbitrary function such as $\exp[X(t_1) + X(t_2)]$. The most useful second order expected values are

Autocorrelation

$$R_X(t_1, t_2) = E\{X(t_1)X(t_2)\} = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \, x_1 x_2 f_X(x_1, x_2; t_1, t_2) \tag{3.5}$$

which is a measure of the linear relation between the values taken by the random process at times t_1 and t_2. Note that $R_X(t, t) = E\{X^2(t)\}$.

Covariance

$$\begin{aligned} \Gamma_X(t_1, t_2) &= E\{[X(t_1) - m_X(t_1)][X(t_2) - m_X(t_2)]\} \\ &= R_X(t_1, t_2) - m_X(t_1) m_X(t_2) \end{aligned}$$

Note that $\Gamma_X(t, t) = \sigma_X^2(t)$ and that if $m_X(t) = 0$, $\Gamma_X(t_1, t_2) = R_X(t_1, t_2)$.

Correlation coefficient

$$\rho_X(t_1, t_2) = \frac{\Gamma_X(t_1, t_2)}{\sigma_X(t_1)\sigma_X(t_2)}$$

which can be shown to satisfy $|\rho_X(t_1, t_2)| \leq 1$.

The concept of a distribution function may be further generalized by defining the *nth order probability distribution function*

$$F_X(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n) = P\{X(t_1) \leq x_1 \cap X(t_2) \leq x_2 \cap \cdots \cap X(t_n) \leq x_n\}$$

for the random process $X(t)$ for all n and any collection of times $t_1, t_2, \ldots, t_n$. In words, the nth order distribution function is the likelihood of occurrence of a sample function whose value $x(t)$ is less than x_1 at time t_1 and less than x_2 at time t_2 and ... and less than x_n at time t_n.

An exactly equivalent description is given by the *nth order density function*

$$f_X(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n) = \frac{\partial^n F_X(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n)}{\partial x_1 \, \partial x_2 \cdots \partial x_n}$$

and, if $g[X(t_1), X(t_2), \ldots, X(t_n)]$ is any function of the values taken by the process at times $t_1, t_2, \ldots, t_n$, then the expected value of this function is defined by

$$E\{g[X(t_1), X(t_2), \ldots, X(t_n)]\} = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \cdots \int_{-\infty}^{\infty} dx_n \, g(x_1, x_2, \ldots, x_n) \times f_X(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n)$$

The description of a random process by its nth order distribution and density functions provides the maximum possible information about the process and will be required by some of the analysis later in this monograph.

Normal random process. A random process $X(t)$ is said to be normal if all its density functions are of the form called Gaussian. In particular. the first is

$$f_X(x; t) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(x - m)^2}{2\sigma^2}\right] \tag{3.6}$$

where $m = m_X(t)$ and $\sigma^2 = \sigma_X^2(t)$, and the second is

$$f_X(x_1, x_2; t_1, t_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-m_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2} + \frac{(x_2-m_2)^2}{\sigma_2^2}\right]\right\} \quad (3.7)$$

where $m_1 = m_X(t_1)$, $m_2 = m_X(t_2)$, $\sigma_1^2 = \sigma_X^2(t_1)$, $\sigma_2^2 = \sigma_X^2(t_2)$, and $\rho = \rho_X(t_1, t_2)$. Note that if $\rho = 0$,

$$f_X(x_1, x_2; t_1, t_2) = f_X(x_1; t_1) f_X(x_2; t_2) \quad (3.8)$$

Similar closed form expressions exist for the density functions of higher orders. Normal random processes are useful because they seem to describe many phenomena that occur in the real world. Further, many of their mathematical properties are quite simple. For example, any linear operation operating on a normal random process yields another normal random process.

Calculus of random processes. A calculus of random processes, called mean square calculus, has been developed based on a concept called mean square convergence. This calculus is a fascinating study in itself and the interested reader is referred to the text by Papoulis.[5] However, for most purposes, it is sufficient to know that all the ordinary operations of calculus, such as differentiation and integration, may be applied to random processes with certain mild restrictions. Only the concept of a limit must be interpreted differently. For example, a random process $X(t)$ is differentiable at $t = t_0$ if its autocorrelation has a second partial derivative where $t_1 = t_2 = t_0$. Likewise, a random process $X(t)$ is integrable over the interval $I = (t_1, t_2)$ if its autocorrelation is integrable over the area $I \times I$.

Consider a random process $Y(t)$ given by the integral of a random process $X(t)$ with respect to some kernel function $K(t; \tau)$, that is,

$$Y(t) = \int_{a(t)}^{b(t)} K(t; \tau) X(\tau)\, d\tau$$

where the limits a and b are arbitrary, but deterministic. Then, the expected value of $Y(t)$ is given by

$$E\{Y(t)\} = \int_{a(t)}^{b(t)} K(t; \tau) E\{X(\tau)\}\, d\tau$$

since the expectation operator, being linear, may be interchanged with integration. This fundamental result will be used repeatedly throughout the monograph.

3.2 Random Variables

Anyone who has read this far in this monograph is undoubtably familiar with the concept of a *random variable*, which varies over a set of numbers which are each assigned to one outcome of a random experiment. For example, the number of dots on the upturned face when a die is rolled is a random variable. There is also an intimate connection between random processes and random variables. Recall that a random process was defined as an ensemble of time functions, one of which was assigned to each outcome of the random experiment. If the time parameter in a random process $X(t)$ is considered fixed, at $t = t_0$ say, then $X(t_0)$ is just a number associated with each outcome and is therefore a random variable. Thus, random variables may be described in the same manner as random processes, having distribution and density functions and expected values.

The normal random variable. A normal random variable X is described by the density function given by equation (3.6) with

$$E\{X\} = m \qquad E\{(X - m)^2\} = \sigma^2$$

Two normal random variables X_1 and X_2 defined on the same experiment $\mathcal{E}$ would have the joint density function given by equation (3.7) with

$$E\{X_1\} = m_1 \qquad E\{(X_1 - m_1)^2\} = \sigma_1^2$$
$$E\{X_2\} = m_2 \qquad E\{(X_2 - m_2)^2\} = \sigma_2^2$$

and

$$\rho = \frac{E\{(X_1 - m_1)(X_2 - m_2)\}}{\sigma_1 \sigma_2}$$

If $\rho = 0$, then the joint density factors as shown in equation (3.8), and the random variables X_1 and X_2 are said to be independent.

The chi-square random variable. An important random variable in understanding the variability of spectral estimation techniques is the chi-square random variable. Suppose that Y_i for $i = 1, 2, \ldots, k$ are independent, normally distributed random variables with zero means

and unit variances, that is $m_i = 0$ and $\sigma_i^2 = 1$. Then the random variable

$$S = Y_1^2 + Y_2^2 + \ldots + Y_k^2 = \sum_{i=1}^{k} Y_i^2 \tag{3.9}$$

is called a chi-square random variable with k degrees of freedom. In other words, the number of degrees of freedom is the number of independent random variables whose squares are added. The mean value of S is

$$E\{S\} = \sum_{i=1}^{k} E\{Y_i^2\} = k$$

and its mean square value is

$$E\{S^2\} = \sum_{i=1}^{k} E\{Y_i^4\} + \sum_{\substack{i=1 \\ i \neq j}}^{k} \sum_{j=1}^{k} E\{Y_i^2\} E\{Y_j^2\}$$

$$= 3k + k^2 - k = 2k + k^2$$

since for a normal random variable,

$$E\{Y_i^4\} = 3E^2\{Y_i^2\}$$

Thus, the variance of the chi-square random variable is

$$\sigma_S^2 = E\{S^2\} - E^2\{S\} = 2k$$

Note that this random variable has the property that

$$\frac{\sigma_S}{E\{S\}} = \sqrt{\frac{2}{k}} \to 0 \quad \text{as} \quad k \to \infty$$

which shows that the variability of this random variable relative to its mean becomes less important as the number of degrees of freedom is increased. The probability density function of the chi-square random variable is given by

$$f_S(s;k) = \frac{1}{2^{k/2}\Gamma(k/2)} s^{k/2-1} e^{-s/2} \qquad (s \geq 0) \tag{3.10}$$

where Γ is the gamma function. From this density, plots of the variation of the chi-square random variable about its mean at various numbers

of degrees of freedom may be produced. Such a graph is included in chapter VIII.

3.3 Jointly Distributed Random Processes

The concepts used in describing a random process may be extended to two random processes defined on the same experiment. For example, consider the experiment of measuring noise transmission through the fuselage of an aircraft. One random process might be the acoustic pressure measured by a microphone placed outside the fuselage near the engine while the other might be the acoustic pressure measured by a microphone inside the fuselage near the pilot's seat.

Description. Consider the two random processes $X(t)$ and $Y(t)$. The *joint distribution function of the two random processes* is defined as

$$\begin{aligned} F_{XY}(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m; t_1, t_2, \ldots, t_{n+m}) \\ = P\{X(t_1) \le x_1 \cap X(t_2) \le x_2 \cap \cdots \cap X(t_n) \le x_n \\ \cap Y(t_{n+1}) \le y_1 \cap Y(t_{n+2}) \le Y_2 \cap \cdots \cap Y(t_{n+m}) \le y_m\} \end{aligned}$$

where there is no importance to the order of the times $t_1, t_2, \ldots, t_{n+m}$. *The joint density function* is then given by

$$\begin{aligned} f_{XY}(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m; t_1, t_2, \ldots, t_{n+m}) \\ = \frac{\partial^{n+m} F_{XY}(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m; t_1, t_2, \ldots, t_{n+m})}{\partial x_1 \partial x_2 \ldots \partial x_n \partial y_1 \partial y_2 \ldots \partial y_m} \end{aligned}$$

Joint expected values may also be defined in a similar manner. The most useful one is

Cross correlation

$$R_{XY}(t_1, t_2) = E\{X(t_1)Y(t_2)\} \tag{3.11}$$

Note that, by convention, the first time goes with the first subscript variable. Thus,

$$R_{YX}(t_1, t_2) = E\{Y(t_1)X(t_2)\}$$

and, in general, $R_{XY}(t_1, t_2) \neq R_{YX}(t_1, t_2)$.

Independence. Two random processes $X(t)$ and $Y(t)$ are said to be independent if

$$\begin{aligned} f_{XY}(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m; t_1, t_2, \ldots, t_{n+m}) \\ = f_X(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n) \\ \times f_Y(y_1, y_2, \ldots, y_m; t_{n+1}, t_{n+2}, \ldots, t_{n+m}) \end{aligned}$$

that is, if their joint density function factors into a product of their individual density functions.

Uncorrelated random processes. Two random processes $X(t)$ and $Y(t)$ are said to be uncorrelated if their cross correlation satisfies

$$R_{XY}(t_1, t_2) = E\{X(t_1)\}E\{Y(t_2)\} = m_X(t_1)m_Y(t_2)$$

Independent random processes are uncorrelated, since

$$\begin{aligned} E\{X(t_1)Y(t_2)\} &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy\, xy f_{XY}(x, y; t_1, t_2) \\ &= \int_{-\infty}^{\infty} dx\, x f_X(x; t_1) \int_{-\infty}^{\infty} dy\, y f_Y(y; t_2) \\ &= m_X(t_1)m_Y(t_2) \end{aligned}$$

for independent processes. However, the converse is not necessarily true.

Complex random processes. Random processes that take complex values also arise from time to time. These are easily handled by writing the complex-valued random process $Z(t)$ as

$$Z(t) = X(t) + i\, Y(t)$$

and considering the real and imaginary parts, $X(t)$ and $Y(t)$, respectively, to be two real-valued random processes defined on the same experiment. For complex-valued random processes, the autocorrelation is defined by

$$R_Z(t_1, t_2) = E\{Z(t_1)Z^*(t_2)\} \tag{3.12}$$

where the asterisk indicates the complex conjugate. This makes

$$R_Z(t, t) = E\{|Z(t)|^2\}$$

which is real and non-negative.

3.4 Stationary Random Processes

A useful subdivision of the class of random processes is based on behavior in time. Some random processes, such as the velocity component measured by a hot-wire anemometer at a point in a turbulent jet running at constant speed, are reasonably independent of the precise value of the time. That is, even though the velocity fluctuates quite rapidly, measurements made at different times are quite similar in their average properties. Other random processes have average properties that vary appreciably with time; for example, the load demand on an electric power generating system depends on whether it is night or day or winter or summer.

Random processes whose statistical properties do not vary with the particular value of time are much more amenable to analysis than those whose statistical properties do. Thus, many more powerful techniques have been developed for extraction of information from them. Such processes are said to be stationary and, in fact, *most of the techniques developed in this monograph will be limited to stationary random processes.*

A random process is said to be strictly stationary if and only if its nth order density function is independent of the origin of time for all n. From this requirement, it can be shown that the first order density is independent of time, that is,

$$f_X(x;t) = f_X(x)$$

Thus, all expected values calculated from this density must also be independent of time, for example,

$$m_X(t) = m_X$$
$$\sigma_X^2(t) = \sigma_X^2$$

That is, the mean and variance of stationary random processes must be constants. For the second order density, independence of the origin of time requires that

$$f_X(x_1, x_2; t_1, t_2) = f_X(x_1, x_2; 0, t_2 - t_1) = f_X(x_1, x_2; \tau)$$

where $\tau = t_2 - t_1$. That is, the second order density depends only on the difference of the two times t_1 and t_2. Thus, all expected values calculated from this density must display the same property.

In particular,

$$R_X(t_1, t_2) = R_X(\tau)$$
$$\Gamma_X(t_1, t_2) = \Gamma_X(\tau)$$
$$\rho_X(t_1, t_2) = \rho_X(\tau)$$

Thus, the autocorrelation, covariance, and correlation coefficient of stationary random processes depend only on the time difference τ. It is this property that aids the analysis of such processes, since the autocorrelation depends on only one independent variable rather than two.

Weak stationarity. Although the mathematical definition of stationarity depends on the density functions, for most of the analysis in this monograph to be valid, only a weaker form of stationarity, based on the expected values, is required. A random process $X(t)$ is said to be weakly stationary if

$$E\{X(t)\} = m_X \qquad E\{X(t_1)X(t_2)\} = R_X(\tau) \tag{3.13}$$

That is, its mean is constant and its autocorrelation depends only on the time difference.

Joint weak stationarity. The concept of weak stationarity may be extended to two random processes defined on the same experiment. Two random processes $X(t)$ and $Y(t)$ are said to be jointly weakly stationary if they are each weakly stationary, and

$$E\{X(t_1)Y(t_2)\} = R_{XY}(\tau) \tag{3.14}$$

Properties of the autocorrelation function of a weakly stationary random process. Suppose that $X(t)$ is a weakly stationary random process with mean zero. (Since the mean value of a weakly stationary random process is constant, if one had a weakly stationary random process $Y(t)$ with mean m_Y, one could merely substract the mean from the data and define

$$X(t) = Y(t) - m_Y$$

which is a weakly stationary random process with mean zero.) Then,

$$E\{X(t_1)X(t_2)\} = R_X(t_2 - t_1) = R_X(\tau)$$

where τ is called the *lag time*, since it is the time by which the second value of the random process $X(t)$ lags the first. Further, since order is immaterial in the expectation,

$$E\{X(t_2)X(t_1)\} = R_X(t_1 - t_2) = R_X(-\tau)$$

Thus, it follows that

$$R_X(\tau) = R_X(-\tau)$$

That is, the autocorrelation must be an even function of τ. Further,

$$R_X(0) = E\{X^2(t)\} \geq 0$$

which says that the value of the autocorrelation at a lag of zero must be non-negative. Also,

$$E\{[X(t_1) \pm X(t_2)]^2\} = 2[R_X(0) \pm R_X(\tau)] \geq 0$$

which implies that

$$R_X(0) \geq |R_X(\tau)|$$

Thus, the absolute value of the autocorrelation at any lag can never be larger than its value at zero lag. Finally, recall that autocorrelation is a measure of the linear relationship between $X(t)$ and $X(t+\tau)$. If $X(t)$ is a *completely random* process (i.e., its mean is zero and it contains no periodic signals), this linear relation weakens as τ increases and, in fact,

$$\lim_{|\tau| \to \infty} R_X(\tau) = 0$$

With this understanding of the properties of the autocorrelation function, it is possible to sketch the autocorrelation function for a typical, completely random process as shown in figure 8. Further if the derivatives exist, it can be seen that

$$\left.\frac{dR_X(\tau)}{d\tau}\right|_{\tau=0} = 0$$

since the autocorrelation is even, and

$$\left.\frac{d^2R_X(\tau)}{d\tau^2}\right|_{\tau=0} \leq 0$$

since the origin must be a maximum. The properties of the autocorrelation function will be important in understanding its Fourier transform, the power spectral density, which will be introduced in the next chapter.

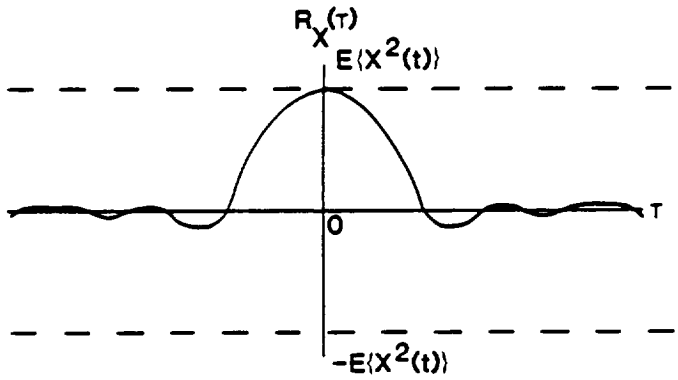


Figure 8. Autocorrelation of a typical, completely random process.

Chapter IV

Power Spectral Analysis

The technique of power spectral analysis of stationary random processes was developed about 50 years ago. Power spectral analysis was first utilized by the electrical engineering community, particularly in the field of communications[7,8]; thus, much of the terminology of the technique comes from electrical engineering. As will be seen, this terminology sometimes creates confusion. In recent years, advances in digital computers and hard-wired spectral analyzers have allowed applications of power spectral analysis to grow exponentially.

There is a fundamental difference between power spectral analysis and ordinary harmonic analysis. Instead of developing a harmonic representation of the sample functions of the random process itself, in power spectral analysis, one develops a harmonic representation of the *autocorrelation* of the random process. The autocorrelation is a well-behaved deterministic function, and from this representation, one can infer average properties of the random process.

The *power spectral density* can be defined as the ordinary Fourier integral transform (eq. (2.3)) of the autocorrelation function of a *stationary* random process, that is,

$$S_X(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R_X(\tau)\, e^{-i\omega\tau}\, d\tau \tag{4.1}$$

By the inversion relation (eq. (2.2)), the autocorrelation can be recovered as

$$R_X(\tau) = \int_{-\infty}^{\infty} S_X(\omega)\, e^{i\omega\tau} d\omega \tag{4.2}$$

This Fourier transform pair is often called the Wiener-Khinchin relations because Wiener and Khinchin derived the pair from a harmonic representation of the random process instead of stating the pair as a definition as done here.

4.1 Properties of Power Spectral Densities

Certain properties of power spectral densities are readily apparent from the Wiener-Khinchin relations. For example, from equation (4.2).

$$R_X(0) = E\{X^2(t)\} = \int_{-\infty}^{\infty} S_X(\omega)\, d\omega \tag{4.3}$$

It is this relation that gave power spectral density its name. Recall that if V and I are the voltage across and current through an electrical element of resistance R, then the power consumed by the element is $I^2R = V^2/R$. In electrical engineering, $X(t)$ is frequently a voltage or current time history, and thus electrical engineers tend to think of $X^2(t)$ as power. For this reason, they called the mean square value, $E\{X^2(t)\}$, the "power" in the random process $X(t)$. This is unfortunate, since the mean square value may have nothing to do with power at all and may be confusing in other fields, such as acoustics, where power has a different definition. However, it is too late at this point to change the terminology.

Returning to equation (4.3), it can be seen that the "power" in the random process $X(t)$ may be obtained by integrating the power spectral density over all frequencies. Thus, the "power spectral density" is the density of power with respect to frequency, or the power per unit frequency, in the signal. One favorite examination question in this field is, what are the units of power spectral density? From the above discussion, it should be clear that the answer is

$$\text{Units of } S_X(\omega) = \frac{[\text{Units of } X(t)]^2}{\text{Units of frequency}}$$

Thus, for example, if $X(t)$ is an acceleration measured in ft/sec^2 and frequency is measured in radians per second, the units of $S_X(\omega)$ are ft^2/sec^3. On the other hand, if $X(t)$ is elevation measured in feet as a function of distance measured in feet, then frequency is measured in radians per foot and the power spectral density has units of ft^3. Furthermore, if t is a *spatial* variable, then $X(t)$ is said to be *homogeneous* in space (as opposed to stationary in time) and the frequency variable ω is called *wave number*, or spatial frequency, and frequently denoted by k or μ.

Now, consider equation (4.1). If the exponential is expanded,

$$S_X(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} R_X(\tau)\cos\omega\tau\, d\tau - \frac{i}{2\pi}\int_{-\infty}^{\infty} R_X(\tau)\sin\omega\tau\, d\tau$$

However, the autocorrelation $R_X(\tau)$ is an even function of τ. Thus, the second integral, being the integral of an odd function over even limits, is zero. Therefore

$$S_X(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} R_X(\tau)\cos\omega\tau d\tau$$

is real.

Further, $\cos\omega\tau$ is an even function of τ and thus $S_X(\omega)$ may be written

$$S_X(\omega) = \frac{1}{\pi}\int_{0}^{\infty} R_X(\tau)\cos\omega\tau d\tau \qquad (4.4)$$

From equation (4.4), it can also be seen that

$$S_X(\omega) = S_X(-\omega)$$

That is, the power spectral density is an even function of ω.

4.2 Problems in Comparing Power Spectral Densities

The freedom inherent in the definition of a Fourier transform pair, as mentioned in chapter II, and the fundamental properties of a power spectral density result in there being no standard definition for power spectral density. This latitude often leads to problems in comparing power spectral densities obtained by different groups utilizing different definitions.

The first ambiguity arises because theoreticians prefer to work in terms of radian frequency ω, defined for both positive and negative frequencies. However, engineers prefer to use cyclic frequency f, where $\omega = 2\pi f$, defined only for positive frequencies. The units of f are cycles per second (hertz). Note that, since the power spectral density is an even function of ω, the mean square value of the process may be obtained from any of the expressions

$$E\{X^2(\tau)\} = \int_{-\infty}^{\infty} S_X(\omega)\,d\omega = 2\int_{0}^{\infty} S_X(\omega)\,d\omega = 4\pi\int_{0}^{\infty} S_X(2\pi f)\,df$$

Thus, engineers prefer to define the *one-sided power spectral density*:

$$G_X(f) = 4\pi S_X(2\pi f) \qquad (f \geq 0)$$

from which the power in the process may be obtained by

$$E\{X^2(\tau)\} = \int_{0}^{\infty} G_X(f)\,df$$

All the modern spectral analyzers compute the function $G_X(f)$.

Second, older analog spectral analyzers use fixed, finite bandwidth filters. These analyzers do not yield a power spectral density at all, but an integrated *power spectrum*, that is,

$$\mathrm{SP}(f^*) = \int_{f_1}^{f_2} G_X(f)\, df \qquad (f_1 \le f^* \le f_2)$$

which amounts to integrating the power spectral density over some finite bandwidth (f_1, f_2), such as a third octave. An integrating analyzer thus assigns the total power in a bandwidth to some frequency f^* within the band, chosen by a committee of workers in the field. To compound matters, many standards for various phenomena are still written in terms of this kind of data.

Finally, in any Fourier transform pair, such as equations (4.1) and (4.2), the placement of the factor $(2\pi)^{-1}$ is completely arbitrary. What such a pair says is that if one starts with an autocorrelation, transforms it into the frequency domain, and then transforms back to the time domain, the autocorrelation is reproduced. Thus, rather than the pair given, no coefficient could be placed in front of the time integral and $(2\pi)^{-1}$ in front of the frequency integral, or $(2\pi)^{-1/2}$ in front of both, or some other combination. Or, if it were preferable to work in terms of cycles per second, one could use the transform pair

$$S_X(f) = \int_{-\infty}^{\infty} R_X(\tau)\, e^{-i2\pi f\tau}\, d\tau$$

$$R_X(\tau) = \int_{-\infty}^{\infty} S_X(f)\, e^{i2\pi f\tau}\, df$$

with no coefficient in front of either integral. Which of these definitions is chosen has absolutely no effect on the reproducibility of the original function, *as long as the definitions are used consistently.* However, if one stops halfway through, at the power spectral density, the amplitude depends on which definition is being employed. The only solution to this problem when trying to compare power spectral densities is to refer to the documentation for the software or to the equipment operator's manual for the hardware and determine what transform pair is being employed. Then, conversion factors which will allow comparison can be developed.

4.3 Interpretation of Power Spectral Densities

A mathematical model that describes most stationary random

signals observed in real world situations is

$$Y(t) = A_0 + \sum_{n=1}^{k} A_n \cos(\omega_n t + \phi_n) + X(t) \tag{4.5}$$

where the A_n's are constants, the ω_n'*s are a set of frequencies which are not necessarily harmonically related*, k may be infinite, the ϕ_n's are independent random phase angles taken to be equally likely to have any value between zero and 2π (i.e., no knowledge of the phase of the periodic functions is assumed), and $X(t)$ is an independent completely random process. That is, most stationary signals can be modeled by a constant, a collection of periodic functions, and a completely random signal. For simplicity, A_0 may also be interpreted as the amplitude of a periodic function of zero frequency.

The mean value of the signal $Y(t)$ is

$$E\{Y(t)\} = A_0$$

and its autocorrelation can be shown to be

$$\boxed{R_Y(\tau) = A_0^2 + \sum_{n=1}^{k} \frac{A_n^2}{2} \cos \omega_n \tau + R_X(\tau)} \tag{4.6}$$

where $R_X(\tau)$ is the autocorrelation of the completely random signal $X(t)$.

Note that if the time history contains periodic functions, the autocorrelation contains periodic functions of the same frequency. This might be surprising, as correlation is a squaring operation which ought to double the frequency (i.e., $2\cos^2 \omega = (1 + \cos 2\omega)$). However, this frequency doubling does not occur. Further, the presence of periodic functions in the signal can readily be detected. Since $\lim_{\tau \to \infty} R_X(\tau) = 0$,

$$\lim_{\tau \to \infty} R_Y(\tau) = A_0^2 + \sum_{n=1}^{k} \frac{A_n^2}{2} \cos \omega_n \tau$$

Thus, if the autocorrelation does not approach zero for large τ, the presence of periodic functions is to be expected.

The power spectral density of the model signal given by equation (4.5) is obtained by Fourier transforming equation (4.6). From

equation (2.6), it can be shown that

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\cos\omega_n\tau\, e^{-i\omega\tau}d\tau = \frac{1}{2}[\delta(\omega-\omega_n)+\delta(\omega+\omega_n)] \tag{4.7}$$

Thus,

$$\boxed{S_Y(\omega) = A_0^2\delta(\omega) + \sum_{n=1}^{k}\frac{A_n^2}{4}[\delta(\omega-\omega_n)+\delta(\omega+\omega_n)] + S_X(\omega)} \tag{4.8}$$

where $S_X(\omega)$ is the power spectral density of the completely random process $X(t)$.

Now since $\lim_{\tau\to\infty} R_X(\tau) = 0$, $S_X(\omega)$ is the Fourier transform of a transient function and represents a continuum of frequency components as shown in figure 3. For example, if

$$R_X(\tau) = \sigma_X^2 e^{-\alpha|\tau|}$$

then

$$S_X(\omega) = \frac{\sigma_X^2}{\pi}\frac{\alpha}{\alpha^2+\omega^2} \tag{4.9}$$

where α is real and positive. Note that $S_X(0) \neq 0$ even though $E\{X(t)\} = 0$. That is, even if there is no dc component in the signal, the power spectral density is nonzero at zero frequency. Thus, in general, the power spectral density of the model stationary signal given by equation (4.5) appears as shown in figure 9, where again the arrows represent the delta functions. This general spectrum consists of a delta function at zero frequency produced by the mean of the process, delta functions at the frequencies $\pm\omega_k$ produced by the periodic components, and a continuous distribution of power produced by the completely random part.

4.4 Relation Between the Power Spectral Density and the Fourier Transform of a Random Process

The relationship between the power spectral density, $S_X(\omega)$, which is a Fourier transform of the autocorrelation function of $X(t)$, and the Fourier transform, $X(\omega)$, of the random process itself is of interest and importance. Here, the full power of generalized harmonic analysis is once again called into play since a stationary random process, not being square integrable, can only be Fourier transformed in the generalized sense.

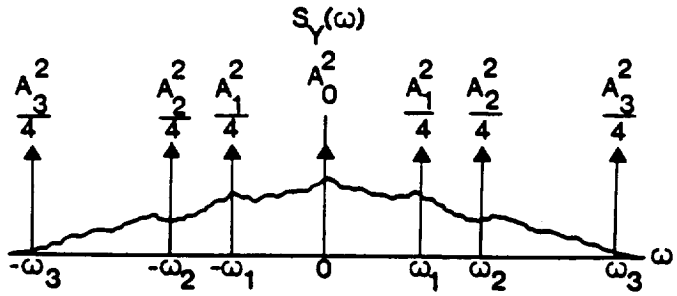


Figure 9. General power spectral density.

A stationary random process $X(t)$ satisfies equation (2.1). Thus, $X(t)$ has a generalized Fourier transform given by

$$X(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(t)\, e^{-i\omega t}\, dt \tag{4.10}$$

Note that $X(\omega)$ is a *complex-valued* random process in frequency. The autocorrelation of $X(\omega)$ is given by (see eq. (3.12))

$$E\{X(\omega)X^*(\omega')\} = \frac{1}{4\pi^2}\int_{-\infty}^{\infty} dt_1 \int_{-\infty}^{\infty} dt_2 R_X(t_2 - t_1)\, e^{-i(\omega t_1 - \omega' t_2)}$$

Introducing the variables $\tilde{t}$ and τ such that

$$\tilde{t} = \frac{t_1 + t_2}{2} \qquad \tau = t_2 - t_1$$

yields

$$\begin{aligned} E\{X(\omega)X^*(\omega')\} &= \frac{1}{2\pi}\int_{-\infty}^{\infty} d\tau\, R_X(\tau)\, e^{i(\omega+\omega')\tau/2} \\ &\quad \times \frac{1}{2\pi}\int_{-\infty}^{\infty} d\tilde{t}\, e^{-i(\omega-\omega')\tilde{t}} \\ &= S_X\left(\frac{\omega+\omega'}{2}\right)\delta(\omega-\omega') \end{aligned}$$

That is, the autocorrelation of $X(\omega)$ is zero except when $\omega = \omega'$. Thus, it follows that

$$S_X(\omega) = \int_{-\infty}^{\infty} E\{X(\omega)X^*(\omega')\}\, d\omega' \tag{4.11}$$

That is, the power spectral density may also be interpreted as the integral of the autocorrelation of the Fourier transform of the random process over all frequencies. Equation (4.11) is closely related to the expression for the power spectral density utilized in the finite Fourier transform approach to be discussed in chapter VII.

4.5 Cross Spectral Density

If $X(t)$ and $Y(t)$ are jointly stationary random processes, it is possible to define the cross spectral density as the Fourier transform of the cross correlation:

$$S_{XY}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} R_{XY}(\tau)\, e^{-i\omega\tau}\, d\tau \tag{4.12}$$

The cross correlation is regained by

$$R_{XY}(\tau) = \int_{-\infty}^{\infty} S_{XY}(\omega)\, e^{i\omega\tau}\, d\omega \tag{4.13}$$

However, the cross correlation is not in general an even function of τ. Thus, if the exponential is expanded to obtain

$$S_{XY}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} R_{XY}(\tau)\cos\omega\tau\, d\tau - \frac{i}{2\pi}\int_{-\infty}^{\infty} R_{XY}(\tau)\sin\omega\tau\, d\tau$$

the second integral does not necessarily vanish, and the cross spectral density is, in general, complex, having a real part

$$\mathrm{Re}\,[S_{XY}(\omega)] = \frac{1}{2\pi}\int_{-\infty}^{\infty} R_{XY}(\tau)\cos\omega\tau\, d\tau$$

which is an even function of ω and is called the *co-spectrum*, and an imaginary part

$$\mathrm{Im}[S_{XY}(\omega)] = -\frac{1}{2\pi}\int_{-\infty}^{\infty} R_{XY}(\tau)\sin\omega\tau\, d\tau$$

which is an odd function of ω and is called the *quad-spectrum*. For plotting purposes, the cross spectral density is usually represented in polar form, that is,

$$S_{XY}(\omega) = |S_{XY}(\omega)|e^{i\theta_{XY}(\omega)}$$

and its magnitude

$$|S_{XY}(\omega)| = \{\mathrm{Re}^2[S_{XY}(\omega)] + \mathrm{Im}^2[S_{XY}(\omega)]\}^{1/2}$$

and phase

$$\theta_{XY}(\omega) = \arctan \frac{\mathrm{Im}[S_{XY}(\omega)]}{\mathrm{Re}[S_{XY}(\omega)]}$$

are plotted as a function of frequency.

Chapter V

Random Processes in Linear Systems

Consider a linear, shift-invariant system with a random process input $X(t)$ as shown in figure 10. The linear system might be a bridge girder subjected to the random loading of automobiles of various weights and speeds arriving at various times, or it might be a Space Shuttle tile subjected to random heat loading during reentry. Any system that can be described by linear equations is linear. Shift invariance implies that the parameters of the system are independent of time.

5.1 Description of the System

The system in figure 10 is uniquely defined by its *impulse response function* $h(t)$, which is the response of the system to an impulsive load $\delta(t)$ at $t = 0$. If the parameter t is actually time, then for all real systems, $h(t) = 0$ for $t < 0$, since there can be no response until the load is applied. A system for which this is true is said to be *causal*. For all real systems, the response also tends to die away with time because of damping and becomes effectively zero for $t > T_r$ where T_r is called the response time. Thus, $h(t)$ is a transient function, as shown in figure 11.

In terms of the impulse response function, the output of the system $Y(t)$ is given by the convolution integral

$$\boxed{Y(t) = \int_{-\infty}^{\infty} h(\alpha) X(t-\alpha)\, d\alpha} \tag{5.1}$$

This fundamental equation, which describes any linear, shift-invariant system, is developed in textbooks on linear system theory,[9] a knowledge of which is assumed in this monograph. If the input $X(t)$ and the impulse response function $h(t)$ each have Fourier transforms by the definition in equation (2.3), then from the time convolution theorem (eq. (2.9)), the Fourier transform of the output $Y(t)$ is given by

Figure 10. Random process input to a linear, shift-invariant system.

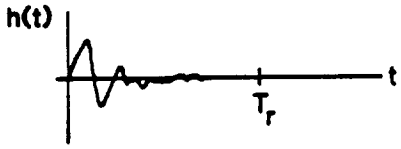


Figure 11. Impulse response function.

the product of the Fourier transforms of the input and impulse response functions with a proportionality factor of 2π. In order to remove this 2π factor and thus simplify the resulting expressions, the following definition of the *frequency response function* $H(\omega)$ for the system is employed in this monograph:

$$H(\omega) = \int_{-\infty}^{\infty} h(t)\, e^{-i\omega t}\, dt \tag{5.2}$$

The frequency response function also provides a unique description of the linear, shift-invariant system since the impulse response function may be recovered by the integral

$$h(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(\omega)\, e^{i\omega t}\, d\omega \tag{5.3}$$

Further, with the use of the definition in equation (5.2), the relation between the Fourier integral transforms $X(\omega)$ and $Y(\omega)$ of the input $X(t)$ and output $Y(t)$, respectively, is the familiar expression

$$Y(\omega) = H(\omega)X(\omega) \tag{5.4}$$

by the time convolution theorem (eq. (2.9)).

5.2 Properties of the Output Random Process

Consider the linear system shown in figure 10 and suppose that the

input $X(t)$ is a weakly stationary random process in time which began at $t = -\infty$ (or in the real world, at $t < -T_r$ so that any starting transients have died away). Then, the output $Y(t)$ is also a random process, which is given by the convolution integral in equation (5.1). The question of interest is then, what are the characteristics of the output random process $Y(t)$?

The mean value of $Y(t)$ is given by

$$\begin{aligned} E\{Y(t)\} &= \int_{-\infty}^{\infty} h(\alpha) E\{X(t-\alpha)\}\, d\alpha \\ &= m_X \int_{-\infty}^{\infty} h(\alpha)\, d\alpha = m_X H(0) \end{aligned}$$

where expectation has been interchanged with integration and $H(0)$ is the frequency response function of the system at zero frequency, or the "dc gain" of the system. Thus, the mean value of $Y(t)$ is constant. Further, its autocorrelation is

$$\begin{aligned} R_Y(t_1, t_2) &= E\{Y(t_1)Y(t_2)\} \\ &= \int_{-\infty}^{\infty} d\alpha_1 \int_{-\infty}^{\infty} d\alpha_2\, h(\alpha_1)h(\alpha_2) E\{X(t_1-\alpha_1)X(t_2-\alpha_2)\} \\ &= \int_{-\infty}^{\infty} d\alpha_1 \int_{-\infty}^{\infty} d\alpha_2\, h(\alpha_1)h(\alpha_2) R_X(\tau - \alpha_2 + \alpha_1) \\ &= R_Y(\tau) \end{aligned} \tag{5.5}$$

which depends only on the time difference τ. The output $Y(t)$ is therefore also a weakly stationary random process, since its mean is constant and its autocorrelation depends only on the time difference.

Likewise, the cross correlation of $X(t)$ and $Y(t)$ depends only on the time difference:

$$R_{XY}(t_1, t_2) = E\{X(t_1)Y(t_2)\} = \int_{-\infty}^{\infty} h(\alpha) R_X(\tau - \alpha)\, d\alpha = R_{XY}(\tau) \tag{5.6}$$

Thus, $X(t)$ and $Y(t)$ are jointly weakly stationary.

Since $Y(t)$ is stationary, its power spectral density is defined. By using equation (5.5) and interchanging the order of integration, it can

be seen that

$$\begin{aligned} S_Y(\omega) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} R_Y(\tau)\, e^{-i\omega\tau}\, d\tau \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} d\tau\, e^{-i\omega\tau} \int_{-\infty}^{\infty} da_1 \int_{-\infty}^{\infty} da_2\, h(a_1)h(a_2)R_X(\tau - a_2 + a_1) \\ &= \int_{-\infty}^{\infty} da_1\, h(a_1)\, e^{i\omega a_1} \int_{-\infty}^{\infty} da_2\, h(a_2)\, e^{-i\omega a_2} \\ &\quad \times \frac{1}{2\pi}\int_{-\infty}^{\infty} d\tau\, R_X(\tau - a_2 + a_1)\, e^{-i\omega(\tau - a_2 + a_1)} \\ &= H^*(\omega)H(\omega)S_X(\omega) \end{aligned}$$

Thus,

$$\boxed{S_Y(\omega) = |H(\omega)|^2 S_X(\omega)} \tag{5.7}$$

which says that the power spectral density of the output signal is just the square of the magnitude of the frequency response function times the input power spectral density. *This simple relation, which is valid for any linear, shift-invariant system, is the fundamental reason for development of power spectral analysis in terms of sines and cosines.* Although other complete sets of orthogonal functions could be used. equivalent relations corresponding to equation (5.7) would not exhibit such simplicity. It might be mentioned that equation (5.7) can also be derived from equation (4.11) with the use of equation (5.4).

Since $X(t)$ and $Y(t)$ are jointly weakly stationary, the cross spectral density is also defined. Using equation (5.6) and the same approach yields

$$S_{XY}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} R_{XY}(\tau)\, e^{-i\omega\tau}\, d\tau = H(\omega)S_X(\omega) \tag{5.8}$$

That is, the cross spectral density is just the frequency response function times the input power spectral density. Note that equations (5.7) and (5.8) imply that

$$S_Y(\omega) = H^*(\omega)S_{XY}(\omega) \tag{5.9}$$

These simple relations are very useful in understanding the response of linear systems to random inputs.

For example, suppose that the system is an ideal band-pass filter with frequency response function as shown in figure 12. Thus,

$$H(\omega) = \begin{cases} 1 & (\omega_1 < |\omega| < \omega_2) \\ 0 & (\text{Otherwise}) \end{cases}$$

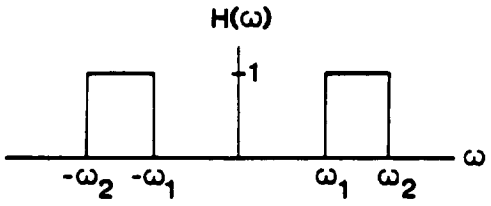


Figure 12. Ideal band-pass filter.

and all frequency components in the range (ω_1, ω_2) are passed without amplification while all others are excluded. For this system, the relation between the input and output spectral densities is given by equation (5.7). Thus,

$$E\{Y^2(t)\} = \int_{-\infty}^{\infty} S_Y(\omega)\, d\omega = \int_{-\infty}^{\infty} |H(\omega)|^2 S_X(\omega)\, d\omega$$
$$= 2\int_{\omega_1}^{\omega_2} S_X(\omega)\, d\omega \geq 0 \tag{5.10}$$

and the total power in the output process can be seen to be just the power in the input process in the frequency band (ω_1, ω_2). Since equation (5.10) must be valid regardless of the values of ω_1 and ω_2, another property of the power spectral density is apparent.

$$S_X(\omega) \geq 0 \tag{5.11}$$

If this were not true, the output power could be negative in some band.

It should be mentioned that the technique for spectral estimation employed in the old analog spectral analyzers, which passed the signal through a bank of fixed bandwidth filters, was based on equation (5.10). Although an ideal band-pass filter for a time signal is not physically realizable since its impulse response function

$$h(t) = \frac{2}{\pi t} \cos \frac{\omega_1 + \omega_2}{2} t \sin \frac{\omega_2 - \omega_1}{2} t$$

is not causal, very good approximations to an ideal band-pass filter can be constructed.

5.3 Determination of Frequency Response Functions

Equations (5.7) and (5.8) can be used to determine the frequency response function of a linear system in a simple manner. A frequency

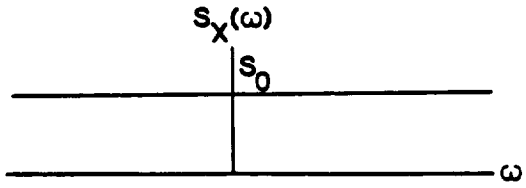


Figure 13. Power spectral density of white noise.

response function is generally complex, that is,

$$H(\omega) = |H(\omega)|e^{i\phi(\omega)}$$

producing both an amplitude and a phase change in the incoming signal. To determine this function, a random process can be input to the system and the input and output spectral densities measured. Then, if only the magnitude of the frequency response function is of interest, equation (5.7) yields

$$|H(\omega)| = \left[\frac{S_Y(\omega)}{S_X(\omega)}\right]^{1/2} \tag{5.12}$$

as long as $S_X(\omega) \neq 0$. To avoid this and to simplify the calculation, the concept of *white noise*, whose power spectral density is a constant as shown in figure 13, is often employed in theoretical analysis. Unfortunately, such noise is not physically realizable since its total power.

$$E\{X^2(t)\} = \int_{-\infty}^{\infty} S_0 \, d\omega$$

is infinite. However, *band-limited white noise* such that

$$S_X(\omega) = \begin{cases} S_0 & (|\omega| < \omega_B) \\ 0 & (\text{Otherwise}) \end{cases}$$

can be generated. One is usually interested in the frequency response function only in some range of frequencies. Thus, if ω_B is larger than the values of ω in which one is interested, this band-limited white noise can conveniently be used experimentally to determine the frequency response function, since the denominator in equation (5.12) will never be zero in the frequency range of interest.

If knowledge of the phase of the frequency response function is also of importance, then the cross spectral density of the input and

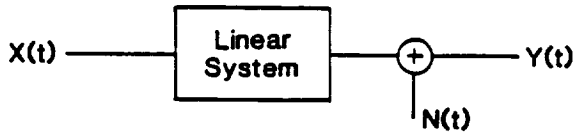


Figure 14. A linear system with noise.

output processes can be measured and the frequency response function determined from equation (5.8), that is,

$$H(\omega) = \frac{S_{XY}(\omega)}{S_X(\omega)} \tag{5.13}$$

These techniques, which can usually determine a frequency response function from a small amount of data, have all but replaced the old time-consuming "sine sweep" method in which the response to each individual frequency input was measured.

5.4 The Coherence Function

In recent years, the *coherence function*

$$\gamma^2(\omega) = \frac{|S_{XY}(\omega)|^2}{S_X(\omega)S_Y(\omega)} \tag{5.14}$$

has come into use. This function has very interesting and useful properties.

Suppose first that $Y(t)$ is obtained by passing $X(t)$ through some linear system. Then from equations (5.7) and (5.8),

$$\gamma^2(\omega) = \frac{|H(\omega)|^2 S_X^2(\omega)}{S_X(\omega)|H(\omega)|^2 S_X(\omega)} = 1$$

That is, if $X(t)$ and $Y(t)$ are related by equation (5.1), the coherence function is unity.

However, suppose now that the output $Y(t)$ contains an additive part which is not linearly related to $X(t)$, as shown in figure 14. The "noise" signal $N(t)$ could be due to nonlinearities in the system and thus depend nonlinearly on the input signal $X(t)$. Or it could be due to other inputs, or contaminating signals. In either case, the noise is assumed to be weakly stationary and uncorrelated with the input $X(t)$. The output $Y(t)$ is then also weakly stationary and can be written

$$Y(t) = \int_{-\infty}^{\infty} h(\alpha)X(t-\alpha)\,d\alpha + N(t)$$

Assuming that the mean of $Y(t)$, if any, is removed is equivalent to assuming that $E\{N(t)\} = 0$ and $R_{XN}(\tau) = 0$. Thus,

$$S_Y(\omega) = |H(\omega)|^2 S_X(\omega) + S_N(\omega) \tag{5.15}$$

where $S_N(\omega)$ is the power spectral density of the noise signal. Since $X(t)$ and $N(t)$ are uncorrelated, from equation (5.6), the cross spectral density

$$S_{XY}(\omega) = H(\omega)S_X(\omega)$$

is unchanged. Thus, the coherence is

$$\begin{aligned}\gamma^2(\omega) &= \frac{|H(\omega)|^2 S_X^2(\omega)}{|H(\omega)|^2 S_X^2(\omega) + S_X(\omega)S_N(\omega)} \\ &= \frac{1}{1 + \dfrac{S_N(\omega)}{|H(\omega)|^2 S_X(\omega)}} \leq 1\end{aligned} \tag{5.16}$$

That is, at those frequencies where the noise spectral density $S_N(\omega)$ is nonzero, the coherence function is less than unity. Thus, a coherence value less than one indicates that the input and output processes are not totally linearly related.

A further interpretation of the coherence function may be obtained by writing equation (5.15) as

$$S_Y(\omega) = S_{YL}(\omega) + S_N(\omega)$$

where

$$S_{YL}(\omega) = |H(\omega)|^2 S_X(\omega)$$

is that portion of the power spectral density of $Y(t)$ which is linearly related to the input $X(t)$. Then equation (5.16) becomes

$$\gamma^2(\omega) = \frac{|H(\omega)|^2 S_X(\omega)}{S_Y(\omega)} = \frac{S_{YL}(\omega)}{S_Y(\omega)}$$

and the linearly related portion of $Y(t)$ may be determined by the relation

$$S_{YL}(\omega) = \gamma^2(\omega)S_Y(\omega)$$

from the measured power spectral density $S_Y(\omega)$.

If there is no linear relation between $X(t)$ and $Y(t)$, such that $R_{XY}(\tau) = 0$, then $S_{XY}(\omega) = 0$ and $\gamma^2(\omega) = 0$. Thus, the coherence function is useful in determining whether a *linear relation* exists between the input and output at any given frequency. It should be emphasized that the absence of a linear relation does not necessarily imply that there is no relation between $X(t)$ and $Y(t)$.

The coherence function may be used in many ways. For example, suppose that a new composite material is developed and one wishes to determine whether it can be characterized as a linear system. The material could be excited at one point by a random signal and its response measured at another point. The coherence between the two signals would then indicate whether the material is linear in its behavior. Another use is to determine whether two time histories, which may have similar power spectral densities, are linearly dependent. For example, many people believe that sunspot activity produces effects observable here on Earth, such as climatic changes evidenced by crop yields. Data on the activity of sunspots going back to the year 1610 exist[10] and their power spectral density exhibits a peak corresponding to a period of approximately 11 years. If one had data on some other phenomenon, such as wheat yields in Iowa, whose power spectral density also exhibited a peak showing an 11 year cycle, then one might suspect a linear relationship between the wheat yields and the sunspot activity. However, the similarity of the spectra might also be merely a coincidence. Coherence analysis of the two records over the same span of years would resolve the question. In one such test of which the author is familiar, coherence analysis showed no linear relation at the 11 year cycle, but did suggest a linear relation with a period near 80 years.

Chapter VI

Estimation Theory

This chapter begins examination of the problem of evaluating the various expressions developed in the previous chapters from data obtained in the real world. Most of these expressions involved ensemble averages over the entire ensemble of time histories that comprise the random process. This entire ensemble is never available for analysis in practical situations. Nor is the probability distribution function of the random process known. Thus, one must be content to *estimate* the quantities of interest, often on the basis of a single realization of the random process.

6.1 Estimation of a Parameter by a Random Variable

Consider an unknown parameter θ which one wishes to estimate by a random variable $\hat{\theta}$. In this monograph, a hat over a quantity indicates an estimate. Why would one want to make such an estimate? Recall that the assumed data comprise a record of length T of a single sample function from a *stationary random process* $X(t)$ as shown in figure 15. On the basis of this single sample function, it is desired to estimate parameters (i.e., moments) of the random process itself, such as m_X, $R_X(\tau)$, and $S_X(\omega)$, which for fixed τ and ω are constants. However, the estimates will be random variables, since they will depend on which particular sample function is used.

The first requirement one would ordinarily desire in an estimate is that

$$E\{\hat{\theta}\} = \theta \tag{6.1}$$

where the expectation is over all possible values of the estimate (i.e., over all possible sample functions that comprise the random process) in order that the average value of the estimate would be the quantity to be estimated. An estimate which satisfies this requirement is said to be *unbiased*. When this is not true, it is possible to define

$$\text{Bias} = E\{\hat{\theta}\} - \theta \tag{6.2}$$

Figure 15. Finite length sample function.

Bias is a systematic error due to the estimation technique employed.

The variation of the estimate $\hat{\theta}$ about its mean value depends on which sample function is employed. In order to keep this uncertainty as small as possible, one should choose the estimate so that it varies as little as possible about its mean value; that is *minimum variance estimation* is desired. This uncertainty is measured by the standard deviation of the estimate:

$$\text{Uncertainty} = (E\{[\hat{\theta} - E(\hat{\theta})]^2\})^{1/2} \tag{6.3}$$

Thus, one ordinarily attempts to choose an estimation technique that is without bias and whose uncertainty is as small as possible.

6.2 Estimation of Mean

Suppose that $X(t)$ is a stationary random process and consider the estimate

$$\hat{m}_X = \frac{1}{T}\int_0^T X(t)\,dt \tag{6.4}$$

Since $\hat{m}_X$ is a number assigned to each outcome of the experiment, it is a random variable. In addition,

$$E\{\hat{m}_X\} = \frac{1}{T}\int_0^T E\{X(t)\}\,dt = \frac{m_X}{T}\int_0^T dt = m_X$$

and, thus, $\hat{m}_X$ is an unbiased estimate of the unknown mean m_X of the random process $X(t)$. Further the variance of $\hat{m}_X$ is

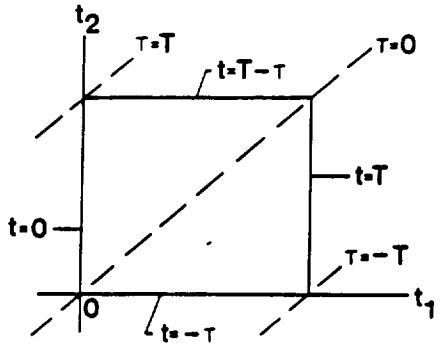


Figure 16. Region of integration.

$$
\begin{aligned}
E\{\dot{m}_X - m_X)^2\} &= E\left\{\left(\frac{1}{T}\int_0^T [X(t) - m_X]\, dt\right)^2\right\} \\
&= E\left\{\frac{1}{T^2}\int_0^T [X(t_1) - m_X]\, dt_1 \int_0^T [X(t_2) - m_X]\, dt_2\right\} \\
&= \frac{1}{T^2}\int_0^T dt_1 \int_0^T dt_2\, \Gamma_X(t_2 - t_1) \\
&= \frac{1}{T}\int_{-T}^T \Gamma_X(\tau)\left(1 - \frac{|\tau|}{T}\right) d\tau \qquad (6.5)
\end{aligned}
$$

where the integral relation

$$
\int_0^T dt_1 \int_0^T dt_2\, F(t_2 - t_1) = \int_{-T}^T F(\tau)(T - |\tau|)d\tau \qquad (6.6)
$$

has been employed. In this equation, $F(\tau)$ is any function of the variable $\tau = t_2 - t_1$. This fundamental relation (eq. (6.6)) is obtained by employing the change of variables $t_1 = t$ and $t_2 = t + \tau$ and noting the region of integration as shown in figure 16. The integral over t may then be evaluated. Upon use of equation (6.5), the uncertainty in the estimate in equation (6.4) of the mean is seen to be

$$
\text{Uncertainty} = \left[\frac{1}{T}\int_{-T}^T \Gamma_X(\tau)\left(1 - \frac{|\tau|}{T}\right) d\tau\right]^{1/2}
$$

which can, in theory, be made as small as desired for a completely random process by obtaining more data (i.e., by letting $T \to \infty$) since $\lim_{\tau \to \infty} \Gamma_X(\tau) = 0$. However, for T fixed, the uncertainty is fixed and nonzero.

6.3 Estimation of Autocorrelation

Consider the autocorrelation estimate defined by

$$\hat{R}_X(\tau) = \frac{1}{T-\tau} \int_0^{T-\tau} X(t)X(t+\tau)\,dt \qquad (0 \le \tau < T) \qquad (6.7)$$

Its expected value is

$$E\{\hat{R}_X(\tau)\} = \frac{1}{T-\tau} \int_0^{T-\tau} R_X(\tau)\,dt = R_X(\tau)$$

and thus, $\hat{R}_X(\tau)$ is an unbiased estimate of $R_X(\tau)$ for $0 \le \tau < T$. Several points should be noted about this estimate:

1. The estimate can be determined only out to a lag of $\tau = T$ from a data record of length T.
2. Equation (6.7) also provides an estimate for $-T < \tau < 0$ since the autocorrelation function is even.
3. Less data contribute to the estimate as τ increases, as can be seen from the limits of integration. Thus one would expect the variance of the estimate to increase with τ. In other words, the estimate of $R_X(\tau)$ becomes more uncertain as τ becomes larger.

An equation for the variance of $\hat{R}_X(\tau)$ involves the expected value of the product of the values taken by the process at four different times and will not be included in this monograph. However, the variance can be shown to depend upon both T and τ and to approach zero as $T \to \infty$. Thus, the uncertainty in this estimate can again be made as small as desired by obtaining more data.

6.4 Estimation of Cross Correlation

In a similar manner, the expressions

$$\hat{R}_{XY}(\tau) = \frac{1}{T-\tau} \int_0^{T-\tau} X(t)Y(t+\tau)\,dt \qquad (0 \le \tau < T) \qquad (6.8)$$

and

$$\hat{R}_{YX}(\tau) = \frac{1}{T-\tau} \int_0^{T-\tau} Y(t)X(t+\tau)\,dt \qquad (0 \le \tau < T) \qquad (6.9)$$

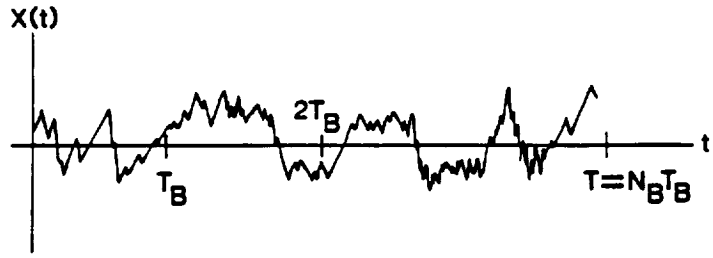


Figure 17. Data broken into blocks.

can be shown to be unbiased estimates of the cross correlations of two jointly weakly stationary random processes. Again, their variances depend on both T and τ and approach zero as $T \to \infty$. Note that the cross correlations are not generally even functions of τ. However, the values for negative lags can be obtained from the relations

$$\hat{R}_{XY}(-\tau) = \hat{R}_{YX}(\tau)$$
$$\hat{R}_{YX}(-\tau) = \hat{R}_{XY}(\tau)$$

6.5 A Test for Stationarity

Because the assumption of stationarity is fundamental to most of the analysis in this monograph, it is important to derive a test to ascertain whether it is reasonable to assume that a particular time history is a sample function from a stationary random process. Fortunately, such a test is readily developed.

Consider a sample function of length T from a random process $X(t)$ and suppose that it is broken into N_B blocks of data of length T_B, as shown in figure 17. Then, an estimate of the mean

$$\hat{m}_X = \frac{1}{T_B} \int_0^{T_B} X(t)\, dt$$

and an estimate of the variance

$$\hat{\sigma}_X^2 = \frac{1}{T_B} \int_0^{T_B} [X(t) - \hat{m}_X]^2\, dt \tag{6.10}$$

can be calculated for each block of data. A test for stationarity is then developed by recalling that the mean and variance of a stationary random process are constants. Thus, the estimates of the mean obtained from the different blocks of data should be equal, except for random variability. In addition, the estimates of the variance obtained

from the different blocks of data should also be equal. If one wants to be sophisticated, there are various statistical tests which will allow one to accept or reject, with a certain level of confidence, the hypotheses that these estimates are all from an underlying random process with constant mean and variance. However, for practical purposes, if these estimates vary more than a few percent for "reasonably sized" blocks of data, one should be cautious in placing much faith in analyses based on stationarity. Here, "reasonable size" is a judgment, based on the frequency content of the data.

It might be mentioned that the estimate of the variance (eq. (6.10)) is an example of a biased estimate. Its mean can be shown to be

$$E\left\{\hat{\sigma}_X^2\right\} = \sigma_X^2 - E\left\{(\hat{m}_X - m_X)^2\right\}$$

where the variance of the estimate of the mean is given by equation (6.5). Thus, the mean value of the estimate of the variance is less than the actual variance. However, this bias does not affect the test for stationarity since all the variance estimates will be biased by the same amount if the process is stationary.

Chapter VII

Estimation of Power Spectral Densities

There are two techniques in common use today for estimation of power spectral densities: the Blackman-Tukey and finite Fourier transform techniques. There are also newer techniques, such as maximum entropy and maximum likelihood, which, though still subjects for research, appear to have certain advantages over the older techniques and may one day become standard. These newer techniques are discussed in chapter XII.

7.1 The Blackman-Tukey Approach

The first practical approach to the estimation of power spectral densities was developed by Blackman and Tukey.[1] Although this approach has been superseded as the standard by the finite Fourier transform approach, it is still valuable for certain applications, particularly when one is analyzing digital data and the number of data points is not a power of two. Further, it illustrates some of the difficulties in spectral estimation more clearly than does the finite Fourier transform. Thus, it will be discussed first.

Recall that the power spectral density is defined as the Fourier transform of the autocorrelation function, that is,

$$S_X(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R_X(\tau)\, e^{-i\omega\tau}\, d\tau$$

From equation (6.7), an estimate of the autocorrelation for $|\tau| < T$ can be obtained as shown in figure 18. Thus, Blackman and Tukey define their estimate as the finite Fourier transform of this function, that is,

$$\hat{S}_X(\omega) = \frac{1}{2\pi} \int_{-T}^{T} \hat{R}_X(\tau)\, e^{-i\omega\tau}\, d\tau \qquad (7.1)$$

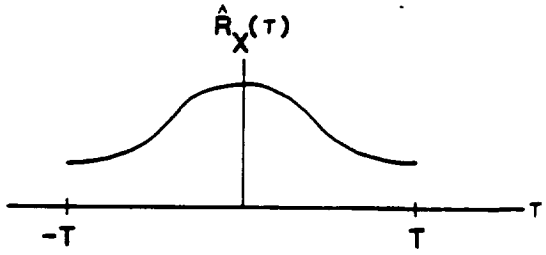


Figure 18. Estimate of autocorrelation.

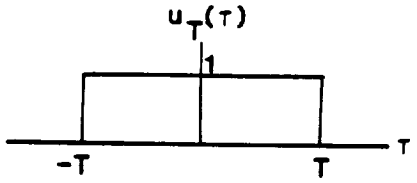


Figure 19. The boxcar function.

In effect, the Blackman-Tukey approach *assumes* that $\hat{R}_X(\tau) = 0$ for $|\tau| > T$. The first question, of course, is how good is this estimate?

Introduce the *boxcar function*

$$u_T(\tau) = \begin{cases} 1 & (|\tau| < T) \\ 0 & (\text{Otherwise}) \end{cases}$$

as shown in figure 19. With the use of this function, the estimate in equation (7.1) may be written as

$$\hat{S}_X(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} u_T(\tau) \hat{R}_X(\tau)\, e^{-i\omega\tau}\, d\tau \tag{7.2}$$

with expectation

$$E\left\{\hat{S}_X(\omega)\right\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} u_T(\tau) R_X(\tau)\, e^{-i\omega\tau}\, d\tau \tag{7.3}$$

since $\hat{R}_X(\tau)$ is an unbiased estimate of $R_X(\tau)$ for $|\tau| < T$. Now, equation (7.3) is the Fourier integral transform of the product of two functions. Thus it may be evaluated by the convolution theorem (eq. (2.8)). That is, since the Fourier transform of the autocorrelation

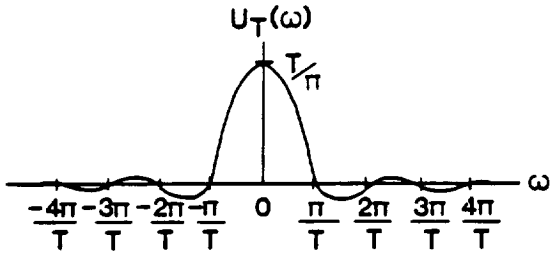


Figure 20. The spectral window $U_T(\omega)$.

is the power spectral density,

$$E\left\{\hat{S}_X(\omega)\right\} = \int_{-\infty}^{\infty} U_T(\omega-\omega')S_X(\omega')\,d\omega' \tag{7.4}$$

where $U_T(\omega)$, called a *spectral window*, is the Fourier transform of the boxcar function and is given by

$$U_T(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} u_T(\tau)\,e^{-i\omega\tau}\,d\tau = \frac{T}{\pi}\,\mathrm{sinc}(\omega T) \tag{7.5}$$

where

$$\mathrm{sinc}\,x = \frac{\sin x}{x} \tag{7.6}$$

is the sinc function, which has the property that $\lim_{x\to 0}(\mathrm{sinc}\,x) = 1$. Although some authors include a π in the definition of the sinc function, equation (7.6) is preferable for time series analysis. The spectral window $U_T(\omega)$ is shown in figure 20. Its main lobe has height T/π and width $2\pi/T$, and its integral is unity since by definition

$$u_T(\tau) = \int_{-\infty}^{\infty} U_T(\omega)\,e^{i\omega\tau}\,d\omega$$

and thus

$$u_T(0) = 1 = \int_{-\infty}^{\infty} U_T(\omega)\,d\omega$$

Therefore, as $T \to \infty$, the spectral window has the characteristics of a delta function, in which case the integral given in equation (7.4) would be $S_X(\omega)$ and the estimate would be unbiased.

However, for finite T, the spectral window $U_T(\omega)$ is not a delta function; it has a main lobe of finite width and both positive and negative side lobes. Thus $\hat{S}_X(\omega)$ is *not* an unbiased estimate of the

power spectral density $S_X(\omega)$. It can be seen from equation (7.4) to be biased because its expectation is a convolution of the actual power spectral density with the spectral window $U_T(\omega)$. Further, the finite width of this window causes a major problem with the frequency resolution of the power spectral density, as will be discussed later in this chapter. There are also other unfortunate aspects to this estimation technique, such as the negative estimates that can be produced at times by the side lobes of the window and the high variability of the estimate, as will be discussed in chapter VIII.

It should be noted, however, that even though this estimate is biased, it is still *power preserving* in the sense that, from equation (7.3),

$$\begin{aligned}\int_{-\infty}^{\infty} E\left\{\hat{S}_X(\omega)\right\} d\omega &= \int_{-\infty}^{\infty} d\tau\, u_T(\tau) R_X(\tau) \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega\tau}\, d\omega \\ &= \int_{-\infty}^{\infty} d\tau\, u_T(\tau) R_X(\tau) \delta(\tau) \\ &= u_T(0) R_X(0) = E\left\{X^2(t)\right\}\end{aligned}$$

since $u_T(0) = 1$. Thus, the mean value of the integral of the spectral estimate is the same as the integral of the actual power spectral density, the total power of the process.

7.2 Windows

In an attempt to improve their estimation technique, Blackman and Tukey developed a new class of estimates

$$\hat{S}_X(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} u(\tau) \hat{R}_X(\tau)\, e^{-i\omega\tau}\, d\tau \tag{7.7}$$

where $u(\tau)$ is a real function called the *lag window* corresponding to the spectral window

$$U(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} u(\tau)\, e^{-i\omega\tau}\, d\tau \tag{7.8}$$

Since equations (7.7) and (7.2) are identical in form, these new estimates have similar mathematical properties to those discussed earlier. Blackman and Tukey's idea was to pick a lag window that led to a spectral window which has smaller side lobes or other desirable properties and which reduces the variability of the estimate.

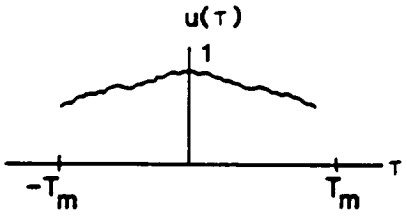


Figure 21. Arbitrary lag window.

All acceptable lag windows must satisfy three requirements:

1. $u(0) = 1$, to preserve power
2. $u(\tau) = u(-\tau)$ so that $\hat{S}_X(\omega)$ is real
3. $u(\tau) = 0$ for $|\tau| > T_m$ where $T_m \leq T$, so that unavailable data are not required

Thus, all lag windows have a shape similar to that shown in figure 21. Since any function that satisfies these three requirements is an even function for $-T_m < \tau < T_m$, it can be expanded in a Fourier cosine series in that region:

$$u(\tau) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos \frac{n\pi\tau}{T_m} \qquad (|\tau| < T_m) \tag{7.9}$$

where

$$u(0) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n = 1$$

Thus, equation (7.9) represents the whole family of windows for different choices of the coefficients a_n. The corresponding spectral window is given by

$$\begin{aligned} U(\omega) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} u(\tau)\, e^{-i\omega\tau}\, d\tau \\ &= \frac{T_m}{2\pi} \sum_{n=-\infty}^{\infty} a_n \operatorname{sinc}(\omega T_m - n\pi) \end{aligned} \tag{7.10}$$

Thus, the arbitrary spectral window consists of a sum of sinc functions occurring at the frequencies $\omega = n\pi/T_m$ as shown in figure 22. Generally, the number of nonzero a_n's determines the effective width

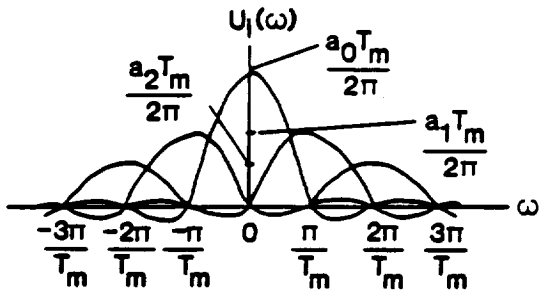


Figure 22. Arbitrary spectral window.

of the main lobe of the window. However, the positive lobes of one sinc function tend to cancel the negative lobes of another.

Some classical windows. Some lag windows have been so widely used that they are worthy of discussion:

1. Boxcar ($a_0 = 2$, $a_n = 0$ for $|n| > 0$):

$$u(\tau) = \begin{cases} 1 & (|\tau| < T_m) \\ 0 & (\text{Otherwise}) \end{cases}$$

This window, which was the original Blackman-Tukey window, is shown in figure 19 and its corresponding spectral window is displayed in figure 20.

2. Hanning ($a_0 = 1$, $a_1 = 1/2$, $a_n = 0$ for $|n| > 1$):

$$u(\tau) = \begin{cases} \frac{1}{2}(1 + \cos \frac{\pi\tau}{T_m}) & (|\tau| < T_m) \\ 0 & (\text{Otherwise}) \end{cases}$$

This window was named for the Austrian meteorologist Julius von Hann[1] and is probably the most widely used window for Blackman-Tukey spectral estimation.

3. Hamming ($a_0 = 1.08$, $a_1 = 0.46$, $a_n = 0$ for $|n| > 1$):

$$u(\tau) = \begin{cases} 0.54 + 0.46 \cos \frac{\pi\tau}{T_m} & (|\tau| < T_m) \\ 0 & (\text{Otherwise}) \end{cases}$$

This window was named for R. W. Hamming.[1] Although it minimizes the height of the side lobes for a two-term series, it has a discontinuity at $|\tau| = T_m$ which is not present in the Hanning window. However, note that the Hamming and Hanning windows differ negligibly.

4. Cosine bell:

$$u(\tau) = \begin{cases} \cos \frac{\pi\tau}{2T_m} & (|\tau| < T_m) \\ 0 & (\text{Otherwise}) \end{cases}$$

Although this window corresponds to a complicated set of a_n's, its transform can be written

$$U(\omega) = \frac{T_m}{2} \frac{\cos \omega T_m}{(\pi^2/4 - \omega^2 T_m^2)}$$

Figure 23 presents a comparison of the cosine bell and Hanning lag windows and figure 24 presents a comparison of the corresponding spectral windows.

5. Bartlett:

$$u(\tau) = \begin{cases} 1 - \frac{|\tau|}{T_m} & (|\tau| < T_m) \\ 0 & (\text{Otherwise}) \end{cases}$$

This window was named for M. S. Bartlett.[11] Again it corresponds to a complicated set of a_n's. However, its transform may be written

$$U(\omega) = \frac{T_m}{2\pi} \operatorname{sinc}^2 \left(\frac{\omega T_m}{2} \right)$$

since it is produced by the convolution of two boxcar windows. As can be seen, the Bartlett spectral window has no negative side lobes. However, its main lobe is very broad.

Cross spectral densities. With the Blackman-Tukey approach, similar estimates of the cross spectral densities may be developed. For example, from equations (6.8) and (6.9), the cross correlation $R_{XY}(\tau)$ may be estimated for $|\tau| < T$. Then

$$\hat{S}_{XY}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} u(\tau) \hat{R}_{XY}(\tau) e^{-i\omega\tau} d\tau \qquad (7.11)$$

yields an estimate of the cross power spectral density.

7.3 The Finite Fourier Transform Approach

The technique for power spectral density estimation in most common use today is based on the Finite Fourier Transform of the data rather than on the transform of the autocorrelation function. Historically, this technique was developed first and produced what was called the

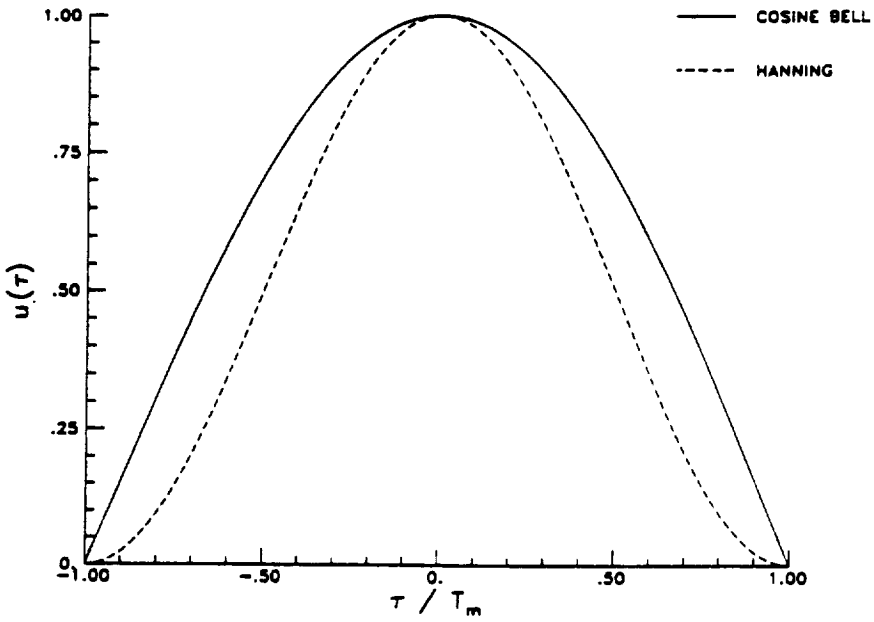


Figure 23. Comparison of cosine bell and Hanning lag windows.

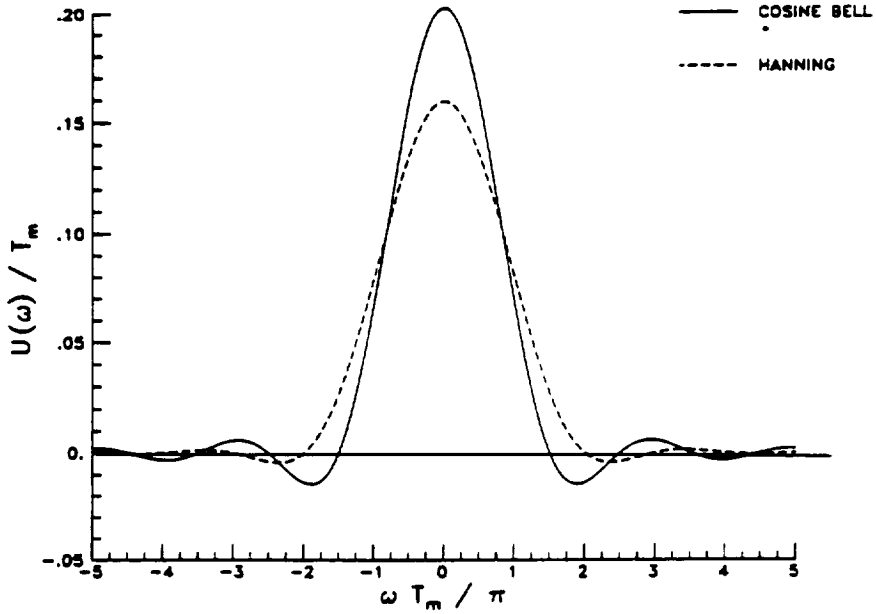


Figure 24. Comparison of cosine bell and Hanning spectral windows.

periodogram.[11] However, it was only with the introduction of the Fast Fourier Transform (to be discussed in chapter X) that the technique became practical for large amounts of data.

From a single sample function of length T of a random process $X(t)$, it is clear that the Fourier integral transform of $X(t)$ cannot be computed because the data are not known for infinite time. However, the finite Fourier transform

$$X_T(\omega) = \frac{1}{2\pi}\int_0^T X(t)\, e^{-i\omega t}\, dt \tag{7.12}$$

can be calculated. Further

$$\begin{aligned} E\left\{X_T^*(\omega)X_T(\omega)\right\} &= \frac{1}{4\pi^2}\int_0^T dt_1 \int_0^T dt_2\, R_X(t_2-t_1)\, e^{-i\omega(t_2-t_1)} \\ &= \frac{T}{2\pi}\left[\frac{1}{2\pi}\int_{-T}^{T} R_X(\tau)\, e^{-i\omega\tau}\, d\tau \right. \\ &\quad \left. - \frac{1}{2\pi T}\int_{-T}^{T} |\tau| R_X(\tau)\, e^{-i\omega\tau}\, d\tau\right] \end{aligned} \tag{7.13}$$

where equation (6.6) has been employed. Now, if $X(t)$ is a completely random process, the second integral in equation (7.13) is finite. Thus, one finds that

$$\lim_{T\to\infty} \frac{2\pi}{T} E\left\{|X_T(\omega)|^2\right\} = S_X(\omega) \tag{7.14}$$

since the second integral is driven to zero by the T^{-1} multiplier and the first integral becomes the power spectral density.

Based on this relation, the class of power spectral estimates

$$\hat{S}_X(\omega) = W_S |X_F(\omega)|^2 \tag{7.15}$$

may be introduced where

$$X_F(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d(t) X(t)\, e^{-i\omega t}\, dt \tag{7.16}$$

is a Fourier transform of the data as seen through the window function $d(t)$. The *data window* is a real function with the property that $d(t) = 0$ for $t < 0$ and $t > T$ so that unavailable data are not required. The correction factor W_S is to be determined.

How good is such an estimate? Again, for ω fixed, this estimate is a random variable with mean

$$E\left\{\hat{S}_X(\omega)\right\} = \frac{W_S}{4\pi^2}\int_{-\infty}^{\infty} dt_1 \int_{-\infty}^{\infty} dt_2\, d(t_1)d(t_2)R_X(t_2-t_1)\, e^{-i\omega(t_2-t_1)}$$
$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{W_S}{2\pi}\int_{-\infty}^{\infty} d(t)d(t-\tau)\, dt\, R_X(\tau)\, e^{-i\omega\tau}\, d\tau \quad (7.17)$$

upon setting $t_1 = t - \tau$ and $t_2 = t$. Note that equation (7.17) is analogous to equation (7.3) with the equivalent lag window

$$u(\tau) = \frac{W_S}{2\pi}\int_{-\infty}^{\infty} d(t)d(t-\tau)\, dt \quad (7.18)$$

which is a convolution of the data window with itself. Thus, the finite Fourier transform and Blackman-Tukey techniques are mathematically equivalent in their expectations as long as equation (7.18) satisfies the conditions for a lag window. The first condition, that $u(0) = 1$, for power preservation requires that

$$u(0) = \frac{W_S}{2\pi}\int_{-\infty}^{\infty} d^2(t)\, dt = 1$$

In other words, the window correction factor must be

$$W_S = \frac{2\pi}{\int_{-\infty}^{\infty} d^2(t)\, dt}$$

Thus, the estimate in equation (7.15) becomes

$$\boxed{\hat{S}_X(\omega) = \frac{2\pi}{\int_{-\infty}^{\infty} d^2(t)\, dt}|X_F(\omega)|^2} \quad (7.19)$$

The second condition, that $u(\tau)$ be an even function of τ, is identically satisfied since

$$u(-\tau) = \frac{W_S}{2\pi}\int_{-\infty}^{\infty} d(t)d(t+\tau)\, dt$$
$$= \frac{W_S}{2\pi}\int_{-\infty}^{\infty} d(t'-\tau)d(t')\, dt' = u(\tau)$$

upon setting $t' = t + \tau$. The third condition, that $u(\tau) = 0$ for $|\tau| > T$ is also identically satisfied since $u(\tau)$ is the convolution of

two data windows that are nonzero only in the range $(0, T)$. Thus, equation (7.18) does satisfy the conditions for a lag window, and it can be seen that the finite Fourier transform and Blackman-Tukey spectral estimation techniques are equivalent in their expectations. That is, *finite Fourier transform estimation with a data window d(t) corresponds to Blackman-Tukey estimation with the lag window given by equation (7.18).* A converse statement is also valid.

One comforting property of the finite Fourier transform technique can be seen by introducing the Fourier transform of the data window

$$D(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d(t)\, e^{-i\omega t}\, dt \tag{7.20}$$

where

$$d(t) = \int_{-\infty}^{\infty} D(\omega)\, e^{i\omega t} d\omega$$

Then, the equivalent spectral window is given by the Fourier transform of equation (7.18). Note that the convolution theorem (eq. (2.9)) cannot be applied to equation (7.18) directly since the convolution in equation (7.18) depends on $t - \tau$ rather than on $\tau - t$. However, the Fourier transform of $f(-t)$ is $F(-\omega)$. Thus, by equation (2.9), the equivalent spectral window is

$$U(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} u(\tau)\, e^{-i\omega\tau}\, d\tau = W_S D(\omega) D(-\omega)$$

Further, since $d(t)$ is real, $D(-\omega) = D^*(\omega)$. Therefore

$$U(\omega) = W_S |D(\omega)|^2 \tag{7.21}$$

and it is seen that the equivalent spectral window is always non-negative. Thus, the finite Fourier transform technique cannot yield negative spectral estimates.

Any of the lag windows introduced previously may be employed as data windows by setting

$$d(t) = u(2t - T) \tag{7.22}$$

which merely amounts to shifting the range from $(-T, T)$ to $(0, T)$. The cosine bell window and the boxcar window modified by cosine tapers[12] on each end are probably the most widely used windows in finite Fourier transform estimation.

Cross spectral density estimation. Again, similar estimates for the cross spectral densities may be developed. For example,

$$\hat{S}_{XY}(\omega) = W_S X_F^*(\omega) Y_F(\omega) \tag{7.23}$$

where $Y_F(\omega)$ is given by equation (7.16) with $X(t)$ replaced by $Y(t)$.

Autocorrelation and cross correlation estimates. If the power spectral density is estimated by the finite Fourier transform, the autocorrelation could then be estimated by Fourier transformation, that is,

$$\hat{R}_X(\tau) = \int_{-\infty}^{\infty} \hat{S}_X(\omega)\, e^{i\omega\tau}\, d\omega \tag{7.24}$$

Again, the question is how good is this estimate? Its expected value is

$$E\left\{\hat{R}_X(\tau)\right\} = \int_{-\infty}^{\infty} E\left\{\hat{S}_X(\omega)\right\} e^{i\omega\tau}\, d\omega$$

where, by equation (7.17),

$$E\left\{\hat{S}_X(\omega)\right\} = \frac{1}{2\pi}\int_{-\infty}^{\infty} u(\tau) R_X(\tau)\, e^{-i\omega\tau}\, d\tau$$

and $u(\tau)$ is given by equation (7.18). Thus,

$$E\left\{\hat{R}_X(\tau)\right\} = u(\tau) R_X(\tau)$$

and the estimate in equation (7.24) is seen to be biased. Even if the boxcar data window is employed,

$$u(\tau) = \frac{1}{T}\int_{-\infty}^{\infty} d(t)d(t-\tau)\, dt = 1 - \frac{|\tau|}{T} \qquad (|\tau| < T)$$

which is a Bartlett lag window, and the estimate is still biased:

$$E\left\{\hat{R}_X(\tau)\right\} = \left(1 - \frac{|\tau|}{T}\right) R_X(\tau)$$

and in order to make the estimate unbiased, it is necessary to define a new estimate

$$\boxed{\hat{R}_X(\tau) = W_R \int_{-\infty}^{\infty} \hat{S}_X(\omega)\, e^{i\omega\tau}\, d\omega} \tag{7.25}$$

where W_R is the window correction factor

$$W_R = \frac{1}{u(\tau)}$$

Similar estimates may be developed for the cross correlations, for example,

$$\hat{R}_{XY}(\tau) = W_R \int_{-\infty}^{\infty} \hat{S}_{XY}(\omega)\, e^{i\omega\tau}\, d\omega \tag{7.26}$$

7.4 Frequency Resolution

The major problem introduced by a finite data record length is that of frequency resolution. Since the finite Fourier transform and Blackman-Tukey spectral estimation techniques have an equivalent bias, this problem occurs in both techniques. Thus, the problem of frequency resolution will be analyzed for the simplest case of a Blackman-Tukey estimate with a boxcar lag window.

If a random process $X(t)$ consists only of a single sinusoid with amplitude A and random phase ϕ, that is,

$$X(t) = A\cos(\omega_1 t + \phi)$$

then, its autocorrelation is

$$R_X(\tau) = \frac{A^2}{2}\cos\omega_1\tau$$

with power spectral density

$$S_X(\omega) = \frac{A^2}{4}\left[\delta(\omega-\omega_1) + \delta(\omega+\omega_1)\right] \tag{7.27}$$

However, if equation (7.27) is substituted in equation (7.4), the expected value of the spectral estimate is

$$E\left\{\hat{S}_X(\omega)\right\} = \frac{A^2}{4}\left[U_T(\omega-\omega_1) + U_T(\omega+\omega_1)\right] \tag{7.28}$$

as shown in figure 25. Thus, the spectral estimate consists of two reproductions of the spectral window. Further, since the window has negative side lobes, the estimate is negative at certain frequencies even though the power spectral density, by definition, must be non-negative! It is only in the limit as $T \to \infty$ that non-negative estimates consisting of two "spikes" at $\omega = \pm\omega_1$ are obtained.

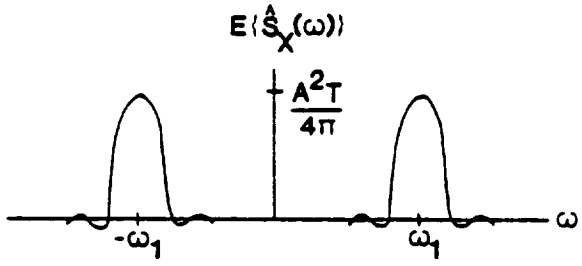


Figure 25. Spectral estimate of sinusoidal signal.

Now, suppose that the signal consists of two equal-amplitude sinusoids at frequencies ω_1 and ω_2, that is,

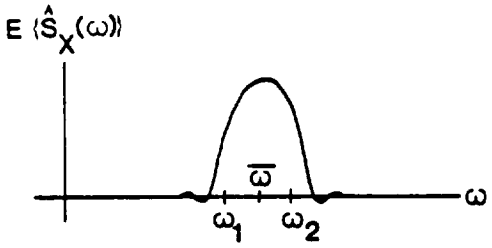
$$X(t) = A\cos(\omega_1 t + \phi_1) + A\cos(\omega_2 t + \phi_2)$$

with random phase angles. Then, the expected value of the spectral estimate is

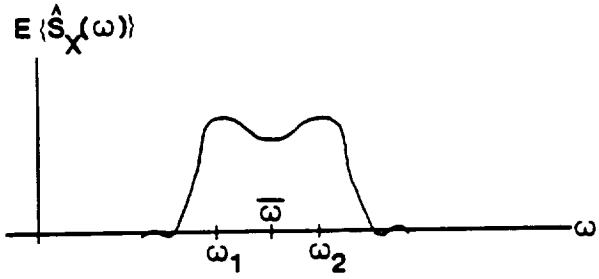
$$E\left\{\hat{S}_X(\omega)\right\} = \frac{A^2}{4}\left[U_T(\omega-\omega_1) + U_T(\omega-\omega_2) + U_T(\omega+\omega_1) + U_T(\omega+\omega_2)\right] \quad (7.29)$$

which is a sum of four window functions, two centered at ω_1 and ω_2 and the other two centered at $-\omega_1$ and $-\omega_2$. Depending on the frequency *separation* $\Delta\omega = (\omega_2 - \omega_1)$ and the characteristics of the spectral window, it may or may not be possible to determine from the estimate that two separate frequencies are present. In the analysis that follows. it is sufficient to consider only the window reproductions at the positive frequencies ω_1 and ω_2. Three cases are possible as shown in figure 26. In figure 26(a), the two frequencies are so close together that the sum of the two window functions that represent them actually peaks at the mean frequency $\bar{\omega} = (\omega_1 + \omega_2)/2$, and no peak is visible at the frequencies ω_1 and ω_2. In figure 26(b), the frequencies are better separated and peaks at ω_1 and ω_2 are visible. However, the two window functions merge into one another. Finally, in figure 26(c), the frequencies are sufficiently separated that the two window functions are distinct.

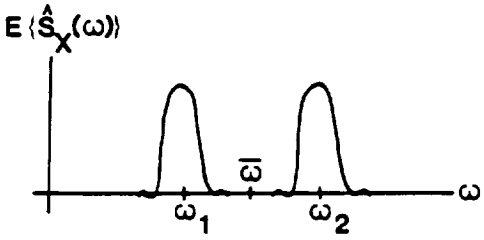
A criterion can be developed that will determine which of these three conditions is present for the spectral window given by equation (7.5)



(a) Unresolved components.



(b) Partially resolved components.



(c) Fully resolved components.

Figure 26. Spectral estimate of signal containing two sinusoidal components.

and various separations of the two frequencies. Define the sum of the two window functions as

$$f(\omega) = U_T(\omega-\omega_1) + U_T(\omega-\omega_2)$$
$$= \frac{2}{\pi}\left[\frac{(\omega-\bar{\omega})\sin(\omega-\bar{\omega})T\cos\left(\frac{\Delta\omega}{2}\right)T - \left(\frac{\Delta\omega}{2}\right)\cos(\omega-\bar{\omega})T\sin\left(\frac{\Delta\omega}{2}\right)T}{(\omega-\bar{\omega})^2 - \left(\frac{\Delta\omega}{2}\right)^2}\right] \tag{7.30}$$

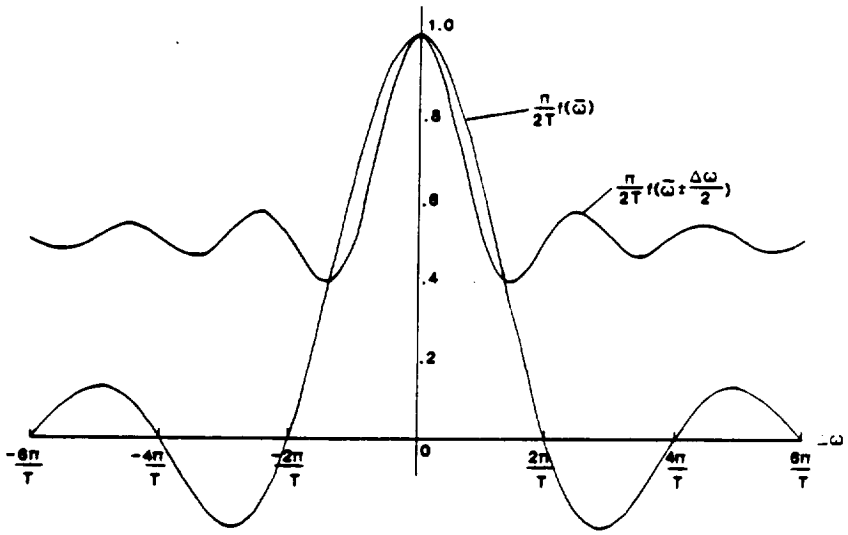


Figure 27. Peak amplitude functions.

where $\Delta\omega = (\omega_2 - \omega_1)$. The value of this function at the mean frequency $\bar{\omega} = (\omega_1 + \omega_2)/2$ is

$$f(\bar{\omega}) = \frac{2T}{\pi} \operatorname{sinc}\left(\frac{\Delta\omega}{2}\right) T \tag{7.31}$$

while its value at the frequencies ω_1 and ω_2 can be obtained from a limiting process as

$$f\left(\bar{\omega} \pm \frac{\Delta\omega}{2}\right) = \frac{T}{\pi}\left[1 + \operatorname{sinc}\left(\frac{\Delta\omega}{2}\right) T \cos\left(\frac{\Delta\omega}{2}\right) T\right] \tag{7.32}$$

In terms of these two values, the criterion for resolution may be given as

Unresolved:	$f(\bar{\omega}) \geq f\left(\bar{\omega} \pm \frac{\Delta\omega}{2}\right)$
Partially resolved:	$f\left(\bar{\omega} \pm \frac{\Delta\omega}{2}\right) > f(\bar{\omega}) > 0$
Fully resolved:	$f(\bar{\omega}) = 0$

Figure 27 is a plot of equations (7.31) and (7.32) as a function of the separation of the two frequencies $\Delta\omega$. From this figure, it can be seen

that partial resolution requires

$$|\Delta\omega| > \frac{4.2}{T}$$

while full resolution requires the two equal-amplitude sinusoids to be separated in frequency by

$$\boxed{|\Delta\omega| = |\omega_2 - \omega_1| \geq \frac{2\pi}{T}} \tag{7.33}$$

where the first zero of $f(\bar{\omega})$ has been utilized.

This analysis of the resolution problem has assumed that the two sinusoids were of equal amplitude. Clearly, the problem of resolving two sinusoids becomes even more difficult if the sinusoids have unequal amplitudes.

The resolution problem also places a lower limit on the frequencies that can be observed in a record of length T. Any frequency ω lower than $2\pi/T$, or

$$\boxed{f < \frac{1}{T}} \tag{7.34}$$

cannot be differentiated from zero frequency. Thus, frequencies f below $1/T$ appear as a nonzero mean or linear trend in the data.

Bias. If the signal $X(t)$ is a completely random process, the preceding discussion of frequency resolution may, on occasion, make matters look worse than they really are. Recalling the definition of the boxcar function and the fact that the autocorrelation is even, it can be seen from equation (7.3) that

$$\begin{aligned} E\left\{\hat{S}_X(\omega)\right\} &= \frac{1}{\pi}\int_0^\infty u_T(\tau) R_X(\tau)\cos\omega\tau\, d\tau \\ &= S_X(\omega) - \frac{1}{\pi}\int_T^\infty R_X(\tau)\cos\omega\tau\, d\tau \end{aligned}$$

and thus

$$\text{Bias} = E\left\{\hat{S}_X(\omega)\right\} - S_X(\omega) = -\frac{1}{\pi}\int_T^\infty R_X(\tau)\cos\omega\tau\, d\tau \tag{7.35}$$

Therefore, the bias depends on the values of the autocorrelation at lags greater than the record length. If $X(t)$ contains periodic signals

or very low frequency components, these autocorrelation values are nonzero and the preceding discussion is relevant. However, if $X(t)$ is a completely random process, the autocorrelation may well become essentially zero before the end of the record length, in which case the estimate is unbiased. It is surprising that this result has received very little attention in the literature, as the conditions for its validity are often fulfilled in practice.

When the estimate is biased, a better understanding of the bias can be obtained by noting that equation (7.4) may also be written

$$E\left\{\hat{S}_X(\omega)\right\} = \int_{-\infty}^{\infty} U_T(\lambda) S_X(\omega - \lambda)\, d\lambda \tag{7.36}$$

Because $U_T(\lambda)$ is highly peaked about $\lambda = 0$, most of the value of the integral comes from this region. The power spectral density may be expanded in a Taylor series about $\lambda = 0$, with primes indicating derivatives with respect to ω:

$$S_X(\omega-\lambda) = S_X(\omega) - \lambda S'_X(\omega) + \frac{\lambda^2}{2} S''_X(\omega) - \cdots$$

and then equation (7.36) can be approximated by integration over only the main lobe of the window. The result is

$$\begin{aligned} E\left\{\hat{S}_X(\omega)\right\} &\approx \int_{-\pi/T}^{\pi/T} U_T(\lambda) S_X(\omega-\lambda)\, d\lambda \\ &\approx S_X(\omega) \int_{-\pi/T}^{\pi/T} U_T(\lambda)\, d\lambda - S'_X(\omega) \int_{-\pi/T}^{\pi/T} \lambda U_T(\lambda)\, d\lambda \\ &\quad + \frac{S''_X(\omega)}{2} \int_{-\pi/T}^{\pi/T} \lambda^2 U_T(\lambda)\, d\lambda \\ &\approx S_X(\omega) + \frac{S''_X(\omega)}{T^2} \end{aligned}$$

since the first integral is nearly unity (i.e., $u_T(0) = 1$) and the second integral vanishes because the window is an even function. Thus, another expression for the bias may be obtained:

$$\text{Bias} = E\left\{\hat{S}_X(\omega)\right\} - S_X(\omega) \approx \frac{S''_X(\omega)}{T^2} \tag{7.37}$$

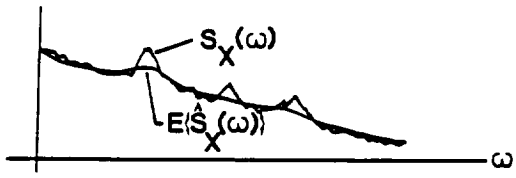


Figure 28. Smoothing produced by estimation.

It can be seen that

$$E\left\{\hat{S}_X(\omega)\right\} > S_X(\omega) \qquad (\text{If } S_X''(\omega) > 0)$$
$$< S_X(\omega) \qquad (\text{If } S_X''(\omega) < 0)$$

That is, the estimate will be too low at maxima of $S_X(\omega)$ and too high at minima and will therefore provide a smoothing of the spectral density as shown in figure 28. This result is to be expected, since the spectral window acts as a moving-average low-pass filter (to be discussed in chapter XI) and suggests a method for attempting to reduce the bias as discussed below.

Prewhitening and postdarkening. One technique that is sometimes used to try to remove the bias caused by estimation is called *prewhitening* and *postdarkening*. Because the bias is proportional to the second derivative of the actual power spectral density, the idea is that if the actual power spectral density were flat or even linear in ω, the bias would be zero.

Suppose that the signal $X(t)$ has a power spectral density with peaks as shown in figure 29. If one could design a filter that has valleys at the frequencies at which the spectrum has peaks, as shown in figure 30, and then pass the signal through the filter, the power spectral density of the output signal $Y(t)$, given by

$$S_Y(\omega) = |H(\omega)|^2 S_X(\omega)$$

would be nearly flat as shown in figure 31. Thus, an estimate of the power spectral density of the random process $Y(t)$ should have little bias. Such a technique is called prewhitening. An estimate of the power spectral density of the original signal $X(t)$ is then obtained by reversing the process, that is,

$$\hat{S}_X(\omega) = \frac{\hat{S}_Y(\omega)}{|H(\omega)|^2}$$

Figure 29. Power spectral density containing peaks.

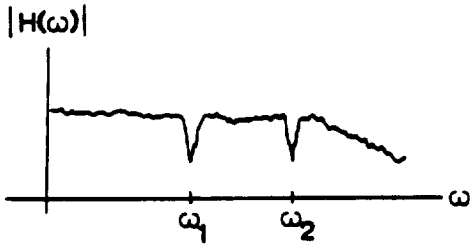


Figure 30. Prewhitening filter.

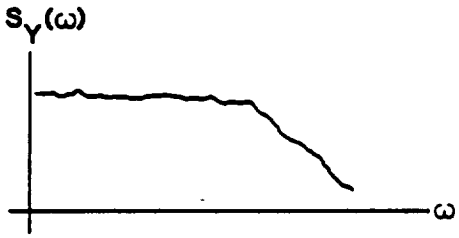


Figure 31. Prewhitened power spectral density.

This technique is called postdarkening.

To actually implement the technique, one would first obtain a raw estimate $\hat{S}_X(\omega)$, use this to design the filter, and then carry out the procedure. It could even be done iteratively if necessary. The process for designing such a filter will be discussed in chapter XI.

Chapter VIII

Uncertainty in Power Spectral Estimates

The spectral estimates discussed in the previous chapter also vary about their mean values, depending on which particular sample function of the random process is employed. This uncertainty can be very large (as will be seen), and thus means for reducing it must be devised. The two common techniques for spectral estimation use completely different approaches to solve this problem.

8.1 Understanding of Uncertainty

An understanding of uncertainty in spectral estimates can be gained from one case which can be worked out completely, because of the assumed normality of the random process. *Suppose that $X(t)$ is a stationary, normal random process with mean zero and variance σ_X^2.* Then by analogy to equation (7.14), consider the random variable (for ω fixed)

$$Z(\omega) = \lim_{T\to\infty} \frac{2\pi}{T} |X_T(\omega)|^2 \tag{8.1}$$

whose expected value is the power spectral density $S_X(\omega)$. Now, from equation (7.12), $Z(\omega)$ may be written

$$Z(\omega) = X_1^2(\omega) + X_2^2(\omega) \tag{8.2}$$

where

$$\left.\begin{aligned} X_1(\omega) &= \lim_{T\to\infty} \frac{1}{\sqrt{2\pi T}} \int_0^T X(t) \cos\omega t \, dt \\ X_2(\omega) &= \lim_{T\to\infty} \frac{1}{\sqrt{2\pi T}} \int_0^T X(t) \sin\omega t \, dt \end{aligned}\right\} \tag{8.3}$$

Since the operations on the normal random process $X(t)$ represented by these expressions are linear, $X_1(\omega)$ and $X_2(\omega)$ are normal random processes, or for ω fixed, normal random variables. (See chapter III.)

Further, it can be shown by taking expectations of equations (8.3) that

$$E\{X_1(\omega)\} = E\{X_2(\omega)\} = 0 \tag{8.4a}$$

and

$$E\left\{X_1^2(\omega)\right\} = E\left\{X_2^2(\omega)\right\} = \frac{S_X(\omega)}{2} \tag{8.4b}$$

Thus, $X_1(\omega)$ and $X_2(\omega)$ are identically distributed. In addition,

$$E\{X_1(\omega)X_2(\omega)\} = 0$$

and so they are uncorrelated; that is, for all ω, the correlation coefficient of these two random variables is

$$\rho = \frac{E\{X_1(\omega)X_2(\omega)\}}{\sigma^2} = 0$$

where $\sigma^2 = S_X(\omega)/2$. Therefore their joint density function (see chapter III) factors, that is,

$$f_{X_1X_2}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$$

Thus, $X_1(\omega)$ and $X_2(\omega)$ are independent random variables and $Z(\omega)$, as given by equation (8.2), is *the sum of squares of two independent, identically distributed normal random variables.*

8.2 Application of the Chi-Square Random Variable to Spectral Estimation

The random variable

$$Z(\omega) = \lim_{T\to\infty} \frac{2\pi}{T}|X_T(\omega)|^2$$

is, except for the limiting process, the power spectral density estimate given by equation (7.19) with a boxcar data window. Further, it has been shown that $Z(\omega)$ is a sum of squares of two independent, normal random variables X_1 and X_2 with means zero and variances $S_X(\omega)/2$. Thus, the random variable

$$Z_0(\omega) = \frac{2Z(\omega)}{S_X(\omega)} = \left[\frac{X_1(\omega)}{\sqrt{S_X(\omega)/2}}\right]^2 + \left[\frac{X_2(\omega)}{\sqrt{S_X(\omega)/2}}\right]^2$$

is a chi-square random variable with two degrees of freedom, as discussed in chapter III, and its mean is

$$E\{Z_0(\omega)\} = 2$$

Figure 32 is a plot of the variation of a chi-square random variable about its mean as a function of degrees of freedom k. Eighty percent of the values taken by the random variable will lie between the bounds shown. Thus, these bounds represent the 80-percent confidence limits on the random variable. Although other limits could have been plotted, these limits have become standard for use in spectral estimation.

Since $Z_0(\omega)$ is a chi-square random variable with two degrees of freedom, it can be seen from figure 32 that

$$P\left\{0.1 < \frac{Z_0(\omega)}{2} < 2.3\right\} = P\left\{0.1 < \frac{Z(\omega)}{S_X(\omega)} < 2.3\right\}$$
$$= P\{0.1S_X(\omega) < Z(\omega) < 2.3S_X(\omega)\} = 0.8 \quad (8.5)$$

which says that if $Z(\omega)$ is viewed as a spectral estimate, 80 percent of the time it will lie between 10 percent and 230 percent of the actual power spectral density. Clearly, this uncertainty is unacceptably large.

8.3 Block Average

The finite Fourier transform spectral estimate given by equation (7.19) is essentially equal to $Z(\omega)$ and thus can be expected to have similar wide variability. Therefore, all spectral estimates by equation (7.19) will be assumed to be essentially chi-square random variables with two degrees of freedom. *This assumption will be employed for all random data, even when the underlying random process is not known to be Gaussian.* Since the Gaussian model fits many real world phenomena, such an assumption may not be unreasonable, especially if one requires only a relative evaluation of uncertainty of one spectral estimate with respect to another.

In order to obtain less variable estimates, suppose that the data are broken into N_B blocks of length T_B such that $N_B T_B = T$, as shown in figure 17. Then a technique similar to the test for stationarity in chapter VI can be employed. A spectral estimate $\hat{S}_X(j)$ for $j = 1, 2, \ldots, N_B$ can be made over each block of data. Then, assuming independence of the blocks, each estimate is a chi-square

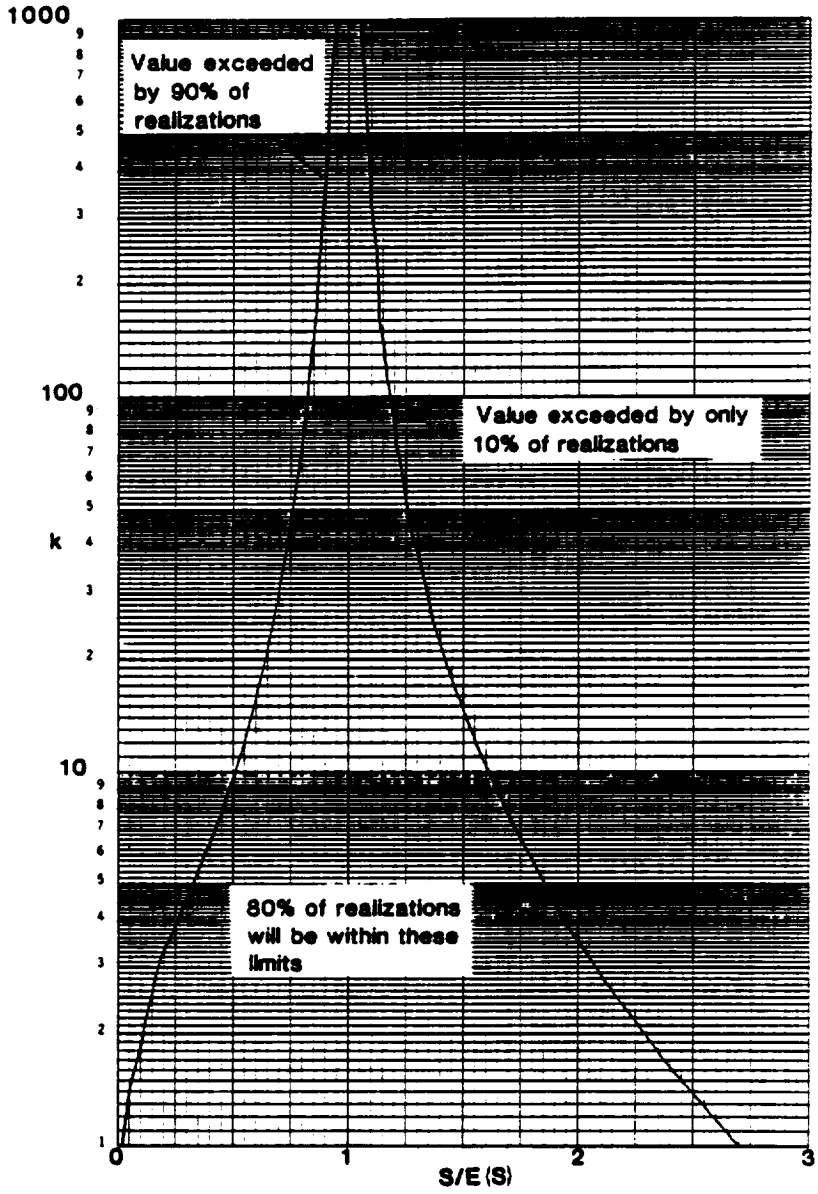


Figure 32. Variation of chi-square random variable.

random variable with two degrees of freedom and their average

$$\overline{\hat{S}_X(\omega)} = \frac{1}{N_B}\sum_{j=1}^{N_B} \hat{S}_X^{(j)}(\omega)$$

is essentially a chi-square random variable with degrees of freedom k given by

$$k = 2N_B$$

As can be seen in figure 32, this greatly reduces the uncertainty of the estimate. In fact, if $a(k)$ is the left bound and $b(k)$ the right bound in figure 32, it can be seen by analogy with equation (8.5) that

$$P\left\{a(k) < \frac{\overline{\hat{S}_X(\omega)}}{S_X} < b(k)\right\} = 0.8$$

or, more meaningfully,

$$P\left\{\frac{\overline{\hat{S}_X(\omega)}}{a(k)} > S_X(\omega) > \frac{\overline{\hat{S}_X(\omega)}}{b(k)}\right\} = 0.8 \qquad (8.6)$$

For instance, if the data are broken into 50 blocks, then $a(100) = 0.82$ and $b(100) = 1.18$. Thus,

$$P\left\{1.22\overline{\hat{S}_X(\omega)} > S_X(\omega) > 0.85\overline{\hat{S}_X(\omega)}\right\} = 0.8$$

or 80 percent of the time the actual spectral density will lie between 85 and 122 percent of the spectral estimate.

Of course, this reduction in variability has not been achieved without cost. Recall that from equation (7.33), the bandwidth of the spectral estimate is given by

$$\Delta\omega = \frac{2\pi}{T} \quad \text{or} \quad \Delta f = \frac{1}{T}$$

If the data are broken into blocks, the effective data length is no longer T, but T_B. Thus, the effective bandwidth of the estimate Δf has increased to

$$\Delta f = \frac{1}{T_B}$$

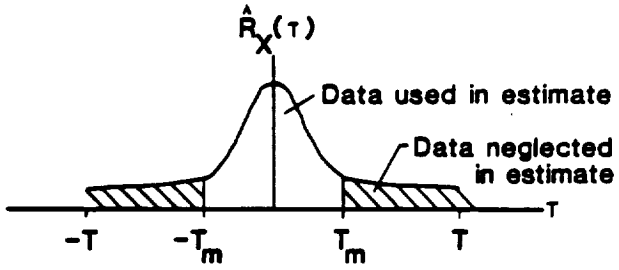


Figure 33. Utilized portion of autocorrelation estimate.

That is, in reducing the variability, the resolution has also been reduced. Writing $k = 2N_B = 2T/T_B$ yields the fundamental relation

$$k = 2\Delta f T \tag{8.7}$$

That is, the degrees of freedom are equal to twice the bandwidth in hertz times the data length. Thus, for T fixed, a tradeoff dilemma.

Reduced variability $\Longleftrightarrow$ Reduced resolution

is apparent. The only way out of this dilemma is to obtain more data (if possible), that is, to increase T.

8.4 Uncertainty Analysis for the Blackman-Tukey Technique

Recall that the Blackman-Tukey spectral estimate is

$$\hat{S}_X(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} u(\tau)\hat{R}_X(\tau)\, e^{-i\omega\tau}\, d\tau$$

where the lag window $u(\tau) = 0$ for $|\tau| > T_m$. By a lengthy analysis, again assuming normality of the data, Blackman and Tukey[1] were able to show that this estimate can also be considered a chi-square random variable with degrees of freedom

$$k = \frac{2T}{T_m}$$

where T_m is the half-width of the lag window. Thus, if the maximum width $T_m = T$ is employed, again the estimate has two degrees of freedom. The degrees of freedom are increased by taking $T_m < T$.

Since the autocorrelation can be estimated out to $|\tau| = T$, taking $T_m < T$ leads to the disquieting result that variability is *decreased* by

not using data at large lag values τ, as shown in figure 33. This result is not so surprising when one recalls that less data went into the estimate of the autocorrelation at large lag values (see eq. (6.7)), thus making the variability of the autocorrelation estimate itself much higher at large lags.

A similar tradeoff between resolution and variability is seen here as well, since the bandwidth of this estimate will now be $\Delta f = 1/T_m$. Thus, again

$$\boxed{k = 2\Delta f T}$$

This relationship yields the number of degrees of freedom regardless of which estimation technique is employed.

Chapter IX

Digital Time Series Analysis

Most time series analysis is now done with the use of digital data, that is, samples of the random process taken at discrete instants of time. Digital computers are, of course, constrained to such data. However, even stand-alone spectral analyzers usually work with digital data now.

The use of sampled data further complicates the estimation of the various statistical quantities of interest, as will be seen. Usually, the data are assumed to be sampled at equal intervals of time Δt, or at a *sampling rate* of $1/\Delta t$ samples per second. However, in recent years, techniques for analyzing data taken at random time intervals have been developed, as will be discussed in chapter XII.

9.1 Shannon's Sampling Theorem

Much of the understanding of the analysis of equally spaced sampled data is based on a marvelous result usually credited to Shannon[13] although its origin is actually much older[14]. An interesting historical aspect of Shannon's work is that, although it was submitted to the journal in which it appeared in 1940, it was not published until 1949, apparently having been caught up in the secrecy surrounding the war effort.

Suppose that $X(t)$ is a stationary random process and that sampled data $X(n\Delta t)$ exist for $-\infty \leq n \leq \infty$. On the basis of these data, it is desired to estimate the values taken by the random process at all times, that is,

$$\hat{X}(t) = \sum_{n=-\infty}^{\infty} a_n(t) X(n\Delta t) \tag{9.1}$$

In other words, one wishes to *interpolate* between the data points to reconstruct the entire time history. An equivalent interpretation of equation (9.1) is that it is an attempt to expand $X(t)$ in terms of a set of special basis functions $a_n(t)$. The data $X(n\Delta t)$ are the coefficients

of the basis functions. The basis functions $a_n(t)$ are chosen such that the mean square error between $X(t)$ and its estimate $\hat{X}(t)$,

$$\overline{e^2} = E\left\{\left[X(t) - \hat{X}(t)\right]^2\right\} \tag{9.2}$$

is minimized. Since

$$\overline{e^2} = E\left\{\left[X(t) - \sum_{n=-\infty}^{\infty} a_n(t)X(n\Delta t)\right]^2\right\}$$

for t fixed, the minimum occurs when

$$\frac{\partial \overline{e^2}}{\partial a_\ell(t)} = -2E\left\{\left[X(t) - \sum_{n=-\infty}^{\infty} a_n(t)X(n\Delta t)\right] X(\ell\Delta t)\right\} = 0 \tag{9.3}$$

or

$$R_X(t - \ell\Delta t) - \sum_{n=-\infty}^{\infty} a_n(t)R_X[(n-\ell)\Delta t] = 0 \tag{9.4}$$

for $-\infty \le \ell \le \infty$.

Now, recall that

$$R_X(\tau) = \int_{-\infty}^{\infty} S_X(\omega)\, e^{i\omega\tau}\, d\omega$$

and thus, equation (9.4) becomes

$$\int_{-\infty}^{\infty} S_X(\omega)\, e^{-i\omega\ell\Delta t}\left[e^{i\omega t} - \sum_{n=-\infty}^{\infty} a_n(t)e^{i\omega n\Delta t}\right] d\omega = 0 \tag{9.5}$$

Now, comparing

$$\sum_{n=-\infty}^{\infty} a_n(t)e^{i\omega n\Delta t}$$

with equation (2.4) for t fixed, it can be seen that this series is the Fourier series for a function $f(\omega)$ that is periodic with period $2\pi/\Delta t$. Thus, for $-\pi/\Delta t < \omega < \pi/\Delta t$, the function $e^{i\omega t}$ may be represented exactly by the series

$$e^{i\omega t} = \sum_{n=-\infty}^{\infty} a_n(t)\, e^{i\omega n\Delta t}$$

where

$$a_n(t) = \frac{\Delta t}{2\pi} \int_{-\pi/\Delta t}^{\pi/\Delta t} e^{i(t-n\Delta t)\omega}\, d\omega = \operatorname{sinc}(\omega_c t - n\pi)$$

and $\omega_c = \pi/\Delta t$.

Now, suppose $S_X(\omega) = 0$ for $|\omega| > \omega_c$. Then equation (9.5) is satisfied because $S_X(\omega)$ causes the integral to be zero for $|\omega| > \omega_c$ and the term in brackets causes the integral to be zero for $|\omega| < \omega_c$. Thus, the best estimate is

$$\hat{X}(t) = \sum_{n=-\infty}^{\infty} X(n\Delta t) \operatorname{sinc}(\omega_c t - n\pi)$$

The mean square error in this estimate is

$$\overline{e^2} = E\left\{\left(X - \hat{X}\right)^2\right\} = E\left\{X^2 - 2X\hat{X} + \hat{X}^2\right\} = E\left\{\left(X - \hat{X}\right) X\right\}$$

since by equation (9.3), $E\left\{\hat{X}^2\right\} = E\left\{X\hat{X}\right\}$. Thus,

$$\begin{aligned}\overline{e^2} &= E\left\{\left[X(t) - \sum_{n=-\infty}^{\infty} a_n(t) X(n\Delta t)\right] X(t)\right\} \\ &= R_X(0) - \sum_{n=-\infty}^{\infty} a_n(t) R_X(n\Delta t - t) \equiv 0\end{aligned}$$

as can be seen from equation (9.4) by letting $t = \ell\Delta t$. Therefore $\hat{X}(t) = X(t)$!

This result shows that, *if* $S_X(\omega) = 0$ *for* $|\omega| > \omega_c$, then a realization of the random process $X(t)$ is perfectly reproduced from the sampled data $X(n\Delta t)$ by

$$\boxed{X(t) = \sum_{n=-\infty}^{\infty} X(n\Delta t) \operatorname{sinc}(\omega_c t - n\pi)} \tag{9.6}$$

This fundamental result is used in secure communications, long distance telephone transmission, digital music systems, and a host of other applications.

Although the derivation given here was for a stationary random process, the result also holds for deterministic and transient functions as long as their frequency content is limited to $|\omega| < \omega_c$, as can easily be checked in two simple cases:

Case I

Let $t = \ell\Delta t$. Then

$$\text{sinc}(\omega_c \ell \Delta t - n\pi) = \text{sinc}[(\ell - n)\pi] = \delta_{\ell,n}$$

where $\delta_{\ell,n}$ is the Kronecker delta function, which is unity when $\ell = n$ and zero otherwise. Thus,

$$X(\ell\Delta t) = \sum_{n=-\infty}^{\infty} X(n\Delta t)\delta_{\ell,n} = X(\ell\Delta t)$$

and the data are reproduced, regardless of the properties of $X(t)$.

Case II

Consider the deterministic, transient function with amplitude A,

$$X(t) = A\,\text{sinc}(\omega_c t)$$

Then

$$X(n\Delta t) = A\,\text{sinc}(\omega_c n\Delta t) = A\,\text{sinc}(n\pi) = A\delta_{n,0}$$

and equation (9.6)

$$X(t) = \sum_{n=-\infty}^{\infty} A\delta_{n,0}\,\text{sinc}(\omega_c t - n\pi) = A\,\text{sinc}(\omega_c t)$$

reproduces the time history.

9.2 The Nyquist Frequency and Aliasing

The cutoff frequency

$$\boxed{\omega_c = \frac{\pi}{\Delta t} \text{ or } f_c = \frac{1}{2\Delta t}}$$

is called the *Nyquist frequency* and is *the highest frequency that can be reproduced from data sampled at equal intervals* Δt. To see this, suppose

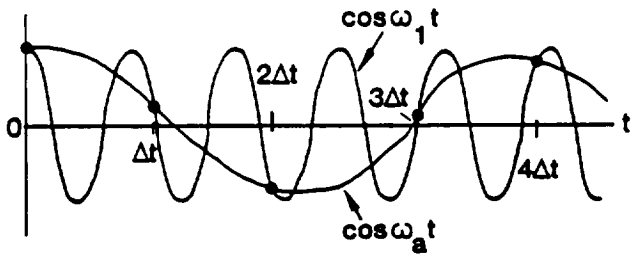


Figure 34. Illustration of aliasing.

that $X(t)$ is a sinusoid of frequency $\omega_1 > \omega_c$, that is,

$$X(t) = A\cos\omega_1 t$$

as shown in figure 34 where the dots represent samples. From the sampled data, the sinusoid of frequency $\omega_1 > \omega_c$ is seen to be indistinguishable from a sinusoid of lower frequency ω_a. Further, $f_c = 1/2\,\Delta t$ corresponds to a sinusoid with period $1/f_c = 2\,\Delta t$. A sinusoid of this frequency would be sampled only twice per period while $\cos\omega_a t$ is seen to be sampled more than twice per period. Thus, $\omega_a < \omega_c$ and the sinusoid of frequency greater than ω_c cannot be distinguished from a sinusoid of frequency less than ω_c. Mathematically, consider the most often encountered case where ω_1 is only slightly larger than ω_c; that is, $\omega_1 = \omega_c + \omega_l$ where $\omega_l < \omega_c$. Then

$$\begin{aligned} X(n\Delta t) &= A\cos\omega_1 n\Delta t = A\cos(\omega_c + \omega_l)n\Delta t \\ &= A\cos(n\pi + \omega_l n\Delta t) \\ &= A(\cos n\pi\cos\omega_l n\Delta t - \sin n\pi\sin\omega_l n\Delta t) \\ &= A\cos(n\pi - \omega_l n\Delta t) \\ &= A\cos(\omega_c - \omega_l)n\Delta t = A\cos\omega_a n\Delta t \end{aligned}$$

where $\omega_a = \omega_c - \omega_l$. Thus, the frequency $\omega_1 = \omega_c + \omega_l$ is indistinguishable in the sampled data from the frequency $\omega_a = \omega_c - \omega_l$. This phenomenon is called *aliasing*, because the frequency ω_1 goes by the new name, or alias, ω_a. Aliasing is the major problem introduced by the use of (equally spaced) sampled data.

The presence of this phenomenon suggests that if one wants to analyze data having a maximum frequency of $f_{\max}$, one must use a sampling rate

$$\boxed{\text{Sampling rate} = \frac{1}{\Delta t} \geq 2f_{\max}} \qquad (9.7)$$

of at least twice the highest frequency in the data. Such data are said to be sufficiently sampled. *To use a much higher sampling rate is wasteful*, since Shannon's sampling theorem implies that there is no more information to be gained from the data. However, in the real world, where few signals are ever truly band limited, some increase is advisable. A $2.5 f_{max}$ sampling rate has been found to give excellent results[12] in most applications.

It should be mentioned that oversampling is sometimes recommended[15] to give more freedom in the reconstruction of sampled data. If the data are sampled at a rate greater than twice the highest frequency, then $S_X(\omega)$ approaches zero at some frequency lower than ω_c. Thus, the term in brackets in equation (9.5) need not be zero between the highest frequency in the data and ω_c. This freedom allows one to choose other functions $a_n(t)$ in the expansion (eq. (9.1)) which converge faster than the sinc functions. These faster converging expansions for $X(t)$ are useful in many applications. such as long distance telephone conversation and photographic image reconstruction.

9.3 Effect of Aliasing on Power Spectral Density

In many applications, one may not know the maximum frequency f_{max}. What happens if one just chooses a sampling rate and estimates the power spectral density?

Let $X(t)$ be a stationary random process with power spectral density

$$S_X(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R_X(\tau)\, e^{-i\omega\tau}\, d\tau \tag{9.8}$$

If the signal is sampled at intervals Δt, then the autocorrelation $R_X(\tau)$ may be estimated (as will be seen later in this chapter) only for $\tau = j\,\Delta t$ for $-\infty < j < \infty$. Thus, the power spectral density estimate, which may include aliasing, is given by the discrete expression (to be shown later in this chapter) of the integral in equation (9.8)

$$\hat{S}_X(\omega) = \frac{\Delta t}{2\pi} \sum_{j=-\infty}^{\infty} \hat{R}_X(j\Delta t)\, e^{-i\omega j \Delta t}$$

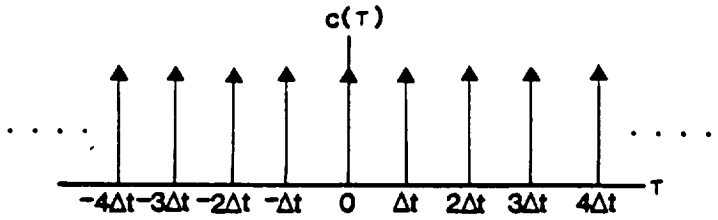


Figure 35. Dirac comb function.

for $|\omega| < \omega_c$. The expected value of this estimate is

$$E\left\{\hat{S}_X(\omega)\right\} = \frac{\Delta t}{2\pi}\sum_{j=-\infty}^{\infty} R_X(j\Delta t)\, e^{-i\omega j\Delta t}$$
$$= \frac{\Delta t}{2\pi}\int_{-\infty}^{\infty} c(\tau) R_X(\tau)\, e^{-i\omega\tau}\, d\tau \qquad (9.9)$$

since $\hat{R}_X(j\Delta t)$, which is estimated from discrete data, will be shown to be an unbiased estimate of $R_X(j\Delta t)$. Here

$$c(\tau) = \sum_{j=-\infty}^{\infty} \delta(\tau - j\Delta t) \qquad (9.10)$$

is the Dirac comb function shown in figure 35.

The Dirac comb function is periodic with period Δt. Thus, it may be expanded in the Fourier series

$$c(\tau) = \frac{1}{\Delta t} + \frac{2}{\Delta t}\sum_{k=1}^{\infty} \cos\frac{2\pi k}{\Delta t}\tau$$

for all τ. The Fourier transform of this series is

$$C(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} c(\tau)\, e^{-i\omega\tau}\, d\tau = \frac{1}{\Delta t}\sum_{k=-\infty}^{\infty} \delta(\omega - 2k\omega_c) \qquad (9.11)$$

another comb function, this time in frequency as shown in figure 36. Thus, the Dirac comb function is one of those unusual functions whose Fourier transform has the same form as the function itself.

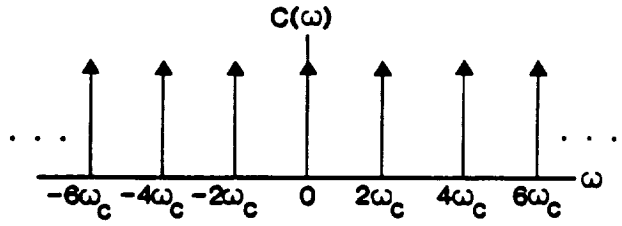


Figure 36. Dirac comb function in frequency.

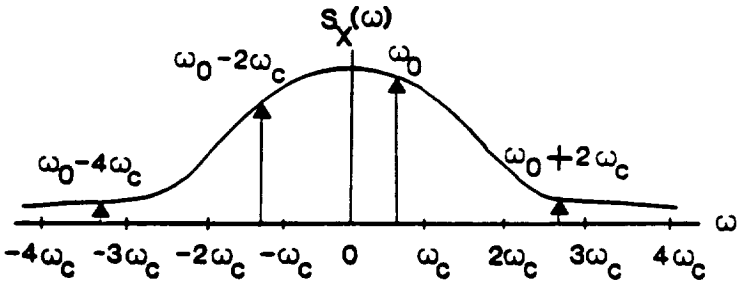


Figure 37. Aliased frequencies in spectral estimation.

Since equation (9.9) is just the Fourier transform of the product of two functions, by the convolution theorem (eq. (2.8)), it may be written

$$E\left\{\hat{S}_X(\omega)\right\} = \Delta t \int_{-\infty}^{\infty} C(\omega) S_X(\omega-\omega')\, d\omega' = \sum_{k=-\infty}^{\infty} S_X(\omega-2k\omega_c) \tag{9.12}$$

which is a sum of the values of the actual power spectral density. In fact, for a power spectral density as shown in figure 37, power at all the frequencies denoted by arrows appears as power at the frequency ω_0 in the aliased spectral density. It should be emphasized that figure 37 represents an atypical case, which would result from very badly undersampled data.

Another way of looking at this phenomenon is to note that since the power spectral density is an even function of ω,

$$S_X(\omega_0-2k\omega_c) = S_X(2k\omega_c-\omega_0)$$

and power is seen to be "folded" about the Nyquist frequency, much as one would fold a fan, with power at odd multiples of ω_c appearing at the frequency ω_c and power at even multiples of ω_c appearing at

the frequency of zero. This situation is illustrated in figure 38. This figure displays two estimates of the power spectral density of sunspot activity obtained from the data mentioned earlier over the years 1610–1960[10]. For the first estimate, the sampling rate was one sample per year yielding a Δt of 1 year and Nyquist frequency of $1/2$ cycle/year. This estimate has 20 degrees of freedom. Note the appearance of a peak near $f = 1/11$ cycle/year corresponding to cyclic behavior in the data with an 11-year period. Also shown in figure 38 is a spectral estimate of the same phenomenon with a Δt of 7 years. Thus, the Nyquist frequency is $1/14$ cycle/year and the sunspot activity is undersampled. In this spectral estimate, the power in the signal at frequencies higher than the Nyquist frequency has been folded about the Nyquist frequency and now appears as power at lower frequencies. In particular, the peak at $f = 1/11$ cycle/year now appears as a peak near $f = 1/21$ cycle/year corresponding to a period of 21 years. If the cyclic behavior was a pure harmonic with frequency $1/11$ cycle/year, which was sampled with a Nyquist frequency of $1/14$ cycle/year, the power at $1/11$ cycle/year would alias back to the frequency $1/14 - (1/11 - 1/14) = 2/14 - 1/11 = 4/77$ cycle/year, which is close to the frequency shown. It might be mentioned that the apparent increase in the amplitude of the peak is due to the power preservation feature of the spectral estimates discussed in chapter VII.

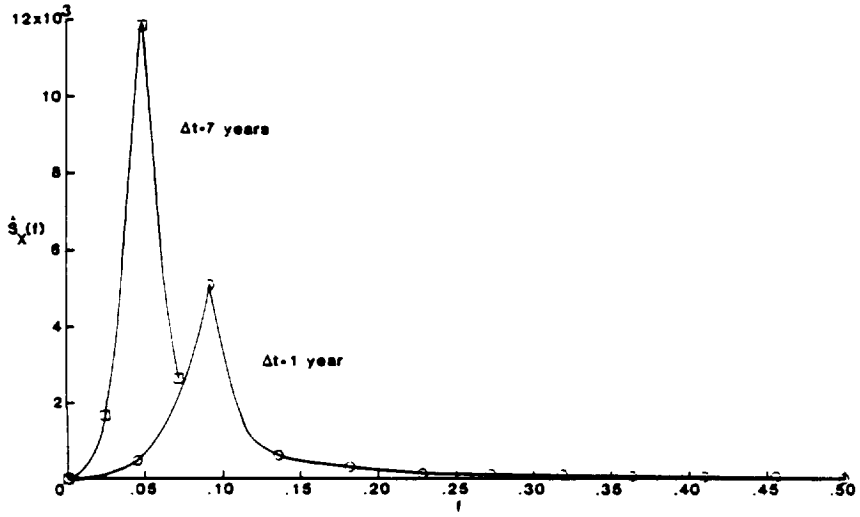


Figure 38. Power spectral density estimates of sunspot activity.

When working with data of unknown frequency content, the only way to be sure of avoiding aliasing is to pass the signal through a

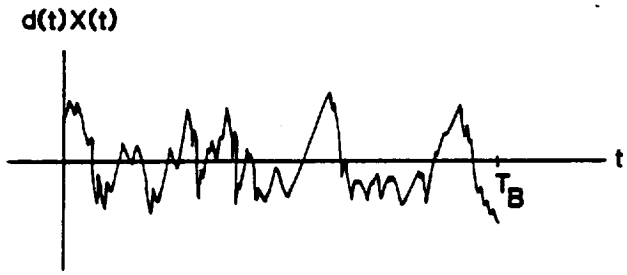


Figure 39. The windowed data function.

low-pass *antialiasing filter*, which filters out all power at frequencies higher than the Nyquist frequency based on the chosen sample interval Δt. Sometimes it is advantageous to do this even when the frequency content is known. For example, if one is interested in only the low-frequency content of a signal that also contains high frequencies, it is possible to reduce the sampling rate and thus the computations required by filtering out the uninteresting higher frequencies. Such *an antialiasing filter must*, of course, *be an analog filter applied to the continuous data before it is digitized*, since after digitization there is no way to distinguish between actual power and aliased power. This point cannot be overemphasized as insufficient sampling will cause the data to be not only worthless but also misleading, as illustrated in figure 38.

9.4 Gibbs' Phenomenon

There is another more subtle way in which aliasing can enter into digital spectral analysis. For such analysis, it is always assumed that the random process $X(t)$ is band limited. However, to use a finite length of data, it is necessary to suppose that the random process $X(t)$ is multiplied by a data window $d(t)$ and to introduce the windowed Fourier transform (eq. (7.16)):

$$X_F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d(t) X(t)\, e^{-i\omega t}\, dt$$

This finite transform is exactly the Fourier integral transform of the windowed data $d(t)X(t)$, which is identically zero outside the region $(0, T_B)$, as shown in figure 39.

Now, it can be proved[4] that a function which is nonzero for only a finite interval of time cannot be band limited in frequency. Thus. the windowed data cannot be band limited and aliasing must result. Unfortunately, nothing can be done about this aliasing as long as one is confined to a finite length of data. However, if T_B is sufficiently large.

the power at high frequencies necessary to reproduce the function is ordinarily so low that it does not destroy the analysis in the frequency range of interest.

Another possible source of aliasing is seen by recalling that the Fourier integral representation (eq. (2.2)),

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

is exact only where $f(t)$ is continuous. At a point of discontinuity t_0, it converges to the average value

$$\frac{f(t_0^+) + f(t_0^-)}{2}$$

of the right- and left-hand limits of $f(t)$ as t approaches t_0, provided that these limits exist. Thus, if there are discontinuities at the beginning and end of the record length, as shown in figure 39, the finite Fourier transform (eq. (7.16)) converges to the average value at these discontinuities. Further, because of Gibbs' phenomenon,[4] if one tries to represent a discontinuous function by a Fourier integral over a finite range of frequencies, the representation produces high-frequency oscillations near the points of discontinuity. Only by allowing unbounded frequencies are these oscillations removed. Thus, the presence of these discontinuities causes more power at the higher frequencies and, thus, more aliasing.

This end-point discontinuity source of aliasing may be reduced by using a data window that is zero at the ends of the record, that is,

$$d(0) = d(T_B) = 0$$

This removes the discontinuities in the record and reduces this cause of aliasing. The Hanning, cosine bell, and Bartlett windows, for example, fulfill this requirement. Some aliasing may still be produced, however, caused by discontinuous derivatives at the ends of the record.

A similar analysis holds for the Blackman-Tukey method of estimation as well. Thus, one ordinarily requires the lag windows to satisfy

$$u(-T_m) = u(T_m) = 0$$

C-2

9.5 Relationship Between Continuous and Discrete Fourier Transforms

Shannon's sampling theorem allows determination of the relationship between Fourier transforms of continuous and discrete data for *band-limited* random processes. To eliminate the effects of aliasing in the following discussion, *it will be assumed that all digital time histories have been passed through an antialiasing filter and are thus band limited.*

Let $X(t)$ be a random process that is limited to the band $|\omega| < \omega_c$, where $\omega_c = \pi/\Delta t$ and Δt is the sample interval. Then, its Fourier integral transform (eq. (2.3)) is given by

$$\begin{aligned} X(\omega) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} X(t)\, e^{-i\omega t}\, dt \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} dt\, e^{-i\omega t} \sum_{n=-\infty}^{\infty} X(n\Delta t)\,\mathrm{sinc}(\omega_c t - n\pi) \\ &= \sum_{n=-\infty}^{\infty} X(n\Delta t)\,\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega t}\,\mathrm{sinc}(\omega_c t - n\pi)\, dt \end{aligned} \tag{9.13}$$

where Shannon's sampling theorem (eq. (9.6)) has been applied.

Now, the Fourier transform of the sinc function is

$$\begin{aligned} \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega t}\,\mathrm{sinc}(\omega_c t - n\pi)\, dt &= \frac{e^{-in\pi\omega/\omega_c}}{2\pi\omega_c}\int_{-\infty}^{\infty} \mathrm{sinc}\, y\, e^{-i\omega y/\omega_c}\, dy \\ &= \begin{cases} \dfrac{e^{-in\pi\omega/\omega_c}}{2\omega_c} & (|\omega| < \omega_c) \\ 0 & (\text{Otherwise}) \end{cases} \end{aligned}$$

upon setting $y = \omega_c t - n\pi$, since

$$\int_{-\infty}^{\infty} \mathrm{sinc}\,\omega\, e^{-i\omega t}\, d\omega = \begin{cases} \pi & (|t| < 1) \\ 0 & (\text{Otherwise}) \end{cases} \tag{9.14}$$

Thus, equation (9.13) becomes

$$\boxed{X(\omega) = \begin{cases} \dfrac{\Delta t}{2\pi} \displaystyle\sum_{n=-\infty}^{\infty} X(n\Delta t)\, e^{-i\omega n\Delta t} & (|\omega| < \omega_c) \\ 0 & (\text{Otherwise}) \end{cases}} \tag{9.15}$$

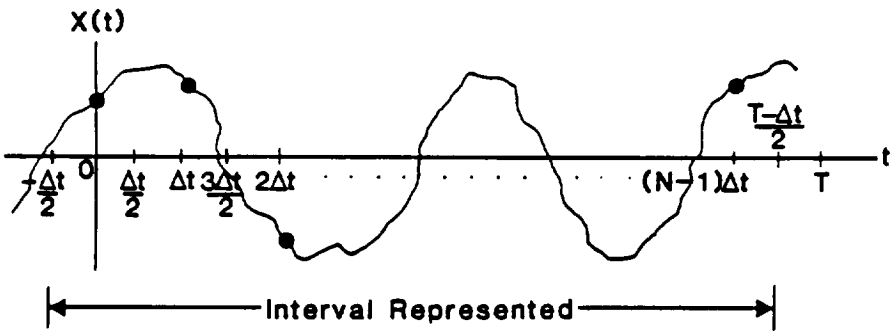


Figure 40. Interval represented by digital data.

since $\omega_c = \pi/\Delta t$. This exact expression would allow calculation of the Fourier integral transform from discrete data $X(n\Delta t)$ for $-\infty < n < \infty$, and is an extension of Shannon's sampling theorem; that is, the Fourier transform of the signal can also be reconstructed from the sampled data.

In actual practice, the amount of data is always finite. Usually, knowledge of the values $X(n\Delta t)$ for $n = 0, 1, 2, \ldots, N-1$ is assumed. These values are taken to represent the sample function $X(t)$ in the interval $(-\Delta t/2, T-\Delta t/2)$ as shown in figure 40 where the dots indicate data points. That is, the data point $X(\ell\Delta t)$ is thought to represent the time history in the interval $((\ell - 1/2)\Delta t, (\ell + 1/2)\Delta t)$.

Now, consider the finite Fourier transform (eq. (7.12)) over this interval

$$X_T(\omega) = \frac{1}{2\pi}\int_{-\Delta t/2}^{T-\Delta t/2} X(t)\, e^{-i\omega t}\, dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} d_0(t) X(t)\, e^{-i\omega t}\, dt \tag{9.16}$$

where $d_0(t)$ is a boxcar function taking the value unity for $-\Delta t/2 < t < T - \Delta t/2$. By the convolution theorem (eq. (2.8)) and substituting equation (9.15),

$$\begin{aligned} X_T(\omega) &= \int_{-\infty}^{\infty} D_0(\omega - \omega') X(\omega') d\omega' \\ &= \frac{\Delta t}{2\pi}\sum_{n=-\infty}^{\infty} X(n\Delta t)\, e^{-i\omega n\Delta t}\int_{\omega-\omega_c}^{\omega+\omega_c} D_0(\omega'')\, e^{i\omega'' n\Delta t}\, d\omega'' \end{aligned}$$

where $D_0(\omega)$ is the Fourier transform of $d_0(t)$. The latter integral, which is to be evaluated only for $|\omega| < \omega_c$, represents an ideal filter

operating on the boxcar function, that is,

$$I(n, N, \omega) = \int_{\omega-\omega_c}^{\omega+\omega_c} D_0(\omega'')\, e^{i\omega'' n\Delta t}\, d\omega'' \approx d_0(n\Delta t)$$

For example, figure 41 presents the real and imaginary parts of the integral for $N = 64$ and $\omega = \omega_c/2$. The real part of the integral very closely approximates the boxcar function, while the imaginary part is negligible. The case shown is typical. The approximation is better as $N \to \infty$ and $|\omega| \to 0$ and worse as $N \to 0$ and $|\omega| \to \omega_c$. From this approximation, equation (9.16) becomes

$$\boxed{X_T(\omega) \approx \frac{\Delta t}{2\pi} \sum_{n=0}^{N-1} X(n\Delta t)\, e^{-i\omega n\Delta t}} \qquad (9.17)$$

and it can be seen that the finite summation on the right-hand side approximately represents the finite Fourier transform of the signal $X(t)$ over the interval $(-\Delta t/2, T - \Delta t/2)$.

It should be noted that because of the assumed stationarity of the random process $X(t)$, the average properties of the random process in the interval $(-\Delta t/2, T - \Delta t/2)$ are the same as those in any interval of length T. Thus, fundamental relation (9.17) will be used as the definition of the discrete finite Fourier transform in this monograph.

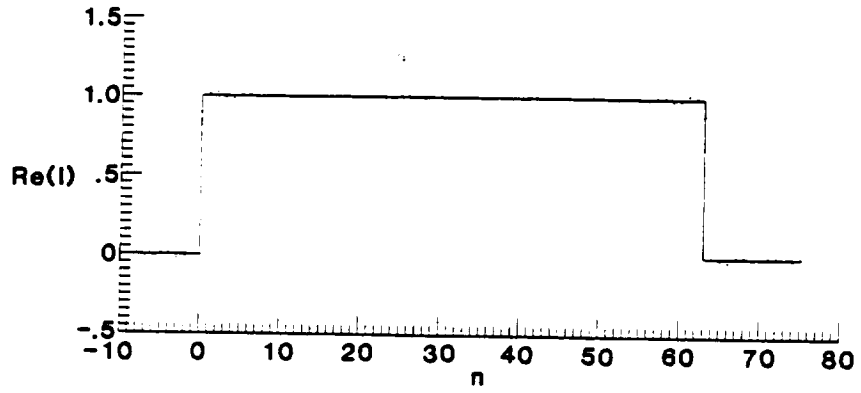
9.6 Digital Blackman-Tukey Estimation

Based on the understanding of digital data analysis developed in the previous sections, it is possible to develop discrete forms for spectral estimation. Suppose that N samples $X(n\Delta t)$ for $n = 0, 1, 2, \ldots, N - 1$ from a stationary random process exist. Then, taking $T = N\Delta t$ and $\tau = j\Delta t$, the autocorrelation estimate analogous to equation (6.7) is

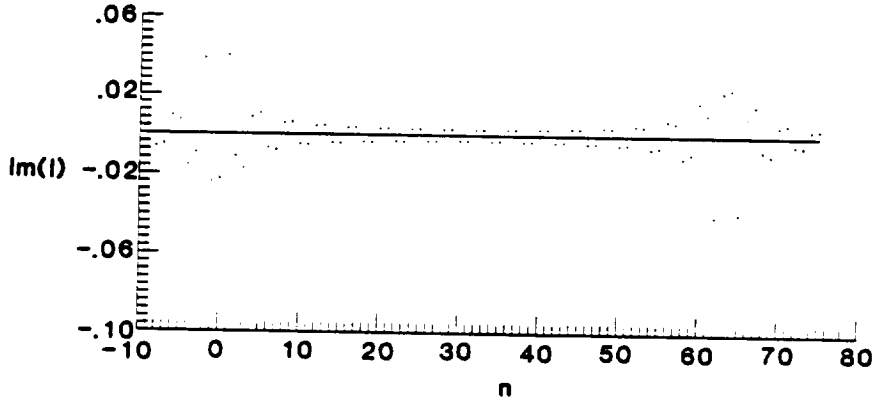
$$\hat{R}_X(j\Delta t) = \frac{1}{N-j} \sum_{n=0}^{N-j-1} X(n\Delta t) X[(n+j)\Delta t] \qquad (9.18)$$

for $j = 0, 1, 2, \ldots, N - 1$. This yields an unbiased estimate of the autocorrelation at these discrete lag values, $j\,\Delta t$.

Likewise, if $T_m = m\,\Delta t$ is the width of the lag window, the power spectral density estimate is defined as the discrete approximation to



(a) Real part of integral.



(b) Imaginary part of integral.

Figure 41. The integral $I(n, N, \omega)$ for $N = 64$ and $\omega = \omega_c/2$.

equation (7.7),

$$\hat{S}_X(\omega) = \frac{\Delta t}{2\pi}\left[u(0)\hat{R}_X(0) + 2\sum_{j=1}^{m} u(j\Delta t)\hat{R}_X(j\Delta t)\cos\omega j\Delta t\right] \quad (9.19)$$

by analysis similar to that which led to equation (9.17).

If data on two jointly stationary random processes exist, similar expressions for the cross correlation and cross spectral density are

$$\hat{R}_{XY}(j\Delta t) = \frac{1}{N-j}\sum_{n=0}^{N-j-1} X(n\Delta t)Y[(n+j)\Delta t] \quad (9.20)$$

$$\hat{R}_{YX}(j\Delta t) = \frac{1}{N-j}\sum_{n=0}^{N-j-1} Y(n\Delta t)X[(n+j)\Delta t] \quad (9.21)$$

for $j = 0, 1, 2, \ldots, N-1$ and

$$\hat{S}_{XY} = \frac{\Delta t}{2\pi}\sum_{j=-m}^{m} u(j\Delta t)\hat{R}_{XY}(j\Delta t)\, e^{-i\omega j\Delta t} \quad (9.22)$$

since the cross correlation is not generally an even function of τ.

9.7 Discrete Finite Fourier Transform Estimation

A similar discrete version of the finite Fourier transform spectral estimate may be developed. If the block size is $T_B = b\Delta t$, then equation (7.16) may be estimated by

$$X_F(\omega) = \frac{\Delta t}{2\pi}\sum_{n=0}^{b-1} d(n\Delta t)X(n\Delta t)\, e^{-i\omega n\Delta t} \quad (9.23)$$

from the discrete data $X(n\Delta t)$ for $n = 0, 1, 2, \ldots, b-1$.

The spectral density estimate is then given by equation (7.15),

$$\hat{S}_X(\omega) = W_S|X_F(\omega)|^2 \quad (9.24)$$

For jointly stationary random processes, the cross spectral density is

$$\hat{S}_{XY}(\omega) = W_S X_F^*(\omega)Y_F(\omega) \quad (9.25)$$

where $\hat{Y}_F(\omega)$ is given by equation (9.23) with $X(n\Delta t)$ replaced by $Y(n\Delta t)$. The window correction factor W_S can usually be determined analytically.

9.8 Frequency Domain Window Insertion

Many times, particularly when one wants to investigate the effects of various windows, it is more efficient to insert the windows in the frequency domain. Expressions by which this may be achieved are quite easily developed. Note first that, although the spectral estimates in equations (9.19), (9.22), (9.24), and (9.25) are given as continuous functions of frequency, the data are sampled, and thus these expressions should not be evaluated at frequencies higher than $\omega_c = \pi/\Delta t$, the Nyquist frequency. Further, since the data have finite length, the resolution criterion (eq. (7.33)) suggests that the estimates should not be evaluated at frequencies that are too close together. The actual resolution depends on the window chosen, of course. However, equation (7.33) provides a reasonable guideline.

Blackman-Tukey estimation. Now, recall from equation (7.2) that the windowed spectral estimate may be written as

$$\hat{S}_X(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} u(\tau)\hat{R}_X(\tau)\, e^{-i\omega\tau}\, d\tau$$

By the convolution theorem, (eq. (2.8)), this estimate becomes

$$\hat{S}_X(\omega) = \int_{-\infty}^{\infty} U(\omega-\omega')\hat{S}_0(\omega')d\omega' \qquad (9.26)$$

where, since $u(\tau) = 0$ for $|\tau| \geq T_m$,

$$\hat{S}_0(\omega) = \frac{1}{2\pi}\int_{-T_m}^{T_m} \hat{R}_X(\tau)\, e^{-i\omega\tau}\, d\tau$$

is the spectral estimate (eq. (7.1)) with the boxcar lag window (i.e., effectively no window at all). Thus, equation (9.26) states that the spectral estimate with an arbitrary lag window is the convolution of the spectral estimate with a boxcar window. This allows one to first estimate the spectral density with a boxcar lag window and then insert various other windows in the frequency domain.

The frequencies at which spectral estimates are evaluated is a matter of choice. The standard choice in Blackman-Tukey estimation is

$$\omega_k = \frac{k\pi}{T_m} = \frac{k\pi}{m\Delta t} \qquad (k = 0, 1, 2, \ldots, m)$$

which yields as many estimates as lag points were used, reaches the Nyquist frequency when $k = m$, and provides a bandwidth of

$$\Delta\omega = \frac{\pi}{T_m} \qquad \text{or} \qquad \Delta f = \frac{1}{2T_m} \tag{9.27}$$

Although this choice does not yield full resolution by equation (7.33), it does have the virtue that it allows simple window insertion. Note that in this case, the discrete analog of equation (9.26) is

$$\hat{S}_X(\omega_k) = \Delta\omega \sum_{j=-m}^{m} U(\omega_k - \omega_j)\hat{S}_0(\omega_j) \tag{9.28}$$

since $\hat{S}_0(\omega_j) = 0$ for $|j| > m$.

Now, from equation (7.10)

$$\begin{aligned} U(\omega_k - \omega_j) &= \frac{T_m}{2\pi} \sum_{n=-\infty}^{\infty} a_n \operatorname{sinc}[(k - j - n)\pi] \\ &= \frac{1}{2\,\Delta\omega} \sum_{n=-\infty}^{\infty} a_n \delta_{k,j+n} \end{aligned}$$

Thus, equation (9.28) becomes

$$\hat{S}_X(\omega_k) = \frac{1}{2} \sum_{n=k-m}^{k+m} a_n \hat{S}_0(\omega_{k-n}) \tag{9.29}$$

which allows window insertion in the frequency domain. Necessary terms for $\omega < 0$ are obtained from $\hat{S}_0(\omega_{-k}) = \hat{S}_0(\omega_k)$. Equation (9.29) shows that windowing is equivalent to applying a low-pass moving-average filter (to be discussed in chapter XI) to the unwindowed spectral estimate. For example, to apply the Hanning window where $a_0 = 1, a_1 = 1/2$, and all other a_n's are zero,

$$\hat{S}_X(\omega_k) = \frac{1}{4}\hat{S}_0(\omega_{k-1}) + \frac{1}{2}\hat{S}_0(\omega_k) + \frac{1}{4}\hat{S}_0(\omega_{k+1})$$

Finite Fourier transform estimation. A similar expression for frequency domain window insertion may be developed for the finite

Fourier transform. Recall equation (7.16):

$$X_F(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d(t)X(t)\,e^{-i\omega t}\,dt$$

Applying the convolution theorem (eq. (2.8)) to this equation results in

$$X_F(\omega) = \int_{-\infty}^{\infty} D(\omega-\omega')X_T(\omega')\,d\omega' \tag{9.30}$$

where $X_T(\omega)$ is the finite Fourier transform (eq. (7.12)) of the data with block length $T_B = b\,\Delta t$,

$$X_T(\omega) = \frac{1}{2\pi}\int_0^{T_B} X(t)\,e^{-i\omega t}\,dt$$

since $X(t) = 0$ for $t < 0$ and $t > T_B$.

Standard practice in the finite Fourier transform estimate is to assume that b is even and to evaluate the estimate at the frequencies

$$\omega_k = \frac{2k\pi}{T_B} = \frac{2k\pi}{b\,\Delta t} \qquad (k = 0, 1, 2, \ldots, b/2) \tag{9.31}$$

This choice yields half as many estimates as data points, reaches the Nyquist frequency when $k = b/2$, and provides a bandwidth of

$$\Delta\omega = \frac{2\pi}{T_B} \qquad or \qquad \Delta f = \frac{1}{T_B} \tag{9.32}$$

Note that, for a similar data length, this is twice the bandwidth of the Blackman-Tukey estimate and provides full resolution by equation (7.33). The frequency choice in equation (9.31) also allows highly efficient computations as will be seen in chapter X.

At these frequencies, the discrete analog of equation (9.30) is

$$X_F(\omega_k) = \Delta\omega \sum_{j=-b/2}^{b/2} D(\omega_k-\omega_j)X_T(\omega_j) \tag{9.33}$$

Now, recalling that the data window is defined only for $0 < t < T_B$, analogous to equation (7.10),

$$D(\omega) = \frac{T_B e^{-i\omega T_B/2}}{4\pi} \sum_{n=-\infty}^{\infty} a_n \operatorname{sinc}\left(\frac{\omega T_B}{2} - n\pi\right)$$

Thus,

$$D(\omega_k - \omega_j) = \frac{T_B(-1)^{k-j}}{4\pi} \sum_{n=-\infty}^{\infty} a_n \operatorname{sinc}[(k-j-n)\pi]$$
$$= \frac{(-1)^{k-j}}{2\,\Delta\omega} \sum_{n=-\infty}^{\infty} a_n \delta_{k,j+n}$$

and equation (9.33) becomes

$$X_F(\omega_k) = \frac{1}{2} \sum_{n=k-b/2}^{k+b/2} (-1)^n a_n X_T(\omega_{k-n}) \tag{9.34}$$

which again allows window insertion in the frequency domain. Necessary terms for $\omega < 0$ are obtained from $X_T(-k) = X_T^*(\omega_k)$. For example, if the Hamming window where $a_0 = 1.08, a_1 = 0.46$, and all other a_n's are 0 is used as the data window,

$$X_F(\omega_k) = -0.23X_T(\omega_{k-1}) + 0.54X_T(\omega_k) - 0.23X_T(\omega_{k+1})$$

Because of the simplicity of equations (9.29) and (9.34), the data are practically always left unaltered in the time domain and any desired window is inserted in the frequency domain.

9.9 Autocorrelation Estimation Via Discrete Fourier Transformation

Estimates from discrete data are not always directly analogous to those from continuous data. For example, the autocorrelation estimate in equation (7.25) is unbiased for continuous data. However, for a discrete set of data $X(n\Delta t)$ for $n = 0, 1, 2, \ldots, b-1$, the discrete Fourier transform (eq. (9.23)), assuming a boxcar data window and evaluation at the frequencies

$$\omega_k = \frac{2\pi k}{b\,\Delta t} \qquad (k = 0, 1, 2, \ldots, b/2)$$

is given by

$$X_F(\omega_k) = \frac{\Delta t}{2\pi} \sum_{n=0}^{b-1} X(n\Delta t)\, e^{-i2\pi kn/b} = X_T(\omega_k) \tag{9.35}$$

Then, the spectral estimate (eq.(7.19)) becomes

$$\hat{S}_X(\omega_k) = \frac{2\pi}{b\,\Delta t}|X_T(\omega_k)|^2 \tag{9.36}$$

which is an even function of ω since $X_T(\omega_{-k}) = X_T^*(\omega_k)$. This can be employed to estimate the autocorrelation, based on equation (7.25), as

$$\hat{R}_X(j\Delta t) = \frac{2\pi W_R\,\Delta\omega}{b\,\Delta t}\sum_{k=-b/2}^{b/2-1}|X_T(\omega_k)|^2 e^{i\omega_k j\Delta t}$$

$$= W_R(\Delta\omega)^2\sum_{k=-b/2}^{b/2-1}|X_T(\omega_k)|^2 e^{i2\pi kj/b}$$

The upper limit of the sum is $b/2-1$ because only one-half of the values at each end of the spectrum is added due to the discontinuity at those points.

Now, note from equation (9.35) that

$$X_T(\omega_{-k}) = X_T(\omega_{b-k})$$

since

$$e^{-i2\pi(b-k)j/b} = e^{-i2\pi(-k)j/b}$$

Thus, the estimate may be written

$$\hat{R}_X(j\Delta t) = W_R(\Delta\omega)^2\sum_{k=0}^{b-1}|X_T(\omega_k)|^2 e^{i2\pi kj/b} \tag{9.37}$$

for $j = 0, 1, 2\ldots, b-1$.

How good is this estimate? Its mean is

$$E\left\{\hat{R}_X(j\Delta t)\right\} = W_R(\Delta\omega)^2\sum_{k=0}^{b-1}E\left\{|X_T(\omega_k)|^2\right\}e^{i2\pi kj/b} \tag{9.38}$$

where, by equation (9.35),

$$E\left\{|X_T(\omega_k)|^2\right\} = \left(\frac{\Delta t}{2\pi}\right)^2\sum_{r=0}^{b-1}\sum_{\ell=0}^{b-1}R_X[(\ell-r)\Delta t]\,e^{-i2\pi(r-\ell)k/b}$$

Thus, equation (9.38) becomes

$$E\left[\hat{R}_X(j\Delta t)\right] = \frac{W_R}{b^2}\sum_{r=0}^{b-1}\sum_{\ell=0}^{b-1} R_X[(\ell - r)\Delta t]\sum_{k=0}^{b-1}\left[e^{i2\pi(j-r+\ell)/b}\right]^k$$

Now, note that for all real or complex values z,

$$\boxed{\sum_{j=0}^{N-1} z^j = \begin{cases} \frac{1-z^N}{1-z} & (z \neq 1) \\ N & (z = 1) \end{cases}} \tag{9.39}$$

This series is of fundamental importance in the analysis of discrete data. Thus,

$$\sum_{k=0}^{b-1}\left[e^{i2\pi(j-r+\ell)/b}\right]^k = \frac{1 - e^{i2\pi(j-r+\ell)}}{1 - e^{i2\pi(j-r+\ell)/b}} = b\left(\delta_{j,r-\ell} + \delta_{j,b+r-\ell}\right) \tag{9.40}$$

The two delta terms arise when $n - r - \ell$ is some integer multiple of b, which occurs only for $j - r + \ell = 0$ and $j - r + \ell = b$ since $-(b-1) < j - r + \ell < 2(b-1)$. Thus,

$$E\left\{\hat{R}_X(j\Delta t)\right\} = \frac{W_R}{b}\left\{(b-j)R_X(j\Delta t) + jR_X[(b-j)\Delta t]\right\}$$

Recall that a boxcar data window results in a Bartlett lag window. Therefore

$$W_R = \frac{1}{u(j\Delta t)} = \left(1 - \frac{j}{b}\right)^{-1}$$

and

$$E\left\{\hat{R}_X(j\Delta t)\right\} = R_X(j\Delta t) + \frac{j}{b-j}R_X[(b-j)\Delta t]$$

Therefore, the estimate is seen to be biased.

9.10 Zero Insertion

The bias in this estimate was caused by the presence of the second term in equation (9.40). This term may be eliminated by the technique of zero insertion. Suppose that the original b data points $X(n\Delta t)$ for $n = 0, 1, 2, \ldots, b-1$ are augmented by b zeros, that is,

$$X(n\Delta t) = 0 \qquad (n = b, b+1, \ldots, 2b-1)$$

The discrete Fourier transform of this entire sequence is then

$$X_T(\omega_k) = \frac{\Delta t}{2\pi} \sum_{n=0}^{2b-1} X(n\Delta t)\, e^{-i2\pi kn/2b}$$
$$= \frac{\Delta t}{2\pi} \sum_{n=0}^{b-1} X(n\Delta t)\, e^{-i\pi kn/b} \qquad (9.41)$$

Note that this has the effect of reducing ω_k by a factor of 2, that is,

$$\omega_k = \frac{\pi k}{b\Delta t} \qquad (k = 0, 1, 2, \ldots, b)$$

with $\Delta\omega = \pi/b\Delta t$. Thus, the spectral estimate (eq. (9.36)) when zeros are inserted has the equivalent bandwidth of a Blackman-Tukey estimate.

Now, analogous to equation (9.37),

$$\boxed{\hat{R}_X(j\Delta t) = 2W_R(\Delta\omega)^2 \sum_{k=0}^{2b-1} |X_T(\omega_k)|^2\, e^{i\pi kj/b}} \qquad (9.42)$$

is the expression used to estimate the autocorrelation. The mean of this estimate is

$$E\left\{\hat{R}_X(j\Delta t)\right\} = 2W_R(\Delta\omega)^2 \sum_{k=0}^{2b-1} E\left\{|X_T(\omega_k)|^2\right\}\, e^{i\pi kj/b}$$

where now

$$E\left\{|X_T(\omega_k)|^2\right\} = \left(\frac{\Delta t}{2\pi}\right)^2 \sum_{r=0}^{b-1}\sum_{\ell=0}^{b-1} R_X[(\ell - r)\Delta t]\, e^{-i\pi(r-\ell)k/b}$$

Thus,

$$E\left[\hat{R}_X(j\Delta t)\right] = \frac{W_R}{2b^2} \sum_{r=0}^{b-1}\sum_{\ell=0}^{b-1} R_X[(\ell - r)\Delta t] \sum_{k=0}^{2b-1} \left[e^{i\pi(j-r+\ell)/b}\right]^\kappa$$

However,

$$\sum_{k=0}^{2b-1} \left[e^{i\pi(j-r+\ell)/b}\right]^k = \frac{1-e^{i2\pi(j-r+\ell)}}{1-e^{i\pi(j-r+\ell)/b}} = 2b\delta_{j,r-\ell}$$

since $j-r+\ell$ never reaches $2b$. Thus,

$$E\left\{\hat{R}_X(j\Delta t)\right\} = \frac{W_R}{b}(b-j)R_X(j\Delta t) = R_X(j\Delta t)$$

using the same b-point window correction factor as before and the estimate in equation (9.42) is unbiased. The same zero insertion technique should also be used for estimation of a cross correlation.

9.11 Digital Spectral Estimation Procedure

Now that all of the required relations have been developed, it is possible to develop a decision-making procedure that one should follow when attempting to estimate power spectral densities with digital data. *This procedure should be thought through before one ever begins to take data.*

Step 1: Are the data stationary? Since all the analysis developed thus far is based on this assumption, these techniques are not valid if the underlying random process is not stationary. Whether or not a particular time history is a sample function of a stationary random process is a decision that must be based primarily on engineering judgment. However, a test of stationarity is discussed in chapter VI. If the data are not stationary, it may be possible to detrend them and make them appear stationary without losing the information of interest. This technique is discussed in chapter XI.

Step 2: What is the maximum frequency of interest $f_{\max}$? This defines the sampling rate since

$$\Delta t \leq \frac{1}{2f_{\max}}$$

Step 3: Does the signal contain power at frequencies higher than $f_{\max}$? If so, the analog data *must* be passed through a low-pass antialiasing filter before it is digitized.

Step 4: What frequency resolution is required? For the finite Fourier transform technique, the required resolution sets the block length T_B, since

$$\Delta f = \frac{1}{T_B} \qquad (T_B = b\,\Delta t)$$

or if zeros are inserted,

$$\Delta f = \frac{1}{2T_B}$$

For the Blackman-Tukey technique, the required resolution sets the width of the lag window T_m, since

$$\Delta f = \frac{1}{2T_m} \qquad (T_m = m\,\Delta t)$$

Step 5: How much accuracy is required? This sets the total length of data to be taken since the degrees of freedom for the finite Fourier transform technique are

$$k = 2N_B$$

where N_B is the number of data blocks. For the Blackman-Tukey technique,

$$k = \frac{2T}{T_m} = \frac{2N}{m}$$

where N is the total number of data points. If it is not possible to obtain this many data points, then one must relax the requirements on either resolution or accuracy or both.

Chapter X

The Fast Fourier Transform

In the late 1960's, the field of time series analysis was completely revolutionized by the introduction of highly efficient techniques for computing the discrete Fourier transform. The application of these *Fast Fourier Transform*, or FFT, techniques on modern digital computers was popularized by Cooley and Tukey[16] in their paper entitled "An Algorithm for the Machine Calculation of Complex Fourier Series." With the use of the FFT, the finite Fourier transform approach to spectral estimation became so computationally efficient that it replaced the Blackman-Tukey approach for most practical applications. Even if one is interested in only the autocorrelation, it is often more efficient to first estimate the power spectral density and then transform to obtain the autocorrelation rather than to use the more direct lagged product approach (eq. (9.18))!

There are actually many FFT algorithms in common use today. In fact, most are tailored to take advantage of the particular architecture of the computer on which they are implemented. However, all are concerned with evaluating the discrete Fourier transform (DFT)

$$Z_k = \sum_{j=0}^{N-1} z_j W^{jk} \qquad (k = 0, 1, 2, \ldots, N-1) \qquad (10.1)$$

where z_j is, in general, a sequence of complex numbers and

$$W = e^{-i2\pi/N}$$

is an Nth root of unity. Note that except for the scale factor $\Delta t/2\pi$, equation (10.1) is precisely the discrete Fourier transform of equation (9.35) if the z_j's are taken to be real. Note also that because of the properties of W, equation (10.1) is periodic with period N and

may thus be considered to be defined for all k; that is, $Z_{k+N} = Z_k$. The inverse of equation (10.1) is

$$z_j = \frac{1}{N} \sum_{k=0}^{N-1} Z_k W^{-jk} \tag{10.2}$$

since substituting equation (10.1) in equation (10.2) results in

$$\frac{1}{N} \sum_{k=0}^{N-1} \sum_{\ell=0}^{N-1} z_\ell W^{-(j-\ell)k} = \frac{1}{N} \sum_{\ell=0}^{N-1} z_\ell \sum_{k=0}^{N-1} \left[e^{-i2\pi(j-\ell)/N} \right]^k$$
$$= \frac{1}{N} \sum_{\ell=0}^{N-1} z_\ell \delta_{j,\ell} = z_j$$

upon use of equation (9.39).

If a single multiply-add operation is taken as a measure of computational work, straightforward calculation of equation (10.1) would take N^2 operations, since for each k, the N data points z_j would have to be multiplied by the appropriate complex exponential and then added to the sum, requiring N operations for each k. However, since $k = 0, 1, 2, \ldots, N-1$, there are N^2 operations for the complete calculation. This discussion assumes, of course, that the complex exponentials W have been previously calculated.

10.1 Theory of the Fast Fourier Transform

The basic idea of the FFT goes back at least to 1903,[17] when Runge noticed that if the number of data points N is not a prime integer. the number of operations can be reduced by splitting the calculation up into parts. Consider the simplest case when $N = AB$. Then the data may be broken up into A subrecords of length B as shown in figure 42, where $a = 0, 1, 2, \ldots, A-1$ is the index of subrecords and $b = 0, 1, 2, \ldots, B-1$ is the index within a subrecord.. Then the time index j in equation (10.1) may be written

$$j = aB + b \tag{10.3}$$

which simply states that b is equal to j modulo B. Likewise, there will be $N = AB$ values of the frequency index k. Suppose the frequency data points Z_k are broken into B subrecords of length A. Then letting

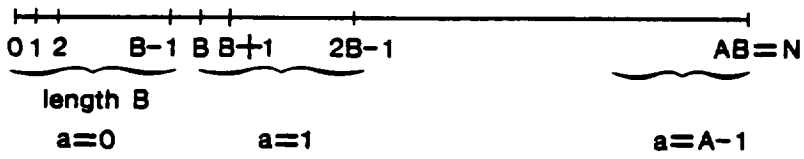


Figure 42. A subrecords of length B.

$c = 0, 1, 2, \ldots, B-1$ be the index of subrecords and $d = 0, 1, 2, \ldots, A-1$ be the index within a subrecord yields

$$k = cA + d \tag{10.4}$$

Thus, equation (10.1) may be written

$$Z_{cA+d} = \sum_{a=0}^{A-1} \sum_{b=0}^{B-1} z_{aB+b} W^{(aB+b)(cA+d)} \tag{10.5}$$

where

$$W^{(aB+b)(cA+d)} = W^{acAB} W^{adB} W^{bcA} W^{bd}$$

However,

$$W^{acAB} = W^{acN} = e^{-i2\pi ac} = 1$$

and thus this term need not enter the computation. Then, equation (10.5) may be written

$$Z_{cA+d} = \underbrace{\sum_{b=0}^{B-1} W^{bcA} \overset{\text{Twiddle factor}}{\overbrace{W^{bd}}} \underbrace{\sum_{a=0}^{A-1} z_{aB+b} W^{adB}}_{\text{1st transform}}}_{\text{2nd transform}} \tag{10.6}$$

This expression has the character of a double Fourier transform. The first transform

$$\sum_{a=0}^{A-1} z_{aB+b} W^{adB}$$

requires A operations and must be done for each b (B of them) and each d (A of them). Thus, A^2B operations are required by the first transform. Each of these transforms (AB of them) must

then be multiplied by the appropriate exponential W^{bd}, called the "twiddle factor" by Gentleman and Sande.[18] This requires another AB operations. Finally, the second transform requiring B operations must be accomplished for each c (B of them) and each d (A of them). Thus, AB^2 operations are required by the second transform. The total number of operations N_{op} is

$$N_{op} = A^2B + AB + AB^2 = AB(A + 1 + B) = N(A + B + 1)$$

compared with N^2 for the direct evaluation. Suppose, for example, that $N = 100 = 10 \times 10$. Then the 10 000 operations required by the direct transform would be reduced to 2100 by splitting the calculation into parts. Further, this reduction is due to only one reduction. If A and/or B is not prime, the technique can be applied again.

The ultimate reduction of course comes when the number of data points is some power of 2, say 2^p. Then the calculation can be reduced to $2^p - 1$ transforms, each of length 2. To see this, note that equation (10.1) may be written

$$\begin{aligned} Z_k &= \sum_{j=0}^{N-1} z_j e^{-i2\pi kj/N} \\ &= \underbrace{\sum_{j=0}^{N/2-1} z_{2j} e^{-i2\pi k(2j)/N}}_{\text{Even terms}} + \underbrace{\sum_{j=0}^{N/2-1} z_{2j+1} e^{-i2\pi k(2j+1)/N}}_{\text{Odd terms}} \\ &= \underbrace{\sum_{j=0}^{N/2-1} z_{2j} e^{-i2\pi kj/(N/2)}}_{\substack{\text{Transform of} \\ N/2 \text{ points}}} + \underbrace{e^{-i2\pi k/N}}_{\substack{\uparrow \\ \text{Twiddle} \\ \text{factor}}} \underbrace{\sum_{j=0}^{N/2-1} z_{2j+1} e^{-i2\pi kj/(N/2)}}_{\substack{\text{Transform of} \\ N/2 \text{ points}}} \end{aligned} \tag{10.7}$$

Thus, by separating the transform into even and odd terms, the calculation has been reduced to two transforms of $N/2$ points. Again note the appearance of the twiddle factor. Now, $N/2 = 2^{p-1}$, so that the procedure may be repeated until the ultimate reduction is achieved. For example, the transform of eight data points would be accomplished as shown in figure 43 where the dots indicate transforms of two data points. This scheme results in a total number of operations equal to $N \log_2 N$ rather than the N^2 for the direct approach. For example, with $N = 128$, the direct approach would require 16 384 operations while the FFT would require only 896.

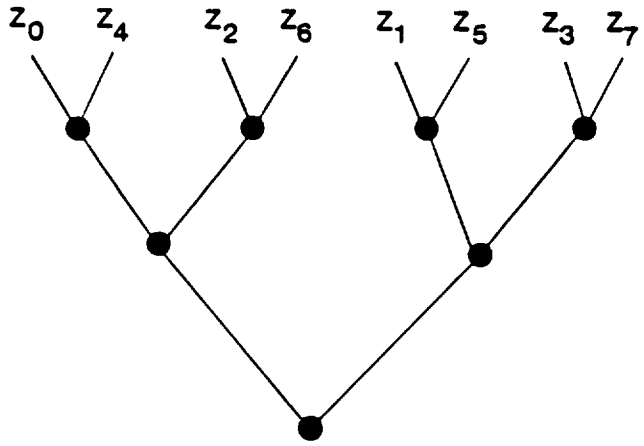


Figure 43. Schematic of eight-point FFT.

The above explanation, found in Nussbaumer,[19] is one of the most easily understandable. For actual coding, binary notation is introduced for the time and frequency indices and the transforms are calculated recursively. To achieve the desired speed, the complex exponentials must be calculated efficiently as well. By using the fact that $W^r = WW^{r-1}$, this calculation can also be done recursively, with evaluation of only W itself required from a series. Great care is also given to the storage of the intermediate results since the data are used in complicated orders as can be seen in figure 43. The k values of the transform also do not remain sequentially ordered. Thus, a massive schuffling of data is required. For this reason, there are many different versions of the FFT, which are very computer specific. Good references to these methods are Otnes and Enochson[12] and Nussbaumer.[19] In spite of the seeming complexity of these techniques, implementation is not overly arduous. For example, Gonzalez and Wintz[20] published a FORTRAN program for an FFT algorithm consisting of only 34 lines of code, including generation of the complex exponentials W!

10.2 Properties of the Discrete Fourier Transform for Real-Valued Data

Although the idea of an FFT can be employed only if the number of data points N is not prime, the discrete Fourier transform (DFT) (eq. (10.1)) always exists. Further, the DFT is sometimes necessary for use in spectral estimation when there are constraints on the number of data points or greater freedom in the selection of frequency resolution is

required. However, it should be emphasized that the DFT is generally computationally less efficient than the Blackman-Tukey technique.

The DFT (of which the FFT is a subset) has some useful properties for real-valued data $z_j = x_j$. In this case, note that equation (10.1) may be written

$$X_k = \sum_{j=0}^{N-1} x_j e^{-i2\pi kj/N} \qquad (k = 0, 1, 2, \ldots, N-1)$$

Now,

$$X_{-k} = \sum_{j=0}^{N-1} x_j e^{i2\pi kj/N} = X_k^*$$

Further,

$$X_{N-k} = \sum_{j=0}^{N-1} x_j e^{-i2\pi(N-k)j/N} = \sum_{j=0}^{N-1} x_j e^{i2\pi kj/N} e^{-i2\pi j}$$
$$= \sum_{j=0}^{N-1} x_j e^{i2\pi kj/N} = X_k^* = X_{-k}$$

since $e^{-i2\pi j} = 1$. Thus, the DFT is periodic with period N since

$$X_{N-k} = X_{-k}$$

Further, only $N/2$ of the points need to be calculated since

$$X_{N-k} = X_k^*$$

(Recall that the Nyquist frequency occurs at $k = N/2$ if N is even.)

For real data, the power of the DFT is not being fully utilized since only half of the transforms are of interest. However, suppose one has two real sequences x_j and y_j to be transformed. Then one can make better use of the DFT by defining complex data

$$z_j = x_j + iy_j$$

Then

$$Z_k = \sum_{j=0}^{N-1} (x_j + iy_j)\, e^{-i2\pi kj/N} = X_k + iY_k$$

and

$$Z^*_{N-k} = \sum_{j=0}^{N-1} (x_j - iy_j)\, e^{-i2\pi kj/N} = X_k - iY_k$$

Thus, the transforms of the real sequences can be recovered by

$$X_k = \sum_{j=0}^{N-1} x_j e^{-i2\pi kj/N} = \frac{Z_k + Z^*_{N-k}}{2}$$

and

$$Y_k = \sum_{j=0}^{N-1} y_j e^{-i2\pi kj/N} = \frac{Z_k - Z^*_{N-k}}{2i}$$

Chapter XI

Digital Filtering

A filter is any physical device or mathematical operation which is applied to a time history in order to change it in some way. Filters are usually broadly classified as *low pass*, passing the low frequencies and attenuating the high frequencies, *high pass*, passing the high frequencies and attenuating the low frequencies, *band pass*, passing frequencies in some band while attenuating both the higher and the lower frequencies, and *band reject*, attenuating the frequencies in some band. Many times, it is of interest to perform such an operation after the time history is digitized. Thus, one is led to the subject of digital filters.

11.1 Linear Filters

The most prevalent filters are linear. Consider the ordinary linear, shift-invariant system shown in figure 44, where $X(t)$ and $Y(t)$ are related by

$$Y(t) = \int_{-\infty}^{\infty} h(\tau) X(t-\tau)\, d\tau \tag{11.1}$$

By definition, $h(t)$ is a filter since the time history $X(t)$ is changed into the time history $Y(t)$. If $X(t)$ is band limited and is known only at discrete times $n\,\Delta t$ for $-\infty < n < \infty$, then by Shannon's sampling theorem (eq. (9.6)) and equation (5.3), equation (11.1) may be written

$$\begin{aligned} Y(t) &= \frac{1}{\omega_c} \sum_{n=-\infty}^{\infty} X(n\Delta t)\, \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega\, H(\omega)\, e^{i\omega(t-n\pi/\omega_c)} \int_{-\infty}^{\infty} \operatorname{sinc} y\, e^{-i\omega y/\omega_c}\, dy \\ &= \frac{\pi}{\omega_c} \sum_{n=-\infty}^{\infty} X(n\Delta t)\, \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} H(\omega)\, e^{i\omega(t-n\pi/\omega_c)}\, d\omega \end{aligned} \tag{11.2}$$

Further, since $X(t)$ is band limited and the system is linear, $Y(t)$ is also band limited because of the convolution theorem (eq. (2.9)). Thus, evaluating equation (11.2) at $t = k\,\Delta t$ results in

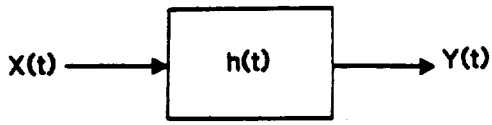


Figure 44. Linear filter.

$$Y(k\Delta t) = \sum_{n=-\infty}^{\infty} h_b[(k-n)\Delta t]X(n\Delta t) = \sum_{j=-\infty}^{\infty} h_b(j\Delta t)X[(k-j)\Delta t] \tag{11.3}$$

where

$$h_b(t) = \frac{\Delta t}{2\pi}\int_{-\omega_c}^{\omega_c} H(\omega)\, e^{i\omega t}\, d\omega$$

Note that $h_b(t)$ is not quite equal to $h(t)\,\Delta t$ since $h(t)$, being a transient function, cannot be band limited. Thus, $h_b(t)/\Delta t$ physically represents a filtered impulse response function.

Based on equation (11.3), a linear digital filter is *defined* to obey the discrete convolution relation

$$Y_k = \sum_{j=-\infty}^{\infty} h_j X_{k-j} \tag{11.4}$$

where $Y_k = Y(k\Delta t)$ and $X_{k-j} = X[(k-j)\Delta t]$. Equation (11.4) represents a simple or nonrecursive digital filter operating on the data $X(n\Delta t)$ and may be implemented by choosing a set of h_j's. Note that for causality, $h_j = 0$ for $j < 0$. However, if the data have been stored on magnetic tape, one need not be constrained by causality. This freedom is a fundamental difference between analog and digital processing which can be exploited to one's advantage at times. A filter of the type represented by equation (11.4), applied in the frequency domain, has already been seen in equation (9.29).

The frequency response of the filter is determined from the frequency response function

$$H(\omega) = \sum_{j=-\infty}^{\infty} h_j e^{-i\omega j\Delta t} \tag{11.5}$$

which can be seen to be a discrete approximation to equation (5.2) utilizing equation (9.15).

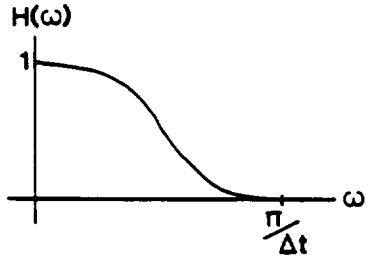


Figure 45. Frequency response of three-point moving average.

One of the most common and useful digital filters is called the *moving average*, which can be used for trend removal. For example, for a three-point moving average, one would take $h_j = 0$ for $|j| > 1$ to yield the noncausal relation

$$Y_k = h_{-1}X_{k+1} + h_0X_k + h_1X_{k-1} \tag{11.6}$$

The frequency response of the filter, from equation (11.5), is

$$\begin{aligned} H(\omega) &= h_{-1}e^{i\omega\Delta t} + h_0 + h_1e^{-i\omega\Delta t} \\ &= h_0 + (h_1 + h_{-1})\cos\omega\Delta t - i(h_1 - h_{-1})\sin\omega\Delta t \end{aligned}$$

For a low-pass filter, the normalization

$$H(0) = h_0 + h_1 + h_{-1} = 1$$

is conventionally applied to ensure that a dc (constant) signal will pass through the filter unchanged. Further, because digital filters operate only in the range $(0, \pi/\Delta t)$ where $\pi/\Delta t$ is the Nyquist frequency, setting

$$H(\pi/\Delta t) = h_0 - (h_1 + h_{-1}) = 0$$

ensures that the frequency response will be low at high frequencies. Then, taking $h_1 = h_{-1}$ to make the frequency response function real (i.e., the phase is zero) and solving these three relations yields

$$h_0 = \frac{1}{2} \qquad h_1 = h_{-1} = \frac{1}{4}$$

Thus,

$$H(\omega) = \frac{1}{2}(1 + \cos\omega\Delta t)$$

which is a simple low-pass filter as shown in figure 45.

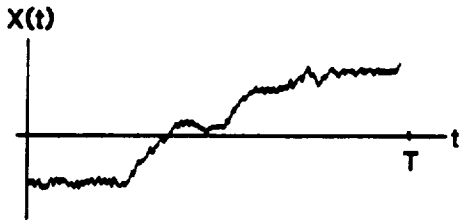


Figure 46. Time history with trend.

Moving-average (low-pass) filters are especially useful for removing trends from data. Suppose, for example, that one had a finite-length digital time history as shown (continuously) in figure 46. The data in this figure are decidely nonstationary, and if this time history were analyzed directly, it would have practically all of its power at very low frequencies. A time history like this might be produced by the price of a stock on a stock exchange (where the values $X(t)$ would all be positive) or the surface temperature of a spacecraft undergoing reentry. for example.

Suppose one were interested in the high-frequency components rather than the long-term trend in the data. If so, the data could be passed through a moving-average filter, that is,

$$Y_k = \frac{1}{4}X_{k+1} + \frac{1}{2}X_k + \frac{1}{4}X_{k-1}$$

and $Y(t)$, again shown continuously, would appear as shown in figure 47. Actually, to achieve this much smoothing, the data might need to be passed through the filter several times in sequence. Such a series of filter operations is equivalent to applying a higher order (i.e., more nonzero h's) filter once. For example, two applications of the operation in equation (11.6) would produce a frequency response function $H^2(\omega) = \frac{1}{4}(1 + \cos\omega\Delta t)^2 = \frac{3}{8} + \frac{1}{2}\cos\omega\Delta t + \frac{1}{8}\cos 2\omega\Delta t$, which might also be produced by one application of a five-point moving average.

After sufficient smoothing has been achieved, the Y_k's could be subtracted from the X_k's to yield

$$Z_k = X_k - Y_k$$

where $Z(t)$ would appear much more stationary, as shown in figure 48. In this way, a much better estimate of the power at the higher frequencies could be obtained. Note that this moving-average operation

Figure 47. Long-term trend in time history.

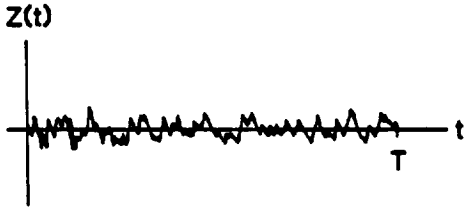


Figure 48. Detrended time history.

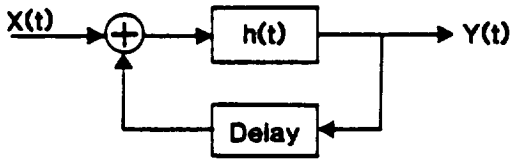


Figure 49. Schematic of recursive filter.

is exactly equivalent to passing the input $X(t)$ through a high-pass moving-average filter with weights $h_0 = 1/2$ and $h_{-1} = h_1 = -1/4$.

11.2 Recursive Filters

A much more computationally efficient filter is one where the output is fed back into the input, as shown in figure 49.

In general, the delay (or *shift* if the independent variable is not time) may be considered a bank of delays where the output is delayed by various delay times before being fed back into the input. Such a filter may, of course, be unstable (i.e., unbounded output) if the feedback is too large, just as a microphone-amplifier-speaker system will begin to screech if the gain is too high. Mathematically, it can be shown by an argument in the complex plane that *such a filter is stable (i.e., bounded*

output) if and only if all the roots of the denominator of its transfer function have positive imaginary parts.

First order recursive filter. Let

$$Y_k = \alpha Y_{k-1} + \beta X_k \tag{11.7}$$

where α and β are real. This is called a first order recursive filter because it uses only a single delay of time Δt. Equation (11.7) is an example of a difference equation and may be solved by defining the generating function

$$Y(z) = \sum_{k=-\infty}^{\infty} Y_k z^k \tag{11.8}$$

where z is complex. If equation (11.7) is multiplied by z^k and summed over all k, the equation

$$Y(z) = \alpha z Y(z) + \beta X(z)$$

results where $X(z)$ is defined similarly to $Y(z)$ in equation (11.8). Thus.

$$Y(z) = \frac{\beta}{1-\alpha z} X(z) = \beta[1 + \alpha z + (\alpha z)^2 + \cdots]X(z) \tag{11.9}$$

assuming that $|\alpha z| < 1$. Thus, equating coefficients yields the solution of equation (11.7),

$$Y_k = \beta \sum_{j=0}^{\infty} \alpha^j X_{k-j} \tag{11.10}$$

which can be seen to be bounded for arbitrary bounded input if and only if $|\alpha| < 1$, since otherwise the magnitude of the coefficients would increase with j.

Consider now the transfer function for this recursive filter. Suppose one takes $z = e^{-i\omega\Delta t}$ which satisfies $|\alpha z| < 1$ for a bounded output filter. Then, equation (11.8) becomes

$$Y(\omega) = \sum_{k=-\infty}^{\infty} Y_k e^{-i\omega k \Delta t}$$

which except for the scale factor is the discrete Fourier transform (eq. (9.15)) of the band limited process $Y(t)$. Further, equation (11.9) becomes

$$Y(\omega) = H(\omega)X(\omega)$$

where

$$H(\omega) = \frac{\beta}{1 - \alpha e^{-i\omega\Delta t}} \tag{11.11}$$

is the transfer function of the first order recursive filter. The root of the denominator of this transfer function is given by $e^{-i\omega\Delta t} = 1/\alpha$ or

$$\omega = \begin{cases} \frac{1}{\Delta t}\ln(\frac{1}{\alpha}) & (\alpha > 0) \\ \frac{-\pi}{\Delta t} + \frac{1}{\Delta t}\ln(-\frac{1}{\alpha}) & (\alpha < 0) \end{cases}$$

and it can be seen that the condition $|\alpha| < 1$ obtained for the boundedness of the equation (11.10) is precisely that required to cause the root of the denominator of the transfer function to have a positive imaginary part as required by the stability criterion. This condition also is intuitive since if it were not true, the output would be amplified before being fed back.

Returning to equation (11.11), the squared magnitude of the transfer function is

$$|H(\omega)|^2 = \frac{\beta^2}{1 + \alpha^2 - 2\alpha\cos\omega\Delta t} \tag{11.12}$$

and it can be seen that

$$|H(0)|^2 = \frac{\beta^2}{(1-\alpha)^2}$$

and

$$|H(\pi/\Delta t)|^2 = \frac{\beta^2}{(1+\alpha)^2}$$

Thus. for $0 < \alpha < 1$, the filter is low pass as shown in figure 50, and for $-1 < \alpha < 0$, the filter is high pass as shown in figure 51. In these figures. the values of β have been chosen to make the gain unity at the respective passband ends.

Second order recursive filter. More freedom in tailoring the shape of the transfer function may be obtained with a second order recursive filter:

$$Y_k = \alpha_1 Y_{k-1} + \alpha_2 Y_{k-2} + \beta X_k \tag{11.13}$$

A similar analysis yields

$$Y(\omega) = \frac{\beta}{1 - \alpha_1 e^{-i\omega\Delta t} - \alpha_2 e^{-i\omega 2\Delta t}} X(\omega)$$

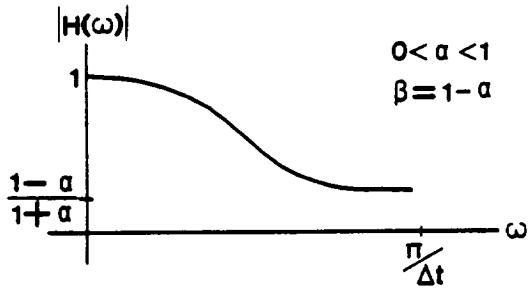


Figure 50. Low-pass first order recursive filter.

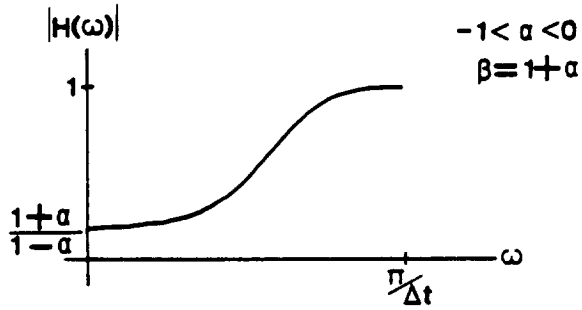


Figure 51. High-pass first order recursive filter.

and the transfer function is seen to be

$$H(\omega) = \frac{\beta}{1 - \alpha_1 e^{-i\omega\Delta t} - \alpha_2 e^{-i\omega 2\Delta t}} \tag{11.14}$$

For stability, the roots of

$$1 - \alpha_1 e^{-i\omega\Delta t} - \alpha_2 e^{-i\omega 2\Delta t} = 0$$

must be examined. For convenience, defining $s = e^{i\omega\Delta t} = 1/z$ and multiplying by s^2 yields

$$s^2 - \alpha_1 s - \alpha_2 = 0$$

with roots

$$s_{1,2} = \frac{\alpha_1 \pm \sqrt{\alpha_1^2 + 4\alpha - 2}}{2} \tag{11.15}$$

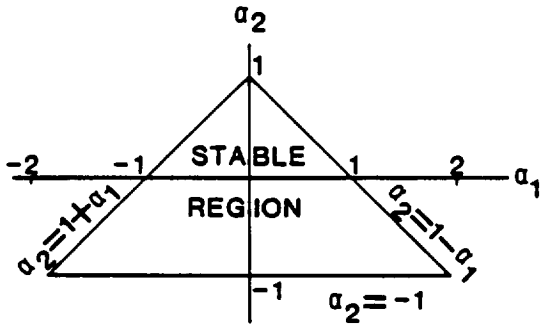


Figure 52. Stability diagram for second order recursive filter.

Therefore, the roots in terms of ω are

$$\omega_{1,2} = \frac{-i}{\Delta t} \ln s_{1,2} = \frac{-i}{\Delta t} (\ln|s_{1,2}| + i\theta_{1,2})$$
$$= \frac{\theta_{1,2}}{\Delta t} - \frac{i}{\Delta t}\ln|s_{1,2}|$$

where $\theta_{1,2}$ are the polar angles of the roots. Thus, an equivalent condition for stability is

$$|s_{1,2}| < 1 \tag{11.16}$$

Now, if $\alpha_1^2 + 4\alpha_2 < 0$ in equation (11.15), then $s_1 = s_2^*$ and

$$|s_{1,2}| = (-\alpha_2)^{1/2}$$

Thus, $(-\alpha_2)^{1/2} < 1$, which implies that $\alpha_2 > -1$. However, if $\alpha_1^2 + 4\alpha_2 > 0$, then the roots are real and unequal. The stability conditions on these roots lead to $\alpha_2 < 1 - \alpha_1$ and $1 + \alpha_1 > \alpha_2$ and therefore to the stability diagram shown in figure 52. In summary, if the parameters α_1 and α_2 are chosen in the triangular region, the filter is stable.

Returning to equation (11.14), the squared magnitude of the transfer function is given by

$$|H(\omega)|^2 = \frac{\beta^2}{(1+\alpha_1^2+\alpha_2^2) - 2\alpha_1(1-\alpha_2)\cos\omega\Delta t - 2\alpha_2\cos 2\omega\Delta t} \tag{11.17}$$

By choosing various values of α_1, α_2, and β, one can tailor this transfer function to have desired characteristics. For example, if $\alpha_2 = 0$, then

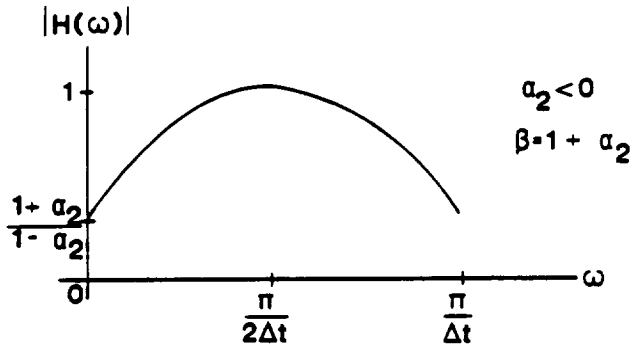


Figure 53. Band-pass recursive filter.

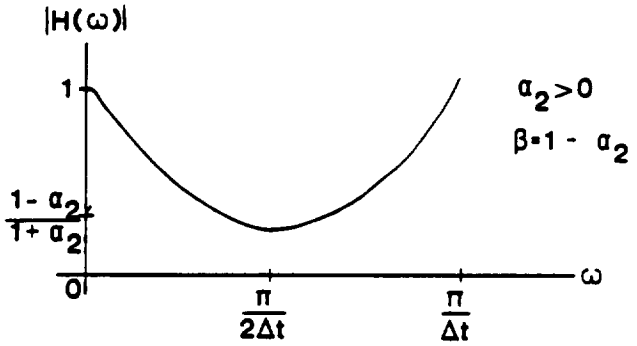


Figure 54. Band-reject recursive filter.

the second order recursive filter reduces to the first order recursive filter studied in the previous section. Thus, $\alpha_1 > 0$ produces a low-pass filter and $\alpha_1 < 0$ produces a high-pass filter. However, if $\alpha_1 = 0$, then

$$|H(0)|^2 = \frac{\beta^2}{(1-\alpha_2)^2} = |H(\pi/\Delta t)|^2$$

while

$$|H(\pi/2\Delta t)|^2 = \frac{\beta^2}{(1+\alpha_2)^2}$$

Thus, for $\alpha_2 < 0$, the filter is band pass as shown in figure 53, while for $\alpha_2 > 0$, the filter is band reject as shown in figure 54. In these figures,

the values of β have again been chosen to make the gain unity in the regions of interest.

By taking both α_1 and α_2 to be nonzero within the stable region of figure 52, various weighted combinations of high-pass, low-pass, band-pass, and band-reject filters may be obtained. More sophisticated filters with even more freedom in their shape may, of course, be developed. A good reference is Otnes and Enockson.[12]

Chapter XII

Special Topics

The field of time series analysis is vast and rapidly changing. In an attempt to provide both complete and current coverage, this chapter presents a potpourri of specialized topics and areas where research is currently in progress.

12.1 The Kendall Series—A Test Case

In the development of computer codes to implement the various estimation techniques discussed in this monograph, it is extremely useful to have a benchmark time history for which the various statistical moments are known theoretically. One particularly simple way of generating such a time history is to use the Kendall series.

The Kendall series Y_n is generated by the recursive relation

$$Y_n = \alpha_1 Y_{n-1} + \alpha_2 Y_{n-2} + X_n \qquad (n = 0, 1, 2, \ldots) \tag{12.1}$$

where X_n is a sequence to be specified later. This series has been studied by Bartlett[11] and can be seen to be a second order recursive filter (eq. (11.13)) operating on the X_n's. Equation (12.1) may be solved by a technique similar to that used in solving equation (11.7) to yield

$$Y_n = \frac{(s_1^{n+2} - s_2^{n+2})}{s_1 - s_2} Y_{-1} - \frac{s_1 s_2 (s_1^{n+1} - s_2^{n+1})}{s_1 - s_2} Y_{-2} + \sum_{k=0}^{n} \frac{(s_1^{k+1} - s_2^{k+1})}{s_1 - s_2} X_{n-k}$$

where Y_{-1} and Y_{-2} are the initial conditions and the s's are the roots

$$s_{1,2} = \frac{\alpha_1 \pm \sqrt{\alpha_1^2 + 4\alpha_2}}{2}$$

which are assumed distinct; that is, $\alpha_1^2 + 4\alpha_2 \neq 0$. The roots $|s_{1,2}|$ are less than unity for stability; thus, as $n \to \infty$, the solution approaches

the steady-state solution

$$Y_n = \sum_{k=0}^{\infty} \frac{(s_1^{k+1} - s_2^{k+1})}{s_1 - s_2} X_{n-k} \tag{12.2}$$

Now, suppose that the terms in the input sequence X_n are independent and identically distributed random variables such that

$$E\{X_n\} = 0$$

and

$$E\{X_n X_{n+m}\} = \sigma_X^2 \delta_{m,0}$$

That is, the input random variables simulate a white noise random process. Then, when steady state has been reached, the mean value of Y_n is

$$E\{Y_n\} = 0$$

and its autocorrelation is

$$R_Y(r) = E\{Y_n Y_{n+r}\} = As_1^r + Bs_2^r \tag{12.3}$$

where

$$A = \frac{\sigma_X^2 s_1}{(s_1 - s_2)(1 - s_1 s_2)(1 - s_1^2)}$$

and

$$B = \frac{-\sigma_X^2 s_2}{(s_1 - s_2)(1 - s_1 s_2)(1 - s_2^2)}$$

Thus, Y_n is a weakly stationary random process. Further,

$$R_{YX}(r) = E\{Y_n X_{n+r}\} = \begin{cases} 0 & (r > 0) \\ \frac{\sigma_X^2 (s_1^{1-r} - s_2^{1-r})}{s_1 - s_2} & (r \leq 0) \end{cases} \tag{12.4}$$

and X_n and Y_n are jointly weakly stationry.

In this case, both the power spectral density

$$\begin{aligned} S_y(\omega) &= \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} R_Y(r) e^{-i\omega r} \\ &= \frac{A}{2\pi} \left(\frac{1}{1 - s_1 e^{-i\omega}} + \frac{1}{1 - s_1 e^{i\omega}} - 1 \right) \\ &\quad + \frac{B}{2\pi} \left(\frac{1}{1 - s_2 e^{-i\omega}} + \frac{1}{1 - s_2 e^{i\omega}} - 1 \right) \end{aligned} \tag{12.5}$$

(taking $\Delta t = 1$) and cross power spectral density

$$S_{YX}(\omega) = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} R_{YX}(r)\, e^{-i\omega r}$$

$$= \frac{\sigma_X^2}{2\pi \left[1 - (s_1 + s_2)e^{i\omega} + s_1 s_2 e^{i2\omega}\right]} \qquad (12.6)$$

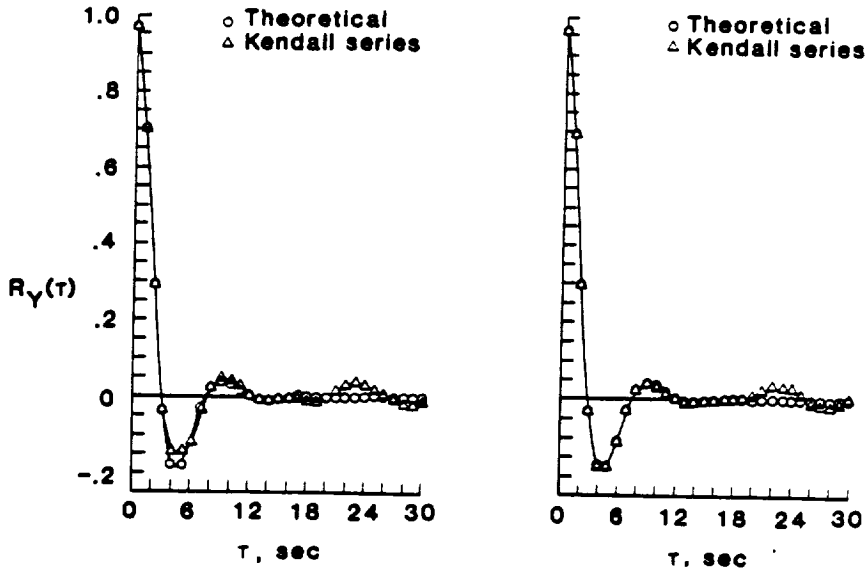
exist for $|\omega| < \pi$. It might be noted that these relations are independent of the distribution function for the random variables X_n.

Most computer systems have standard software that generates white noise data with mean zero, any desired variance σ_X^2, and many different distribution functions. Thus, one may choose values of α_1 and α_2 (such that the filter is stable according to fig. 52), use the initial conditions $Y_{-1} = Y_{-2} = 0$ (thus starting off with the steady-state solution), and recursively compute equation (12.1) to yield a set of digital random data of any desired length. Any program to analyze such data can then be exercised and the results compared with the theoretical expressions (12.3), (12.4), (12.5), and (12.6).

For example, figure 55 shows a comparison of the theoretical expression (12.3) for the autocorrelation with estimates of the autocorrelation obtained from the classical lag product approach (eq. (9.18)).[21] The data used in the estimate shown in figure 55(a) were generated by the series in equation (12.1) with $\alpha_1 = 1.1$, $\alpha_2 = -0.5$, $\sigma_X^2 = 0.333$, and a uniform distribution of the random variables X_n. For figure 55(b), the conditions were the same except that the random variables were normally distributed. Note that the autocorrelation estimates are nearly the same in both cases and agree well with the theoretical expression except near the lag value of 22, that is $\tau = 22\,\Delta t$. This difference is apparently due to spurious correlation induced by the white noise generator. Such spurious correlation is often seen as it is very difficult to generate truly uncorrelated random variables.[22]

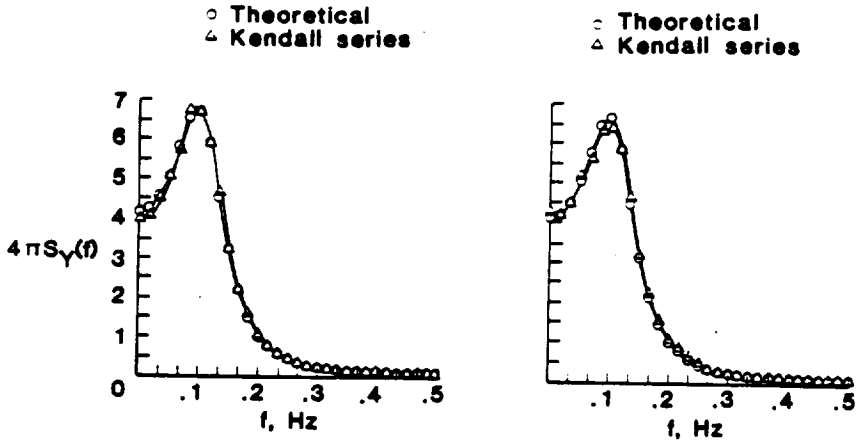
Figure 56 shows a similar comparison of power spectral density estimates with the theoretical expression (12.5). These estimates were obtained by transforming the autocorrelation estimates shown in figure 55 according to the standard Blackman-Tukey technique (eq. (9.19)) with a Hanning lag window. The number of degrees of freedom was 1000 for this estimation technique. Note the good agreement between the theoretical expression and the estimates in both cases.

In figure 57, estimates of the magnitude and phase of the cross power spectral density are compared with the theoretical expression (12.6). The data utilized in this study[23] were generated with the



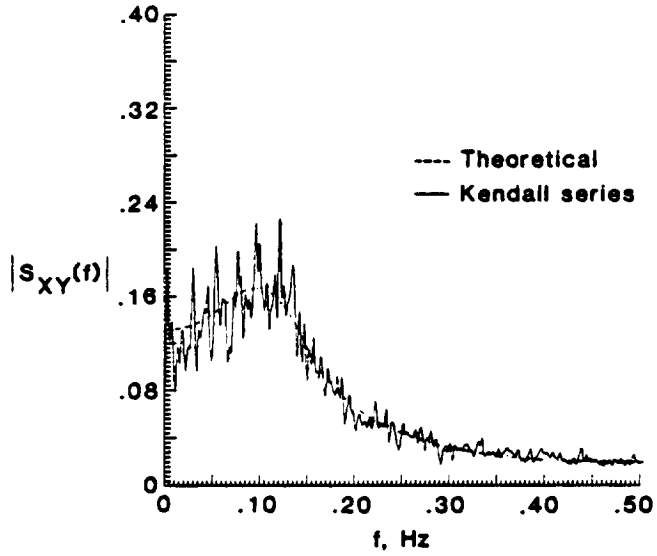
(a) Uniform distribution. (b) Normal distribution.

Figure 55. Comparison of theoretical and estimated autocorrelations.

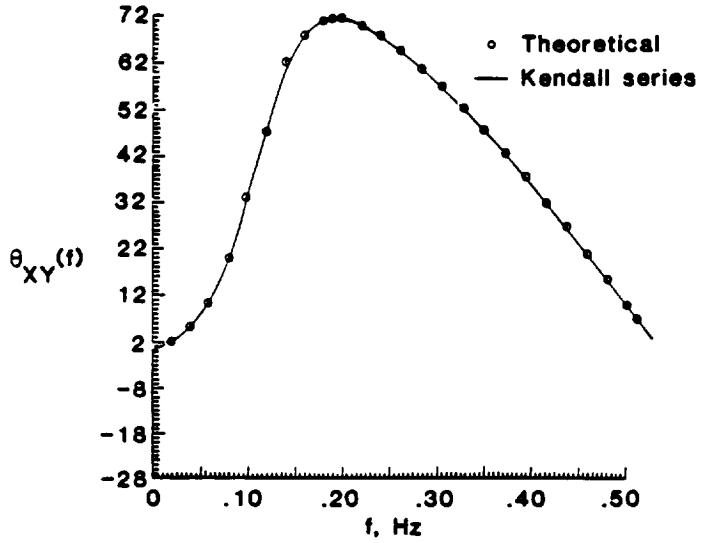


(a) Uniform distribution. (b) Normal distribution.

Figure 56. Comparison of theoretical and estimated power spectral densities.



(a) Cross spectral magnitude.



(b) Cross spectral phase.

Figure 57. Comparison of theoretical and estimated cross spectral densities.

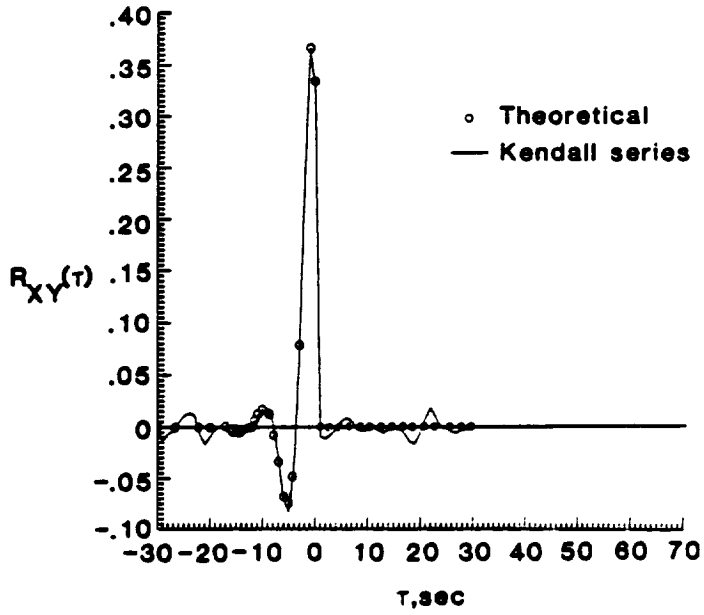


Figure 58. Comparison of theoretical and estimated cross correlations.

same parameters as before and normally distributed random variables. However, the cross spectral density was estimated directly using the finite Fourier transform technique (eq. (9.25)) with a Hanning data window. This estimation technique had only 64 degrees of freedom. Note that the cross spectral magnitude generally follows the trend of the theoretical expression. However, there is much random variability because of the small number of degrees of freedom. The cross spectral phase estimates, however, agree exactly with the theoretical expression.

The cross correlation of these data was estimated by transforming the cross spectral estimate shown in figure 57 using the finite Fourier transform technique (eq. (9.42)) adapted for a cross correlation. This estimate is compared in figure 58 with the theoretical expression (12.4). Note that reasonably good agreement is achieved although there is again a deviation near a lag of 22.

12.2 AR, MA, and ARMA Models

Results like that of the previous section, in which the passage of a white noise signal through a second order recursive filter leads to an output random process whose statistical characteristics can be determined analytically, have stimulated much interest in attempts to model arbitrary random processes by the passage of white noise through

various types of filters. Such modeling is the topic of considerable current research.

The Kendall series considered in the previous section is a simple example of an autoregressive (AR) model of a random process, that is,

$$Y_n = a_1 Y_{n-1} + a_2 Y_{n-2} + \cdots + a_p Y_{n-p} + X_n \tag{12.7}$$

where X_n is a white noise process. Equation (12.7) is called an autoregressive model of order p. Similarly, a moving average (MA) model is defined by

$$Y_n = b_0 X_n + b_1 X_{n-1} + b_2 X_{n-2} + \cdots + b_q X_{n-q}$$

where q is the order of the model. These models are special cases of the more general relation

$$\begin{aligned} Y_n =& a_1 Y_{n-1} + a_2 Y_{n-2} + \cdots + a_p Y_{n-p} \\ &+ b_0 X_n + b_1 X_{n-1} + \cdots + b_q X_{n-q} \end{aligned} \tag{12.8}$$

which is called an autoregressive–moving-average (ARMA) model of order (p,q).

When a random process $Y(t)$ may be represented by such a model, it admits solution just as did the Kendall series and thus its moments are well understood. For example, multiplying equation (12.8) by $e^{i\omega n \Delta t}$ and summing yields

$$Y(\omega) = \frac{b_0 + b_1 z + b_2 z^2 + \cdots + b_q z^q}{1 - a_1 z - a_2 z^2 - \cdots - a_p z^p} X(\omega) = H(\omega) X(\omega)$$

where $z = e^{-i\omega\Delta t}$ and $Y(\omega)$ and $X(\omega)$ are defined by equation (11.8). Thus, $X(t)$ is the output of a linear filter with frequency response function $H(\omega)$.

Assume that $H(\omega)$ is a stable filter and note that the input $X(t)$ is a weakly stationary random process with

$$E\{X_n\} = 0$$

and

$$E\{X_n X_{n+m}\} = \sigma_X^2 \delta_{m,0}$$

Then, as $n \to \infty$, $Y(t)$ is a weakly stationary random process with

$$E\{Y(t)\} = 0$$

and

$$S_Y(\omega) = |H(\omega)|^2 S_X(\omega) = \sigma_X^2 |H(\omega)|^2$$

Thus, such models can represent any mean zero random process whose power spectral density can be expressed as

$$S_Y(\omega) = \sigma_X^2 \left| \frac{b_0 + b_1 z + b_2 z^2 + \cdots + b_q z^q}{1 - a_1 z - a_2 z^2 - \cdots - a_p z^p} \right|^2$$

in which $p+q+2$ arbitrary constants are available to match the power spectral density.

This result has led to much interest in representing real data sequences in terms of such models. Many different techniques have been developed for choosing an appropriate set of a's and b's. For example, for an AR model, multiplying equation (12.7) by $Y_{n-\tau}$ for some integer τ greater than or equal to 0, taking expectation, and using the fact that $R_Y(-\tau) = R_Y(\tau)$ yields

$$R_Y(\tau) = a_1 R_Y(\tau-1) + a_2 R_Y(\tau-2) + \cdots + a_p R_Y(\tau-p) + \sigma_X^2 \delta_{\tau,0} \quad (12.9)$$

since $E\{X_n Y_n\} = E\{X_n^2\} = \sigma_X^2$. This recursive equation may be used to extrapolate the autocorrelation from known values for an AR process. However, for the purpose of estimating the coefficients, note that if equation (12.9) is evaluated for $\tau = 1, 2, \ldots, p$, the matrix equation

$$\begin{bmatrix} R_Y(0) & R_Y(1) & \cdots & R_Y(p-1) \\ R_Y(1) & R_Y(0) & \cdots & R_Y(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ R_Y(p-1) & R_Y(p-2) & \cdots & R_Y(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} R_Y(1) \\ R_Y(2) \\ \vdots \\ R_Y(p) \end{bmatrix}$$

results. Thus, if estimates $\hat{R}_Y(j\Delta t)$ for $j = 0, 1, 2, \ldots, p$ of the autocorrelation of a random process are obtained from equation (9.18), time may be considered to be nondimensionalized by Δt and appropriate coefficients for representing that process by an AR model can be obtained by solving the matrix equation.

12.3 Data Adaptive Spectral Estimation Techniques

Spectral estimation techniques such as the Blackman-Tukey or finite Fourier transform are said to be nonadaptive in the sense that the characteristics of these techniques are the same for all sets of data. That is, the algorithms for their implementation treat all

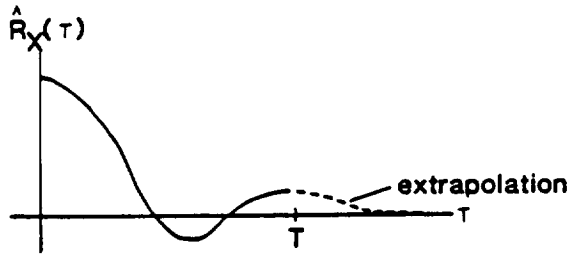


Figure 59. Extrapolation of autocorrelation.

sets of data in exactly the same way. Recently, two new spectral estimation techniques, the maximum entropy method (MEM) and maximum likelihood method (MLM) have been developed. These methods are said to be adaptive since their design is data dependent. The characteristics of these techniques adapt to the particular data being analyzed. *Such techniques allow much higher resolution than the nonadaptive techniques* and are particularly useful when resolution is limited by short data lengths.

Maximum entropy method (MEM). This technique was introduced by Burg.[24] The *basic idea is to choose as the spectral estimate the power spectral density of the most random (i.e., maximum entropy) time series whose autocorrelation agrees with the known values.* It can be shown that this amounts to extrapolating the autocorrelation to larger lag values than can be estimated from data of length T as shown in figure 59. The extrapolation is done in such a way that the entropy is maximized; that is, as little information as possible is added. Since no lag window exists in this case, the *resolution is theoretically unbounded and the estimate is unbiased.*

Suppose that discrete equally spaced data $X(n\Delta t)$ for $n = 1, 2, \ldots, N$ from a stationary random process $X(t)$ exists. Shannon[25] has defined the entropy of the random process as

$$H_N = -\int_{-\infty}^{\infty} f_X(\mathbf{x}) \ln \left[c^{2N} f_X(\mathbf{x})\right] d\mathbf{x} \qquad (12.10)$$

where $f_X(\mathbf{x}) = f_X(x_1, x_2, \ldots, x_N; \Delta t, 2\Delta t, \ldots, N\Delta t)$ is the Nth order density function of the random process $X(t)$ and c is an arbitrary constant.

Now, suppose that $X(t)$ is a normal random process with mean zero (see chapter III). Then

$$f_X(\mathbf{x}) = \frac{1}{(2\pi)^{N/2}|\Lambda|^{1/2}} \exp\left(-\frac{1}{2|\Lambda|}\sum_{m=1}^{N}\sum_{n=1}^{N}|\Lambda|_{mn}X_mX_n\right)$$

where $X_m = X(m\Delta t)$ and $|\Lambda|$ is the determinant and $|\Lambda|_{mn}$ is the mnth cofactor of the matrix of correlations,

$$\Lambda = \begin{bmatrix} R_X(0) & R_X(\Delta t) & \cdots & R_X[(N-1)\Delta t] \\ R_X(\Delta t) & R_X(0) & \cdots & R_X[(N-2)\Delta t] \\ \vdots & \vdots & \ddots & \vdots \\ R_X[(N-1)\Delta t] & R_X[(N-2)\Delta t] & \cdots & R_X(0) \end{bmatrix}$$

In this case, for proper choice of the constant c,

$$H_N = \frac{1}{2}\ln|\Lambda|$$

As $N \to \infty$, $H_N \to \infty$. However, it is possible to define the entropy *rate* to be

$$h = \lim_{N\to\infty}\frac{H_N}{N} = \lim_{N\to\infty}\frac{1}{2}\ln|\Lambda|^{1/N}$$

Further, since the autocorrelation depends on the power spectral density, it can be shown that

$$\lim_{N\to\infty}|\Lambda|^{1/N} = \frac{\omega_c}{\pi}\exp\left\{\frac{1}{2\omega_c}\int_{-\omega_c}^{\omega_c}\ln[2\pi S_X(\omega)]\,d\omega\right\}$$

where $\omega_c = \pi/\Delta t$ is the Nyquist frequency and it has been assumed that $S_X(\omega) = 0$ for $|\omega| > \omega_c$. Thus,

$$h = \frac{1}{2}\ln\left(\frac{\omega_c}{\pi}\right) + \frac{1}{4\omega_c}\int_{-\omega_c}^{\omega_c}\ln[2\pi S_X(\omega)]\,d\omega$$

Recall that

$$S_X(\omega) = \frac{\Delta t}{2\pi}\sum_{m=-\infty}^{\infty}R_X(m\Delta t)\,e^{-i\omega m\Delta t}$$

and thus

$$h = \frac{1}{4\omega_c} \int_{-\omega_c}^{\omega_c} \ln \left[\sum_{m=-\infty}^{\infty} R_X(m\Delta t)\, e^{-i\omega m\Delta t} \right] d\omega \qquad (12.11)$$

From the data, estimates of the autocorrelation $\hat{R}_X(m\Delta t)$ for $m = 0, 1, 2, \ldots, N-1$ can be obtained. The desire is to extrapolate these values to obtain other estimates $\hat{R}_X(m\Delta t)$ for $m \geq N$ in such a way that the entropy rate (eq. (12.11)) is maximized. Thus

$$\frac{\partial h}{\partial \hat{R}_X(m\Delta t)} = 0 \qquad (m \geq N)$$

or

$$\int_{-\omega_c}^{\omega_c} \frac{e^{-i\omega n\Delta t}}{\sum_{m=-\infty}^{\infty} \hat{R}_X(m\Delta t)\, e^{-i\omega m\Delta t}} d\omega = 0 \qquad (12.12)$$

Thus, if the new unbiased estimate of the spectral density is defined as

$$\hat{S}_X(\omega) = \frac{\Delta t}{2\pi} \sum_{m=-\infty}^{\infty} \hat{R}_X(m\Delta t)\, e^{-i\omega m\Delta t}$$

in terms of the estimated and extrapolated autocorrelation values, then equation (12.12) may be solved to yield

$$\hat{S}_X(\omega) = \frac{\sum_{m=0}^{N-1} a_m \hat{R}_X(m\Delta t)}{2\omega_c \left| 1 + \sum_{m=1}^{N} a_m e^{-i\omega m\Delta t} \right|^2} \qquad (12.13)$$

where

$$a_m = \begin{cases} 1 & (m = 0) \\ -h_m & (m = 1, 2, \ldots, N) \end{cases}$$

and the h's are given by the solution of the matrix equation

$$\Lambda \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_N \end{bmatrix} = \begin{bmatrix} \hat{R}_X(0) \\ \hat{R}_X(1) \\ \vdots \\ \hat{R}_X[(N-1)\Delta t] \end{bmatrix}$$

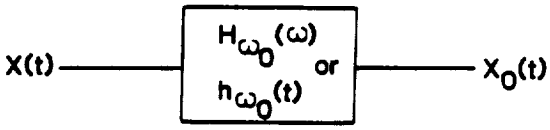


Figure 60. MLM filter.

Maximum likelihood method (MLM). The maximum likelihood method was introduced by Capon.[26] *The basic idea is to develop a minimum variance unbiased estimator of the spectral components of the time history by designing a filter at each particular frequency that passes that frequency undistorted and rejects all other frequencies in an optimal manner.*

Consider a record of length T of a stationary random process $X(t)$. It is desired to estimate the power spectral density of this random process at a particular frequency ω_0. This can be done by designing an optimal causal filter for the particular frequency as shown in figure 60. Note that the filter impulse response need not be restricted to be real. In this case, the output $X_0(t)$ can also be complex. The spectral estimate is then the power output by this filter:

$$\hat{S}_X(\omega_0) = |X_0(t)|^2$$

where if the impulse response has duration T_r,

$$X_0(t) = \int_0^{T_r} h_{\omega_0}(\alpha) X(t-\alpha) d\alpha \tag{12.14}$$

is the time history passed by the filter.

In order for the filter to be optimal, it should have unit gain at the frequency of interest, that is,

$$H_{\omega_0}(\omega_0) = \int_0^{T_r} h_{\omega_0}(t)\, e^{-i\omega_0 t}\, dt = 1$$

Further, in order to reject other frequencies in an optimal manner, the filter should minimize the output power when the input process has an autocorrelation that agrees with the known data over the range $(-T, T)$. By equation (12.14), the average output power of the filter is

$$E\left\{|X_0(t)|^2\right\} = \int_0^{T_r} d\alpha_1 \int_0^{T_r} d\alpha_2 h_{\omega_0}(\alpha_1) h^*_{\omega_0}(\alpha_2) R_X(\alpha_2 - \alpha_1)$$

Minimizing this expression subject to the constraint that $H_{\omega_0}(\omega_0) = 1$ yields

$$h_{\omega_0}(\tau_1)h^*_{\omega_0}(\tau_2) = E\left\{|X_0(t)|^2\right\} Q(\tau_1, \tau_2) \tag{12.15}$$

where

$$\int_0^{T_r} Q(t,\tau)R_X(\tau - t')\,d\tau = \delta(t - t')$$

Now, multiply equation (12.15) by $e^{-i\omega_0(\tau_1-\tau_2)}$ and integrate over τ_1 and τ_2 to obtain

$$E\left\{|X_0(t)|^2\right\} \int_0^{T_r} d\tau_1 \int_0^{T_r} d\tau_2\, Q(\tau_1, \tau_2) e^{-i\omega_0(\tau_1-\tau_2)}$$
$$= \int_0^{T_r} d\tau_1\; h_{\omega_0}(\tau_1)\; e^{-i\omega_0\tau_1} \int_0^{T_r} d\tau_2\; h^*_{\omega_0}(\tau_2)\; e^{i\omega_0\tau_2} = 1$$

Thus, since the spectral estimate is taken to be the output power,

$$\hat{S}_X(\omega_0) = \left[\int_0^{T_r} d\tau_1 \int_0^{T_r} d\tau_2\, Q(\tau_1, \tau_2)\; e^{-i\omega_0(\tau_1-\tau_2)}\right]^{-1} \tag{12.16}$$

where $Q(t,\tau)$ may be estimated from the autocorrelation estimate by

$$\int_0^{T_r} Q(t,\tau)\hat{R}_X(\tau - t')\,d\tau = \delta(t - t')$$

for $t > 0$ and $t' < T$. For discrete data, this can again be written as a matrix equation.

Although these techniques provide higher resolution, they do so at the cost of increased computational effort. Basically, an additional matrix equation must be solved. Other techniques could, of course, be developed. Recall that

$$R_X(\tau) = \int_{-\infty}^{\infty} S_X(\omega)\, e^{i\omega\tau}\, d\omega$$

The data $\hat{R}_X(n\Delta t)$ for $n = 0, 1, 2, \ldots, N-1$ provide a set of N constraints

$$\hat{R}_X(n\Delta t) = \int_{-\infty}^{\infty} S_X(\omega)\, e^{i\omega n\Delta t}\, d\omega$$

on the possible form of the power spectral density. Many methods for estimating $S_X(\omega)$ within these constraints might be devised.

12.4 Spectral Analysis of Randomly Sampled Signals

All the digital analysis that has been covered so far in this monograph has assumed that the time history is sampled at *equal* intervals Δt. However, in recent years, certain applications have stimulated interest in data sampled at random intervals. For example, the laser Doppler anemometry technique, which is used to measure velocity components in flows, amounts to setting up a control volume within the flow by crossing laser beams from different angles. The flow is then seeded with some particulates and a velocity measurement is obtained whenever a particle happens to pass through the control volume. *Spectral estimates obtained from such data are comparatively free from aliasing but have higher variability than corresponding estimates from equally spaced data.* This is because the unequally spaced sample times eliminate the ambiguity associated with equally spaced samples that leads to aliasing, while the uncertainty of the random sample times leads to further uncertainty in the estimate.

In both methods to be developed, the sampling times are assumed to be Poisson distributed[5] such that

$$P\{\text{Sample in time interval } (t, t+\Delta t)\} = \lambda\,\Delta t + o(\Delta t)$$

where λ is the average rate at which samples occur and $o(\Delta t)$ indicates a term that approaches zero faster than Δt does as $\Delta t \to 0$. In the absence of any better information, this model is preferred in the sense that it fits many real world phenomena where events occur randomly in time.

Method 1. This method was developed by Gaster and Roberts[27] in 1975. Consider a stationary random process $X(t)$ that is sampled at Poisson distributed random times t_i for $i = 1, 2, \ldots$ and define

$$C(n) = E\{X(t_i)X(t_{i+n})\}$$

Note that this is similar to an autocorrelation, being the expected value of the product of a sample with another sample n samples later. However, here the expectation is not only over the ensemble comprising the random process $X(t)$ but also over the random sampling times t_i.

From the concept of conditional probability,[5]

$$C(n) = \int_0^\infty E\{X(t_i)X(t_{i+n})|t_{i+n} - t_i = \tau\} P\{t_{i+n} - t_i = \tau\} d\tau$$
$$= \int_0^\infty R_X(\tau)p_n(\tau) d\tau \qquad (12.17)$$

where $E\{|\}$ is the conditional expectation and $p_n(\tau)$ is the probability density function for the time interval between the ith and $(i+n)$th samples

$$p_n(\tau) = \frac{\lambda^n \tau^{n-1} e^{-\lambda\tau}}{(n-1)!} \qquad (\tau \geq 0; n \geq 1) \qquad (12.18)$$

which is called the gamma density function.[5]

Recall that

$$R_X(\tau) = \int_{-\infty}^\infty S_X(\omega) e^{i\omega\tau} d\omega$$

Then equation (12.17) may be written

$$C(n) = \int_{-\infty}^\infty S_X(\omega)\phi_n(\omega) d\omega \qquad (12.19)$$

where

$$\phi_n(\omega) = \int_0^\infty p_n(\tau) e^{i\omega\tau} d\tau = \left(\frac{\lambda}{\lambda - i\omega}\right)^n$$

Thus, equation (12.19) becomes the integral equation

$$C(n) = \int_{-\infty}^\infty S_X(\omega) \left(\frac{\lambda}{\lambda - i\omega}\right)^n d\omega \qquad (12.20)$$

for the power spectral density $S_X(\omega)$.

Integral equation (12.20) has been solved for $S_X(\omega)$ by Shapiro and Silverman.[28] The unique solution is

$$S_X(\omega) = \frac{1}{\pi} \sum_{n=1}^\infty b(n)\psi_n(\omega) \qquad (12.21)$$

where

$$\psi_n(\omega) = \mathrm{Re}\left[-\sqrt{2\lambda}\frac{(i\omega+\lambda)^{n-1}}{(i\omega-\lambda)^n}\right]$$

$$b(n) = \sqrt{\frac{2}{\lambda}}\sum_{k=0}^{n-1}(-2)^k\binom{n-1}{k}C(k+1)$$

and

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

is the binomial coefficient.

In practice, with N data points $X(t_i)$ for $i = 1, 2, \ldots, N$ from a single sample function, the function $C(n)$ is estimated by

$$\hat{C}(n) = \frac{1}{N-n}\sum_{i=1}^{N-n} X(t_i)X(t_{i+n}) \tag{12.22}$$

which can be seen to be an unbiased estimate. This estimate is then used in equation (12.21) to yield the spectral estimate $\hat{S}_X(\omega)$.

Method 2. In a later analysis (1977), Gaster and Roberts[29] estimated the spectral density more directly. Recall the discrete Fourier transform spectral estimate (eq. (7.15)):

$$\hat{S}_X(\omega) = W_S|X_F(\omega)|^2 \tag{12.23}$$

where

$$\begin{aligned} X_F(\omega) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} d(t)X(t)\,e^{-i\omega t}\,dt \\ &= \frac{1}{2\pi}\int_0^T d(t)X(t)\,e^{-i\omega t}\,dt \end{aligned} \tag{12.24}$$

Now, if N randomly sampled data points $X(t_i)$ for $i = 1, 2, \ldots, N$ occur in the interval $(0, T)$, an approximation to integral equation (12.24) may be obtained from

$$X_F(\omega) \approx \frac{1}{2\pi}\sum_{i=1}^{N} d(t_i)X(t_i)\,e^{-i\omega t_i}\,\Delta t_i \tag{12.25}$$

where $\Delta t_i = t_i - t_{i-1}$ and $t_0 = 0$. Further, if samples occur at the average sampling rate λ, then

$$\overline{\Delta t_i} = \frac{1}{\lambda}$$

and equation (12.25) may be further approximated by

$$X_F(\omega) \approx \frac{1}{2\pi\lambda} \sum_{i=1}^{N} d(t_i) X(t_i)\, e^{-i\omega t_i} \tag{12.26}$$

This expression could be used in equation (12.23) to provide the spectral estimate. However, upon doing so, Gaster and Roberts found this estimate to be biased by a false constant shift, which they removed by the use of the estimate

$$\hat{S}_X(\omega) = \frac{W_S}{4\pi^2\lambda^2} \left[\left| \sum_{i=1}^{N} d(t_i) X(t_i)\, e^{-i\omega t_i} \right|^2 - \sum_{i=1}^{N} d^2(t_i) X^2(t_i) \right] \tag{12.27}$$

Although this technique is computationally more efficient than method 1, more data are necessary to achieve the same level of accuracy.

12.5 Cepstrum Analysis

In the past few years, the use of cepstrum analysis[30] has come into prominence. The name is derived by inverting the first four letters in spectrum. This type of analysis is particularly useful for time histories involving a signal that is delayed and then added to itself, such as an echo, or for noise transmission by different paths, such as airborne and structure-borne sound. If $X(t)$ is a stationary random process, then its *power cepstrum* is defined by

$$C_p(\tau) = \int_{-\infty}^{\infty} \ln[S_X(\omega)]\, e^{i\omega\tau}\, d\omega \tag{12.28}$$

which is the Fourier integral transform of the natural logarithm of the power spectral density. Since the power spectral density is real and an even function of frequency, the power cepstrum can be seen to be real and an even function of the variable τ. The reason for the use of the logarithm here is that any product term in the power spectral density appears as a summation in the cepstrum.

The definition in equation (12.28) assumes that the power spectral density is never zero. When working with digital data, equation (12.28)

is approximated by the integral from $-\omega_c$ to ω_c, where ω_c is the Nyquist frequency. For real world random data, the occurrence of an identically zero value in this range is unlikely.

Although the idea of inverting the first four letters in spectrum may have been a good one, this idea has been carried too far in cepstrum analysis. For example, the variable τ in equation (12.28) is called "quefrency," paraphrasing frequency, which is unfortunate since τ is a timelike variable. Clearly, if it were not for the logarithm, equation (12.28) would be just the autocorrelation and τ would be the lag time. Other examples of this type of paraphrastic excess will be noted.

It is also possible to define the *complex cepstrum* to be

$$C_c(\tau) = \int_{-\infty}^{\infty} \ln[X(\omega)]\, e^{i\omega\tau}\, d\omega \tag{12.29}$$

where $X(\omega)$ is the Fourier integral transform of $X(t)$:

$$X(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(t)\, e^{-i\omega t}\, dt = |X(\omega)| e^{i\phi(\omega)} \tag{12.30}$$

Note that, for $X(t)$ real,

$$X(-\omega) = X^*(\omega)$$

and thus

$$|X(\omega)| = |X(-\omega)|$$

and

$$\phi(\omega) = -\phi(-\omega)$$

Thus, from equation (12.30)

$$\begin{aligned} C_c(\tau) &= \int_{-\infty}^{\infty} [\ln|X(\omega)| + i\phi(\omega)](\cos\omega\tau + i\sin\omega\tau)\, d\omega \\ &= \int_{-\infty}^{\infty} [\ln|X(\omega)|\cos\omega\tau - \phi(\omega)\sin\omega\tau]\, d\omega \\ &\quad + i\int_{-\infty}^{\infty} [\ln|X(\omega)|\sin\omega\tau + \phi(\omega)\cos\omega\tau]\, d\omega \end{aligned}$$

The second integral vanishes since it is an odd function. Thus, the complex cepstrum is real! However, it is not an even function of τ.

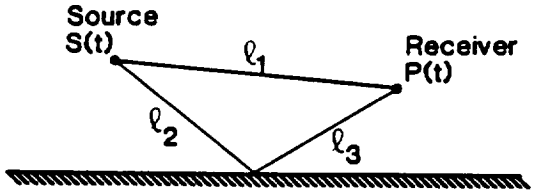


Figure 61. Sound source above flat surface.

A third type of cepstrum, the *real cepstrum*,

$$C_r(\tau) = \int_{-\infty}^{\infty} \ln |X(\omega)|^2 e^{i\omega\tau}\, d\omega \tag{12.31}$$

is both real and an even function of τ and may prove to be even more convenient for transient functions.

As an example of the use of this type of analysis, consider the geometry shown in figure 61. An acoustic source is producing a stationary acoustic signal $S(t)$, which is being received by a microphone. Sound can reach the microphone by the direct path of length ℓ_1 or by echoing off the surface, resulting in a path length of $\ell_2+\ell_3$. The pressure signal recorded by the microphone is thus

$$P(t) = \frac{S(t-\ell_1/c)}{\ell_1} + \alpha\frac{S[t-(\ell_2+\ell_3)/c]}{\ell_2+\ell_3} \tag{12.32}$$

where c is the speed of sound and α is the fraction of the incoming energy that is reflected by the surface. If the reflecting surface was not present, the microphone would record the signal

$$Q(t) = \frac{S(t-\ell_1/c)}{\ell_1} \tag{12.33}$$

which is equivalent to choosing $\alpha = 0$ in equation (12.32).

From equation (12.32), the autocorrelation of the microphone signal is

$$\begin{aligned} R_P(\tau) &= E\{P(t)P(t+\tau)\} \\ &= \beta R_S(\tau) + \gamma[R_S(\tau-\tau_0) + R_S(\tau+\tau_0)] \end{aligned} \tag{12.34}$$

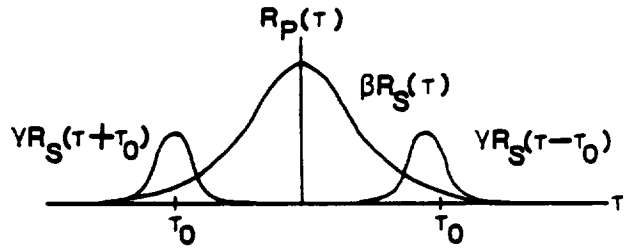


Figure 62. Autocorrelation of microphone signal.

where $R_S(\tau)$ is the autocorrelation of the source signal,

$$\beta = \frac{1}{\ell_1^2} + \frac{\alpha^2}{(\ell_2 + \ell_3)^2}$$

$$\gamma = \frac{\alpha}{\ell_1(\ell_2 + \ell_3)}$$

and

$$\tau_0 = \frac{\ell_2 + \ell_3 - \ell_1}{c}$$

Thus, the autocorrelation of the microphone signal appears as the sum of the three curves shown in figure 62, and separating the directly radiated sound from the echo would be very difficult because of the distributed nature of the autocorrelation $R_S(\tau)$.

However, the power spectral density of the microphone signal is, from equation (12.34),

$$S_P(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} R_P(\tau)\, e^{-i\omega\tau}\, d\tau = [\beta + 2\gamma\cos\omega\tau_0]S_S(\omega) \qquad (12.35)$$

where $S_S(\omega)$ is the power spectral density of the source signal $S(t)$. Thus, the source and receiver signals are related by the standard linear system relation

$$S_P(\omega) = |H(\omega)|^2 S_S(\omega)$$

where the squared frequency response function of the equivalent linear filter is

$$|H(\omega)|^2 = \beta + 2\gamma\cos\omega\tau_0$$

Thus, if the propagation paths and reflection coefficient were known, the source spectrum could be recovered by

$$S_S(\omega) = \frac{S_P(\omega)}{\beta + 2\gamma \cos \omega\tau_0}$$

Generally, the propagation paths and reflection coefficient(s) are not known, in which case equation (12.35) may be written

$$S_P(\omega) = \beta S_S(\omega) \left(1 + \frac{2\gamma}{\beta} \cos \omega\tau_0\right)$$

Thus,

$$\ln S_P(\omega) = \ln S_Q(\omega) + \ln(\beta\ell_1^2) + \ln \left(1 + \frac{2\gamma}{\beta} \cos \omega\tau_0\right)$$

where $S_Q(\omega) = S_S(\omega)/\ell_1^2$ is the power spectral density of the echo-free signal (eq. (12.33)), and the power cepstrum of the microphone signal is

$$C_p(\tau) = \int_{-\infty}^{\infty} \ln[S_P(\omega)]\, e^{i\omega\tau}\, d\omega = \int_{-\infty}^{\infty} \ln[S_Q(\omega)]\, e^{i\omega\tau}\, d\omega + \ln \beta\ell_1^2 \int_{-\infty}^{\infty} e^{i\omega\tau}\, d\omega + \int_{-\infty}^{\infty} \ln \left(1 + \frac{2\gamma}{\beta} \cos \omega\tau_0\right) e^{i\omega\tau}\, d\omega \quad (12.36)$$

Here, the first integral is the power cepstrum of the echo-free signal and the second is $2\pi\delta(\tau)$ by equation (2.6). Further, for $|x| < 1$

$$\ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots$$

Since

$$\frac{2\gamma}{\beta} = \frac{2\left(\frac{\alpha\ell_1}{\ell_2+\ell_3}\right)}{1 + \left(\frac{\alpha\ell_1}{\ell_2+\ell_3}\right)^2}$$

whose magnitude is less than unity, the third integral in equation (12.36) may be written

$$\int_{-\infty}^{\infty} \ln\left(1+\frac{2\gamma}{\beta}\cos\omega\tau_0\right) e^{i\omega\tau}\,d\omega$$
$$= \frac{2\gamma}{\beta}\int_{-\infty}^{\infty} \cos\omega\tau_0\, e^{i\omega\tau}\,d\omega - \frac{1}{2}\left(\frac{2\gamma}{\beta}\right)^2 \int_{-\infty}^{\infty} \cos^2\omega\tau_0\, e^{i\omega\tau}\,d\omega + \cdots$$
$$= \left(\frac{2\gamma}{\beta}\right)\pi[\delta(\tau-\tau_0)+\delta(\tau+\tau_0)]$$
$$-\frac{1}{4}\left(\frac{2\gamma}{\beta}\right)^2 \pi[2\delta(\tau)+\delta(\tau-2\tau_0)+\delta(\tau+2\tau_0)]+\cdots$$

upon noting that $\cos^2 x = 1/2(1+\cos 2x)$ and using equation (4.7). Thus, the power cepstrum of the microphone signal is

$$C_p(\tau) = C_Q(\tau) + \pi\left[2\ln\beta\ell_1^2 - \frac{1}{2}\left(\frac{2\gamma}{\beta}\right)^2+\cdots\right]\delta(\tau)$$
$$+\pi\left[\left(\frac{2\gamma}{\beta}\right)+\cdots\right][\delta(\tau-\tau_0)+\delta(\tau+\tau_0)]$$
$$-\pi\left[\frac{1}{4}\left(\frac{2\gamma}{\beta}\right)^2+\cdots\right][\delta(\tau-2\tau_0)+\delta(\tau+2\tau_0)]+\cdots \quad (12.37)$$

where $C_Q(\tau)$ is the power cepstrum of the echo-free signal. Further. it can be shown that the coefficient of the $\delta(\tau)$ term is identically zero. Thus, the power cepstrum is as shown in figure 63, where the echo shows up as delta functions (called "rahmonics") at multiples of the delay time τ_0.

This cepstrum can theoretically be readily filtered (or "liftered") by interpolating the continuous function at the positions where the delta functions occur leaving only the echo-free receiver cepstrum. The echo-free receiver spectrum is then recovered by inverse Fourier transforming, that is,

$$S_Q(\omega) = \exp\frac{1}{2\pi}\int_{-\infty}^{\infty} C_Q(\tau)\, e^{-i\omega\tau}\,d\tau \quad (12.38)$$

It should be mentioned, however, that programming the liftering operation may be difficult, particularly when more than one echo is present.

12.6 Zoom FFT

In recent years, the manufacturers of stand-alone spectral analyzers have developed a new feature called the zoom FFT,[31] which allows

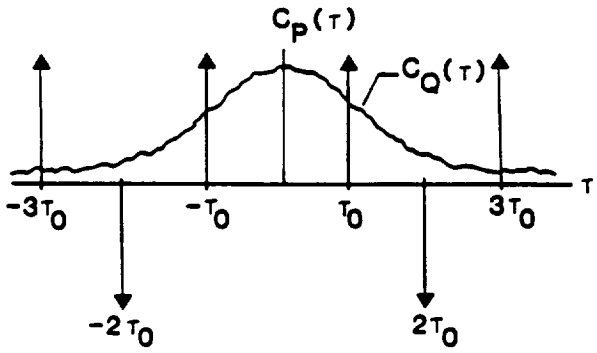


Figure 63. Power cepstrum of microphone signal.

one to greatly increase the resolution in a small portion of the spectrum. However, although the technique involves clever hardware implementation, it does not violate the fundamental resolution constraint (eq. (7.33)) that

$$|\Delta\omega| \geq \frac{2\pi}{T}$$

where T is the length of the record.

The method involves two assumptions:

1. That data storage is limited by the memory of the spectral analyzer to N real data points
2. That access to additional data is essentially unlimited

Consider N data points $x(n\Delta t)$ for $n = 0, 1, 2, \ldots, N-1$. The FFT of these data is

$$X(k\Delta\omega) = \sum_{n=0}^{N-1} x(n\Delta t)\, e^{-i2\pi kn/N} \qquad (12.39)$$

where $\Delta\omega = 2\pi/N\,\Delta t$. Thus, the bandwidth of the FFT must satisfy the fundamental constraint

$$|\Delta\omega|\,\Delta t = \frac{2\pi}{N}$$

If N is fixed by memory restrictions, the only way to increase the resolution (i.e., reduce $\Delta\omega$) is to increase Δt. However, increasing Δt would lower the Nyquist frequency and introduce aliasing into the spectral estimate. The way out of this dilemma is to use the old technique that electrical engineers call heterodyning. That is, multiply the signal by

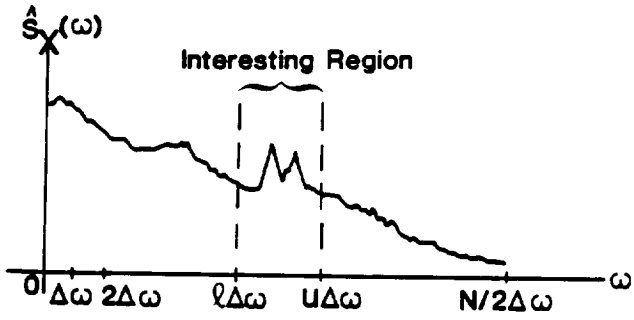


Figure 64. Interesting region in spectral estimate.

another signal to generate sum and difference frequencies, filter out the sum frequencies, and analyze only the difference frequencies. These frequencies are low and can thus be sampled at a much larger Δt without introducing aliasing. However, since the record length is $T = N\,\Delta t$, the record length must increase in order to obtain the same number of data points at the larger Δt.

The implementation of the zoom FFT technique involves the following steps. From the FFT (eq. (12.39)) of the original N data points, the spectral density may be estimated in the region $0 < \omega < \omega_c = N/2\,\Delta\omega$ as shown in figure 64, where $\Delta\omega = 2\pi/N\Delta t$. Suppose that the range $\omega_\ell = \ell\,\Delta\omega < \omega < \omega_u = u\,\Delta\omega$ is of particular interest. Then, let

$$m = \frac{\ell + u}{2}$$

and multiply the original data $X(n\Delta t)$ by the complex exponential $e^{-i2\pi mn/N}$ to obtain the series of complex data points

$$z(n\Delta t) = e^{-i2\pi mn/N} x(n\Delta t)$$

The FFT of this new series would be

$$\begin{aligned} Z(k\Delta\omega) &= \sum_{n=0}^{N-1} z(n\Delta t) e^{-i2\pi kn/N} \\ &= \sum_{n=0}^{N-1} x(n\Delta t) e^{-i2\pi(k+m)n/N} \\ &= X[(k+m)\Delta\omega] \end{aligned}$$

Thus, the power in the random process $Z(t)$ at the frequency of zero is precisely the power in the random process $X(t)$ at the frequency

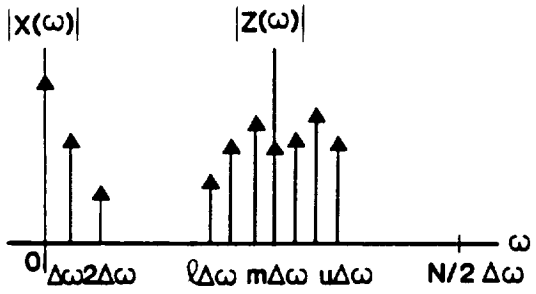


Figure 65. Change in frequency origin.

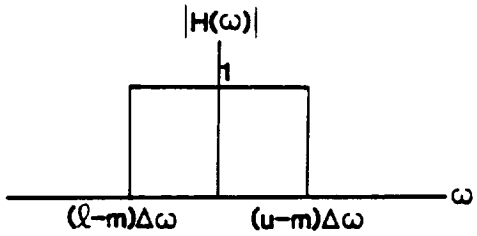


Figure 66. Low-pass zoom filter.

$m\,\Delta\omega$ and the frequency origin has effectively been moved as shown in figure 65.

Then, suppose that the random process $Z(t)$ is passed through a low-pass digital filter with frequency response function as shown in figure 66 to yield a new random process $Y(t)$ containing only low frequencies such that $|\omega| < (u-m)\Delta\omega$. Since $Y(t)$ contains only low frequencies, the Nyquist criterion (eq. (9.7)) then says that the data are being sampled more frequently than necessary. In fact, one may now use a new Δt given by

$$\Delta t_{\text{new}} = \frac{\pi}{(u-m)\Delta\omega} = \frac{N}{2(u-m)}\Delta t \qquad (12.40)$$

For example, if $N = 128$ and $u - m = 2$, then $\Delta t_{\text{new}} = 32\,\Delta t$ and the same information can be obtained by keeping only every 32nd value of $Y(t)$.

Suppose one keeps only every Δt_{new} data point and still fills up the memory. The memory now holds $N/2$ complex data points $y(n\Delta t_{\text{new}})$ for $n = 0, 1, 2, \ldots, N/2 - 1$. The FFT of these data

$$Y(k\Delta\omega_{\text{new}}) = \sum_{n=0}^{N/2-1} y(n\Delta t_{\text{new}})\, e^{-i2\pi kn/(N/2)}$$

then yields a spectral estimate only in the region of interest with resolution

$$\Delta\omega_{\text{new}} = \frac{2\pi}{(N/2)\Delta t_{\text{new}}} = \frac{4(u-m)}{N}\Delta\omega \tag{12.41}$$

Thus, for the same example, $N = 128$ and $(u - m) = 2$,

$$\Delta\omega_{\text{new}} = \frac{\Delta\omega}{16}$$

and the resolution is increased by a factor of 16. However, the total record length required to do this is

$$\frac{N}{2}\Delta t_{\text{new}} = \frac{N}{4(u-m)}(N\,\Delta t)$$

or 16 times more data for the example cited.

12.7 Digital Spectral Analysis of Periodic Signals

Often one is interested in analyzing a periodic signal, which may or may not be contaminated by random noise. For example, the noise produced by a helicopter is largely due to the periodic motion of the rotor. However, there are also other sources of helicopter noise that are more random. In this case, many of the problems seen in the analysis of random data, as well as some new ones, arise.

Recall that if $f(t)$ is a periodic signal with period p, then it may be represented by the Fourier series (eq. (2.4))

$$\begin{aligned} f(t) &= \frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n \cos\omega_n t + b_n \sin\omega_n t) \\ &= \sum_{n=-\infty}^{\infty} F_n e^{i\omega_n t} \end{aligned} \tag{12.42}$$

where $\omega_n = 2\pi n/p$ are harmonics of the fundamental radian frequency $\omega_1 = 2\pi/p$ of the signal. The Fourier integral transform of this signal is (eq. (2.7))

$$F(\omega) = \sum_{n=-\infty}^{\infty} F_n \delta(\omega - \omega_n)$$

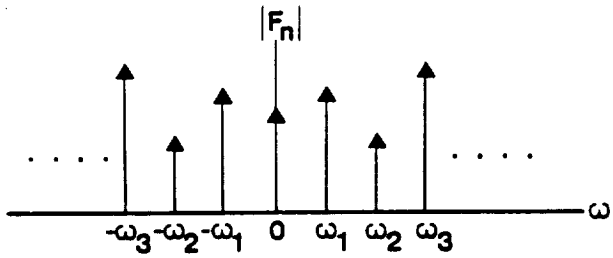


Figure 67. Amplitude spectrum of periodic signal.

Such harmonic analyses are ordinarily represented by the *amplitude spectrum* as shown in figure 67. In terms of the more familiar a's and b's,

$$|F_n| = \frac{\sqrt{a_n^2 + b_n^2}}{2}$$

Note that this is a much different representation from a power spectral density since $|F_n|^2$ *is the finite power in the signal at the discrete frequency* ω_n. One is ordinarily interested in obtaining accurate estimates of these amplitudes $|F_n|$ and the frequencies ω_n.

For the digital analysis of such data, it might intuitively be expected that one ought to analyze data of length

$$T = \nu p \tag{12.43}$$

where ν is an integer. That is, the data length ought to be some whole number of periods of the signal. Equation (12.43) can, in fact, be shown to be true mathematically.

Suppose that the periodic signal is sampled at equal intervals Δt yielding data $f(j\Delta t)$ for $j = 0, 1, 2, \ldots, N-1$ for a total sample length of $T = N\,\Delta t$. The finite Fourier transform of this signal is

$$F(k\Delta\omega) = \sum_{j=0}^{N-1} f(j\Delta t)\, e^{-i2\pi kj/N} \tag{12.44}$$

which is evaluated at the frequencies

$$\omega_k = k\,\Delta\omega = \frac{2\pi k}{T} \qquad (k = 0, 1, 2, \ldots, N/2)$$

The signal (eq. (12.42)) has power occurring at the frequencies $\omega_n = 2\pi n/p$. Thus, the frequencies at which equation (12.44) is

evaluated may not be the frequencies at which equation (12.42) has power! For n fixed, there is a k such that $\omega_k = \omega_n$ if and only if

$$\frac{2\pi k}{T} = \frac{2\pi n}{p}$$

or

$$T = \frac{k}{n} p$$

If one wishes this to be true for all n such that ω_n is less than the Nyquist frequency, $\omega_c = \pi/\Delta t$, then

$$T = \nu p$$

where ν is an integer and the k corresponding to a given n is

$$k = \nu n$$

Thus, *the data length T must be some integer number of periods of the signal.*

Unfortunately, it is not always possible to satisfy the criterion in equation (12.43) that the data length be some integer number of periods because either (1) The period is not known a priori or (2) use of the FFT requiring $N = 2^\ell$ does not make it convenient.

Then, what happens if one proceeds with the analysis when the frequencies at which the discrete Fourier transform is evaluated are not equal to the frequencies at which the periodic signal has power? This case has been studied extensively by Burgess.[32]

Suppose $T = \nu p$ where ν is *not* an integer and data $f(j\Delta t)$ for $j = 0, 1, 2, \ldots, N-1$ from a periodic process $f(t)$ exist. The discrete Fourier transform of these data is

$$\begin{aligned} F(k\Delta\omega) &= \sum_{j=0}^{N-1} f(j\Delta t)\, e^{-i2\pi kj/N} \\ &= \sum_{j=0}^{N-1} \sum_{n=-\infty}^{\infty} F_n e^{i\omega_n j\Delta t} e^{-i2\pi kj/N} \\ &= \sum_{n=-\infty}^{\infty} F_n \sum_{j=0}^{N-1} \left[e^{i(\omega_n \Delta t - 2\pi k/N)}\right]^j \end{aligned}$$

where

$$\omega_n \Delta t - \frac{2\pi k}{N} = \frac{2\pi}{N}(n\nu - k)$$

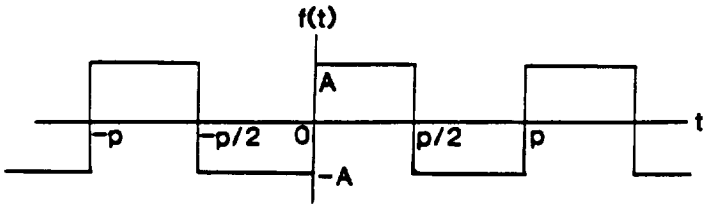


Figure 68. Square wave of amplitude A.

since $T = N\,\Delta t = \nu p$. Thus, defining

$$w_{nk}(\nu) = \sum_{j=0}^{N-1} \left[e^{i2\pi(n\nu-k)/N}\right]^j = \frac{1 - e^{i2\pi(n\nu-k)}}{1 - e^{i2\pi(n\nu-k)/N}} \tag{12.45}$$

and noting that division by N is necessary for equation (12.44) to approximate equation (2.5), the estimated Fourier coefficients are

$$\hat{F}_k = \frac{F(k\Delta\omega)}{N} = \frac{1}{N} \sum_{n=-\infty}^{\infty} F_n w_{kn}(\nu) \tag{12.46}$$

which is a weighted sum of all the Fourier coefficients with respect to the very complicated weighting function (eq. (12.45)). Physically, $w_{kn}(\nu)$ represents a transfer of that power which in the periodic signal would appear at the frequency $\omega_n = 2\pi n/p$ to the estimated Fourier coefficient at the frequency $\omega_k = 2\pi k/T$. This phenomenon is called *leakage* and can result in substantial errors in both the amplitude and the frequency of the Fourier coefficients, basically dependent on how close $n\nu$ ever gets to k.

Periodic signals may also have problems with aliasing, if they contain power at frequencies higher than the Nyquist frequency, just as in random signals. In fact, aliasing is much more evident in periodic signals. A simple example that can be readily analyzed is to consider a square wave of amplitude A as shown in figure 68. The Fourier coefficients are given by equation (2.5):

$$F_n = \frac{1}{p}\int_0^p f(t)\, e^{-i\omega n t}\, dt$$
$$= \frac{A}{p}\int_0^{p/2} e^{-i\omega n t}\, dt - \frac{A}{p}\int_{p/2}^{p} e^{-i\omega n t}\, dt$$
$$= \frac{iA}{n\pi}[(-1)^n - 1]$$

Now, suppose that the signal is sampled at equal intervals Δt for one period, that is, $N\Delta t = p$, and that N is even. Then. by equation (12.44), noting that $f(0) = f(p/2) = 0$ because of the jump discontinuity,

$$F(k\Delta\omega) = \sum_{j=0}^{N-1} f(j\Delta t)\, e^{-i2\pi kj/N}$$
$$= A\sum_{j=0}^{N/2-1}\left(e^{-i2\pi k/N}\right)^j - A\sum_{j=N/2+1}^{N-1}\left(e^{-i2\pi k/N}\right)^j$$
$$= iA[(-1)^k - 1]\cot\frac{\pi k}{N}$$

Thus, the approximate Fourier coefficients are

$$\hat{F}_n = \frac{F(n\Delta\omega)}{N} = \frac{iA}{N}[(-1)^n - 1]\cot\frac{\pi k}{N}$$
$$= \frac{iA}{N}[(-1)^n - 1]\left[\frac{N}{n\pi} - \sum_{k=1}^{\infty}\frac{2^{2k}|B_{2k}|}{(2k)!}\left(\frac{\pi n}{N}\right)^{2k-1}\right]$$

upon expansion of the cotangent. Here B_{2k} is a Bernoulli number. Note that the first term is the exact Fourier coefficient and the sum represents aliased terms. Thus, periodic signals must also be filtered to avoid aliasing.

12.8 Spectral Analysis of Nonstationary Random Processes

Practically all the analysis discussed thus far in this monograph has assumed that the random process of interest is at least weakly stationary. When this is not true, two major problems arise:

1. Since the statistics of the random process vary with time, time averages cannot be used to reduce variability. Thus, an ensemble of sample functions must be collected, analyzed, and averaged.

2. Since second order moments depend on two variables instead of one, interpretation is more difficult.

For example, the autocorrelation of a nonstationary random process $X(t)$, $R_X(t_1, t_2)$, depends on two variables t_1 and t_2. The corresponding power spectral density may be defined as

$$\begin{aligned} S_X(\omega_1, \omega_2) &= \frac{1}{4\pi^2} \int_{-\infty}^{\infty} dt_1 \int_{-\infty}^{\infty} dt_2 R_X(t_1, t_2)\, e^{i(\omega_1 t_1 - \omega_2 t_2)} \\ &= E\{X^*(\omega_1) X(\omega_2)\} \end{aligned} \tag{12.47}$$

where

$$X(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(t)\, e^{-i\omega t}\, dt$$

The autocorrelation may then be recovered by the inversion relation:

$$R_X(t_1, t_2) = \int_{-\infty}^{\infty} d\omega_1 \int_{-\infty}^{\infty} d\omega_2\, S_X(\omega_1, \omega_2)\, e^{-i(\omega_1 t_1 - \omega_2 t_2)} \tag{12.48}$$

However, the power in the process at time t is

$$E[X^2(t)] = R_X(t, t) = \int_{-\infty}^{\infty} d\omega_1 \int_{-\infty}^{\infty} d\omega_2 S_x(\omega_1, \omega_2) e^{-i(\omega_1 - \omega_2)t}$$

Thus, the power spectral density (eq. (12.47)) does not admit a simple interpretation in terms of power per unit frequency. If the random process were stationary, then $S_X(\omega_1, \omega_2) = S_X(\omega_1)\delta(\omega_2 - \omega_1)$. Thus, it has been suggested[33] that the spread of the values of the spectrum $S_X(\omega_1, \omega_2)$ about the line $\omega_1 = \omega_2$ is a measure of the nonstationarity of the random process. One virtue of the definition in equation (12.47) is that if $X(t)$ is input to a linear system, the power spectral density of the output $Y(t)$ is given by

$$S_Y(\omega_1, \omega_2) = H^*(\omega_1) H(\omega_2) S_X(\omega_1, \omega_2)$$

However, although this definition is relatively straightforward, it has proven to be of limited practical use.

A more useful definition is accomplished by introducing the variables

$$\tilde{t} = \frac{t_1 + t_2}{2} \qquad \tau = t_2 - t_1$$

called the mean time and time difference. Then, the autocorrelation may be written

$$R_X(t_1, t_2) = R_X(\bar{t}-\tau/2, \bar{t}+\tau/2)$$

and one can define the *time varying power spectral density* as

$$S_X(\bar{t}, \omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} R_X(\bar{t}-\tau/2, \bar{t}+\tau/2)\, e^{-i\omega\tau}\, d\tau \tag{12.49}$$

with inverse

$$R_X(\bar{t}-\tau/2, \bar{t}+\tau/2) = \int_{-\infty}^{\infty} S_X(\bar{t}, \omega)\, e^{i\omega\tau}\, d\omega \tag{12.50}$$

Although this definition does not satisfy a simple relation for a linear system, the power is

$$E\{X^2(t)\} = R_X(\bar{t}, \bar{t}) = \int_{-\infty}^{\infty} S_X(\bar{t}, \omega)\, d\omega$$

Thus, $S(t, \omega)$ admits interpretation as the power per unit frequency in the signal at time t. This definition has proven useful, probably being most widely used in the field of voice analysis and identification.

A class of nonstationary random processes that have been fairly widely studied are called "pseudostationary" random processes. Here

$$Y(t) = A(t)X(t) \tag{12.51}$$

where $X(t)$ is a stationary random process and $A(t)$ is a deterministic modulation signal that is assumed to vary much more slowly than $X(t)$. For example, figure 69 presents the acoustic pressure time history measured by a stationary microphone as an aircraft flies over it. Such data can be represented by equation (12.51), although the Doppler shift in frequency inherent in this measurement technique must also be taken into account. For processes that can be described by equation (12.51),

$$\begin{aligned} R_Y(\bar{t}-\tau/2, \bar{t}+\tau/2) &= E\{Y(\bar{t}-\tau/2)Y(\bar{t}+\tau/2)\} \\ &= A(\bar{t}-\tau/2)A(\bar{t}+\tau/2)E\{X(\bar{t}-\tau/2)X(\bar{t}+\tau/2)\} \\ &\approx A^2(\bar{t})R_X(\tau) \end{aligned}$$

and thus

$$S_Y(\bar{t}, \omega) \approx A^2(\bar{t})S_X(\omega)$$

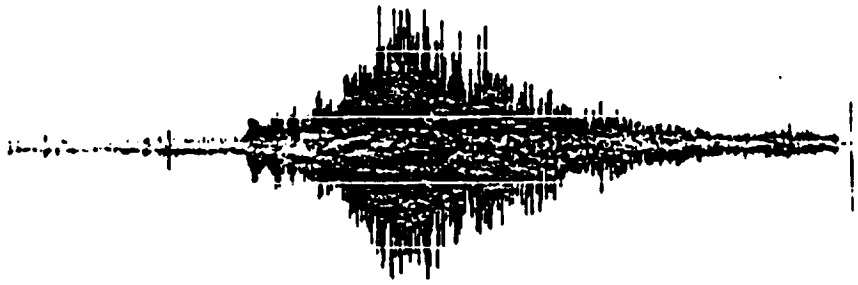


Figure 69. Acoustic pressure time history produced by aircraft flyover.

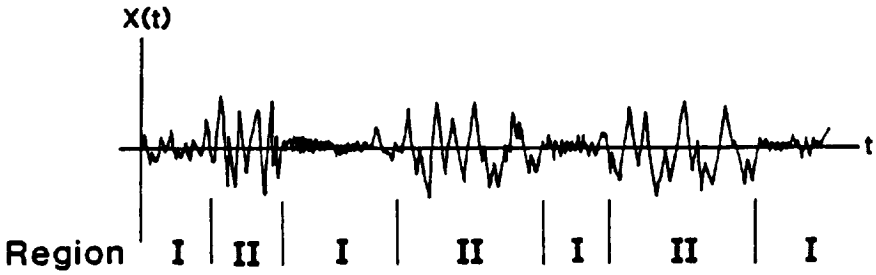


Figure 70. Identifiably nonstationary signal.

where $S_X(\omega)$ is the ordinary power spectral density of $X(t)$.

Another type of nonstationary signal that seems to occur fairly frequently is shown in figure 70. This type might be called identifiably nonstationary. In such signals, the time history consists of two (or more) regions where clearly different phenomena are occurring. Such a record might be produced, for example, by a velocity sensor mounted on an aircraft flying in and out of thunderstorms, by a microphone measuring the noise level inside a train traversing sections of smooth and rough track, or by a seismometer recording periods of more and less seismic activity. For such records, it seems reasonable to break the time history into blocks corresponding to the different regions and analyze the records in the blocks from each region as if they were produced by a single stationary random process.

While such an approach undoubtably produces useful information, it requires much engineering judgment. First, the various regions must be identified and the break points between them determined. Then, the reasonability of treating the sections of a given region as stationary must be evaluated. For this purpose, the test for stationarity given in chapter VI is useful. Sometimes the means may not pass such a test, although the variances do. In this case, the mean of each block may be subtracted from the data in that block before analysis. Other times,

detrending, as was discussed in chapter XI, may have to be applied to each individual block before the stationarity test is satisfied.

Assuming satisfactory passage (which again requires engineering judgment), spectral estimates for each block can then be computed and averaged together, just as in block averaging, to reduce the statistical variability. However, the blocks may be so short that the needed resolution cannot be obtained with standard techniques. In this case, it may be necessary to apply techniques such as the maximum entropy or maximum likelihood methods.

In analyzing such data, one should use good engineering judgment at every step along the way and should always be mindful in interpreting the final results that many assumptions have gone into the analysis.

A way in which one can make any random process appear more stationary is to "normalize" it. Suppose $X(t)$ is a nonstationary random process with mean $m_X(t)$ and variance $\sigma_X^2(t)$. Then, the new random process

$$Z(t) = \frac{X(t) - m_X(t)}{\sigma_X(t)}$$

is such that $E\{Z(t)\} = 0$ and $\sigma_Z^2 = 1$. Thus, $Z(t)$ is much closer to being stationary than $X(t)$ was. Further, $X(t)$ may be written

$$X(t) = m_X(t) + \sigma_X(t)Z(t)$$

which is just the sum of a deterministic signal $m_X(t)$ plus a random process that looks something like equation (12.51). This type of analysis has often been applied to transient phenomena such as vibration during a Space Shuttle liftoff or noise measurements during an exhalation of breath, where $m_X(t)$ and $\sigma_X(t)$ may be estimated by ensemble averages over repeated experiments.

The analysis of nonstationary processes is not in a very satisfactory state and may, in fact, never be, although they are the topic of considerable current research. The difficulty lies in the fundamental importance of the assumption of stationarity in the analysis and interpretation of random data. Basically, the state of the art is that one tries to make the nonstationary signal look enough like a stationary signal that stationary techniques may be used. Recently, [34] this approach has been placed on a firm foundation by a unified theory that considers more general types of invariance under transformation in addition to independence of the origin of time, which led to the concept of stationarity.

References

1. Blackman, R. B.; and Tukey, J. W.: *The Measurement of Power Spectra.* Dover Publ., Inc., 1959.
2. Bracewell, Ron: *The Fourier Transform and Its Applications.* McGraw-Hill Book Co., c.1965.
3. Jones, D. S.: *Generalised Functions.* McGraw-Hill Book Co., Inc., c.1966.
4. Papoulis, Athanasios: *The Fourier Integral and Its Applications.* McGraw-Hill Book Co., Inc., 1962.
5. Papoulis, Athanasios: *Probability, Random Variables, and Stochastic Processes*, Second ed. McGraw-Hill Book Co., c.1984.
6. Von Mises, Richard: *Probability, Statistics, and Truth*, Second rev. English ed. Dover Publ., Inc., 1981.
7. Wiener, Norbert: *Extrapolation, Interpolation, and Smoothing of Stationary Time Series With Engineering Applications.* Technology Press, M.I.T., and John Wiley & Sons, Inc., 1949.
8. Rice, S. O.: Mathematical Analysis of Random Noise. *Selected Papers on Noise and Stochastic Processes,* Nelson Wax, ed., Dover Publ., Inc., c.1954, pp. 133–294.
9. Kailath, Thomas: *Linear Systems.* Prentice-Hall, Inc., c.1980.
10. Waldmeier, N.: *The Sunspot-Activity in the Years 1610–1960.* Zürich Schulthess & Co AG, 1961.
11. Bartlett, M. S.: *An Introduction to Stochastic Processes With Special Reference to Methods and Applications.* Cambridge at the Univ. Press, 1966.
12. Otnes, Robert K.; and Enochson, Loren: *Digital Time Series Analysis.* John Wiley & Sons, Inc., c.1972.
13. Shannon, Claude E.: Communication in the Presence of Noise. *Proc. IRE*, vol. 37, no. 1, Jan. 1949, pp. 10–21.
14. Higgins, J. R.: Five Short Stories About the Cardinal Series. *Bull. American Math. Soc.*, vol. 12, no. 1, Jan. 1985, pp. 45–89.
15. Petersen, Daniel P.; and Middleton, David: Sampling and Reconstruction of Wave-Number-Limited Functions in N-Dimensional Euclidean Spaces. *Inf. & Control*, vol. 5, 1962, pp. 279–323.
16. Cooley, James W.; and Tukey, John W.: An Algorithm for the Machine Calculation of Complex Fourier Series. *Math. Comput.*, vol. 19, no. 90, Apr. 1965, pp. 297–301.
17. Runge, C.: Uber die Zerlegung empirisch gegebener Periodscher Funktionen in Sinnuswellen. *Z. Math. Phys.*, vol. 48, 1903, pp. 443–456.

18. Gentleman, W. M.; and Sande, G.: Fast Fourier Transforms—For Fun and Profit. *AFIPS Conference Proceedings, Volume 29—Fall Joint Computer Conference*, Spartan Books, Inc., 1966. pp. 563–578.
19. Nussbaumer, Henri J.: *Fast Fourier Transform and Convolution Algorithms*, Second corr. & updated ed. Springer-Verlag, 1982.
20. Gonzalez, Rafael C.; and Wintz, Paul: *Digital Image Processing*. Addison-Wesley Pub. Co., 1977.
21. Brown, T. J.; and Hardin, J. C.: A Note on Kendall's Autoregressive Series. *J. Appl. Probab.*, vol. 10, 1973, pp. 475–478.
22. Kennedy, William J., Jr.; and Gentle, James E.: *Statistical Computing*. Marcel Dekker, Inc., 1980.
23. Hardin, Jay C.; and Brown, Thomas J.: Further Results on Kendall's Autoregressive Series. *J. Appl. Probab.*, vol. 12, 1975. pp. 180–182.
24. Burg, John Parker: Maximum Entropy Spectral Analysis. *Modern Spectrum Analysis*, Donald G. Childers, ed., John Wiley & Sons. Inc., c.1978, pp. 34–41.
25. Shannon, C. E.: A Mathematical Theory of Communication. *Bell Syst. Tech. J.*, vol. XXVII, no. 3, July 1948, pp. 379–423 and no. 4. Oct. 1948, pp. 623–656.
26. Capon, J.: High-Resolution Frequency-Wavenumber Spectrum Analysis. *Proc. IEEE*, vol. 57, no. 8, Aug. 1969, pp. 1408–1418.
27. Gaster, M.; and Roberts, J. B.: Spectral Analysis of Randomly Sampled Signals. *J. Inst. Math. & Its Appl.*, vol. 15, no. 2, Apr. 1975. pp. 195–216.
28. Shapiro, Harold S.; and Silverman, Richard A.: Alias-Free Sampling of Random Noise. *J. Soc. Ind. & Appl. Math.*, vol. 8, no. 2, June 1960, pp. 225–248.
29. Gaster, M.; and Roberts, J. B.: The Spectral Analysis of Randomly Sampled Records by a Direct Transform. *Proc. R. Soc. (London)*, ser. A, vol. 354, no. 1676, Apr. 21, 1977, pp. 27–58.
30. Childers, Donald G.; Skinner, David P.; and Kemerait, Robert C.: The Cepstrum: A Guide to Processing. *IEEE Proc.*, vol. 65, no. 10, Oct. 1977, pp. 1428–1443.
31. Thrane, N.: ZOOM-FFT. *Tech. Rev.*, no. 2, B & K Instruments. Inc., 1980, pp. 3–41.
32. Burgess, John C.: On Digital Spectrum Analysis of Periodic Signals. *J. Acoust. Soc. America*, vol. 58, no. 3, Sept. 1975, pp. 556–567.
33. Trevino, George: The Frequency Spectrum of Nonstationary Random Processes. *Time Series Analysis: Theory and Practice 2*, O. D. Anderson, ed., North-Holland Pub. Co., 1982, pp. 237–246.

34. Trevino, George: Topological Invariance in Stochastic Analysis. *Mathematical Modelling in Science and Technology,* Xavier J. R. Avula, Rudolf E. Kalman, Anthanasios I. Liapis, and Ervin Y. Rodin, eds., Pergamon Press, 1983, pp. 278–283.

1. Report No. NASA RP-1145	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Introduction to Time Series Analysis		5. Report Date March 1986
		6. Performing Organization Code 532-06-13-06
7. Author(s) Jay C. Hardin		8. Performing Organization Report No. L-15958
9. Performing Organization Name and Address NASA Langley Research Center Hampton, VA 23665-5225		10. Work Unit No.
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration, Washington, DC 20546-0001		13. Type of Report and Period Covered Reference Publication
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

The field of time series analysis is explored from its logical foundations to the most modern data analysis techniques. The presentation is developed, as far as possible, for continuous data, so that the inevitable use of discrete mathematics is postponed until the reader has gained some familiarity with the concepts. The monograph seeks to provide the reader with both the theoretical overview and the practical details necessary to correctly apply the full range of these powerful techniques. In addition, the last chapter introduces many specialized areas where research is currently in progress.

17. Key Words (Suggested by Author(s)) Time series analysis Power spectral analysis Data analysis	18. Distribution Statement Unclassified–Unlimited Subject Category 71		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 178	22. Price A09

For sale by the National Technical Information Service, Springfield, Virginia 22161

NASA-Langley, 1986