Richard A. Poisel

Target

Acquisition

in Communication Electronic Warfare Systems

For a listing of recent titles in the *Artech House Information Warfare Library,*
turn to the back of this book.

Target Acquisition in Communication Electronic Warfare Systems

Richard A. Poisel

AH

Artech House
Boston • London
www.artechhouse.com

Cover design by Gary Ragaglia

10 9 8 7 6 5 4 3 2 1

To my parents: my earliest and best teachers

Contents

Preface

Target acquisiton in communication electronic warfare (EW) systems is a requisite function for proper system functioning. In general it consists of determining the presence of signals at particular frequencies and whether those signals are targets or not. Measurements of signal parameters are required so such determination can be accomplished. These measurements are automated to the maximum extent possible.

This book is intended for technical practitioners, such as engineers and scientists, who are interested in learning the basics of how to acquire target signals in communication EW systems. The educational level required is that of a baccalaureate degree in an appropriate technical discipline. The material is suitable for one new to the EW field and contains references to additional, more advanced material for those motivated to pursue such investigations.

Chapter 1 provides a brief introduction to the tagret acquisition problem. The communication system model is introduced as well as what is meant by electronic warfare in a communication system setting.

There are two generalized classifications for signals: random and nonrandom. The latter is usually termed *deterministic* in the sense that if the value of the signal is known at any point in time, then its future (and past) values can be determined. In the former, one or more parameters associated with the signal are random, and therefore knowing the signal's value does not allow for ascertaining its value at other times. These signals are called *stochastic*. Such signals can only be described in probabilistic terms.

Chapter 2 presents background material on statistical processes and contrasts deterministic versus stochastic processes. For the most part, the signals of concern for communication EW systems are determinsistic with one or more random parameters. The theory of communications has developed along certain paths, and the design of modern communication systems utilizes these concepts. For example, the oscillators used to transform signals from one frequency regime to another (frequency conversion) are sinusoidal. The characteristics of sinusoids are well known, and except perhaps for unknown (and maybe random) amplitude and phase, they are deterministic. A review of Fourier transforms is presented to remind the reader of the fundamental principles involved. A thorough coverage of such transforms is certainly beyond the scope of this book, however.

Chapter 3 provides a brief introduction to radio frequency (RF) spectrum search techniques. Searching the RF spectrum is required when the targets of interest do not necessarily occupy the same frequency all the time. For static situations, the targets remain fixed and searching is simple. In fact, searching may not be necessary at all—just tune receivers to the known frequencies where the targets of interest are located and leave them there. In most cases of interest, however, this will not be true and some form of searching is required. This is dictated by the limited resources available compared to the number of targets of concern.

Chapter 4 introduces the reader to the notion of hypothesis testing applied to target detection. Hypothesis testing is a statistical technique used to ascertain whether particular conditions are true or not in an optimal way.

Estimation of target parameters is important so that determination of whether the signal is one of interest or not can be made. The fundamentals of this area as they apply to EW target acquisition are introduced in Chapter 5. The target parameters of interest extend beyond just the frequency of active signals. They include signal modulation type and power levels, to mention just a few.

Perhaps the most important parameters of interest are the frequencies of target signals. The methods for estimating the spectrum of a target are introduced in Chapter 6. The methods discussed produce relatively low frequency resolution, however. Resolution in this sense is the ability to distinguish two signals that are closely spaced in frequency. In many practical scenarios where EW systems are applied, the RF spectrum is very crowded and many signals are present. The ability to separate signals in those cases becomes very important.

Chapter 7 introduces the reader to methods for detection of deterministic signals. As mentioned, for the most part a considerable amount is known about the signals of interest from communication targets. Typically one or more parameters are random in nature, but for the most part the signals are a sinusoid of some type. Therefore, detection methods for deterministic signals are important.

Chapter 8 is the counterpart to Chapter 7 for detection techniques for stochastic signals. Normally, the less that is known about the signal to be detected, the poorer the detection performance. When the signal is completely known the best technique for detection is the matched filter. Such filters can be designed to detect known signals in an optimum sense; what is meant by optimality is described in Chapter 7. When one or more parameters about the signal are not known, then the matched filter is no longer optimal and some other methods must be employed for detection.

As mentioned above, classical methods for signal detection lead to relatively low resoluton in signal separation, which can be detrimental in crowded RF environments. High-resolution techniques have been devised to deal with this problem, and such methods are introduced in Chapter 9. Two methods are

described in detail to give the reader a flavor of what is involved. Other techniques are discussed in general terms and a table is presented that compares most of the known methods of high-resolution spectrum estimation.

Chapter 10 introduces two artificial intelligence methods for automation or partial automation of the signal identification process. The chapter is not intended to provide a comprehensive treatise on the subject, but instead to provide a few techniques as an introduction to the subject, providing an existence proof for such techniques.

The process of conducting electronic warfare in any real situation is extremely complicated, and there are never enough resources to thoroughly cover all targets. Chapter 11 contains an introduction to methods for resource allocation and techniques for evaluating system performance when there are limited resources.

This book is not intended to be a textbook on target acquisition for electronic warfare, but rather a reference that can be consulted on specific topics as they arise in system design. It can also serve as a reference for short courses on the subject.

I would like to acknowledge two colleagues at I2WD. The efforts of John Kosinski who read the manuscript in detail are appreciated. I would also like to note the actions of Maria Wright to get the timely security release.

Chapter 1

Introduction

Electronic warfare (EW) is one of the tools available to force commanders conducting military operations. It is similar to artillery fire in the sense that it is an indirect method for application of its effects. It is *indirect* because direct line of site with a target is not required for its use. All that is required is that the target be within radio line of sight, the range of which is often much longer than visual line of sight.

The direct effects of the application of EW are temporary—they go away when the EW energy is removed. This is not to say that the overall effects are temporary. Indeed, the application of EW techniques and methods can have significant impact on the outcome of hostile activities.

This chapter serves as a general introduction to EW and its main components. The remainder of the book focuses on a specific aspect of communication EW that is important enough to deserve such extensive coverage.

1.1 Electronic Warfare

The appellation EW applies to any attempt to conduct adverse actions on electronic equipment. It consists of three main components: *electronic support* (ES), *electronic attack* (EA), and *electronic protect* (EP) [1]. The contents of this book apply primarily to one of the components of ES: target detection.

ES is a supporting function for EA. It consists of extracting information about relevant signals to be attacked, so that the attack will be more effective. The four main functions performed for ES are all passive and are:

- Searching for signals;
- Intercepting and categorizing *signals of interest* (SOI);
- Direction finding/emitter location;

- Analysis of the ensuing information.

This information consists of technical information, such as frequency and modulation, as well as operational information, such as enemy intent and high-value target determination. Herein, communication ES will be the focus.

EA refers to attacking an electronic device in some fashion short of physical destruction. It generally refers to emitting energy into receiving components so that such receivers will not correctly decode an intended signal. Such signals can be communication, radar, or telemetry signals, or any other radiated energy, man-made or natural. The principal EA activity is active, which is jamming.

EP consists of those activities conducted to prevent adversaries from being effective at EA and ES of our own electronic devices, and is, in a sense, the counterpart to EA and ES. EP can be either active or passive. Examples of passive EP are:

- Physical siting of the EW system;
- Shielding of the electronic emissions in the direction of hostile EW systems;
- *Emission control* (EMCON), which controls when friendly communications occur;
- Directional antennas that transmit and receive signals in perferred direcions;
- Frequency management;
- Deception.

Examples of active EP are:

- Encryption;
- Low probability of intercept (LPI) modulations;
- *Antijam* (AJ) communications.

Nothing more will be said of EP herein except to the extent that the EA and ES topics discussed can be used for friendly EP.

1.2 Communications and EW

Communication is the exchange of information between two or more entities. It could be physical, as in the case of a transmitter and receiver in separate locations, or it can be temporal, as in the storage of informaiton on a tape recorder, for
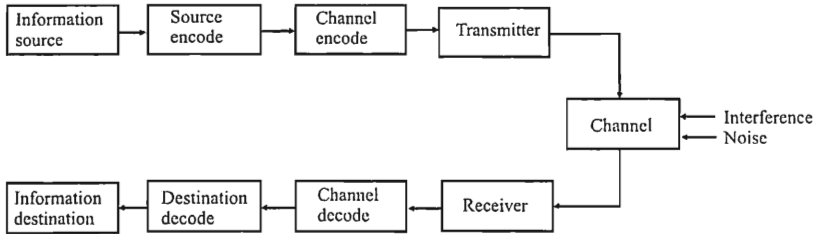
Figure 1.1 Communication channel model.

example. The communication channel of concern here is illustrated in Figure 1.1 [2–7].

The source generates information that it wishes to transfer to the destination. That source information is encoded for the purpose of removing redundancy in the baseband data. The source-encoded data is subsequently sent to the channel encoder. The purpose of this function is to add redundancy to the data to combat channel noise. The two encoding operations are shown as separate functions because, typically, they are different physical functions and the techniques and rationale for doing the two are different. The block labeled *transmitter* is the part that modulates the baseband data onto a carrier, amplifies the resultant signal, and sends the carrier out onto the channel. Not much will be contained herein on those functions, as they are covered extensively in other books. The channel is the physical medium over which the information is exchanged. In the case of a transmitter and receiver, the channel would normally be the open ether between them. In the case of storage of information for later retrieval, it would be the storage device. The channel is typically perturbed with noise and interference, the latter of which could be intentional or unintentional. At the destination, the receiver takes in the signal from the channel, converts that signal to a lower frequency, and demodulates the results. This baseband data is sent to the channel decoder, where the channel coding is removed. Next, the source coding is removed in an attempt to recreate as accurate a replication of the source information as possible. If there is no noise or interference, then this replication is perfect, with no errors. More typically, however, the replication is an approximation to the original data.

A countercommunication EW system, or *jammer* for short, creates intentional interference for the channel. This interference could be in the form of additional noise or one or more tones that may or may not be modulated. The purpose of this interference is, of course, to make it more difficult for the communication system to exchange information [8].

Although there are many types of channels, the one of primary interest herein is the *additive white Gaussian noise* (AWGN) channel. In such a channel the noise present, generated by thermal and galactic sources, is added to the signal. That is the case whether the noise is background noise or jammer noise [9–12].

One type of channel that is particularly important for wireless communications, the type of communications of interest here, is the *fading channel* [13–16]. There are several sources of fading, but probably the one most relevant for wireless communications is multipath reflection. As a transmitter or receiver moves, or obstacles enter the path between them, the signal strength of multipath reflections changes, thus causing fading. Although fading is important for analysis of communication effectiveness, as well as analysis of jamming effectiveness in fading situations, nothing more will be mentioned here.

1.3 Signal Detection

The first step in signal processing in communication EW systems is to determine the presence of the signal. Herein that is known as the signal detection problem. To determine if a detected signal is a signal of interest, further processing of the signal is frequently required to determine or estimate parameters associated with the signal.

Detecting the presence of a signal is a different problem from estimating the value of some parameter. Detecting presence is a binary problem, or at least a problem of limited dimension—detecting the presence of one of m types of signals, for example. The variate takes on only certain values and there is a finite number of them. In parameter estimation, on the other hand, the variate can, in general, take on any value from a range of values—an uncountably infinite number of possibilities.

The target set for communication EW systems consists of all types of communication systems of interest to the users of the EW system. It is frequently not known ahead of time what all those signals will be, so the signals arriving at the receiver belonging to the EW system likely have an unknown form until they are processed further [17]. Known-signal searching is the most applicable approximation for searching for signals in communication EW systems because the processing to make the determination if the signal is present is usually performed prior to demodulating the signal. For most modulating schemes, there is a substantial carrier component to the signal. That carrier originates from an oscillator at the transmitter that generates sinusoidal waves in most cases of interest here. Therefore, the signal being sought is a sinusoidal signal with some form of modulation on it. The form of the carrier is therefore known.

Typically, in the spectrum of interest to an EW system, there will be several signals present. These signals may be close together—indeed they can be in the same frequency channel. This leads to the question of how well two (or more) signals can be separated when they are close together and what the impacts are of signal amplitude on this resolvability. Chapter 5 presents what are more traditional methods of signal detection, couched in the terminology of spectral estimation. *Spectral estimation* is determining (estimating) the parameters of a signal such as its frequency and amplitudes (or power levels). These techniques typically have low resolution compared with some other techniques.

Reliably detecting the presence of a signal at a frequency when little or nothing is known about that signal is a difficult and challenging problem. When the radio frequency (RF) environment is composed of many communication networks, there can be many signals to sort through, especially when many of these signals belong to friendly forces and are therefore normally of little interest. In such environments the resolution provided by simple search schemes may not be fine enough. Techniques have been devised to increase this resolution, and some of these techniques are presented in Chapter 8.

Cochannel interference [18] is also a significant problem in crowded environments. There is only so much RF spectrum available to all sides in an adversarial relationship. What is available must be shared, and between hostile parties this sharing is not cooperative. The same frequencies are used by all sides. Cochannel interference results where two (or more) signals use the same frequency channel at the same time. This certainly affects the communication systems if a comunication receiver hears both transmitters. It also significantly affects the ability of an EW system to intercept the signals.

The most straightforward way to do spectral analysis of a signal is to present the signal to a bank of narrowband filters, either analog or digital. This configuration produces an approximation to the *short-term Fourier transform* (STFT) of the signal. The characteristics of such an approach are limited only by the practicality of its implementation.

The resolution of such an approach is limited to the number of practical filters, which, by implication, determines the bandwidth of each filter. In addition, the time the signal must be present at the input to the filter bank is proportional to the bandwidth of the filters—for finer resolution (narrow bandwidth) the signal must be present for a longer duration, and vice versa, for low-frequency resolution, the signal need only be present for a short period.

Presenting the signal to the filters for only a finite time creates a window effect in the frequency domain. Limiting the sample time to T seconds places a rectangular window around the time function. In the frequency domain this creates the convolution of the signal spectrum $S(f)$ with the spectrum of the sample function. This, in turn, creates false responses in the frequency domain. Windows
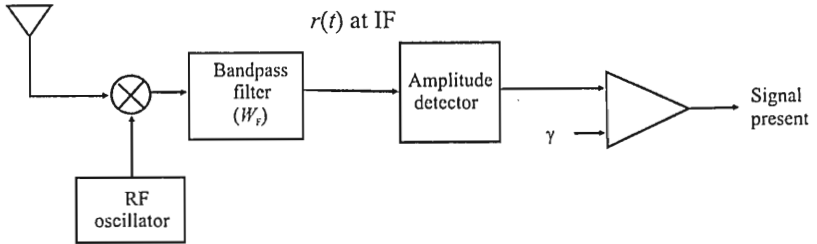
Figure 1.2 Time domain signal detector.

other than the rectangular pulse are possible and ameliorate this effect to a certain degree. Windowing a time series always produces some distortion in the frequency spectrum. Energy from the channel being examined appears in other channels. When this occurs, it is called *leakage*. This leakage can appear in close channels or channels far away. Windowing also always broadens the peak response width. This is called *smearing*. Windows are discussed in Chapter 5.

In general, it is possible to detect signals in the time domain or the frequency domain. Simplistically, time domain detection is possible by applying the received signal to an AM detector as illustrated in Figure 1.2. Irrespective of the type of modulation the signal has (assuming a signal is present), it has some amplitude. The triangle-shaped element in Figure 1.2 is a comparator with the transfer characteristic

$$\text{output} = \begin{cases} 1, & \text{input} \geq \gamma \\ 0, & \text{input} < \gamma \end{cases} \tag{1.1}$$

Therefore, if the output of the detector exceeds the threshold given by γ, then the output of the comparator will be a logical 1 and signal presence is indicated. Unfortunately, this type of detection is normally not optimum in that better *signal-to-noise ratio* (SNR) performance is possible with more sophisticated techniques.

To detect signals in the frequency domain, the technique illustrated in Figure 1.3 is intuitive. In this case, the signal is converted into the frequency domain by computing the *Fourier transform* (FT). Each resulting frequency bin is examined simultaneously for energy above some threshold γ. The last module is an OR gate whose output is a logical 1 if any one or more of its inputs is a logical 1. If any bin contains sufficient energy, its corresponding comparator output will be a logical 1, as will the OR output, and a signal present is declared. The FT essentially forms parallel filter banks and is usually implemented digitally—that is, samples of the signal are obtained and the *discrete Fourier transform* (DFT) is used to compute the FT.
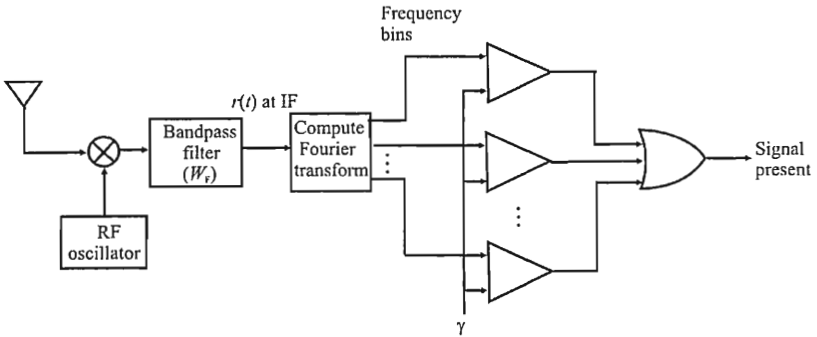
Figure 1.3 Frequency domain signal detector.

It can be shown that a Fourier receiver, that is, a receiver in which the FT of the time series followed by incoherent integration is used for signal detection, is the optimum configuration for a significant set of signals, including many communication signals of interest herein [19]. This set consists of narrowband signals with random phase, duration longer than the reciprocal of the bandwidth, with constant amplitude, and unknown frequency. This receiver is optimum according to the Bayes criterion (discussed later).

Spectral estimation for use in EW systems is for the purpose of determining the distribution in frequency of the power of signals that are random processes. This estimate is useful for determining the presence of signals, their power content, and their frequencies. Spectral estimation may be classified as either nonparametric or parametric. The nonparametric approaches such as the periodogram, average periodogram, *Blackman-Tukey* (BT), and *minimum variance spectral estimation* (MVSE) require no assumptions about the data other than *wide-sense stationarity* (wss, defined in Chapter 2). The parametric spectral estimators are based on rational transfer functions of *linear, time-invariant* (LTI) filters or time series models of the data. We assume that these filters are driven by a Gaussian noise source, and the parameters of the filters are adjusted so that the observed time series emerges at the output. The advantage of the parametric spectral estimator is that when applicable they yield more accurate spectral estimates. The drawback is that it is critical that the model is correct, and any departure from the model will result in a bias error in the spectral estimate and hence the parameters sought. Compared with nonparametric methods, parametric methods have restricted applicable areas; however, if applicable, usually more accurate estimates result.

Table 1.1 lists some of the ways the *power spectral density* (psd) is estimated. More will be presented later. The two classical methods listed in Table 1.1 are the fastest and most efficient, as indicated by the $N \log_2 N$ complexity. The psd

Table 1.1 Spectral Estimation Methods

Class	Model	Method	Complexity	Resolution Control
Classical	Preselected sines	Periodogram	$N \log_2 N$	Window
Classical	None	Correlogram	$N \log_2 N$	Window
Parametric	Autoregressive	Yule-Walker	N^2	AR order
Parametric	Autoregressive	Burg	N^2	AR order

estimates are proportional to the power in whatever sinusoids are present in the data stream. The main disadvantage is the distortion caused by windowing the data. Windowing the data naturally occurs when the input data is limited to a finite time window, which is always necessary. Weighting of the input data can reduce the effect of sidelobes at the expense of decreasing the spectral resolution. Also, the statistics associated with the input data tend to vary with subsequent sample windows (stationarity assumption is violated). This statistical instability can be minimized by ensemble averaging; however, this occurs with further reduction of the spectral resolution. Generally, the resolution of the classical methods can be no better than the reciprocal of the total time window length, independent of the characteristics of the data. The methods are therefore not well suited for short data records.

Alternative methods, such as using parametric models, may improve or maintain high resolution without sacrificing much statistical stability. The transformation relationship between the autocorrelation function and the psd by Fourier transformation is a nonparametric description of a second-order statistic. A parametric description may be devised by assuming a time series model of a random process. The psd of such a model will then be dependent on the model parameters and not on the original autocorrelation data.

The parametric approach to spectral estimation involves three steps: select a parametric time series model to represent the measured data, either *autoregressive* (AR), *moving average* (MA), or combined *autoregressive moving average* (ARMA); estimate the parameters of the model; and finally insert the parameters into the theoretical psd expression for that model. The characteristics of these methods are:

- Parametric methods have a good frequency resolution even on small data sets (this is their main advantage).
- Spurious peaks are generated if a large model order is selected relative to the number of data samples, so model order selection is very important.

The nonlinear relationship of spectral power to actual power exhibited by most parametric methods makes absolute power measurements meaningless, so only relative power measurements are possible.

There are several factors that need to be taken into consideration in detection theory. Roberts [20] points out the following:

- If available, increased observation time will almost always be useful, assuming that the signal persists.
- Another point of view, another location, even another system might improve the situation. For example, in antisubmarine warfare, is it better to listen for submarines on dry land, from hydrophones near shore, or from hydrophones at sea with telephone lines to shore, or is it better to fly airplanes or sail ships to the submarines' operating areas and provide ways for them to listen?
- Signal processing exists to exploit differences in the nature of signals and noise. If differences exist, signal processing can transform both to some appropriate domain (e.g., frequency domain) where they are orthogonal and where they may be separated more distinctly. But signal processing cannot create what is not there; it cannot increase the information content. (This is a manifestation of the data processing theorem from the field of information theory [21].)
- To utilize signal processing, engineering knowledge of signals and noise is necessary. Noise measurements can come from ordinary engineering experiments, but if the signal source is uncooperative (such as is the case for communication EW systems) then information relating to it will need to come from uncontrolled available measurements and subsequent deductions and calculations.
- With consideration of the consequence, it is important to evaluate how good or bad the decision will be, once made, in some average sense. Such evaluation might, for example, indicate the need for more equipment, or operators, or training.

1.4 Signal Searching

Typically a communication EW system operates in an RF environment that is hostile and noncooperative. Sometimes some of the parameters of the signals of

interest, such as frequency and modulation type, are available, but often not entirely. These parameters must be detemined when they are unavailable. Searching for signals of interest can be accomplished in several ways, such as rapidly tuning a receiver through the spectrum. When little or nothing is known about the target environment, the problem is different from when most of the targets have been identified. The former requires searching over a region of the spectrum in a continuous and contiguous fashion, looking for energy corresponding to unknown (a priori) frequencies. In the latter, the frequency locations of target signals are generally known—the problem is to find out if such signals are either still at that location or have moved, presumably because it is known that they stopped transmitting for some period. These search problems are discussed in Chapter 3.

1.5 Notation

Herein, scalars are denoted as italic plain text letters, while small bold letters denote vectors, and large bold letters denote matrices. Thus, θ is a scalar while $\boldsymbol{\theta}$ is a vector and $\boldsymbol{\Theta}$ is a matrix. A vector $\boldsymbol{\theta}$ normally consists of scalars but could also consist of other vectors or matrices. Thus, normally, $\boldsymbol{\theta} = \begin{bmatrix} \theta_1 & \theta_2 & \cdots & \theta_n \end{bmatrix}^{\mathrm{T}}$.

At various points throughout the presentation, sections are cordoned off with the title "Scholium." This refers to material that is important and interesting, but is not central to the development being presented.

The end of a section describing a singular thought, a theorem and its proof, for example, is delineated by the symbol ■ so that the reader knows the preceding discussion has been concluded.

1.6 Concluding Remarks

The literature on signal detection and estimation is vast. Perhaps one of the most cited references is the series of books by Van Trees [22–24]. Another early source is due to Helstrom [25]. An excellent overview of the history of the development of the field is given in [26]. A very readable introduction to the theory of signal detection in noise for both communication and radar signals is presented in [27]. Root provides an introduction to the theory of signal detection in noise [28].

The topics covered in this work are certainly focused on a narrow field— communication EW system theory. It will be established, however, that there are fundamental principles involved with this topic, and the operation of these systems is based on sound first principles.

References

[1]   Poisel, R. A., *Introduction to Communication Electronic Warfare Systems*, Norwood, MA: Artech House, 2002, p. 3.

[2]   Ziemer, R. E., and W. H. Tranter, *Principles of Communications Systems Modulation and Noise, 5th Ed.*, New York: John Wiley & Sons, 2002.

[3]   Ziemer, R. E., and R. L. Peterson, *Introduction to Digital Communication, 2nd Ed.*, Upper Saddle River, NJ: Prentice Hall, 2001.

[4]   Gagliardi, R. M., *Introduction to Communications Engineering, 2nd Ed.*, New York: John Wiley & Sons, 1988.

[5]   Proakis, J. G., *Digital Communications, 3rd Ed.*, New York: McGraw-Hill, 1995.

[6]   Simon, M. K., S. M. Hinedi, and W. C. Lindsey, *Digital Communication Techniques: Signal Design and Detection*, Upper Saddle River, NJ: Prentice Hall, 1995.

[7]   Das, J., S. K. Mullick, and P. K. Chatterlee, *Principles of Digital Communication: Signal Representation, Detection, Estimation, and Information Coding*, New York: John Wiley & Sons, 1986.

[8]   Poisel, R. A., *Introduction to Communication Electronic Warfare Systems*, Norwood, MA: Artech House, 2002, Ch. 7.

[9]   Root, W. L., "Communication Through Unspecified Additive Noise," *Information and Control*, Vol. 4, 1961, pp. 15–29.

[10] Gagliardi, R. M., *Introduction to Communications Engineering, 2nd Ed.*, New York: John Wiley & Sons, 1988, pp. 148–154.

[11] Ziemer, R. E., and W. H. Tranter, *Principles of Communications: Systems, Modulation, and Noise, 5th Ed.*, New York: John Wiley & Sons, 2002, pp. 270–275.

[12] *Reference Data for Radio Engineers, 6th Ed.*, Indianapolis: Howard W. Sams & Co., Inc., 1975, Ch. 29.

[13] Simon, M. K., and M. S. Alouini, *Digital Communication over Fading Channels*, New York: John Wiley & Sons, Inc., 2000.

[14] McGuffin, B. F., "Distributed Jammer Performance in Rayleigh Fading," *Proceedings IEEE MILCOM 2002*, Anaheim, CA, October 7–11, 2002.

[15] Ericson, T., "A Gaussian Channel with Slow Fading," *IEEE Transactions on Information Theory*, May 1970, pp. 353–355.

[16] Biglieri, E., J. Proakis, and S. Shamai, "Fading Channels: Information-Theoretic and Communication Aspects," *IEEE Transactions on Information Theory*, Vol. 44, No. 6, October 1998, pp. 2619–2692.

[17] McNicol, D., *A Primer of Signal Detection Theory*, Sydney: George Australian Publishing Company, 1972.

[18] Spooner, C. M., "Classification of Co-channel Communication Signals Using Cyclic Cumulants," *Proceedings of ASILOMAR-29*, 1996, pp. 531–536.

[19] Williams, J. R., and G. G. Ricker, "Signal Detectability Performance of Optimum Fourier Receivers," *IEEE Transactions on Audio and Electroacoustics*, Vol. AU-20, No. 4, October 1972, pp. 264–270.

[20] Roberts, L. R., *Signal Processing Techniques*, Anaheim, CA: Interstate Electronics Corporation, 1981, p. 7-2.

[21] Gallager, R. G., *Information Theory and Reliable Communication*, New York: John Wiley & Sons, 1968, p. 80.

[22]   Van Trees, H. L., *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory*, New York: John Wiley & Sons, 1968.
[23]   Van Trees, H. L., *Detection, Estimation, and Modulation Theory, Part II: Nonlinear Modulation Theory*, New York: John Wiley & Sons, 1971.
[24]   Van Trees, H. L., *Detection, Estimation, and Modulation Theory, Part III: Radar-Sonar Signal Processing and Gaussian Signals in Noise*, New York: John Wiley & Sons, 1971.
[25]   Helstrom, C. W., *Statistical Theory of Signal Detection*, New York: Pergamon Press, 1960.
[26]   Kailath, T., and H. V. Poor, "Detection of Stochastic Processes," *IEEE Transactions on Information Theory,* Vol. 44, No. 6, October 1998, pp. 2230–2259.
[27]   DiFranco, J. V., and W. L. Rubin, *Radar Detection,* Dedham, MA: Artech House, 1980, Chs. 5–8.
[28]   Root, W. L., "An Introduction to the Theory of the Detection of Signals in Noise," *Proceedings of the IEEE,* Vol. 58, No. 5, May 1970, pp. 610–623.

Chapter 2

Deterministic and Stochastic Processes

A signal or function is *deterministic* if its future value can be predicted when the current value is known. A signal or function is random, or *stochastic*, if its future value can only be described statistically; the precise future value cannot be determined even if the current value is known.

Communication signals in noise can be considered as exemplars of *stochastic processes*. Sometimes the signals themselves, without the noise, are modeled as stochastic processes. An example of this is general searching (defined in Chapter 3), where little or nothing is known about the signals received except, perhaps, that there is energy at some frequency. Such signals can be modeled as random, stochastic processes. In general, stochastic processes can be scalar- or vector-valued quantities that vary with time. Thus, if two samples of the process were measured at the same time, the measured value would likely be different.

Spectral analysis is a commonly used tool for analysis of signals, to include communication signals of interest for EW systems. The mathematical foundations of spectral analysis date back to the French mathematician Jean-Baptiste Fourier (1768–1830) who discovered the relationship between a deterministic time domain function $s(t)$ and its unique representation in the frequency domain, commonly referred to as its *spectrum*, or *Fourier spectrum* in honor of the mathematician.

2.1 The Fourier Transform

If $s(t)$ is a signal in the time domain then its FT, denoted by $S(f)$, in the frequency domain is given by

$$S(f) = \int_{-\infty}^{\infty} s(t)e^{-j2\pi ft} dt \qquad (2.1)$$

Similarly, if $S(f)$ is the FT of a signal $s(t)$, then $s(t)$ is given by

$$s(t) = \int_{-\infty}^{\infty} S(f)e^{j2\pi f t} \, df \qquad (2.2)$$

This relationship is denoted by

$$s(t) \leftrightarrow S(f) \qquad (2.3)$$

In general, the FT is a complex function, with a real part denoted by $\mathrm{Re}\{S(f)\}$ and an imaginary part denoted by $\mathrm{Im}\{S(f)\}$. Equivalently, $S(f)$ has a magnitude (argument)

$$|S(f)| = \sqrt{\mathrm{Re}^2\{S(f)\} + \mathrm{Im}^2\{S(f)\}} \qquad (2.4)$$

and phase angle

$$\phi(f) = \tan^{-1} \left[\frac{\mathrm{Im}\{S(f)\}}{\mathrm{Re}\{S(f)\}} \right] \qquad (2.5)$$

It is usually the magnitude of the FT that is called the spectrum of the signal. As will be discussed at length later in this section, the FT does not exist for all functions.

Scholium

Notice that

- The value of $S(0)$ is equal to the area under the graph of $s(t)$:

$$S(0) = \int_{-\infty}^{\infty} s(t) \, dt \qquad (2.6)$$

- Similarly, the value of $s(0)$ is equal to the area under $S(f)$:

$$s(0) = \int_{-\infty}^{\infty} S(f) \, df \qquad (2.7)$$

These are general results and are useful for checking.
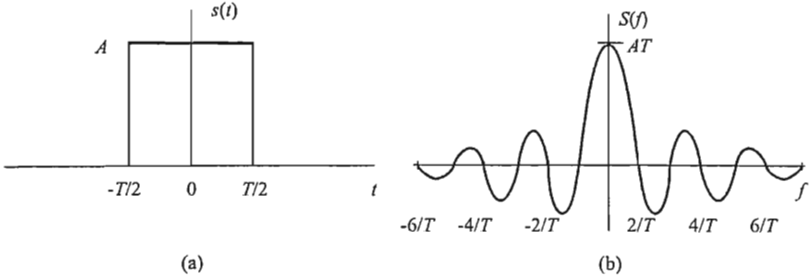
Figure 2.1 (a) Rectangular pulse and (b) its FT.

2.1.1 Important Fourier Transforms

Several examples of time waveforms and their FTs that are common in EW systems analysis and design are presented in this section. In addition, some of the more important properties of FTs and their associated time waveforms are presented. More complete lists can be found in [1].

2.1.1.1 The Rectangular Pulse/Time Window

The *rectangular pulse, boxcar,* or *time window* function is

$$s(t) = A\Pi_T(t) = \begin{cases} A, & |t| < T/2 \\ 0, & \text{otherwise} \end{cases} \tag{2.8}$$

which is illustrated in Figure 2.1(a). The FT is given by

$$S(f) = \int_{-T/2}^{T/2} Ae^{-j2\pi ft}dt = A\left[\frac{e^{-j2\pi ft}}{-j2\pi f}\right]_{-T/2}^{T/2}$$

$$= A\left[\frac{e^{-j2\pi f\frac{T}{2}} - e^{j2\pi f\frac{T}{2}}}{-j2\pi f}\right] = A\frac{\sin(\pi fT)}{\pi f}$$

$$= AT\text{sinc}(fT) \tag{2.9}$$
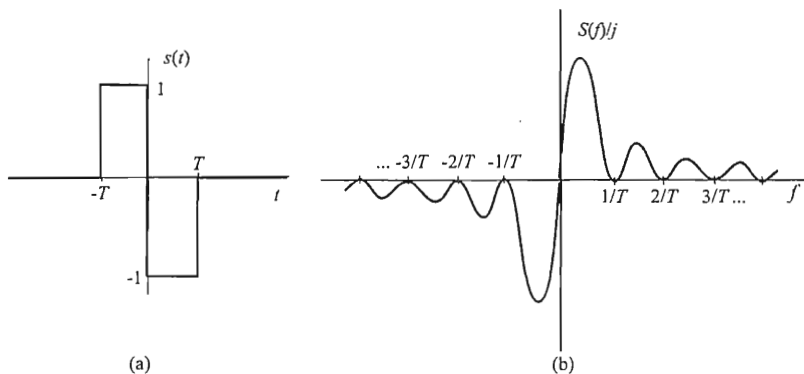
Figure 2.2 (a) Two rectangular pulses and (b) their FT.

where

$$\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x} \tag{2.10}$$

This FT is shown in Figure 2.1(b).

Scholium

Notice that the zeros of $S(f)$ are at integer multiples of $1/T$. Thus, the wider the pulse in the time domain, the narrower the transform in the frequency domain.

2.1.1.2 Double Pulses

The signal shown in Figure 2.2(a) can be written as

$$s(t) = \begin{cases} 1, & -T < t < 0 \\ -1, & 0 < t < T \\ 0, & \text{otherwise} \end{cases} \tag{2.11}$$

as well as

$$s_d(t) = \Pi_T\left(t + \frac{T}{2}\right) - \Pi_T\left(t - \frac{T}{2}\right) \tag{2.12}$$

Using linearity, the shifting theorem, and the FT of the pulse (2.9),

$$S(f) = Te^{j2\pi f \frac{T}{2}} \operatorname{sinc}(fT) - Te^{-j2\pi f \frac{T}{2}} \operatorname{sinc}(fT)$$
$$= \frac{2j}{\pi f} \sin^2(\pi fT) \tag{2.13}$$

This FT is shown in Figure 2.2(b).

■

2.1.1.3 Triangular Pulse

The triangular pulse shown in Figure 2.3(a) can be written as

$$s(t) = \begin{cases} \dfrac{T+t}{T}, & -T < t \leq 0 \\ \dfrac{T-t}{T}, & 0 < t < T \\ 0, & \text{otherwise} \end{cases} \tag{2.14}$$

This is $1/T$ multiplied by the integral of the double pulses. Since

$$\text{when } s(t) \leftrightarrow S(f)$$
$$\text{then } \int_{-\infty}^{t} s(\tau)d\tau \leftrightarrow \frac{1}{2\pi jf} S(f) + S(0)\delta(0) \tag{2.15}$$

Now the area under the double pulse function is zero:
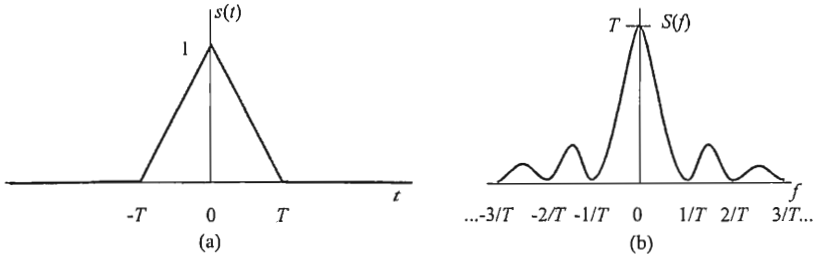
Figure 2.3 (a) Triangular pulse and (b) its FT.

$$\int_{-\infty}^{\infty} s_d(t) = 0 \tag{2.16}$$

Using (2.6) and (2.13), it can be concluded that

$$\begin{aligned}
S(f) &= \frac{1}{T}\frac{1}{j2\pi f}\frac{2j}{\pi f}\sin^2(\pi fT) + 0\delta(0) \\
&= T\frac{1}{(\pi fT)^2}\sin^2(\pi fT) \\
&= T\operatorname{sinc}^2(fT)
\end{aligned} \tag{2.17}$$

The FT is shown in Figure 2.3(b).

∎

2.1.1.4 The Exponential Pulse

The exponential pulse illustrated in Figure 2.4(a) is given by

$$s(t) = e^{-\alpha t}u(t) \tag{2.18}$$

where $u(t)$ is the unit step function. The FT of the exponential pulse is given by

$$\begin{aligned}
S(f) &= \int_{-\infty}^{\infty} e^{-\alpha t}u(t)e^{-j2\pi ft}dt = \int_{0}^{\infty} e^{-(\alpha+j2\pi f)t}dt \\
&= \frac{1}{\alpha + j2\pi f} \\
&= \frac{1}{\sqrt{\alpha^2 + 4\pi^2 f^2}}e^{-j\tan^{-1}\left(\frac{2\pi f}{\alpha}\right)}
\end{aligned} \tag{2.19}$$

The magnitude of this function is plotted in Figure 2.4(b) and the phase is shown in Figure 2.4(c) for $\alpha = 1$. The magnitude exhibits a sharp peak at $f = 0$, while the phase function transitions from $\pi/2$ to $-\pi/2$ radians quickly in the same region. With larger values of α the time function falls faster and the magnitude of the spectrum is less sharp, while the phase still changes from $\pi/2$ to $-\pi/2$, but slower.

(a)

(b)

(c)
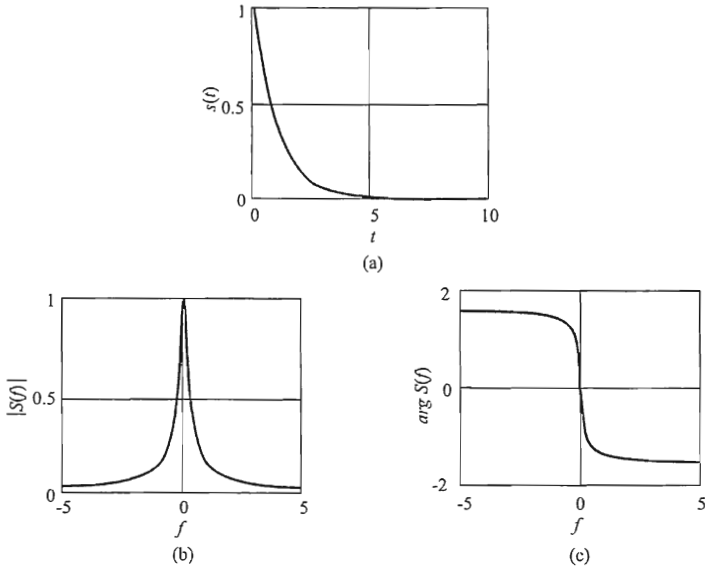
Figure 2.4 (a) An exponential pulse and (b) the magnitude and (c) phase of its FT when $\alpha = 1$.

■

2.1.1.5 The Gaussian Function

The Gaussian function is given by

$$s(t) = e^{-\alpha t^2} \tag{2.20}$$

which is illustrated in Figure 2.5(a). Calculation of the FT yields

$$S(f) = \int_{-\infty}^{\infty} e^{-\alpha t^2} e^{-j2\pi ft} dt = \int_{-\infty}^{\infty} e^{-\alpha\left(t^2 + j\frac{2\pi ft}{\alpha}\right)} dt \tag{2.21}$$

Adding the necessary term in the exponent to complete the square yields

$$S(f) = \int_{-\infty}^{\infty} e^{-\alpha\left(t^2 + j\frac{2\pi ft}{\alpha} - \frac{\pi^2 f^2}{\alpha^2}\right)} e^{-\alpha\frac{\pi^2 f^2}{\alpha^2}} dt = e^{-\frac{\pi^2 f^2}{\alpha}} \int_{-\infty}^{\infty} e^{-\alpha\left(t + j\frac{\pi f}{\alpha}\right)^2} dt \tag{2.22}$$
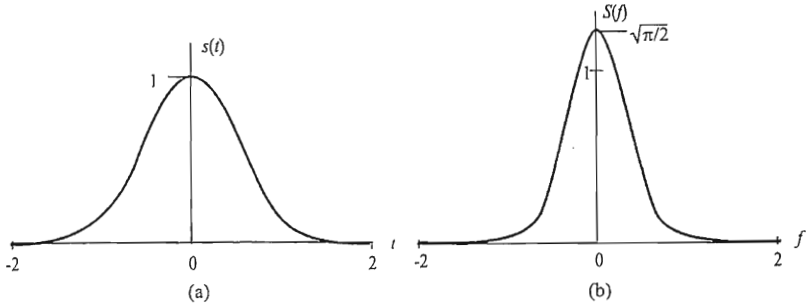
Figure 2.5 (a) The Gaussian function and (b) its FT when $\alpha = 2$.

Make the variable substitution $u = t + j\pi f/\alpha$, then $dt = du$ and

$$S(f) = e^{-\frac{\pi^2 f^2}{\alpha}} \int_{-\infty + j\pi f/\alpha}^{\infty + j\pi f/\alpha} e^{-\alpha u^2} du \qquad (2.23)$$

which is a contour integral in the complex plane. The integral over the rectangular contour is

$$\int_{-\beta + j\pi f/\alpha}^{\beta + j\pi f/\alpha} e^{-\alpha u^2} du + \int_{\beta + j\pi f/\alpha}^{\beta R} e^{-\alpha u^2} du + \int_{\beta}^{-\beta} e^{-\alpha u^2} du + \int_{-\beta}^{-\beta + j\pi f/\alpha} e^{-\alpha u^2} du \qquad (2.24)$$

where in each integral, the straight-line path between the limits is taken. Since the integrand is analytic throughout the complex plane, Cauchy's theorem states that the integral over any closed contour is zero [2]. As β becomes large, the integrals along the lines $(\beta + j\pi f/\alpha, \beta)$ and $(-\beta, -\beta + j\pi f/\alpha)$ become small since the integrand falls off rapidly while the length of the contour stays fixed. Hence, in the limit as $\beta \rightarrow \infty$, these integrals vanish leaving

$$\int_{-\infty + j\pi f/\alpha}^{\infty + j\pi f/\alpha} e^{-\alpha u^2} du + \int_{\infty}^{-\infty} e^{-\alpha u^2} du = 0 \qquad (2.25)$$

or

$$\int_{-\infty+j^{\pi f}/\alpha}^{\infty+j^{\pi f}/\alpha} e^{-\alpha u^2}\, du = \int_{-\infty}^{\infty} e^{-\alpha u^2}\, du = \sqrt{\frac{\pi}{\alpha}} \tag{2.26}$$

Therefore,

$$S(f) = \sqrt{\frac{\pi}{\alpha}} e^{-\frac{\pi^2 f^2}{\alpha}} \tag{2.27}$$

which is shown in Figure 2.5(b). In general, as the width of the pulse in the time domain broadens, the width of that in the frequency domain narrows, and vice versa.

∎

Property: Shannon/Nyquist Sampling Theorem [3, 4]

Let $x(t)$ be a signal with FT $X(f)$ such that $X(f) = 0$ for $|f| \geq f_c$. Then

$$x(t) = \sum_{k=-\infty}^{\infty} x(kT_s)\, \mathrm{sinc}\left[2\pi f_c\left(t - kT_s\right)\right] \tag{2.28}$$

where $f_s = 2f_c$ and $T_s = 1/f_s$.

∎

Thus, the time function that is band limited to a finite frequency region has infinite extent in the time domain. Furthermore, that time function can be exactly reproduced with appropriately scaled sinc functions.

Property: Cosine Function

If

$$x(t) = A\cos(2\pi f_0 t) \tag{2.29}$$

then

$$X(f) = \frac{A}{2}\delta(f - f_0) + \frac{A}{2}\delta(f + f_0) \tag{2.30}$$

∎

The FT of a cosine function of infinite extent consists of two positive Dirac delta functions located at the positive and negative frequencies of the function.

Property: Sine Function

If

$$x(t) = A\sin(2\pi f_0 t) \tag{2.31}$$

then

$$X(f) = j\frac{A}{2}\delta(f + f_0) - j\frac{A}{2}\delta(f - f_0) \tag{2.32}$$

■

The FT of an infinite sine function consists of two Dirac delta functions located at the positive and negative frequencies of the function shifted in phase. The positive delta function is shifted $+\pi/2$ radians and the negative delta function is shifted $-\pi/2$ radians.

2.2 Deterministic Signals

As mentioned, a deterministic signal is known for all time given its value at a single time. Examples of deterministic signals are the cos( ) and sin( ) functions. Such functions are, of course, useful for system analysis. Many practical signals of interest in communications EW analysis and design are, however, stochastic or at least have some parameters associated with them that are stochastic. These signals will be considered in the next section.

Parseval's relationship says that if

$$s(t) \leftrightarrow S(f) \tag{2.33}$$

then the energy, $e(t)$, in the time domain is related to that in the frequency domain as

$$E_f = \int_{-\infty}^{\infty} |s(t)|^2 \, dt = \int_{-\infty}^{\infty} |S(f)|^2 \, df \tag{2.34}$$

This simply says that the total energy in a signal $s(t)$ is equal to the area under the square of the magnitude of its FT.[1] $|S(f)|^2$ is typically called the *energy density, or energy spectral density* function, and $|S(f)|^2 \, df$ describes the density of signal energy contained in the differential frequency band from f to $f + df$.

Scholium

The average power in electronic circuits is given by $P = i^2R$ or $P = v^2/R$. As is commonly assumed, when considering the analysis of signals it will here be assumed that this resistive value is unity, allowing the power to be expressed simply as $s^2(t) = v^2(t)$ or $i^2(t)$. This allows units of power to be expressed as the square of the signal amplitudes and the units of energy measured as volts²-second (amperes²-second).

∎

The average power, P_{avg}, over a time interval t_2 to t_1 is obtained by integrating $s^2(t)$ from t_1 to t_2 and then dividing the result by $T = t_2 - t_1$ or

$$P_{avg} = \frac{1}{T} \int_0^T s^2(t)dt \tag{2.35}$$

where T is the period of the signal. Total energy can thus be expressed in terms of power, $p(t)$,

$$E = \int_{-\infty}^{\infty} s^2(t)dt$$

---

[1] Recall the electrical energy storage elements (assuming zero initial energy):
  • Inductor
$v(t) = Ldi(t) / dt$
$$e(t) = \int_0^T v(t)i(t)dt = \int_0^T L\frac{di(t)}{dt}i(t)dt = L\int_0^T i(t)di(t) = \frac{1}{2}Li^2(T)$$
  • Capacitor
$i(t) = Cdv(t) / dt$
$$e(t) = \int_0^T v(t)i(t)dt = \int_0^T v(t)C\frac{dv(t)}{dt}dt = C\int_0^T v(t)dv(t) = \frac{1}{2}Cv^2(T)$$

$$= \int_{-\infty}^{\infty} p(t)dt \qquad (2.36)$$

In general, the quantities discussed here are complex and therefore have both real parts and imaginary parts, or amplitudes and phase functions.

2.2.1 Energy and Power in Deterministic Signals

2.2.1.1 Finite Energy Signals

All deterministic signals can be divided into finite energy signals or infinite energy signals. Most real signals exist for only a finite amount of time and, generally, have a limited frequency range of interest. Theoretically, no signal can have both finite duration and finite frequency extent, however.

The total energy in the signal $s(t)$ is given by (2.34). Signals such that $E_f < \infty$ are called *finite energy signals* and are also referred to as $\mathcal{L}^2$ functions. If $s(t)$ is a finite energy signal, then $\lim_{t \to \pm\infty} s(t) = 0$.

If $s(t)$ is passed through an ideal bandpass filter of bandwidth B, with transfer function $H(f)$ given by

$$H(f) = \begin{cases} 1, & |f| < \dfrac{B}{2} \\ 0, & \text{otherwise} \end{cases} \qquad (2.37)$$

producing $y(t)$ at the output, then the energy in $y(t)$ is given by

$$E_y = \int_{-\infty}^{\infty} |Y(f)|^2 \, df = \int_{-\infty}^{\infty} |S(f)H(f)|^2 \, df$$

$$= \int_{-B/2}^{B/2} |S(f)|^2 \, df \qquad (2.38)$$

2.2.1.2 Finite Power Signals

Not all signals have finite energy, that is, there is an infinite amount of energy in them. Typical examples are the sin and cos functions, as well as impulse functions.

For this type of function, the power is the important parameter. The *average power* in $s(t)$ is given by

$$P_s = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} |s(t)|^2 dt \qquad (2.39)$$

and functions for which $P_f < \infty$ are called *finite power* functions.

2.3 Stochastic Processes

This section introduces the basic notions behind stochastic processes.

2.3.1 Ensembles

Consider the hypothetical communication system shown in Figure 2.6. All of the transmitters are perfectly collocated and are identical. They all transmit the exact same signal to the single receiver at exactly the same time. Noise is added to each of these signals, which causes the signals arriving at the receiver to be different, even though the transmitted signals were the same. At any specific instant in time, say $t = 5$, the amplitude of any given one of these received signals will have some value, but this amplitude is unpredictable. This collection of signals is called an *ensemble*. The statistical descriptions for stochastic processes always refer to such an ensemble of waveforms. Any one of these waveforms is referred to as a *realization* of the stochastic process.

A specific example of an ensemble of time functions consisting of four elements is shown in Figure 2.7. The statistics of this random process are calculated across ensemble elements. Thus the average value at $t = t_1$ is

$$s_{avg}(t_1) = \frac{1}{4}[s_1(t_1) + s_2(t_1) + s_3(t_1) + s_4(t_1)] = 2 \qquad (2.40)$$

An ensemble represents the complete and exhaustive set of possible waveforms that can be received. The value actually received in one experiment with this system at a particular time, say x_5, is thus a *random variable* (rv) with, for example, an associated *probability density function* (pdf), *cumulative probability distribution function*, mean value, variance, and standard deviation. Here, the pdf will be denoted by $p(s)$, and when a particular time is important, say $t = 5$, then $p_5(s)$. The *cumulative probability density function*, cdf, will be denoted
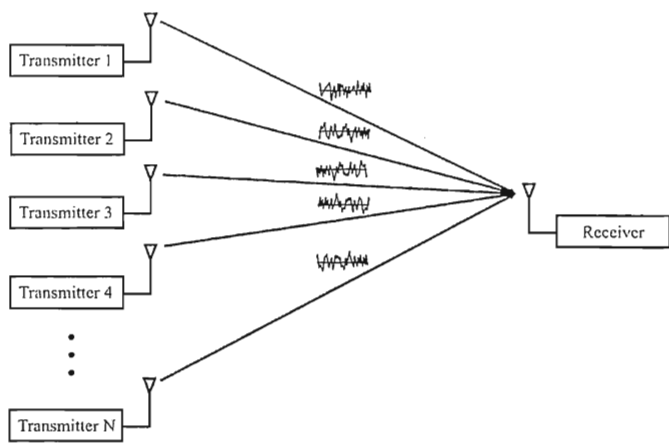
Figure 2.6 Example of an ensemble of a statistical process.
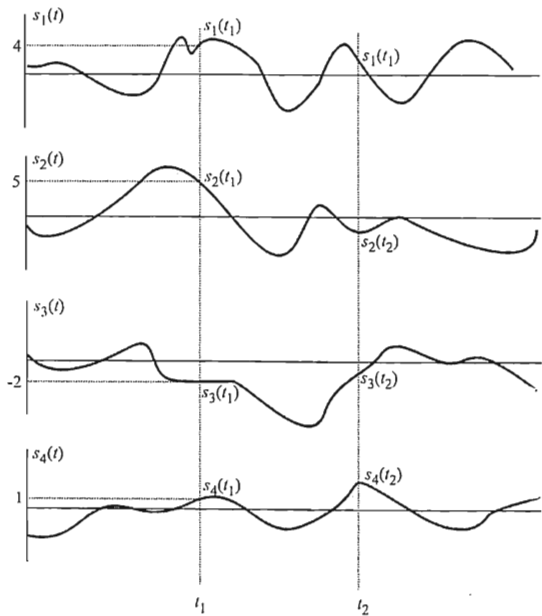
Figure 2.7 Ensemble consisting of four time functions.

$P(s)$ with similar additions when time is important. The mean value of s_5 is given by

$$\mathcal{E}\{s_s\} = \int_{-\infty}^{\infty} s p_s(s) ds \qquad (2.41)$$

where $\mathcal{E}\{.\}$ refers to the *ensemble expected value function*. Likewise, the variance of s_5 is given by

$$\mathcal{E}\{(s - \mathcal{E}\{s_s\})^2\} = \int_{-\infty}^{\infty} (s - \mathcal{E}\{s_s\})^2 p_s(s) ds$$

$$= \int_{-\infty}^{\infty} s^2 p_s(s) ds - \left[\int_{-\infty}^{\infty} s p_s(s) ds \right]^2$$

$$= \mathcal{E}\{s_s^2\} - \mathcal{E}^2\{s_s\} \qquad (2.42)$$

The mean ($\mathcal{E}\{s_5\}$) and variance ($\mathcal{E}\{s_5^2\}$) are two examples of the *moments* of the rv. The pth moment of an rv is calculated at $t = 5$ as

$$\mathcal{E}\{s_5^p\} = \int_{-\infty}^{\infty} s^p p_s(s) ds \qquad (2.43)$$

2.3.2 Power Spectral Densities

The FT of a time signal is a representation of the frequency content of the signal. There are various ways of using this transform to ascertain this frequency content. Each of these approaches has strengths and weaknesses. The ones examined in detail here are based on parametric models of the underlying stochastic time series that emerges from digitizing the predetection signal from a receiver [5]. Power spectral densities and autocorrelation functions for statistical processes are related by the FT when the process is continuous. The analogous transform when the time series is discrete is the z transform.

The *correlation* of two sample functions $x(t)$ and $y(t)$ is given by

$$\gamma_{xy}(\tau) \triangleq \mathcal{E}\{x(t+\tau) y(t)\} \qquad (2.44)$$

If $x(t) \leftrightarrow X(f)$, the Wiener-Khinchin theorem states that the autocorrelation function of $x(t)$ is related to its FT, $X(f)$, as

$$\gamma_{xx}(\tau) \leftrightarrow |X(f)|^2 \tag{2.45}$$

where τ is time delay and f is frequency. Thus, the absolute value of the FT of the autocorrelation function is the psd of that function.

For discrete time processes x_n and y_n, where n is a running index such that the sample times are $t_n = nT$ when T is the time between samples (assumed constant), the correlation function is given by

$$\gamma_{xy,m} = \mathcal{E}\{x_{n+m}y_n\} \tag{2.46}$$

This value is to be estimated using N values of the functions as $x_0, x_1, \ldots, x_{N-1}$ and $y_0, y_1, \ldots, y_{N-1}$.

One estimate of the autocorrelation function is given by

$$\gamma_{N,m} = \frac{1}{N}\sum_{n=0}^{N-1-|m|} x_{n+|m|}x_n, \qquad |m| \leq N-1 \tag{2.47}$$

which is called the *sample autocorrelation function*. The expected value of the sample autocorrelation function is given by

$$\mathcal{E}\{\gamma_{N,m}\} = \frac{1}{N}\sum_{n=0}^{N-1-|m|} \mathcal{E}\{x_{n+|m|}x_n\} = \frac{N-|m|}{N}\gamma_{xx,m} \tag{2.48}$$

so the expected value of the sample is not equal to the actual expected value of the series. When this occurs, the estimate is said to be *biased*.

The calculation of an estimate is *consistent* if the estimate converges to the true value as N approaches infinity. This will be true if the bias as well as the variance both converge to zero as N approaches infinity. This is true for $\gamma_{N,m}$ as well as the variance when $x_n \sim \mathcal{N}(0, \sigma^2)$.

2.3.3 Mean, Autocorrelation, and Autocovariance Functions

Correlation functions of a stochastic process can be defined that, in general, indicate how well the rv in one realization from the ensemble at a point in time is similar to another realization. The mean of the process is defined as

$$\mu_s(t) = \mathcal{E}\{s(t)\} = \int_{-\infty}^{\infty} s p_t(s) ds \qquad (2.49)$$

It is worth pointing out again that there is a mean value for each t, since the averaging taking place is over the ensemble at each t, as opposed to over time.

The similarity of the stochastic process at two distinct times is given by the *autocorrelation* function defined as

$$\gamma_{ss}(t_1, t_2) = \mathcal{E}\{s^*(t_1) s(t_2)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} s_1^* s_2 p(s_1, t_1; s_2, t_2) ds_1 ds_2 \qquad (2.50)$$

while the *autocovariance* function at two different times is an indication of the spread of the rv around the mean value. It is given by

$$\kappa_{ss}(t_1, t_2) = \mathcal{E}\{[s^*(t_1) - \mu_s(t_1)][s(t_2) - \mu_s(t_2)]\} \qquad (2.51)$$

These definitions can be extended to two distinct sample functions s_1 and s_2 from the random process as $\gamma_{s_1 s_2}$ and $\kappa_{s_1 s_2}$, respectively.

There is no limit on the number of times that can be included in these statistics, so the joint moment at n time instants is given by

$$\gamma_{ss}(t_1, t_2, \cdots, t_n) \triangleq \mathcal{E}\{s^*(t_1) s(t_2) \cdots s(t_n)\} \qquad (2.52)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} s_1^* s_2 \cdots s_n p(s_1, t_1; s_2, t_2; \cdots; s_n, t_n) ds_1 ds_2 \cdots ds_n \qquad (2.53)$$

2.3.4 Stationary and Wide-Sense Stationary Processes

There are two important definitions of stationarity for stochastic processes. Both imply characteristics of the behavior of the process with time.

2.3.4.1 Strict-Sense Stationarity

If the statistics of a stochastic process are independent of the choice of the time origin, the process is said to be *strict-sense stationary*. Therefore, instead of

having a pdf for each time instant t, $p(s; t)$, a stationary process has a single probability density independent of t given by $p(s)$. The mean of a stationary random process is a single value rather than a function of time. The n-time joint probability density is given by

$$p\left(s_1,t_1;s_2,t_2;\cdots;s_n,t_n\right)= p\left(s_1,t_1 +\tau;s_2,t_2 +\tau;\cdots;s_n,t_n +\tau\right) \qquad (2.54)$$

for any τ. The autocorrelation and autocovariance functions then depend only on the time difference between the two sample times and

$$\gamma_{ss}\left(\tau\right)\triangleq \mathcal{E}\left\{s^*\left(t\right)s\left(t+\tau\right)\right\}= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s_1^* s_2 p\left(s_1,t;s_2,t+\tau\right)ds_1 ds_2 \qquad (2.55)$$

$$\kappa_{ss}\left(\tau\right)= \mathcal{E}\left\{\left[s^*\left(t\right)-\mu_s\right]\left[s\left(t+\tau\right)-\mu_s\right]\right\} \qquad (2.56)$$

A strict sense stationary process requires that the autocorrelation functions of all orders be independent of absolute time.

2.3.4.2 Wide-Sense Stationarity

If it is only known that the mean and two-time autocorrelation function are independent of the time origin, the process is said to be *wide-sense stationary* (wss). It is frequently only realistic to establish that a signal is wss. Establishing stationarity in general is a difficult problem to solve.

2.3.5 Ergodic Processes

Ergodic processes are those for which the ensemble averages can be replaced with the time averages over a single realization of the process. Ergodic processes are always stationary, but the reverse is not true, although a stationary process may be ergodic.

2.3.6 Cyclostationary Processes

If the statistics of a process repeat after a period of time, then the process is said to be *cyclostationary* [6, 7].

2.4 Stochastic Signals

The discussion in Section 2.2 on deterministic signals does not apply directly when the signals under consideration are stochastic. There are two reasons for this. First, $S(f)$ is a random variable, since, for any fixed f, each sample would be represented by a different value of the ensemble of possible sample functions. Hence, it is not a frequency representation of the process but only of one member of the process. The second reason for not using the $S(f)$ of (2.1) is that, for stationary processes, it almost never exists. One of the conditions for a time function to have an FT is that it be absolutely integrable so that

$$\int_{-\infty}^{\infty} |s(t)| \, dt < \infty \tag{2.57}$$

A sample from a stationary random process can never satisfy this condition (with the exception of generalized functions inclusive of impulses and so forth) because if a signal has nonzero power, then it has infinite energy and if it has finite energy then it has zero average power. The class of functions having no Fourier integral, due to (2.57), but whose average power is finite can be described by statistical means.

Let $s(t)$ be a realization of a stochastic process. Define the truncated version of the function $s_T(t)$ as

$$s_T(t) = \begin{cases} s(t), & |t| \leq T \\ 0, & |t| > T \end{cases} \tag{2.58}$$

and ·

$$s(t) = \lim_{T \to \infty} s_T(t) \tag{2.59}$$

If $s(t)$ is a power signal, then the transform of $s(t)$ is not defined, but the transform of $s_T(t)$ is defined because

$$\int_{-\infty}^{\infty} |s_T(t)| \, dt < \infty \tag{2.60}$$

The FT pair of the truncated function $s_T(t)$ can thus be found using (2.1) and (2.2). Since $s(t)$ is a power signal, there must be a power spectral density function

associated with it and the total area under this density must be the average power despite the fact that $s(t)$ is non-Fourier transformable.

Equation (2.34) using the truncated function $s_T(t)$ is

$$\int_{-\infty}^{\infty} |s_T(t)|^2 \, dt = \int_{-\infty}^{\infty} |S_T(f)|^2 \, df \tag{2.61}$$

Dividing both sides by $2T$ yields

$$\frac{1}{2T} \int_{-\infty}^{\infty} |s_T(t)|^2 \, dt = \frac{1}{2T} \int_{-\infty}^{\infty} |S_T(f)|^2 \, df \tag{2.62}$$

The left side of (2.62) is proportional to the average power of the realization in the time interval $-T$ to T. This assumes $s_T(t)$ is a voltage (current) associated with a unit resistance. It is the square of the effective value of $s_T(t)$ and for an ergodic process approaches the mean-square value of the process as T approaches infinity.

Since $S_T(f)$ is nonexistent in the limit, the limit as T approaches infinity cannot be taken. Recall, though, that $S_T(f)$ is a random variable with respect to the ensemble of sample functions from which $s(t)$ was taken. The limit of the expected value of

$$\frac{1}{2T} |S_T(f)|^2 \tag{2.63}$$

does exist, since its integral, (2.62), is never negative and exists. If the expectations of both sides of (2.62) are determined

$$\mathcal{E}\left\{ \frac{1}{2T} \int_{-\infty}^{\infty} |s_T(t)|^2 \, dt \right\} = \mathcal{E}\left\{ \frac{1}{2T} \int_{-\infty}^{\infty} |S_T(f)|^2 \, df \right\} \tag{2.64}$$

then interchanging the integration and expectation[2] yields

---

[2] If $f(t)$ is a nonrandom time function and $s(t)$ a realization of a random process, then

$$\mathcal{E}\left\{ \int_{t_1}^{t_2} s(t)f(t)dt \right\} = \int_{t_1}^{t_2} \mathcal{E}\{s(t)\}f(t)dt$$

is allowed under the conditions

$$\frac{1}{2T}\int\limits_{-\infty}^{\infty}\mathcal{E}\left\{\left|s_T(t)\right|^2\right\}dt = \frac{1}{2T}\int\limits_{-\infty}^{\infty}\mathcal{E}\left\{\left|S_T(f)\right|^2\right\}df \qquad (2.65)$$

Taking the limit as $T \to \infty$,

$$\lim_{T\to\infty}\frac{1}{2T}\int\limits_{-\infty}^{\infty}\overline{x^2}(t)dt = \lim_{T\to\infty}\frac{1}{2T}\int\limits_{-\infty}^{\infty}\mathcal{E}\left\{\left|S_T(f)\right|^2\right\}df \qquad (2.66)$$

results in

$$\left\langle\overline{s^2(t)}\right\rangle = \int\limits_{-\infty}^{\infty}\lim_{T\to\infty}\frac{1}{2T}\mathcal{E}\left\{\left|S_T(f)\right|^2\right\}df \qquad (2.67)$$

where $\left\langle\overline{s^2}(t)\right\rangle$ is defined as the mean-square value ($\overline{x}$ denotes ensemble averaging and $<x>$ denotes time averaging).

As long as the process under consideration is stationary then the time average of the mean-square value is equal to the mean-square value, and (2.67) becomes

$$\overline{s^2}(t) = \int\limits_{-\infty}^{\infty}\lim_{T\to\infty}\frac{1}{2T}\mathcal{E}\left\{\left|S_T(f)\right|^2\right\}df \qquad (2.68)$$

The integrand of the right side of (2.68), similar to (2.34), is called the psd function of a stochastic process and is denoted by $S(f)$; thus,

$$S(f) = \lim_{T\to\infty}\frac{1}{2T}\mathcal{E}\left\{\left|S_T(f)\right|^2\right\} \qquad (2.69)$$

It is important to note again that letting $T \to \infty$ is not possible before finding the expectation in (2.69).

- $\int\limits_{t_1}^{t_2}\mathcal{E}\{s(t)\}f(t)dt < \infty$;

- $s(t)$ is bounded on the interval t_1 to t_2. Note that t_1 and/or t_2 may be infinite. Also, $s(t)$ may be stationary or nonstationary.

If $s(t)$ is a voltage (current) associated with a 1Ω resistance, $s^2(t)$ is the average power dissipated in that resistor and $S(f)$ is the average power associated with a bandwidth of 1 Hz centered at f Hz.

$S(f)$ has the units volts^2-second and its integral, (2.68), leads to the mean-square value, hence,

$$\overline{s^2}(t) = \int_{-\infty}^{\infty} S(f)df \tag{2.70}$$

Since $S_T(f)$ is the FT of $s_T(t)$, assuming a nonstationary process, from (2.69),

$$S(f) = \lim_{T \to \infty} \frac{1}{2T} \mathcal{E}\left\{ \int_{-\infty}^{\infty} s_T(t_1)e^{j2\pi f t_1}dt_1 \int_{-\infty}^{\infty} s_T(t_2)e^{j2\pi f t_2}dt_2 \right\} \tag{2.71}$$

The subscripts of t_1 and t_2 have been introduced to keep the variables of integration distinct. So,

$$S(f) = \lim_{T \to \infty}\left[\frac{1}{2T} \mathcal{E}\left\{ \int_{-\infty}^{\infty} dt_2 \int_{-\infty}^{\infty} e^{-j2\pi f(t_2-t_1)} s_T(t_1)s_T(t_2)dt_1 \right\}\right]$$

$$= \lim_{T \to \infty}\left[\frac{1}{2T} \int_{-\infty}^{\infty} dt_2 \int_{-\infty}^{\infty} \mathcal{E}\left\{ s_T(t_1)s_T(t_2)\right\} e^{-j2\pi f(t_2-t_1)}dt_1 \right] \tag{2.72}$$

The expectation $\mathcal{E}\left\{ s_T(t_1)s_T(t_2)\right\}$ is the autocorrelation function, $\gamma_{ss}(t_1, t_2)$, of the truncated process where

$$\mathcal{E}\left\{ s_T(t_1)s_T(t_2)\right\} = \gamma_{ss}(t_1,t_2), \qquad |t_1|,|t_2| \leq T \tag{2.73}$$

Substituting

$$\begin{aligned} t_2 - t_1 &= \tau \\ dt_2 &= d\tau \end{aligned} \tag{2.74}$$

(2.72) becomes

$$S(f) = \lim_{T \to \infty} \left[\frac{1}{2T} \mathcal{E} \left\{ \int\limits_{-\infty}^{\infty} d\tau \int\limits_{-\infty}^{\infty} \gamma_{ss}(t_1, t_1 + \tau) e^{-j2\pi f \tau} dt_1 \right\} \right] \qquad (2.75)$$

or

$$S(f) = \left\{ \int\limits_{-\infty}^{\infty} \lim_{T \to \infty} \frac{1}{2T} \int\limits_{-T}^{T} \gamma_{ss}(t_1, t_1 + \tau) dt_1 \right\} e^{-j2\pi f \tau} d\tau \qquad (2.76)$$

Thus, the spectral density is the FT of the time average of the autocorrelation function. The relationship of (2.76) is valid even for nonstationary processes.

For the stationary process, the autocorrelation function is independent of time origin, and therefore

$$\langle \gamma_{ss}(t_1, t_1 + \tau) \rangle = \gamma_{ss}(\tau) \qquad (2.77)$$

It has just been shown that for a stationary random process the autocorrelation function is the inverse FT of the spectral density function. This cannot be said for nonstationary processes, however.

The FT of the autocorrelation function is called the psd, or the *power spectrum* of $s(t)$, denoted by $S(f)$:

$$S(f) = \int\limits_{-\infty}^{\infty} \gamma_{ss}(\tau) e^{-j2\pi ft} dt \qquad (2.78)$$

and

$$\gamma_{ss}(\tau) = \int\limits_{-\infty}^{\infty} S(f) e^{j2\pi ft} df \qquad (2.79)$$

The *average power* of a stationary stochastic process is defined as

$$P_s = \mathcal{E} \left\{ |s(t)|^2 \right\} \qquad (2.80)$$

Likewise, the *autocorrelation* of $s(t)$ is given by

$$\gamma_{ss}(\tau) = \mathcal{E}\{s^*(t)s(t+\tau)\} \tag{2.81}$$

and the *cross-correlation* function of $s(t)$ and $x(t)$ is given by

$$\gamma_{sx}(\tau) = \mathcal{E}\{s^*(t)x(t+\tau)\} \tag{2.82}$$

If a stationary stochastic process $s(t)$ is input to an LTI filter with impulse response $h(t)$ producing output $y(t) = (s * h)(t)$, the cross-correlation of the input and output is

$$
\begin{aligned}
\gamma_{sy}(t_1,t_2) &= \mathcal{E}\{s^*(t_1)y(t_2)\} \\
&= \mathcal{E}\left\{s^*(t_1)\int_{-\infty}^{\infty} h(v)s(t_2-v)dv\right\} \\
&= \int_{-\infty}^{\infty} h(v)\gamma_{ss}(t_2-t_1-v)dv \\
&= \int_{-\infty}^{\infty} h(v)\mathcal{E}\{s^*(t_1)s(t_2-v)\}dv \\
&= (\gamma_{ss}*h)(t_2-t_1)
\end{aligned}
\tag{2.83}
$$

Thus, $\gamma_{sy}(.)$ is not a function of time but a function of only the time difference $\tau = t_2 - t_1$, so

$$\gamma_{sy}(\tau) = (\gamma_{ss}*h)(\tau) \tag{2.84}$$

By a derivation similar to (2.83),

$$\gamma_{yy}(\tau) = (\gamma_{sy}*\tilde{h})(\tau) = (\gamma_{ss}*h*\tilde{h})(\tau) \tag{2.85}$$

where $\tilde{h}(t) = h^*(-t)$. Note that the psd is an even function of frequency.

The autocorrelation and cross-correlation functions are defined for finite power stochastic functions as

$$\gamma_{ss} = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} s^*(t) s(t+\tau) dt \tag{2.86}$$

and

$$\gamma_{s_1 s_2} = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} s_1^*(t) s_2(t+\tau) dt \tag{2.87}$$

respectively. When $\tau = 0$,

$$\gamma_{ss}(0) = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} s^*(t) s(\tau) dt = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} |s(t)|^2 dt$$

$$= P_s \tag{2.88}$$

By derivations similar to that leading up to (2.84), it can be shown that the autocorrelation of the output of an ideal LTI filter is given by

$$\gamma_{yy}(\tau) = \left(\gamma_{ss} * \tilde{h} \right)(\tau) \tag{2.89}$$

with corresponding FT

$$S_{yy}(f) = S_{ss} |H(f)|^2 \tag{2.90}$$

If a finite power signal is passed through an ideal bandpass filter of bandwidth B, the average power of the output, calculated in a similar manner to the above, is

$$\gamma_{yy}(0) = \int_{-B/2}^{B/2} S_{ss}(f) df \tag{2.91}$$

which justifies calling $S_{ss}(f)$ the psd of $s(t)$.

The cross-correlation function of the input and output is similarly given by

$$\gamma_{sy}(\tau) = (\gamma_{ss} * h)(\tau) \tag{2.92}$$

with corresponding FT

$$S_{sy}(f) = S_{ss}H(f) \tag{2.93}$$

Assuming that y is wss, the psd $\Phi_{yy}(e^{j\omega})$ can be shown to be $\left|Y(e^{j\omega})\right|^2$ as follows.

$$
\begin{aligned}
\Phi_{yy}(e^{j\omega}) &= \sum_{m=-\infty}^{\infty} \phi_{yy}(m)e^{-jm\omega}, \qquad |\omega| < \pi \\
&= \sum_{m=-\infty}^{\infty} \mathcal{E}\left\{y_{n+m}y_n^*\right\}e^{-jm\omega} \\
&= \mathcal{E}\left\{y_n^* \sum_{m=-\infty}^{\infty} y_{n+m}e^{-jm\omega}\right\} \\
&= X^*(e^{j\omega})X(e^{j\omega}) \\
&= \mathcal{E}\left\{y_n^* e^{jn\omega}X(e^{j\omega})\right\} \tag{2.94} \\
&= \left|X(e^{j\omega})\right|^2 \tag{2.95}
\end{aligned}
$$

where $\Phi_{yy}(e^{j\omega})$ is the psd of y and $\phi_{yy}(m)$ is the autocorrelation sequence of y. Equation (2.94) is true by the time shifting property of the FT. This is known as the Weiner-Khinchin theorem.

If $y(k)$ represents a specific instance of the wss ergodic random process Y_N, a finite-length sequence can be defined as

$$y_N(k) = \begin{cases} y(k), & -N \le k \le N \\ 0, & \text{otherwise} \end{cases} \tag{2.96}$$

From this an estimate $\hat{r}_{yy}(m)$ of the autocorrelation of y can be generated as

$$\hat{r}_{yy}(m) = \frac{1}{2N+1} \sum_{k=-N}^{N} y_N(k+m)y_N^*(k) \tag{2.97}$$

Thus,

$$
\begin{aligned}
\hat{\Phi}_{yy}(e^{j\omega}) &= \sum_{k=-N}^{N} \hat{r}_{yy}(k)e^{-jk\omega}, && |\omega| < \pi \\
&= \sum_{k=-N}^{N}\left[\frac{1}{2N+1}\sum_{m=-N}^{N} y_N(m+k)y_N^*(m)\right]e^{-jk\omega} \\
&= \frac{1}{2N+1}\sum_{m=-N}^{N} y_N^*(m)\left[\sum_{k=-N}^{N} y_N(k+m)\right]e^{-jk\omega} \\
&= \frac{1}{2N+1}Y_N\left(e^{j\omega}\right)\sum_{m=-N}^{N} y_N^*(m)e^{jk\omega} && (2.98) \\
&= \frac{1}{2N+1}\left|Y_N\left(e^{j\omega}\right)\right|^2 \\
&= \frac{1}{2N+1}\left|\sum_{k=-N}^{N} y_N(k)e^{-jk\omega}\right|^2 && (2.99)
\end{aligned}
$$

Again, (2.98) is possible due to the time shifting property of the FT.

The true psd is the expected value of $\hat{\Phi}_{yy}(e^{j\omega})$ as $N \to \infty$:

$$
\begin{aligned}
\Phi_{yy}(e^{j\omega}) &= \lim_{N\to\infty} \mathcal{E}\left\{\hat{\Phi}_{yy}(e^{j\omega})\right\} \\
&= \lim_{N\to\infty} \mathcal{E}\left\{\frac{1}{2N+1}\left|\sum_{k=-N}^{N} y_N(k)e^{-jk\omega}\right|^2\right\} && (2.100)
\end{aligned}
$$

.

2.5 White Noise

If the psd of a stationary stochastic process $x(t)$ is a constant value with frequency it is said to be *white* (see Figure 2.8). This corresponds to the naturally occurring thermal noise in the atmosphere as well as all electronic devices. The psd can be specified either over the entire frequency spectrum of $-\infty < f < \infty$, when it is given by $N_0 / 2$ W/Hz, or equivalently, over just the positive frequency range of $0 \le f < \infty$, where it is specified by N_0 W/Hz.
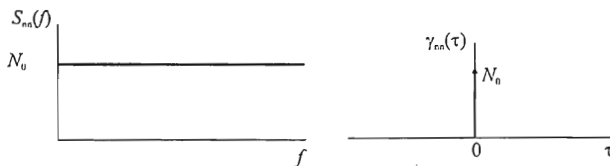
Figure 2.8 White noise.

Considering the positive frequency range only, the autocorrelation function is given by

$$\gamma_{nn}(\tau) = \int_{-\infty}^{\infty} N_0 e^{j2\pi ft} df$$
$$= N_0 \delta(\tau) \tag{2.101}$$

which is also illustrated in Figure 2.8.

Thus, the values of the process at any two distinct times are uncorrelated. Since $P_f = \gamma_{xx}(0)$, it also means that instantaneous power is infinite so that the value of the process at each time can be arbitrarily large. In most cases of practical interest, the noise is modeled with a flat spectral density over the range of frequencies of interest. All real noise sources tend to zero due to natural causes as the frequency gets large.

2.5.1 Signals in Noise

In many cases of interest in communication EW system analysis, the signal is embedded within a stationary noise process. If $s(t)$ is the deterministic signal and $n(t)$ represents a realization of a noise process, the received signal $r(t)$ is given by

$$r(t) = s(t) + n(t) \tag{2.102}$$

The SNR, denoted here by ν, is defined as

$$\nu = \frac{\text{Average signal power}}{\text{Total noise power}} \tag{2.103}$$

For example, if the signal is given by

$$s(t) = \sqrt{2S} \sin(2\pi ft) \tag{2.104}$$

where S is the average power in the signal, the noise is specified as N_0 W/Hz, and the total noise power in a double-sided bandwidth B is $P_N = BN_0$, then the SNR is

$$\nu = \frac{S}{BN_0} = \frac{S}{BN_0} \tag{2.105}$$

2.6 Concluding Remarks

The fundamental properties of deterministic and stochastic processes were presented in this chapter. Processes in the context important here are signals and the mechanisms that generate them.

Deterministic signals are idealizations of realistic signals. Most signals of importance to communications EW are stochastic in some sense. There are one or more parameters about them that are random, and therefore these signals can only be described by probability functions and the resulting statistical properties.

References

[1]   Bracewell, R., *The Fourier Transform and Its Applications,* New York: McGraw-Hill, 1965.

[2]   Kaplan, W., *Introduction to Analytic Functions,* Reading, MA: Addison-Wesley, 1966, p. 42.

[3]   Tretter, S. A., *Introduction to Discrete-Time Signal Processing,* New York: John Wiley & Sons, 1976, pp. 14–15.

[4]   Cover, T. M., and J. A. Thomas, *Elements of Information Theory,* New York: John Wiley & Sons, 1991, pp. 248–249.

[5]   Tretter, S. A., *Introduction to Discrete-Time Signal Processing,* New York: John Wiley & Sons, 1976, Chapter 11.

[6]   Gardner, W. A., *Statistical Spectral Analysis: A Nonprobabilistic Theory,* Englewood Cliffs, NJ: Prentice Hall, 1988, Part II.

[7]   Gardner, W. A., *Introduction to Random Processes with Applications to Signals and Systems,* New York: Macmillan, 1986, Chapter 12.

Chapter 3

Target Search Methods

In communication EW problems, signal detection refers to establishing the presence or absence of a signal at a frequency by searching the frequency spectrum [1]. There are three distinct types of such searching: *general search* (GS), *directed search* (DS), and signal verification. This last category is not searching in the same sense as the other two—it consists of verifying that the target signal is still present at the frequency channel that is being jammed [2]. Characteristics of these search schemes are examined in this chapter [3].

GS is the type of search when little or nothing is known about the signal environment and the frequencies of the active targets must be established—this is an ES function. The search receiver scans the frequency spectrum from some start frequency to some stop frequency (or multiple bands of these) and measures the energy present at each channel. If energy is detected, then that frequency is tagged as active and the receiver moves on. Alternatively, other actions can occur upon detection of an active channel, such as notifying an operator or automated processing equipment to begin making measurements on the signal. In GS searching, the target types, in general, are unknown, although there can be exceptions to this. Therefore it is best assumed that the signals have at least some random parameters. These parameters are the signal's amplitude, time of detection, and phase of the carrier. The frequency is known because it is either assumed to be the center of the channel being searched, or it can be measured as the position of the centroid of the spectrum or the highest peak of the spectrum. The two types of signals of interest then are:

- Targets with random amplitude and phase at known frequency;
- Modulated sinusoid with known frequency, random modulation, random amplitude, and random phase.
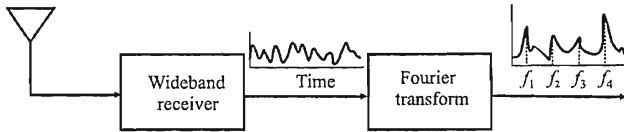
**Figure 3.1** Wideband receiver for signal searching.

DS assumes that most or all of the frequencies of the targets of interest are known. In addition, other characteristics of the signals may be known, such as the modulation type. The search receiver tunes to each discrete channel, the priorities of which can be taken into account. The search only covers those frequencies that are known active at some point in the recent past. In addition, if an operator or automation has determined that the signal at a particular frequency is a high-priority target, that channel could be revisited more often than some other channels that may be of just general interest. With DS searching, there is a much higher probability of knowing the target type associated with the frequency. Thus, specific targets can be modeled as:

- Random targets at known frequency with known amplitude, known modulation, and unknown phase;
- Modulated sinusoidal target with known frequency, known modulation, known amplitude, and random phase.

Combinations of GS and DS are also possible. This scenario would be used when the target environment is partially known (DS), but a significant portion is not (GS).

As mentioned, verification is ensuring that the target being jammed at a frequency is still active at that frequency. This is also called *look-through*. During the look-though period, the jamming signal must be blanked to not damage the sensitive ES equipment. In this case, the signal type is known, including all the relevant parameters. In that case, the hypothesis testing approach for signal detection, discussed later, is probably not required. The high-resolution signal detection capabilities are required, which is true of all the search strategies discussed here.

Searching is typically accomplished by converting the signal output of a relatively wideband receiver into the frequency domain and determining the frequencies where the peaks occur, as illustrated in Figure 3.1. There are two fundamental types of spectral estimation methods that can be used for signal detection. The first is called the traditional method and is used when relatively low resolution is adequate with resolution defined as indicated in Figure 3.2. Two equal amplitude signals are resolved when their peak levels are separated by a
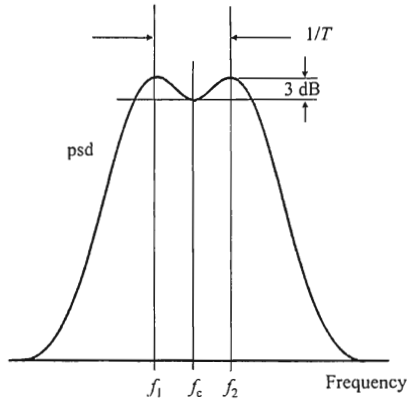
Figure 3.2 Frequency resolution. The signals are just resolvable when the dip between two signals is 3 dB below the peak levels. This frequency separation is approximately $1/T$ for low-resolution spectrum estimation. An alternate definition is used in Chapter 9 that allows for differing power levels of the signals.

frequency span defined by the 3-dB drop from the peaks values. In Figure 3.2, this span is given by $f_2 - f_1$. When this amount of resolution is not adequate, the second technique provides for higher resolution signal detection at the expense of a higher computational burden. Two traditional techniques are presented in Chapter 4: the periodogram and the Blackman-Tukey method. The higher resolution methods are presented in Chapter 8.

All spectral estimation techniques are imperfect in that false detections occur and detections are sometimes missed, as discussed in this chapter.

3.1 General Search

As indicated in Chapter 1, general searching for signals of interest is used to establish where signals are located in the RF environment. The signal environment in GS is assumed to consist of deterministic signals with random parameters, as well as random signals with unknown parameters as the most general case. The background noise is assumed here to consist of AWGN.

When little or nothing is known about the signal being detected, the optimum detector is the radiometer. The radiometer estimates the amount of energy and noise present in the signal. Detailed discussions about the radiometer are presented in Chapter 5.

The parameters for GS must be provided by tasking prior to a mission. The most general tasking for GS is to search a single frequency band from start to finish, repetitively. The frequency band list, of course, can be modified during a mission. Once a signal is detected, system assets are tasked to determine as much about the signal as possible for subsequent use in DS or for other purposes, such as tasking set-on receivers. Databases could also exist that are provided by external sources or determined on prior missions as to frequencies in use by SOIs.

When energy is detected, signal detection is declared. Rarely is this sufficient to declare the signal as being associated with a target, even though the frequency is the same. Further processing is almost always required for target declaration. This further processing could be determining the type of modulation present (if any)

Another possibility is to obtain a *line of bearing* (LOB) on the target and/or task other EW systems to do the same. An LOB is the direction from which the signal arrives at the EW system. The resultant data could then be used to calculate a *position fix* (PF). A signal classifier could be tuned to the frequency where the energy is detected, making appropriate measurements to ascertain the modulation of the signal. A set-on receiver could be tuned to the frequency so that the signal can be intercepted. These are but a few of the possibilities, but they all start with the fundamental requirement to measure the energy to detect the presence of signals.

Two or more signals can be occupying the same frequency channel. In general these signals can be targets, friendly interferers, or gray interferers. Based on the detected energy alone, the distinction is difficult to ascertain. Other means must then be employed to determine the status of the signals. This assumes that it is possible to determine the presence of more than one signal, which is a dubious assumption at best. In the tactical military communication frequency bands, the channelization is fairly narrow—10 kHz in the HF band, 25 kHz in the low VHF band (30–90 MHz), and 50 kHz in the low UHF band (110–400 MHz). If a signal is using a channel in one of these bands, it will likely use the whole channel, and if two or more signals are present, they will almost assuredly substantially overlap in frequency content.

In some cases it is possible to remove one (or more) of the interfering signals when cochannel interference occurs. The system shown in Figure 3.3 is one such technique for FM signals. The first PLL locks onto the stronger of the signals at the input and forms an identical signal that is π radians out of phase with it (as indicated by the negative sign at the adder in Figure 3.3). That signal is then added to the incoming signal that effectively removes the stronger signal. The resultant signal is the weaker of the signals at the input. This could be repeated a number of times.
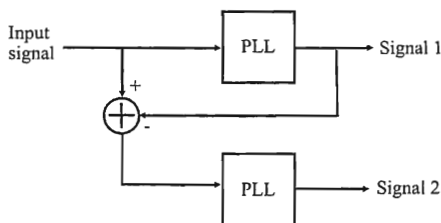
Figure 3.3 Coupled phase-locked loops (PLL) for extracting two cochannel FM or PM signals.

The same technique can also be used as an indication of the presence of cochannel interference. If there is little signal left after subtracting the stronger component, it is an indication that it was the only signal there.

Modern military communications tend to standardize on the type of modulation depending on frequency band. In the HF band, analog AM is popular for voice and *binary frequency shift key* (BFSK) is popular for teletype signals. In the low VHF band, analog FM is popular for voice communications, and *frequency shift key* (FSK) and *phase shift key* (PSK) are present as well for digital communications. In the low UHF range, analog AM is popular for airborne communications. Radars can show up as interferers in all of these bands as well. Therefore, in general, it is highly unlikely that the correct interfering modulations will occur, whereas for specific segments of the RF band the probability is considerably higher.

Modern radio transmitters, implemented with largely digital components, are better at retaining proper channelization than older analog radios. If the transmitter drifts, this also can cause interference problems, even if the channel is not completely overlapped by the interference. With the large thrust toward commercial use of the RF spectrum for communications as exemplified by the proliferation of *personal communication systems* (PCS) and cellular phone systems around the world, the drift problem is likely to be less of a concern in the future.

In ground-based EW systems, target signals are likely to be weaker than friendly interferers most of the time. That is because they are likely to be further away. This, of course, does not apply to airborne EW systems. In particular, for *unattended aerial vehicle* (UAV)-based systems for example, the UAV can fly closer to the targets to cause them to be larger in amplitude. Also, in conflicts in urban terrain, the targets are totally mixed with friendly and gray signals, so this comment does not apply. In general, signal amplitude is a bad sort criterion for determining whether a signal is a target or not.

Sometimes the angle of arrival of a signal can be used as an indication of whether a detected signal is a target or not. If the general region of the target array
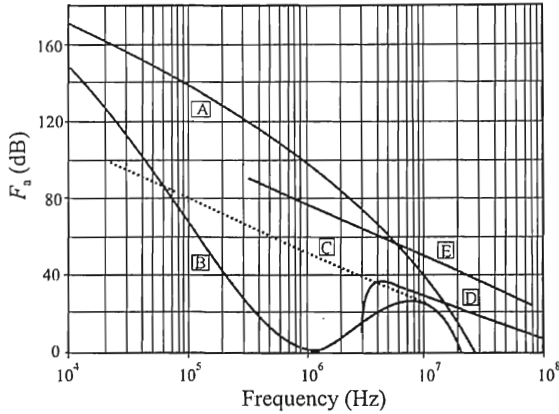
Figure 3.4 Typical noise characteristics in the low RF spectrum: A: atmospheric noise, value exceeded 0.5% of the time; B: atmospheric noise, value exceeded 99.5% of the time; C: man-made noise, quiet receiving site; D: galactic noise; and E: median business area man-made noise; minimum noise level expected. (*From:* [4], © 1982 CRC Press, Inc. Reprinted with permission.)

is known, even approximately, then signals not coming from that region may be assumed to be friendly or gray.

Digital signals are normally easier to classify than analog voice. The telltale characteristics of digital signals are easier to isolate than those associated with voice.

The frequency resolution is inversely proportional to dwell time for the most common type of spectrum estimation technique, the periodogram. However, the longer the dwell time, producing better frequency resolution, the lower the revisit rate possible to other channels for scanning implementations where the search receiver moves from one set of channels to the next. For implementations where a staring receiver is used, revisit time is not an issue.

Staring is possible over limited portions of the tactical communication frequency range. Digital staring receivers can cover the entire HF band, although noise variations would typically require some degree of channelization. The noise levels below 100 MHz are typified by that shown in Figure 3.4 [4], and to stare at the entire band would require a dynamic range in the receiver beyond that available today. The noise alone over 1 to 30 MHz varies by about 50 dB minimum. That is before any signals are present, which in the HF band can vary by 100 dB or more. In Figure 3.4, noise levels are shown in decibels above kTB, where k is Boltzmann's constant, $k = 1.38 \times 10^{-23}$ W/K/Hz, T = temperature (Kelvin), and B = noise bandwidth (Hz). Portions of the VHF (30–300 MHz) band can be covered with staring receivers when the requirement for high dynamic

range is considered. The same is true for the UHF band, although the important PCS bands can be covered with adequate dynamic range, typically 15 MHz around 400 MHz, 70 MHz around 900 MHz, and 140 MHz around 1.9 GHz (1.850–1.990 GHz).

The results of GS are a list of signals with their parameters. These parameters could be:

- Frequency;
- Power level;
- Modulation type;
- Geolocation/LOB.

and other, more detailed information.

3.2 Directed Search

Directed search is used when most of the parameters associated with the targets are known. Most importantly, it is known that the signal at a DS frequency is a true, or at least suspected, target. The DS frequencies are known, and the EW system moves from one frequency to the next, measuring the energy at each frequency channel to determine the presence or absence of a signal. Just as for GS, if energy is detected, several subsequent actions are possible, and whether any of them are executed depends on the particular mission and the signal.

The data required for directed searching, such as frequencies, signal types, amplitudes, and so forth, are provided for directed search. This information could be provided by other sources, prior to or during the mission, or it could be the result of GS discussed in Section 3.1. The more information that is provided, of course, the   higher the probability is that a detected signal is a SOI.

For scanning implementations, the time it takes to revisit a DS frequency impacts on the probability of detecting that target on its next transmission, which may be important for some missions.

3.3 Concluding Remarks

Determining the presence or absence of signals is the purpose of spectral searching. If little is known about the target environment then, GS is implemented to establish where the signals are and to determine as much about them as

possible. If the target environment is fairly well known, then DS can be used. In DS, the frequencies are known, as are several other threat signal parameters.

The result of GS is a list of frequencies where potential threat signals may be located. These results are used for subsequent DS when targets are further exploited.

In practice, both GS and DS are conducted simultaneously. With limited system assets, the two search methods might share the same receiver, alternating between the two search types with that receiver. In larger systems, separate subsystems can be used for the two search modes. GS can be running separately, categorizing the target environment, while DS is used to keep track of known targets.

References

[1]   Frater, M. R., and M. Ryan, *Electronic Warfare for the Digitized Battlefield,* Norwood, MA: Artech House, 2001, pp. 100–112.

[2]   Poisel, R. A., *Introduction to Communication Electronic Warfare Systems,* Norwood, MA: Artech House, 2002, Chapter 14.

[3]   Adamy, D., *EW101: A First Course in Electronic Warfare,* Norwood, MA: Artech House, 2001, p. 110.

[4]   Volland, H., ed., *Handbook of Atmospherics, Volume 1,* Boca Raton, FL: CRC Press, 1982, pp. 295–296.

Chapter 4

Hypothesis Testing for Signal Detection

There is a firm theoretical background for optimal detection of signals in noise available with hypothesis testing [1–3]. Hypothesis testing is one method used to optimally or suboptimally determine whether a signal is present or not [2, 4–8].

4.1 Hypothesis Testing

The simplest case to consider is when all of the parameters of a deterministic signal are known and the channel provides AWGN, whether the signal is present or not. Suppose that the receiver in the EW system outputs a predetection (not demodulated) waveform $x = \{x(t) : t \in [0, T]\}$ over the observation time interval $[0, T]$. This signal may have been produced by noise alone or by a received signal of known form plus noise. These two hypotheses can be expressed as

$$
\begin{aligned}
H_0 &: x = \text{noise}; \\
H_1 &: x = \text{signal} + \text{noise}
\end{aligned}
\tag{4.1}
$$

Stated another way, there are two possible hypotheses:

H_0: there is no signal and the time series consists of noise only $\sim p_0$
H_1: the time series consists of the sum of the signal and noise $\sim p_1$

That is, hypothesis H_0 has a priori probability distribution p_0 and H_1 has a priori probability distribution p_1. The noise associated with the reception of a signal causes this to be a stochastic problem, and even though the characteristics of the signal are completely known, the received signal can only be described statistically. The observation upon which the decision as to which hypothesis is to

51

be selected is denoted by Y. H_0 is referred to as the *null hypothesis* and H_1 is called the *alternative hypothesis*. H_0 and H_1 are called *simple hypotheses* when the pdf of x under H_0 and H_1 does not depend on any unknown parameters such as signal amplitude or phase. When such unknown parameters are involved, these hypotheses are called *composite hypotheses*. The unknown parameters are called *nuisance parameters*. Their values are uninteresting with regard to the problem at hand, yet their presence influences the ability to correctly decide the presence or absence of a signal.

An example of a simple (binary) hypothesis test is

H_0: $Y = Y_0$
H_1: $Y = Y_1$

where there are only two possible outcomes and they are completely specified. An example of a composite hypothesis test is

H_0: $Y = Y_0$
H_1: $Y \neq Y_0$

so that in H_1, more than one outcome is possible.

Consider Table 4.1. Ignoring the first row in Table 4.1, the remaining table entries are denoted by d_{ij} $j = 1,2$. This table indicates, for each condition of having the signal present or absent, the possible outcomes of making a decision, and each of these outcomes has an associated probability. Since one of the decisions must be made, the sum of the probabilities for any one column must total to unity. For decision (1, 1), (decide the signal is present when it is present), denoted by d_{11}, the outcome of the decision is the correct one. The probability in this case is called the *probability of detection*, denoted by P_d. For d_{21}, the signal is there but it was not detected. The corresponding probability for this decision is called the *probability of miss*, and denoted by P_m. Thus,

$$P_d = 1 - P_m \qquad (4.2)$$

For d_{12}, the detector thought the signal was there and it actually was not. This is known as a *false alarm*, with the probability denoted by P_{fa}. Lastly, for d_{22}, the

Table 4.1 Possible Decisions

Signal Present	Signal Absent
Decide signal present	Decide signal present
Decide signal absent	Decide signal absent

signal was declared absent when it actually was not there. This is a correct decision, but its probability does not carry a particular designation; this particular event is of little direct interest.

To estimate which hypothesis is true, the signal can be compared to a threshold and H_1 declared true only if the signal is larger than the threshold. The higher the threshold value, the lower the chance that a signal will be declared present when it is not present. On the other hand, the lower the threshold value, the lower the chance that a signal will be missed if it is in fact present. Thus, there are two types of errors possible: (1) the error of declaring the signal present when it is absent (declare H_1 when H_0 is true), and (2) the error of missing the signal when it is present (declare H_0 when H_1 is true). The first of these is called a *Type I error* as well as a false alarm. The second of these is called a *Type II error* as well as a missed detection. In all realistic cases of signal detection, then, there is a trade-off between setting the threshold high to lower P_{fa} and setting the threshold low to lower P_m. To ascertain the quantitative effects of this trade-off, it is necessary to specify the pdf for x in (4.1) under each of the hypotheses H_0 and H_1.

Depending on the situation, one of the error types can be more important than the other. False alarms corresponding to type I errors task system resources unnecessarily and thereby slow down system throughput. If there are enough system resources to adequately handle the signal environment with significant false alarms, then this is not a significant problem. On the other hand, rejecting the presence of a signal when in fact there is one present (missed detection) can cause critical signals to be missed.

A *decision rule* (or hypothesis test) δ partitions the observation space Γ into two regions given by Γ_1 and Γ_0 such that if $Y \in \Gamma_0$, hypothesis H_0 is chosen and if $Y \in \Gamma_1$, hypothesis H_1 is chosen. The set Γ_1 is referred to as the *rejection region* and the set Γ_0 is called the *acceptance region*. Such a decision rule is illustrated in Figure 4.1 for arbitrary probability densities. The decision rule, δ, is thus defined as a function of the observation Y as

$$\delta(Y) = \begin{cases} 1, & Y \in \Gamma_1 \\ 0, & Y \in \Gamma_0 \end{cases} \quad (4.3)$$
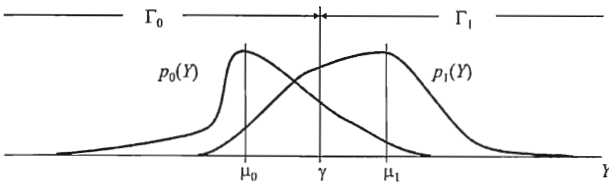

Figure 4.1 Decision regions for decision rule δ.

In statistics, Γ_1 is known as the *critical region* because it is in this region within which the null hypothesis is rejected. Alternatively, if the observed value is not in the critical region (that is, it is in Γ_0), the conclusion is to not reject the null hypothesis. The *critical value* for a hypothesis test is a threshold that determines whether the null hypothesis is rejected. In addition, in statistics, the *significance level* is the probability of false alarm—that is, the probability of incorrectly rejecting the null hypothesis. The *power* of a statistical hypothesis test is the probability of detection—that is, the probability of correctly rejecting the null hypothesis.

4.2 Receiver Operating Characteristic

The threshold for x is denoted by γ. The critical region, denoted by $\mathcal{R}_c$, is that range for x for which $x(t) > \gamma$. That is,

$$\mathcal{R}_c = \{x : x(t) > \gamma\} \tag{4.4}$$

This region then specifies the conditions on x for which the detector declares the signal to be present. Based on the critical region, then, the probability of detection, P_d, is given by

$$P_d = P(\mathcal{R}_{H_1} | H_1) \tag{4.5}$$

$$P_{fa} = P(\mathcal{R}_{H_1} | H_0) \tag{4.6}$$

and

$$P_m = 1 - P_d \tag{4.7}$$

where $P(A|B)$ denotes the probability of event A given hypothesis B.

The two-dimensional graph of $P_d(\gamma)$ versus $P_{fa}(\gamma)$ specifies the error performance of the detector for various values of $P_d(\gamma)$ and $P_{fa}(\gamma)$, an example of which is shown in Figure 4.2 for a particular value of γ. When the plots are included so that $-\infty < \gamma < \infty$ are contained on the graph, or some limited range for γ of interest, it is referred to as the *receiver operating characteristic* (ROC), an example of which for $\gamma \geq 1$ is shown in Figure 4.3.
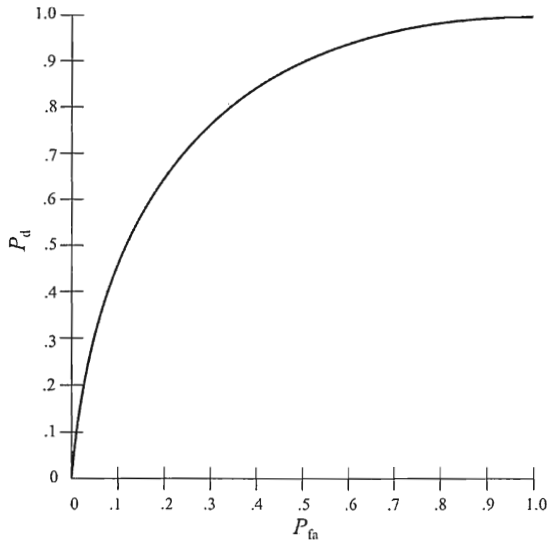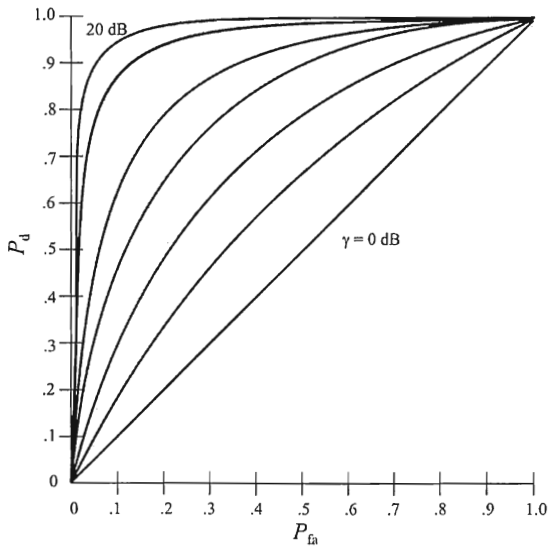
Figure 4.2 ROC curve.

Figure 4.3 ROC curves.

The ROC curve for a good detector is characterized as being convex, the curve is monotonic as P_{fa} increases, and the slope of the curve at (0, 0) is large.

In the above case it was assumed that the signal characteristics and noise statistics were completely known. That resulted in simple hypotheses, and an optimal detector can be found such that the ROC curve for any other detector will be less than the ROC curve for the optimal detector. "Less than" means that for any given P_{fa}, the P_d for these other detectors will be less than the P_d for the optimum detector. Thus, the ROC curve for the optimal detector forms an upper bound on the possible ROC curves. Such an optimal detector is called the *uniformly most powerful* (UMP) test.

It is much more common if the signal or the noise (or both) are associated with some nuisance parameters so that at least one of the hypotheses is composite. In that case there will be a different ROC curve for different values of the parameters. This results in the paucity of cases when the UMP test is available.

4.3 Likelihood Ratio

Let θ denote an unknown parameter from parameter space Θ. Partitions Θ_0 and Θ_1 are nonempty subsets of Θ which separate Θ into two regions. These sets are disjoint such that $\Theta_0 \cap \Theta_1 = \varnothing$ and exhaustive such that $\Theta_0 \cup \Theta_1 = \Theta$. Then $p(x|\theta)$ denotes the conditional pdf of x given the occurrence of θ, and we assume that this pdf is known. The hypothesis test (4.1) can then be rewritten as

$$
\begin{aligned}
H_0 &: x \sim p(x|\theta), \theta \in \Theta_0 \\
H_1 &: x \sim p(x|\theta), \theta \in \Theta_1
\end{aligned}
\tag{4.8}
$$

For any pair of parameters $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta_1$, P_d and P_{fa} can be found by integrating the pdf over Γ_1,

$$
P_d = \int_{x \in \Gamma_1} p(x|\theta_1)dx
\tag{4.9}
$$

and

$$
P_{fa} = \int_{x \in \Gamma_1} p(x|\theta_0)dx
\tag{4.10}
$$

If the two hypotheses as expressed by (4.8) are simple (not composite), then $\Theta = \{\theta_0, \theta_1\}$. In that case, $\Theta_0 = \{\theta_0\}$ and $\Theta_1 = \{\theta_1\}$, each consisting of a single element. The Neyman-Pearson theorem [8] states that in this case there exists a UMP test such that P_d is maximized subject to a set level of P_{fa}; that is, $P_{fa} \leq \alpha$. The UMP test is called the *likelihood ratio test* and is given by

$$\lambda(x) \triangleq \frac{p(x|\theta_1)}{p(x|\theta_0)} \underset{H_0}{\overset{H_1}{\underset{<}{>}}} \gamma \qquad (4.11)$$

where γ is a threshold value determined by $\alpha = P_{fa}$. The value of γ is determined by

$$\alpha = \int_{\gamma}^{\infty} p(y|\theta_0) dy \qquad (4.12)$$

where $p(y|\theta)$ is the pdf of the likelihood ratio statistic $\lambda(x)$.

When one of the hypotheses is composite, then the Neyman-Pearson theorem does not apply. In that case the *generalized LRT* (GLRT) is often used. This is an ad hoc extension to the LRT and is given by

$$\lambda_G(x) \triangleq \frac{\max_{\theta_1 \in \Theta_1} p(x|\theta_1)}{\max_{\theta_0 \in \Theta_0} p(x|\theta_0)} \underset{H_0}{\overset{H_1}{\underset{<}{>}}} \eta \qquad (4.13)$$

That is, the GLRT replaces the unknown parameters with their maximum likelihood estimates in the pdfs. There is no underlying theory that guarantees that the GLRT will yield optimum results, but it generally produces good results in practice.

4.4 Hypothesis Tests

In general, the specific parameter of interest may not be measurable directly. In that case, one or more quantities related to the parameter of interest are measured instead. Such measurements and the ensuing judgments are called *tests*. Two of the more common types of tests are the aforementioned UMP test and the *locally most powerful* (LMP) test. Formally,

Definition. Let $w_i(\kappa)$ be the pdf for κ for hypothesis H_i. Then

$$p_i\left(y|H_i\right) = \int p_\kappa\left(y\right)w_i\left(\kappa\right)d\kappa \tag{4.14}$$

The UMP *test criterion*, given

$$
\begin{aligned}
H_0 &: P_\kappa \text{ for } \kappa \in \Gamma_0 \\
H_1 &: P_\kappa \text{ for } \kappa \in \Gamma_1
\end{aligned} \tag{4.15}
$$

is

$$\max_\delta P_d\left(\delta,\kappa\right) \qquad \forall \kappa \in \Gamma_1 \tag{4.16}$$

subject to

$$P_{fa}\left(\delta,\kappa\right) \le \alpha, \qquad \forall \kappa \in \Gamma_0 \tag{4.17}$$

∎

Hence, UMP tests are independent of the particular κ.

Definition. Given that

$$
\begin{aligned}
H_0 &: P_\kappa \text{ for } \kappa = \kappa_0 \\
H_1 &: P_\kappa \text{ for } \kappa \in \Gamma_1 = \left(\kappa_0,\infty\right)
\end{aligned} \tag{4.18}
$$

the LMP test is

$$\max_\delta \left. \frac{\partial}{\partial \kappa} P_d\left(\delta,\kappa\right)\right|_{\kappa=\kappa_0} \quad \text{subject to } P_{fa}\left(\delta\right) \le \alpha \tag{4.19}$$

so the likelihood ratio becomes

$$\lambda\left(y\right) = \frac{\left.\dfrac{\partial}{\partial \kappa} p_\kappa\left(y\right)\right|_{\kappa=\kappa_0}}{p_0\left(y\right)} \tag{4.20}$$

∎

For the generalized likelihood ratio test, the likelihood ratio is

$$\lambda_G(y) = \frac{\max\limits_{\kappa \in \Lambda_1} p_\kappa(y)}{\max\limits_{\kappa \in \Lambda_0} p_\kappa(y)} \tag{4.21}$$

4.4.1 Bayes Criterion

Given the observation Y, how is the hypothesis selected? One approach assigns cost functions to each of the possible decisions. These costs are denoted by C_{00}, C_{01}, C_{10}, and C_{11} corresponding to d_{00}, d_{01}, d_{10}, and d_{11}, respectively. These cost functions, denoted by C_{ij}, are defined as

$$C_{ij} = \text{cost of chosing } H_i \text{ when } H_j \text{ is true} \tag{4.22}$$

This approach was first investigated by Thomas Bayes (1702–1761), and carries his name. The costs are referred to as Bayesian costs. The Bayes rule is used to minimize Bayesian costs. Let π_j denote the a priori probability that H_j is true.
Let

$$P_j(\Gamma_i) = \text{probability of choosing } H_i \text{ given that } H_j \text{ is true} \tag{4.23}$$

then the conditional risk is given by

$$r_j(\delta) = C_{1j} P_j(\Gamma_1) + C_{0j} P_j(\Gamma_0) \tag{4.24}$$

Let the pdfs of H_0 and H_1 be denoted by $p_0(Y)$ and $p_1(Y)$, respectively. Then the probabilities of error are given by

$$P_0(\Gamma_1) = \int_\gamma^\infty p_0(Y) dY \tag{4.25}$$

and

$$P_1(\Gamma_0) = \int_{-\infty}^\gamma p_1(Y) dY \tag{4.26}$$

For example, if Y is a normal random variable, then

$$p_0(Y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(Y-\mu_0)^2}{2\sigma^2}\right] \qquad (4.27)$$

and

$$p_1(Y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(Y-\mu_1)^2}{2\sigma^2}\right] \qquad (4.28)$$

The Bayes (average) risk or cost is given by

$$
\begin{aligned}
r(\delta) &= \pi_0 r_0(\delta) + \pi_1 r_1(\delta) \\
&= \pi_0 \left[C_{00}P_0(\Gamma_1) + C_{10}P_0(\Gamma_1)\right] \\
&\quad + \pi_1 \left[C_{01}P_1(\Gamma_0) + C_{11}P_1(\Gamma_1)\right] \\
&= \pi_0 \left\{C_{00}\left[1 - P_0(\Gamma_0)\right] + C_{10}P_0(\Gamma_1)\right\} \\
&\quad + \pi_1 \left\{C_{01}\left[1 - P_1(\Gamma_1)\right] + C_{11}P_1(\Gamma_1)\right\} \\
&= \pi_0 \left[C_{00} - C_{00}P_0(\Gamma_1) + C_{10}P_0(\Gamma_1)\right] \\
&\quad + \pi_1 \left[C_{01} - C_{01}P_1(\Gamma_1) + C_{11}P_1(\Gamma_1)\right] \\
&= \pi_0 \left[C_{00} + P_0(\Gamma_1)(C_{10} - C_{00})\right] \\
&\quad + \pi_1 \left[C_{01} + P_1(\Gamma_1)(C_{11} - C_{01})\right] \\
&= \pi_0 C_{00} + \pi_0 (C_{10} - C_{00})P_0(\Gamma_1) \\
&\quad + \pi_1 C_{01} + \pi_1 (C_{11} - C_{01})P_1(\Gamma_1) \\
&= \pi_0 C_{00} + \pi_0 (C_{10} - C_{00})\int_{\Gamma_1} p_0(y)\mu(dy) \\
&\quad + \pi_1 C_{01} + \pi_1 (C_{11} - C_{01})\int_{\Gamma_1} p_1(y)\mu(dy)
\end{aligned}
\qquad (4.29)
$$

The optimal decision rule is known as the Bayes rule and is determined by choosing set Γ_1 to minimize $r(\delta)$. Thus, choose Γ_1 to be

$$\Gamma_1 = \left\{ x \in \Gamma \middle| \begin{array}{l} \pi_0 (C_{10} - C_{00}) p_0 (x) \\ \qquad + \pi_1 (C_{11} - C_{01}) p_1 (x) \leq 0 \end{array} \right\}$$

$$= \left\{ x \in \Gamma \middle| \begin{array}{l} \pi_1 (C_{11} - C_{01}) p_1 (x) \\ \qquad \leq \pi_0 (C_{00} - C_{10}) p_0 (x) \end{array} \right\} \tag{4.30}$$

It is normally the case that the costs of the mistakes are larger than the cost associated with making a right decision. Therefore, $C_{11} < C_{01}$ and $C_{00} < C_{10}$, so Γ_1 can be written as

$$\Gamma_1 = \left\{ Y \in \Gamma \middle| p_1 (Y) / p_0 (Y) \geq \gamma_B \right\} \tag{4.31}$$

where

$$\gamma_B = \frac{\pi_0 (C_{10} - C_{00})}{\pi_1 (C_{01} - C_{11})} \tag{4.32}$$

The Bayes rule is given by the likelihood test

$$\lambda(Y) = \frac{p_1(Y)}{p_0(Y)} \begin{cases} > \gamma_B \text{ choose } H_1 \\ < \gamma_B \text{ choose } H_0 \end{cases} \tag{4.33}$$

and

$$\delta_B(Y) = \begin{cases} 1, & \lambda(Y) > \gamma_B \\ 0, & \lambda(Y) < \gamma_B \end{cases} \tag{4.34}$$

Note that the region defined by $\left\{ Y \in \Gamma \middle| p_1 (Y) = \gamma_B p_0 (Y) \right\}$ does not contribute to the risk, so therefore it can be deleted or only part of it retained. Therefore, δ_B is not necessarily uniquely defined.

The conditional probability that H_j is true, given that the observation X takes on the value x, is given by

$$\pi_j(X) = P\big(H_j \text{ true}|Y = y\big) = \frac{p_j(Y)\pi_j}{p(Y)} = \frac{p_j(Y)\pi_j}{p_0(Y)\pi_0 + p_1(Y)\pi_1} \quad (4.35)$$

The probabilities $\pi_0(Y)$ and $\pi_1(Y)$ are called the *a posteriori probabilities* of the hypotheses. The Bayes decision rule can be written in terms of the a posteriori probabilities as

$$\delta_B(Y) = \begin{cases} 1, & \pi_1(Y) \geq \pi_0(Y) \\ 0, & \pi_1(Y) < \pi_0(Y) \end{cases} \quad (4.36)$$

Thus, the minimum probability of error decision rule chooses the hypothesis with the maximum a posteriori probability of occurrence, given that $Y = y$. A decision rule such as this is referred to as a *maximum a posteriori* (MAP) decision rule.

Suppose the hypothesis in question is given by

$H_0: Y \sim N(\mu_0, \sigma^2)$
$H_1: Y \sim N(\mu_1, \sigma^2)$

where μ_0 and μ_1 are two fixed numbers with $\mu_1 > \mu_0$. The likelihood ratio for this test is given by

$$\lambda(Y) = \frac{p_1(Y)}{p_0(Y)} = \frac{\dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y-\mu_1)^2}{2\sigma^2}}}{\dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y-\mu_0)^2}{2\sigma^2}}}$$

$$= \exp\left[\frac{\mu_1 - \mu_0}{\sigma^2}\left(x - \frac{\mu_1 + \mu_0}{2}\right)\right] \quad (4.37)$$

with the corresponding Bayes test

$$\delta_B(Y) = \begin{cases} 1, & \exp\left[\dfrac{\mu_1 - \mu_0}{\sigma^2}\left(x - \dfrac{\mu_1 + \mu_0}{2}\right)\right] > \gamma_B \\ 0, & \text{otherwise} \end{cases} \quad (4.38)$$

Since the log is a monotonically increasing function of its argument, the test parameter can be put into a different form as

$$\delta_B(Y) = \begin{cases} 1, & \dfrac{\mu_1 - \mu_0}{\sigma^2}\left(x - \dfrac{\mu_1 + \mu_0}{2}\right) > \log \gamma_B \\ 0, & \text{otherwise} \end{cases}$$

$$= \begin{cases} 1, & x > \dfrac{\sigma^2}{\mu_1 - \mu_0} \log \gamma_B + \dfrac{\mu_1 + \mu_0}{2} \\ 0, & \text{otherwise} \end{cases} \tag{4.39}$$

If the priors are equal and the costs are uniform, then

$$\delta_B = \begin{cases} 1, & Y > +\dfrac{\mu_1 + \mu_0}{2} \\ 0, & \text{otherwise} \end{cases} \tag{4.40}$$

which is illustrated in Figure 4.4.
In this case, $P_0(\Gamma_1) = P_1(\Gamma_0)$, so

$$r(\delta_B) = \frac{1}{2}P_1(\Gamma_0) + \frac{1}{2}P_0(\Gamma_1) = P_0(\Gamma_1)$$

$$= \int_{\frac{\mu_1 + \mu_0}{2}}^{\infty} p_0(Y)dY$$

$$= \frac{1}{\sqrt{2\pi}} \int_{\frac{\mu_0 + \mu_1}{2} - \mu_0}^{\infty} e^{-\frac{Y^2}{2}} dY$$
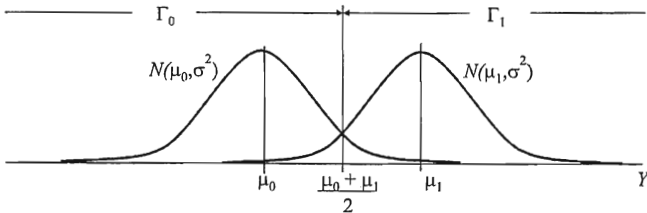

Figure 4.4 Binary hypothesis test with normal pdfs, uniform costs, and equal priors.
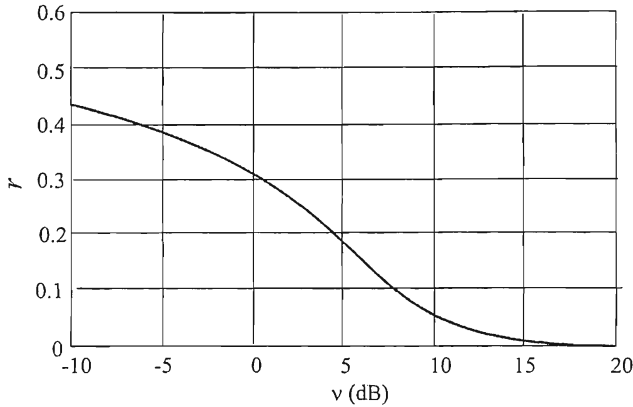
Figure 4.5 Bayes cost for normal pdfs, uniform costs, and equal priors.

$$= 1 - \Phi\left(\frac{\mu_1 - \mu_0}{2\sigma}\right)$$

$$= 1 - \Phi\left(\frac{\xi}{2}\right) \tag{4.41}$$

where Φ is the distribution function for the standard normal random variable $\mathcal{N}(0,1)$ and

$$\nu = \frac{(\mu_1 - \mu_0)^2}{\sigma^2} \tag{4.42}$$

is the SNR. The Bayes risk in this case as a function of the SNR is illustrated in Figure 4.5.

4.4.2 Minimax Criterion

In many practical cases of interest, the a priori probabilities are not known or, perhaps, not knowable. The minimax criterion is useful in this case. The minimax risk is given by

$$r(\pi, \delta) = \pi r_0(\delta) + (1 - \pi) r_1(\delta) \tag{4.43}$$

When δ is fixed, $r(\pi, \delta)$ is an affine function[1] of π, so

$$\max_{0 \leq \pi \leq 1} r(\pi, \delta) = \max \left\{ r_0(\delta), r_1(\delta) \right\} \tag{4.44}$$

The goal is to find the decision rule δ^* that minimizes (4.43); that is,

$$\max_{0 \leq \pi \leq 1} r(\pi, \delta^*) = \min_{\delta} \max_{0 \leq \pi \leq 1} r(\pi, \delta) \tag{4.45}$$

In order to find this decision rule, assuming it exists, we assume that there is an associated probability π^* such that

$$\max_{0 \leq \pi \leq 1} r(\pi, \delta^*) \leq r(\pi^*, \delta^*) \leq \min_{\delta} r(\pi^*, \delta) \tag{4.46}$$

Such a π^* is the *least favorable* probability for δ^*, while δ^* is a Bayes rule for π^*.
  Now, for any pair (π, δ) it is necessarily true that

$$\min_{\delta} r(\pi^*, \delta) \leq r(\pi^*, \delta^*) \leq \max_{0 \leq \pi \leq 1} r(\pi, \delta^*) \tag{4.47}$$

Therefore, for the pair that satisfies (4.46),

$$\max_{0 \leq \pi \leq 1} r(\pi, \delta^*) = r(\pi^*, \delta^*) = \min_{\delta} r(\pi^*, \delta) \tag{4.48}$$

Therefore,

$$r(\pi^*, \delta^*) = \min_{\delta} r(\pi^*, \delta)$$
$$\leq \max_{0 \leq \pi \leq 1} \min_{\delta} r(\pi, \delta)$$

---

[1] An affine function or transformation is one that is invariant to a linear transformation followed by a translation. In one dimension, an affine function is of the form $y = ax + b$; that is, it is composed of a linear function plus a constant and the graph of y is a straight line. In two dimensions it is of the form $z = ax + by + c$, and the graph is still a straight line. This can be extended to any number of dimensions. The properties of such a function are:

- If three points are on a straight line before the transformation, they will still be on the line after the transformation;
- Parallel lines remain parallel;
- The ratio of the length of line segments remains constant after the transformation;
- Quadratic functions remain the same after the transformation—that is, parabolas remain parabolas, ellipses remain ellipses.

$$\leq \min_{\delta} \max_{0 \leq \pi \leq 1} r(\pi, \delta)$$

$$\leq \max_{0 \leq \pi \leq 1} r(\pi, \delta^*) = r(\pi^*, \delta^*)$$

or

$$\max_{0 \leq \pi \leq 1} r(\pi, \delta^*) = r(\pi^*, \delta^*) = \min_{\delta} \max_{0 \leq \pi \leq 1} r(\pi, \delta) \qquad (4.49)$$

which means that δ^* is a minimax rule. In addition,

$$\max_{0 \leq \pi \leq 1} \min_{\delta} r(\pi, \delta) = \min_{\delta} r(\pi^*, \delta) \qquad (4.50)$$

which states that π^* is a *least favorable probability distribution* in terms of the minimum Bayes risk.

Property: Minimax

If π^* is a least favorable distribution associated with δ^*, and one of the following conditions is true:

    (1) $r_0(\delta^*) = r_1(\delta^*)$
    (2) $\pi^* = 0$ and $r_0(\delta^*) \leq \delta^*$
    (3) $\pi^* = 1$ and $r_0(\delta^*) \geq r_1(\delta^*)$

then δ^* is a minimax rule.

    Let

$$V(\pi) = \min_{\delta} r(\pi, \delta) = \text{minimum Bayes risk for } \pi \qquad (4.51)$$

$V(\pi)$ is a concave and continuous function and, given that $C_{ij} \geq C_{jj}$ for $i \neq j$, $V(0) = C_{11}$ and $V(1) = C_{00}$.

The minimax rule is then given by

$$\min_{\delta} \max_{0 \leq \pi \leq 1} \left[r_0(\delta), r_1(\delta) \right] = \min_{\delta} \max_{0 \leq \pi \leq 1} r(\pi, \delta)$$

$$= \max_{0 \leq \pi \leq 1} \min_{\delta} r(\pi, \delta) = \max_{0 \leq \pi \leq 1} V(\pi) \qquad (4.52)$$

The minimax rule is the Bayes rule for π^*, where

$$\pi^* = \arg \max_{0 \le \pi \le 1} V(\pi) \tag{4.53}$$

4.4.3 Neyman-Pearson Criterion

If neither the a priori probabilities nor costs are available, then the Bayes criterion cannot be applied to ascertain which hypotheses to adopt. In that case, the Neyman-Person (NP) criteria can be used. The NP criterion maximizes P_d while simultaneously limiting the value of P_{fa}.

As indicated above, the power of a test is given by the probability of detection $P_d = P_1(\Gamma_1)$. The significance of that test, α, is the maximum probability of false alarm, $P_{fa} = P_1(\Gamma_0)$, that is tolerable.

The *power function* is a graph of P_d versus SNR for a fixed α. An example of a power function is illustrated in Figure 4.6.

The randomized decision rule is a generalization of the decision rule discussed above,
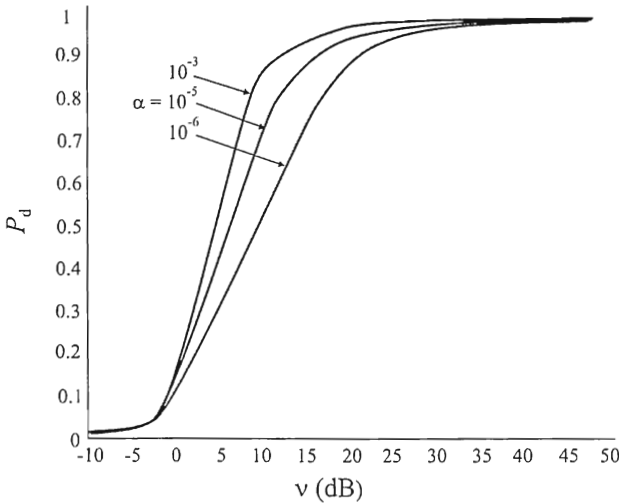

Figure 4.6 Example of a statistical power function.

$$\delta(y) = \begin{cases} 1 \text{ if } \lambda(y) > \gamma \\ \eta \text{ if } \lambda(y) = \gamma \\ 0 \text{ if } \lambda(y) < \gamma \end{cases} \tag{4.54}$$

Then

$$P_{fa}(\delta) = \mathcal{E}_0\{\delta(Y)\} = \int_\Gamma \delta(y) p_0(y) \mu(dy)$$
$$= P_0[\lambda(Y) > \gamma] + \eta P_0[\lambda(Y) = \gamma] \tag{4.55}$$

and

$$P_d(\delta) = \mathcal{E}_1\{\delta(Y)\} = \int_\Gamma \delta(y) p_1(y) \mu(dy)$$
$$= P_1[\lambda(Y) > \gamma] + \eta P_1[\lambda(Y) = \gamma] \tag{4.56}$$

The NP rule is a likelihood ratio test where the threshold γ_0 and the randomization η_0 are chosen such that $P_{fa} = \alpha$. That is, the false alarm probability is fixed at some desired value. Such testing is also called *constant false alarm rate* (CFAR) testing.

4.5 Multiple Measurements

The more usual case is when there are several measurements that can be combined and used in some fashion to make the decision. For a time series of samples, for example, perhaps 64 sequential samples of the signal are available.

For example, given that H_1 is true, let D_i be a 1 if the detector correctly detects the signal present during test i and 0 if the detector detects incorrectly during test i. Then, if there are M frames of data,

$$\hat{P}_d = \frac{1}{M} \sum_{i=1}^{M} D_i \tag{4.57}$$

is an estimate of the probability of detection. Generating an estimate for the probability of false alarm, $\hat{P}_{fa}$ is done in the same way. Let $D_i = 1$ if the detector incorrectly detects the signal during test i and 0 if the detector correctly rejects the signal during test i. Then, for the M frames of data

$$\hat{P}_{\text{fa}} = \frac{1}{M} \sum_{i=1}^{M} D_i \qquad (4.58)$$

4.6 Multiple Hypotheses

The notions introduced in prior sections assumed that there was only one unknown parameter and that parameter could assume only one value. With testing of multiple hypotheses, that parameter can take on many values, with a hypothesis associated with each one. In that case, there is more than one hypothesis associated with each decision region. Symbolically,

$$H_{i,\kappa} \sim P_{\kappa}, \qquad \kappa \in \Lambda_i \qquad (4.59)$$

That is, hypothesis $H_{i,\kappa}$ has distribution P_{κ}, where κ parameterizes the distribution and κ is drawn from set Λ_i.

4.7 Concluding Remarks

Hypothesis testing for signal detection is the process of making guesses (hypotheses) about whether a signal is present based on the data series received. A threshold is established based on whichever technique is being used. If some aspect of the data series exceeds that threshold, signal presence is declared. If it does not exceed that threshold, then signal absence is declared.

There are two types of errors that can occur in this process: declaring that a signal is not present when it is (missed detection) and declaring that a signal is present when it is not (false alarm).

Optimal signal detectors can be designed based on the criteria presented in this chapter. The definition of optimum depends on the type of detection technique used. The Bayes criterion minimizes the overall cost (risk) in the detection process. The Bayes technique, however, requires that the a priori probabilities are known or can be estimated. When that is not the case, the minimax criterion can be applied. The NP detection criteria assures that the probability of detection is maximized while keeping the probability of false alarm below a prescribed level.

References

[1]   Whalen, A. D., *Detection of Signals in Noise,* New York: Academic Press, 1971, Chapters 5–9.

[2]   Roberts, L. R., *Signal Processing Techniques,* Anaheim, CA: Interstate Electronics Corporation, 1981, Chapter 4.

[3]   Van Trees, H. L., *Detection, Estimation, and Modulation Theory,* Part I, New York: John Wiley & Sons, 1968, Chapter 2.

[4]   Neyman, J., and E. Pearson, "On the Problem of the Most Efficient Tests of Statistical Hypotheses," *Philadelphia Transactions of the Royal Society,* Series A, 231, 1933, pp. 289–337.

[5]   Helstrom, C. W., *Statistical Theory of Signal Detection,* New York: Pergamon Press, 1960, Chapter 4.

[6]   Whalen, A. D., *Detection of Signals in Noise,* New York: Academic Press, 1971, Chapter 4.

[7]   Johnson, D. H., and D. E. Dudgeon, *Array Signal Processing Concepts and Techniques,* Upper Saddle River, NJ: Prentice-Hall, 1993, Chapter 4.

[8]   Hoel, P. G., S. C. Port, and C. J. Stone, *Introduction to Statistical Theory,* New York: Houghton Mifflin, 1971, pp. 56–67.

Chapter 5

Target Parameter Estimation

The process of recognizing targets in communication EW systems generally consists of two steps:

1. Extract parameters (also called features) from the intercepted signal by measurement;
2. Process these parameters by using some applicable technique to ascertain whether the signal is one of interest, referred to as an SOI.

This process is illustrated in Figure 5.1, which also lists some of the parameters used when recognizing communication EW targets.

5.1 Signal Parameter Estimation

Parameter estimation is the process of recovering an estimate of some feature of a signal by measurements of some sort. Signals are typically corrupted by noise, and any measurements are therefore corrupted. As opposed to the signal detection problem discussed in Chapters 7 and 8, where the results of the detection process result in a finite number of answers (in the binary case, for example, the answer takes on only two values), the parameters of interest can take on an infinite number of values. These values, however, are normally confined to a certain range [1–3].

Parameter estimation is important in EW systems for several reasons. Parameters of interest include the angle of arrival of the signal, or the time of arrival for the purpose of establishing the geolocation of a target. The modulation type is also of interest so that signal sorting based on targets of interest can be performed. The phase angle of a signal can be important as well. The amplitude, or equivalently, the power of a signal, is important. These and other parameters

71
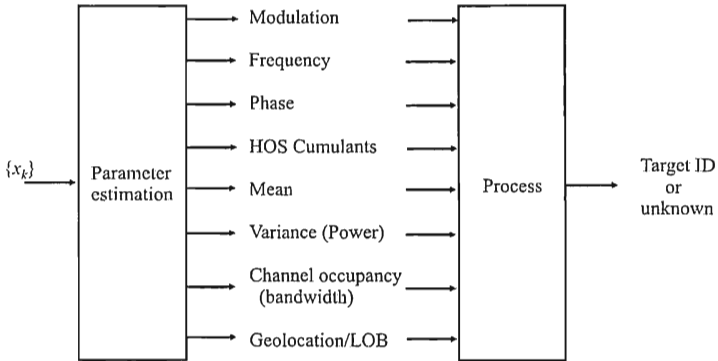
Figure 5.1 The communication signal target recognition process.

typically must be determined essentially in real time across a broad frequency range.

The signal $x(t)$ for which parameters are to be estimated is assumed to be sampled at regular intervals generating the sample set $\{x_k\}$. Signal $x(t)$ could be a predetection signal (prior to demodulation) or could be demodulated.

The concept of composite hypothesis testing introduced in this chapter can be used for the purpose of signal classification, recognition, or otherwise estimating some parameters associated with target signals.

The *Fisher information matrix* and its inverse, the *Cramer-Rao lower bound* (CRLB), quantify how well a specific set of parameters can be estimated when there are other, unknown nuisance parameters present that influence the parameters of interest. The entries in the Fisher information matrix consist of the Fisher information, which is a measure of the information content of the measured signal relative to a particular parameter. The Cramer-Rao bound is a lower bound on the error variance of an unbiased best estimator for estimating this parameter when the noise is AWGN.

For the multivariate parameter estimation problem, the information available consists of θ, the unknown parameters—that is, it is known which parameters are to be estimated; x, the measurements or observed variables; $p(x; \theta)$, the pdf x parameterized by θ (often called the *likelihood function*); and $p(x|\theta)$ the conditional pdf of x given θ. The goal is to find a function of the data that provides the best "guess" of θ from x. This function is denoted either as $\hat{q}$ or $\hat{q}(x)$.

As an example of an estimator, consider the navigation systems on an airplane. The three dimensional forces on accelerometers are the measurements, while the

parameters to be estimated are the three dimensional location of the aircraft. By Newton's second law of mechanics, $\mathbf{F} = m\mathbf{a}$, where $\mathbf{F}$ are the forces, m is the mass of the accelerometer, and $\mathbf{a}$ is the acceleration vector. Since the velocity is the time integral of the acceleration, it can be estimated by measuring these forces. Likewise, the location is given by the time integral of the velocity, so it too can be estimated.

As in detection theory, there are several approaches to estimation. Bayesian estimation is in common use, where the underlying first principle is the Bayes theorem from probability theory. In Bayesian estimation, it is assumed that θ consists of random variables whose pdfs, $p(\theta)$, are known; the conditional densities $p(\mathbf{x}|\mathbf{q})$ are also known. Bayesian estimation also requires cost functions $C(\theta, \hat{\theta})$, which are the costs, or penalties incurred, for estimating θ with $\hat{\mathbf{q}}$. Usually (but not always), these cost functions are the squared error in the estimate. When the cost functions are the squared error, the optimal Bayesian estimates are $\mathcal{E}\{\theta|\mathbf{x}\}$, the conditional means of the unknown parameters given the observed data.

Minimum mean square error estimation (MMSE) leads to the Weiner and Kalman filters, which optimally estimate (in a mean square error sense) signal parameters. In MMSE the performance measure that is optimized is the mean square error of the estimate from the true value of the parameter.

Discussed in detail here is *maximum likelihood estimation* (MLE). In MLE, it is assumed that θ is unknown and nonrandom. The likelihood function is the pdf of $\mathbf{x}$ parameterized by θ.

5.2 The Cramer-Rao Bound

The *Cramer-Rao lower bound* (CRLB) provides a lower bound on the variance of an unbiased estimator. It takes on different forms depending on whether there is one or more than one unknown parameter to be estimated. In the former case, the CRLB is a scalar, and in the latter, the CRLB is a vector.

An estimator is called *efficient* if it achieves the CRLB. An estimator is *consistent* if the estimate $\hat{\theta}$ approaches θ in some probabilistic sense. The *bias* of an estimator is defined as $\mathcal{E}\{\hat{\mathbf{q}}\}$. $\hat{\theta}$ is an *unbiased estimator* if its bias is zero: $\mathcal{E}\{\hat{\theta}\} - \theta = 0$, or $\mathcal{E}\{\hat{\theta}\} = \theta$.

The *estimate variance* of $\hat{\theta}$ is

$$\sigma^2 = \mathcal{E}\{\hat{\theta}^2\} - \mathcal{E}^2\{\hat{\theta}\} \tag{5.1}$$

Example:

The estimate variance of the sample mean

$$\hat{\theta} = \frac{1}{N}\sum_{k=0}^{N-1} x_k \tag{5.2}$$

is found as follows. The mean square value is

$$
\begin{aligned}
\mathcal{E}\{\hat{\theta}^2\} &= \frac{1}{N^2}\mathcal{E}\left\{\left(\sum_{k=0}^{N-1} x_k\right)^2\right\} \\
&= \frac{1}{N^2}\left[\sum_{i=1}^{N-1}\mathcal{E}\{x_i^2\} + \sum_{i=0}^{N-1}\sum_{\substack{j=0 \\ j\neq i}}^{N-1}\mathcal{E}\{x_i\}\mathcal{E}\{x_j\}\right] \\
&= \frac{1}{N^2}\left[N\mathcal{E}\{x_n^2\} + \sum_{i=0}^{N-1}\mathcal{E}\{x_i\}\sum_{\substack{j=0 \\ j\neq i}}^{N-1}\mathcal{E}\{x_j\}\right] \\
&= \frac{1}{N}\mathcal{E}\{x_n^2\} + \mu_x^2\frac{N-1}{N} \tag{5.3}
\end{aligned}
$$

where μ_x is the mean value of x_i, assumed the same for all i. Thus, the estimate variance is

$$
\begin{aligned}
\sigma^2 &= \mathcal{E}\{\hat{\theta}^2\} - \mathcal{E}^2\{\hat{\theta}_x\} \\
&= \frac{1}{N}\mathcal{E}\{x_n^2\} + \mu_x^2\frac{N-1}{N} - \mathcal{E}^2\{\hat{\theta}_x\} \\
&= \frac{1}{N}\left[\mathcal{E}\{x_n^2\} - \mu_x^2\right] \\
&= \frac{1}{N}\sigma_x^2 \tag{5.4}
\end{aligned}
$$

■

This example illustrates the fact that as the number of observations, N, increases, the variance of the sample mean decreases. Since the bias is zero, the sample mean is a consistent estimator.

Example:

Suppose the mean value, θ, is known and the variance is to be estimated. Then

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=0}^{N-1}(x_i - \theta)^2 \tag{5.5}$$

which is a consistent estimator.

∎

Example:

Suppose that both the mean value and the variance are to be estimated. Then

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=0}^{N-1}(x_i - \hat{\theta})^2 \tag{5.6}$$

where $\hat{\theta}$ is the sample mean.

∎

The only difference between (5.5) and (5.6) is that (5.5) uses the true mean value while (5.6) uses the sample mean. The mean value of the variance estimate is determined as

$$
\begin{aligned}
\mathcal{E}\{\hat{\sigma}^2\} &= \frac{1}{N}\sum_{i=0}^{N-1}[\mathcal{E}\{x_i\} - \mathcal{E}\{\hat{\theta}\}]^2 \\
&= \frac{1}{N}\sum_{i=0}^{N-1}\left[\mathcal{E}\{x_i^2\} - 2\mathcal{E}\{x_i\hat{\theta}\} + \mathcal{E}\{\hat{\theta}^2\}\right] \\
&= \frac{1}{N}\sum_{i=0}^{N-1}\mathcal{E}\{x_i^2\} - \frac{2}{N^2}\sum_{i=0}^{N-1}\left[\sum_{j=0}^{N-1}\mathcal{E}\{x_i x_j\}\right] + \frac{1}{N^2}\sum_{i=0}^{N-1}\left[\sum_{j=0}^{N-1}\mathcal{E}\{x_i x_j\}\right] \\
&= \frac{1}{N}\sum_{i=0}^{N-1}\mathcal{E}\{x_i^2\} - \frac{2}{N^2}\left[\sum_{i=0}^{N-1}\mathcal{E}\{x_i^2\} + \sum_{i=0}^{N-1}\left(\sum_{\substack{j=0\\j\neq i}}^{N-1}\mathcal{E}\{x_i\}\mathcal{E}\{x_j\}\right)\right] \\
&\quad + \frac{1}{N^2}\left[\sum_{i=0}^{N-1}\mathcal{E}\{x_i^2\} + \sum_{i=0}^{N-1}\left(\sum_{j=0}^{N-1}\mathcal{E}\{x_i\}\mathcal{E}\{x_j\}\right)\right]
\end{aligned}
$$

$$= \frac{1}{N} N\mathcal{E}\{x_i^2\} - \frac{2}{N^2}\left[N\mathcal{E}\{x_i^2\} + N(N-1)\theta^2\right]$$

$$+ \frac{1}{N^2}\left[N\mathcal{E}\{x_i^2\} + N(N-1)\theta^2\right]$$

$$= \frac{1}{N} N\mathcal{E}\{x_i^2\} - \frac{2N}{N^2}\mathcal{E}\{x_i^2\} - \frac{2N(N-1)}{N^2}\theta^2$$

$$+ \frac{N}{N^2}\mathcal{E}\{x_i^2\} + \frac{N(N-1)}{N^2}\theta^2$$

$$= \frac{1}{N} N\mathcal{E}\{x_i^2\} - \frac{2}{N}\mathcal{E}\{x_i^2\} - \frac{2(N-1)}{N}\theta^2$$

$$+ \frac{1}{N}\mathcal{E}\{x_i^2\} + \frac{N-1}{N}\theta^2$$

$$= \frac{1}{N}\left[N - \mathcal{E}\{x_i^2\}\right] - \frac{1}{N}\mathcal{E}\{x_i^2\} - \frac{N-1}{N}\theta^2$$

$$= \frac{N-1}{N}\mathcal{E}\{x_i^2\} - \frac{N-1}{N}\theta^2$$

$$= \frac{N-1}{N}\sigma^2 \tag{5.7}$$

Thus, the mean of the sample variance is biased away from the true variance by a factor of $(N-1)/N$. As N becomes large, however, this factor approaches unity, and therefore the sample variance is asymptotically unbiased.

To determine whether this estimator is consistent, assume the mean is zero for convenience. Let $\zeta = \hat{\sigma}^2$. Then

$$\zeta = \frac{1}{N}\sum_{i=1}^{N-1} x_i^2$$

and

$$\mathcal{E}\{\zeta^2\} = \frac{1}{N^2}\sum_{i=0}^{N-1}\left[\sum_{j=0}^{N-1}\mathcal{E}\{x_i^2 x_j^2\}\right]$$

$$= \frac{1}{N^2}\left[N\mathcal{E}\{x_n^4\} + N(N-1)\mathcal{E}^2\{x_n^2\}\right]$$

$$= \frac{1}{N}\left[\mathcal{E}\{x_n^4\}+(N-1)\mathcal{E}^2\{x_n^2\}\right]$$

but

$$\mathcal{E}\{\zeta\} = \mathcal{E}\{x_n^2\}$$

so

$$
\begin{aligned}
\text{var}(\hat{\sigma}^2) &= \mathcal{E}\{\zeta^2\}-\mathcal{E}^2\{\zeta\} \\
&= \frac{1}{N}\left[\mathcal{E}\{x_n^4\}-\mathcal{E}^2\{x_n^2\}\right]
\end{aligned}
\tag{5.8}
$$

Therefore, based on (5.7) and (5.8), as N gets large the sample variance produces a consistent estimate. ∎

5.2.1 CRLB for Signals in AWGN

5.2.1.1 Scalar CRLB

The scalar CRLB is an indication of how well a single parameter of a signal, such as its amplitude, frequency, phase, or direction of arrival, can be estimated. The signal is given by

$$x_k = s_{\theta k} + n_k, \qquad k = 0,1,\cdots,N-1 \tag{5.9}$$

where θ is unknown, $n_k \sim \mathcal{N}(0,\ \sigma^2)$, and the x_k are *independent and identically distributed* (i.i.d.).

The scalar CRLB is given by the negative expected value of the inverse of the second derivative of the natural logarithm of the pdf of the x_k, parameterized on the parameter to be estimated. This falls out of the definition of Fisher information. Denote the CRLB with $\text{var}(\hat{\theta})$. Then

$$\text{var}(\hat{\theta}) = -\mathcal{E}\left\{\left[\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2}\right]^{-1}\right\} \tag{5.10}$$

For a signal in AWGN, the CRLB is determined as follows.

$$p(x_0,\cdots,x_{N-1};\theta) = \frac{1}{(2\pi)^{N/2}\sigma^N} \exp\left[-\frac{1}{2\sigma^2}\sum_{k=0}^{N-1}(x_k - s_{\theta k})^2\right]$$

$$\ln p(x_0,\cdots,x_{N-1};\theta) = -\ln(2\pi)^{N/2}\sigma^N - \frac{1}{2\sigma^2}\sum_{k=0}^{N-1}(x_k - s_{\theta k})^2$$

$$\frac{\partial \ln p(x_0,\cdots,x_{N-1};\theta)}{\partial \theta} = \frac{1}{\sigma^2}\sum_{k=0}^{N-1}(x_k - s_{\theta k})\frac{\partial s_{\theta k}}{\partial \theta}$$

$$\frac{\partial^2 \ln p(x_0,\cdots,x_{N-1};\theta)}{\partial \theta^2} = \frac{1}{\sigma^2}\sum_{k=0}^{N-1}(x_k - s_{\theta k})\frac{\partial^2 s_{\theta k}}{\partial \theta^2} - \left(\frac{\partial s_{\theta k}}{\partial \theta}\right)^2$$

Now

$$\mathcal{E}\left\{\frac{\partial^2 \ln p(x_0,\cdots,x_{N-1};\theta)}{\partial \theta^2}\right\} = \mathcal{E}\left\{\frac{1}{\sigma^2}\sum_{k=0}^{N-1}(x_k - s_{\theta k})\frac{\partial^2 s_{\theta k}}{\partial \theta^2} - \left(\frac{\partial s_{\theta k}}{\partial \theta}\right)^2\right\}$$

$$= \frac{1}{\sigma^2}\sum_{k=0}^{N-1}\left[(\mathcal{E}\{x_k\} - \mathcal{E}\{s_{\theta k}\})\frac{\partial^2 s_{\theta k}}{\partial \theta^2} - \mathcal{E}\left\{\left(\frac{\partial s_{\theta k}}{\partial \theta}\right)^2\right\}\right] \quad (5.11)$$

$$= \frac{1}{\sigma^2}\sum_{k=0}^{N-1}\left[(\mathcal{E}\{x_k\} - \mathcal{E}\{s_{\theta k}\})\frac{\partial^2 s_{\theta k}}{\partial \theta^2} - \mathcal{E}\left\{\left(\frac{\partial s_{\theta k}}{\partial \theta}\right)^2\right\}\right]$$

but

$$\mathcal{E}\{x_k\} = \mathcal{E}\{s_k + n_k\} = \mathcal{E}\{s_k\} + \mathcal{E}\{n_k\} = \mathcal{E}\{s_k\} = \mu_x \quad (5.12)$$

and the last term in the summation in (5.11) is given by

$$\mathcal{E}\left\{\left(\frac{\partial s_{\theta k}}{\partial \theta}\right)^2\right\} = \left(\frac{\partial s_{\theta k}}{\partial \theta}\right)^2 \quad (5.13)$$

so

$$\mathcal{E}\left\{\frac{\partial^2 \ln p(x_0,\cdots,x_{N-1};\theta)}{\partial \theta^2}\right\} = \frac{1}{\sigma^2}\sum_{k=0}^{N-1}\left[(\mu_x - \mu_x)\frac{\partial^2 s_{\theta k}}{\partial \theta^2} - \left(\frac{\partial s_{\theta k}}{\partial \theta}\right)^2\right]$$

$$= -\frac{1}{\sigma^2} \sum_{k=0}^{N-1} \left(\frac{\partial s_{\theta k}}{\partial \theta} \right)^2 \tag{5.14}$$

So the CRLB for signals in AWGN is

$$\text{var}\left(\hat{\theta}\right) \geq \frac{\sigma^2}{\sum_{k=0}^{N-1} \left(\frac{\partial s_{\theta k}}{\partial \theta} \right)^2} \tag{5.15}$$

An unbiased estimator that is an efficient estimator exists if and only if:

$$\frac{\partial \ln p(x;\theta)}{\partial \theta} = F(\theta)\left[g(x) - \theta\right] \tag{5.16}$$

If this is true,

$$\hat{\theta} = g(x) \tag{5.17}$$

∎

Example:

Suppose the measurement sample is given by

$$x_0 = \theta + n_0 \tag{5.18}$$

where θ is an unknown constant that is to be estimated and $n_0 \sim \mathcal{N}(0, \sigma^2)$. The variance of the unbiased estimator with the smallest variance is found as

$$p(x_0;\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x_0 - \theta)^2 \right]$$

$$\ln p(x_0;\theta) = -\ln \sqrt{2\pi\sigma^2} - \left[\frac{1}{2\sigma^2}(x_0 - \theta)^2 \right]$$

$$\frac{\partial \ln p(x_0;\theta)}{\partial \theta} = \frac{1}{\sigma^2}(x_0 - \theta) \tag{5.19}$$

$$\frac{\partial^2 \ln p(x_0;\theta)}{\partial \theta^2} = -\frac{1}{\sigma^2}$$

$$-\mathcal{E}\left\{\frac{\partial^2 \ln p(x_0;\theta)}{\partial \theta^2}\right\} = \frac{1}{\sigma^2}$$

so

$$\text{var}\left(\hat{\theta}\right) \geq \sigma^2 \tag{5.20}$$

The efficient estimator is determined from (5.16) and (5.19). If $F(\theta) = 1/\sigma^2$ and $g(x_0) = x_0$, then the estimator is

$$\hat{\theta} = x_0 \tag{5.21}$$

∎

Example:

Suppose there are N i.i.d. samples with an exponential density with parameter θ. The CRLB for this parameter can be estimated as follows. The exponential density is given by

$$p(x_0,\cdots,x_{N-1};\theta) = \theta^N \exp\left[-\theta\sum_{k=0}^{N-1} x_k\right] \tag{5.22}$$

so

$$\ln p(x_0,\cdots,x_{N-1};\theta) = N \ln \theta - \theta\sum_{k=0}^{N-1} x_k$$

$$\frac{\partial \ln p(x_0,\cdots,x_{N-1};\theta)}{\partial \theta} = \frac{N}{\theta} - \sum_{k=0}^{N-1} x_k \tag{5.23}$$

$$\frac{\partial^2 \ln p(x_0,\cdots,x_{N-1};\theta)}{\partial \theta^2} = -\frac{N}{\theta^2}$$

and finally

$$\text{var}\left(\hat{\theta}\right) \geq \frac{1}{N}\theta^2 \tag{5.24}$$

In this case, the CRLB depends on the value of the unknown θ that is being estimated. Also, (5.23) cannot be factored into $F(\theta)$ and $g(x_0)$, so the estimator cannot be stated easily.

■

Example:

Suppose the signal is a cosine wave and that the amplitude is not known. Then

$$x_k = \theta \cos(2\pi f_0 k) + n_k \tag{5.25}$$

and θ is to be estimated. In this case, the frequency f_0 is known. The CRLB is found as

$$s_{\theta k} = \theta \cos(2\pi f_0 k)$$
$$\frac{\partial s_{\theta k}}{\partial \theta} = \cos(2\pi f_0 k)$$

and

$$\text{var}(\hat{\theta}) \geq \frac{\sigma^2}{\displaystyle\sum_{k=0}^{N-1} \cos^2(2\pi f_0 k)} \tag{5.26}$$

The denominator of (5.26) is plotted in Figure 5.2 for f_0 = 0.1, 0.2, and 0.3. The curves are virtually identical for all values of N considered. Furthermore, the denominator is well approximated by $N/2$. Thus,
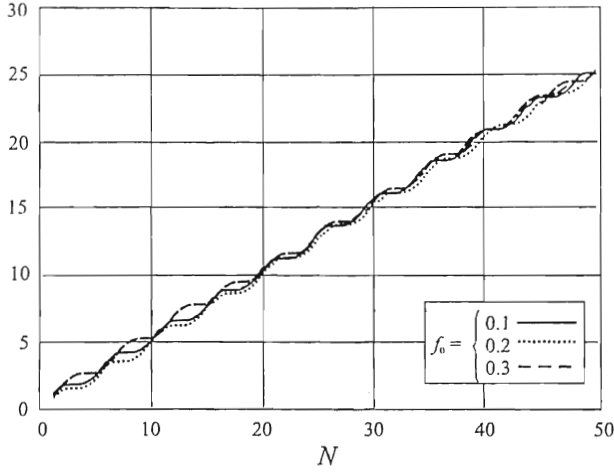
$$\text{var}(\hat{\theta}) \geq \frac{\sigma^2}{N/2} \tag{5.27}$$

■

Example:

In this case, the amplitude is known but the frequency is not. Then

$$x_k = \sqrt{2S} \cos(2\pi\theta k) + n_k \tag{5.28}$$

where S is the average power in the signal. In this case,

Figure 5.2 Denominator of the variance for cosine wave with unknown amplitude. The three curves are for $f_0 = 0.1$, 0.2, and 0.3.

$$s_{\theta k} = \sqrt{2S} \cos(2\pi\theta k) \qquad (5.29)$$

$$\frac{\partial s_{\theta k}}{\partial \theta} = \sqrt{2S}\, 2\pi k \sin(2\pi\theta k) \qquad (5.30)$$

and

$$\begin{aligned}
\mathrm{var}(\hat{\theta}) &\geq \frac{\sigma^2}{\displaystyle\sum_{k=0}^{N-1}\left(\sqrt{2S}\,2\pi k\right)^2 \sin^2(2\pi\theta k)} \\
&= \frac{1}{2S/\sigma^2}\frac{1}{\displaystyle\sum_{k=0}^{N-1}(2\pi k)^2 \sin^2(2\pi\theta k)}
\end{aligned} \qquad (5.31)$$

Note that this CRLB is a function of θ. The CRLB is plotted in Figure 5.3 when $\nu = 2S/\sigma^2 = 1$ and 10 with $N = 10$ as a function of θ.
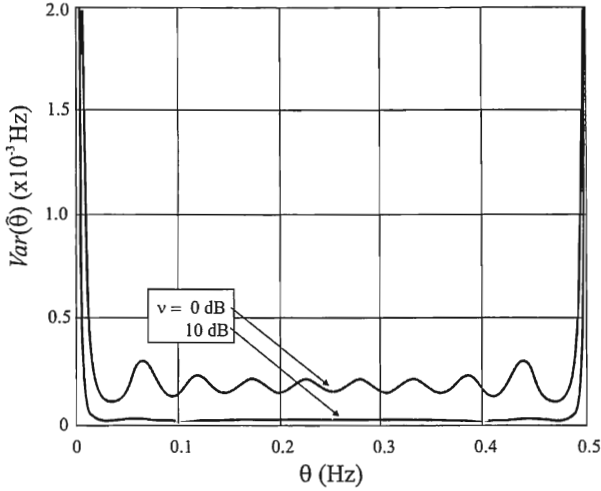
■

Figure 5.3 CRLB for frequency estimation when $N = 10$ and $\nu = 0$ dB and 10 dB.

5.2.1.2 Vector CRLB

In this case, let θ denote a vector of unknown parameters that are to be estimated. $\hat{\theta}$ is the estimate of θ, while $\hat{\theta}_i$ is the ith element of $\hat{\theta}$ and is a scalar. This is the most common case for communication EW target detection. The unknowns are typically the center frequency, mean value, power (variance), and phase.

Definition: The CRLB of the components of θ is given by

$$\text{var}\left(\hat{\theta}_i\right) \geq \left[\mathbf{F}^{-1}\left(\theta\right)\right]_{ii} \tag{5.32}$$

where $\left[\mathbf{F}^{-1}\left(\mathbf{q}\right)\right]_{ii}$ is the iith element in the inverse of Fisher information matrix $\mathbf{F}(.)$, defined as

$$\left[\mathbf{F}\left(\theta\right)\right]_{ij} = -\mathcal{E}_{p(x;\theta)}\left[\frac{\partial^2 \ln \mathbf{p}\left(\mathbf{x};\theta\right)}{\partial \theta_i \partial \theta_j}\right] \tag{5.33}$$

■

An unbiased estimator that attains the bound (i.e., an efficient estimator) exists if and only if $\partial \ln p(\mathbf{x};\theta)/\partial\theta$ can be factored as

$$\frac{\partial \ln p(x;\theta)}{\partial \theta} = F(\theta)[g(x)-\theta] \tag{5.34}$$

If this is true,

$$\hat{\theta} = g(x) \tag{5.35}$$

The Fisher information matrix can be expressed as

$$F_\theta = \mathcal{E}_\theta\{VV^\mathrm{T}\} \tag{5.36}$$

where

$$V = \left[\left.\frac{\partial}{\partial \theta_1}\ln p(x|\theta)\right|_{\theta=\hat{\theta}_{ML}} \quad \cdots \quad \left.\frac{\partial}{\partial \theta_m}\ln p(x|\theta)\right|_{\theta=\hat{\theta}_{ML}}\right]^\mathrm{T} \tag{5.37}$$

Under the $\mathcal{N}(0,\sigma^2)$ process assumption, (5.36) reduces to

$$F(q) = \frac{1}{\sigma^2}G^\mathrm{T}G, \qquad G^\mathrm{T} = \frac{\partial x^\mathrm{T}(\theta)}{\partial \theta} \tag{5.38}$$

The matrix G is called a *sensitivity matrix*.

Assume that the parameter vector θ is partitioned into two sets so that $\theta = [\theta_1^\mathrm{T} \quad \theta_2^\mathrm{T}]^\mathrm{T}$. The first set of these parameters θ_1^T will be those quantities to be estimated. The other set θ_2^T are nuisance parameters because they influence the measured data but whose values are not desired. Most of the time, this second set of parameters negatively impacts the estimation of the parameters in the first set.

By partitioning the matrix G of (5.38) according to the partitioning into θ_1 and θ_2,

$$G = [G_1 \quad G_2], \quad G_1 = \frac{\partial x(\theta)}{\partial q_1}, \quad G_2 = \frac{\partial x(\theta)}{\partial q_2} \tag{5.39}$$

The inverse of the Fisher information matrix of (5.38) is given by

$$\mathbf{F}^{-1}(\mathbf{q}) = \sigma^2 \begin{bmatrix} \left[\mathbf{G}_1^T \mathbf{Q}_{\mathbf{G}_2}^\perp \mathbf{G}_1\right]^{-1} & 0 \\ 0 & \left[\mathbf{G}_2^T \mathbf{Q}_{\mathbf{G}_1}^\perp \mathbf{G}_2\right]^{-1} \end{bmatrix} \tag{5.40}$$

where

$$\mathbf{Q}_{\mathbf{G}_i}^\perp = \mathbf{I} - \mathbf{G}_i \left[\mathbf{G}_i^T \mathbf{G}_i\right]^{-1} \mathbf{G}_i^T \tag{5.41}$$

is a projection matrix projecting onto the space orthogonal to the space spanned by the matrix $\mathbf{G}_i$. $\mathbf{I}$ is the identity matrix.

5.2.1.3 CRLB on the Modulation Parameters of Phase Shift Key Signals

PSK modulation is a common method for digital signaling. In particular, it is by far the most popular modulation for direct sequence spread spectrum signals. In PSK modulation, the phase of the carrier is changed to carry the selected symbol during the time intervals. Binary PSK (BPSK) is also a popular modulation for digital communications with low-speed modems. In BPSK, there are two possible phase shifts, usually 0 and π or π and $-\pi$ radians, but others are possible.

Ho derived the CRLB on the modulation parameters of PSK signals [4]. The signal model in quadrature form is given by

$$s(t) = \alpha_I(t) \cos(2\pi f_0 t) - \alpha_Q \sin(2\pi f_0 t) \tag{5.42}$$

where the α_i's, $i \in (I, Q)$, are the in-phase and quadrature amplitudes given by

$$\alpha_I(t) = \sum_n \sqrt{S} \cos \phi_n g(t - nT_s) \tag{5.43}$$

and

$$\alpha_Q(t) = \sum_n \sqrt{S} \sin \phi_n g(t - nT_s) \tag{5.44}$$

when the samples are indexed by n. Furthermore,

f_0 = center frequency
T_s = symbol time

S = average power in each channel

$$\phi_n \in \left\{ \frac{2\pi m}{M}, m = 0,1,\cdots,M-1 \right\}$$

M = order of the modulation [BPSK, M = 2; quaternary phase shift key (QPSK),

    M = 4; and so forth]

$g(t)$ = pulse shaping function.

The signal intercepted by the EW system is corrupted by AWGN and so is given by

$$r(t) = s(t) + n(t), \qquad 0 < t \leq L \tag{5.45}$$

where $n(t) \sim \mathcal{N}(0,\sigma_n^2)$. The PSK signal parameters are given by $\theta = [f_0, T_s, S]^\mathsf{T}$, for which the CRLB is to be determined.

Because $n(t)$ is AWGN, the pdf for x is given by

$$\mathbf{p}(x,t,\theta) = K \exp\left\{ -\frac{1}{2\sigma_n^2} \int\limits_0^L [x(t) - s(t,\theta)]^2 dt \right\} \tag{5.46}$$

with the natural log given by

$$\ln[\mathbf{p}(x,t,\theta)] = \ln K - \frac{1}{2\sigma^2} \int\limits_0^L [x(t) - s(t,\theta)]^2 dt \tag{5.47}$$

BPSK

For BPSK,

$$\alpha_I(t) = \sqrt{2S} \sum_n d_n \phi_n g(t - nT_s), \qquad \alpha_Q(t) = 0 \tag{5.48}$$

where $d_n \in \{+1, -1\}$ with equal probability. The derived elements of the Fisher information matrix are (see [4] for details)

$$F_{11} = \frac{4\pi^2 S}{\sigma_n^2} \int\limits_0^L t^2 \sin^2(2\pi f_0 t) \sum_n g(t - nT_s)^2 dt \tag{5.49}$$

$$F_{22} = \frac{S}{\sigma_n^2} \int_0^L \cos^2(2\pi f_0 t) \sum_n n^2 g(t - nT_s)^2 \, dt \tag{5.50}$$

$$F_{33} = \frac{1}{\sigma_n^2} \int_0^L \cos^2(2\pi f_0 t) \sum_n g(t - nT_s)^2 \, dt \tag{5.51}$$

$$F_{12} = F_{21} = \frac{\pi S}{\sigma_n^2} \int_0^L t \sin(4\pi f_0 t) \sum_n n g(t - nT_s) g'(t - nT_s) \, dt \tag{5.52}$$

$$F_{13} = \frac{-\pi \sqrt{S}}{\sigma_n^2} \int_0^L t \sin(4\pi f_0 t) \sum_n g(t - nT_s)^2 \, dt \tag{5.53}$$

and

$$F_{23} = F_{32} = \frac{-\sqrt{S}}{\sigma_n^2} \int_0^L \cos^2(2\pi f_0 t) \sum_n n g(t - nT_s) g'(t - nT_s) \, dt \tag{5.54}$$

where $g'(t)$ is the time derivative of $g(t)$.

M-ary PSK

For higher forms of PSK, the amplitude modulations on the two channels are given by

$$\alpha_I(t) = \sqrt{S} \sum_n d_{I,n} \phi_n g(t - nT_s), \qquad d_{I,n} = \cos\left(\frac{2\pi m}{M}\right) \tag{5.55}$$

and

$$\alpha_Q(t) = \sqrt{S} \sum_n d_{Q,n} \phi_n g(t - nT_s), \qquad d_{Q,n} = \sin\left(\frac{2\pi m}{M}\right) \tag{5.56}$$

where $0 \leq m \leq M-1$, each with probability $1/M$. In this case, the elements of the Fisher information matrix are

$$F_{11} = \frac{2\pi^2 S}{\sigma_n^2} \int_0^L t^2 \sum_n g(t - nT_s)^2 \, dt \tag{5.57}$$

$$F_{22} = \frac{S}{2\sigma_n^2} \int_0^L \sum_n n^2 g'(t - nT_s)^2 dt \tag{5.58}$$

$$F_{33} = \frac{1}{2\sigma_n^2} \int_0^L \sum_n g(t - nT_s)^2 dt \tag{5.59}$$

$$F_{12} = F_{21} = F_{13} = F_{31} = 0 \tag{5.60}$$

and

$$F_{23} = F_{32} = \frac{-\sqrt{S}}{2\sigma_n^2} \int_0^L \sum_n g(t - nT_s)g'(t - nT_s)dt \tag{5.61}$$

Clearly, the pulse shaping functions affect the accuracy with which these parameters can be estimated. In [4], simulations were performed using typical pulse shaping functions. The CRLB for 2^M PSK, for all $M > 2$, are equivalent. BPSK and 2^M PSK are not the same, but very close. These results are illustrated in Figure 5.4.

The CRLB for estimating the center frequency with a resistive/capacitive (RC) shaping function is shown in Figure 5.4(a). This CRLB is quite low at about –38 dB, as is the CRLB for estimating the symbol time [Figure 5.4(b)]. Power estimation is not as good, with the CRLB at about –6 dB. When using a lowpass pulse shaping function, the CRLBs are somewhat larger than when using the RC function. These are shown in Figure 5.4(d)–(f).

5.3 Maximum Likelihood Estimation

When the outcome of an experiment is observed and probabilities are calculated based on these outcomes a model is implicitly assumed that generates the experimental outcomes. For example, observing events such as the outcome of a toss of a coin, a model is assumed that produces heads one-half the time and tails the other times. In the case of a coin, the model would state that there is a fixed probability for the particular outcomes. This model has one *parameter*, θ, the probability of the coin landing on heads. If the coin is fair, then $\theta = 0.5$.

If the probability of an event x dependent on model parameter θ is written
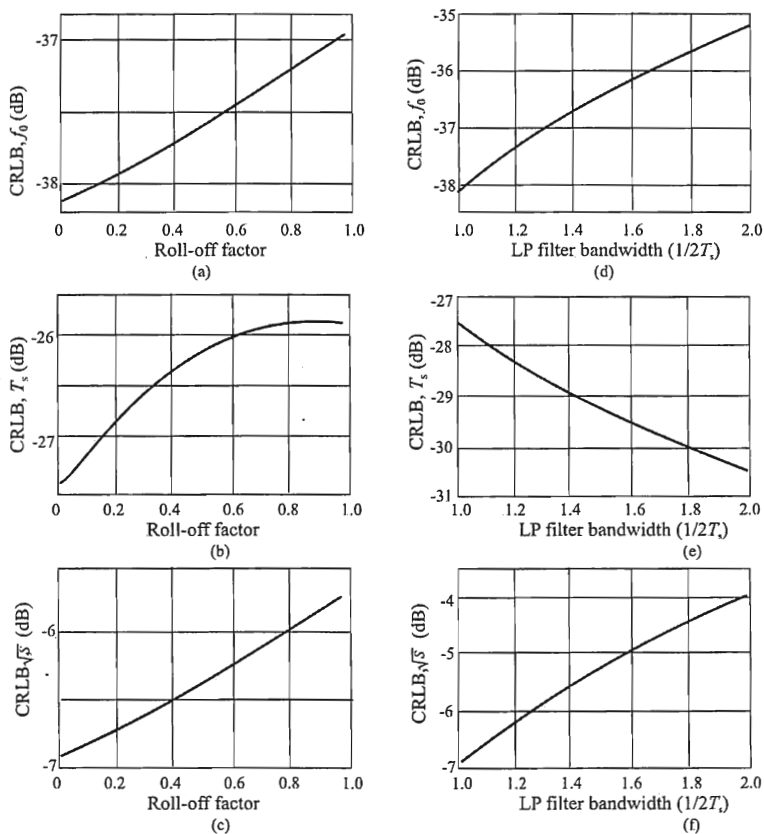
Figure 5.4 CRLB for PSK signals: (a), (b), and (c) RC pulse shaping function; (d), (e), and (f) lowpass pulse shaping function. The CRLB are normalized by the noise variance and expressed in decibels relative to that value. (*From:* [4]. © 1999 IEEE. Reprinted with permission.)

$$\Pr(x|\theta) \tag{5.62}$$

then the corresponding likelihood is

$$\lambda(\theta|x) \tag{5.63}$$

that is, the likelihood of the parameters given the data.

In most cases, certain values are more probable than others. The underlying principle of maximum likelihood estimation is to find the parameter value(s) that makes the observed data most likely. This is because the likelihood of the parameters given the data is defined to be equal to the probability of the data given the parameters.

When predicting outcomes based on a set of assumptions, then probabilities are of interest—the probabilities of certain outcomes occurring or not occurring. In the case of *data analysis*, however, the data are already available—there is no randomness to it. Once the data have been observed, they are fixed, there is no "probabilistic" part to them anymore. In this case, the likelihood of the model parameters are of interest that underlie the fixed data.

The two approaches can be summarized as

Determining probabilities: The parameters are known, estimate the probabilities;

Determining likelihood: Having observed the data, estimate the parameters.

The MLE of θ given x is

$$\hat{\theta} = \arg\max_{\theta} p(\mathbf{x}|\theta) \tag{5.64}$$

The argmax function returns the value of θ that maximizes $p(\mathbf{x}|\theta)$; $\hat{\theta}$ is the "most likely" value for θ given the observed values x. Normally, $p(\mathbf{x}|\theta)$ is thought of as a function of x with θ as a fixed parameter. In this context, $p(\mathbf{x}|\theta)$ is thought of as a probability density. For MLE, $p(\mathbf{x}|\theta)$ is thought of as a function of θ with x as a fixed parameter. In this context, $p(\mathbf{x}|\theta)$ is thought of as a likelihood function.

For large sample sets, MLEs are asymptotically optimal. Also, for large N, MLEs are asymptotically unbiased, normally distributed with minimum variance.

$\hat{\theta}$ is usually determined by the classical technique from elementary calculus of differentiating the likelihood function (or the logarithm of the likelihood function if such a logarithm is a nondecreasing function of the parameters) by θ, setting the derivative to zero, and solving for θ in terms of $\mathbf{x}$.

Let the unknown parameters of a given system be denoted by the vector θ as

$$\theta = [\theta_1 \quad \theta_2 \quad \cdots \quad \theta_P]^T \tag{5.65}$$

where the noiseless measurement is some vector function of these parameters. All actual measurement in any real circumstance will be corrupted by noise. Let the noisy measurement be given by $\mathbf{x}(\theta)$. Assume the noise is $\mathcal{N}(0, \sigma^2)$. The ability, on the average, to estimate θ is bounded by the Cramer-Rao bound discussed in Section 5.1.

For a scalar, the likelihood equation is

$$\left. \frac{\partial}{\partial \theta} \log p(x|\theta) \right|_{\theta=\hat{\theta}_{ML}} = 0 \tag{5.66}$$

If $p(x|\theta)$ is from an *exponential family* (defined below), then there is a unique solution to the likelihood equation. However, in general, there may be one, many, or no solutions. If $\hat{\theta}$ achieves the CRLB, then it is a solution to the likelihood equation.

The MLE equation in vector form is

$$\left. \frac{\partial}{\partial \theta_1} \log p(\mathbf{x}|\theta) \right|_{\theta=\hat{\theta}_{ML}} = 0$$
$$\vdots \tag{5.67}$$
$$\left. \frac{\partial}{\partial \theta_m} \log p(\mathbf{x}|\theta) \right|_{\theta=\hat{\theta}_{ML}} = 0$$

Example:

Suppose $N = 1$ and $x_0 \sim \mathcal{N}(\theta, \sigma^2)$, where θ is unknown and σ^2 is known. The MLE of θ is found as

$$p(x_0;\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_0-\theta)^2}{2\sigma^2}\right]$$

$$\ln p(x_0;\theta) = -\ln\sqrt{2\pi\sigma^2} - \frac{(x_0-\theta)^2}{2\sigma^2}$$

$$\frac{\partial \ln p(x_0;\theta)}{\partial \theta} = \frac{1}{\sigma^2}(x_0-\theta) = 0 \tag{5.68}$$
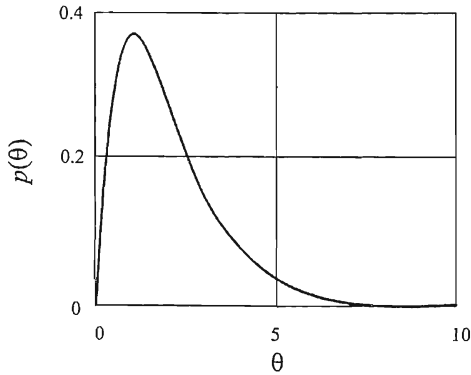
Therefore,

$$x_0 = \hat{\theta} \tag{5.69}$$

∎

Example:

Suppose $N = 1$, x_0 is drawn from an exponential density with parameter θ illustrated in Figure 5.5:

$$p(x_0;\theta) = \begin{cases} \theta e^{-\theta x_0}, & x_0 \geq 0 \\ 0, & x_0 < 0 \end{cases} \tag{5.70}$$

The MLE of θ in terms of x_0 is determined as follows. For $x_0 < 0, \hat{\theta} = 0$. For $x_0 \geq$ 0,

Figure 5.5 Example of exponential density when $N = 1$.

$$\ln p(x_0;\theta) = \ln\theta + \ln e^{-\theta x_0} = \ln\theta - \theta x_0$$

$$\frac{\partial \ln p(x_0;\theta)}{\partial\theta} = \frac{1}{\theta} - x_0 = 0$$

so

$$\hat{\theta} = \frac{1}{x_0} \qquad (5.71)$$

■

Example:

Suppose observations x_0 through x_{N-1} are i.i.d. exponential with parameter θ. The MLE of θ in terms of x_0 through x_{N-1} is determined by

$$p(x_0, x_1, \cdots, x_{N-1};\theta) = \theta^N \exp\left(-\theta\sum_{k=0}^{N-1} x_k\right) \qquad (5.72)$$

$$\ln p(x_0, x_1, \cdots, x_{N-1};\theta) = N\ln\theta + \ln\left[\exp\left(-\theta\sum_{k=0}^{N-1} x_k\right)\right]$$

$$= N\ln\theta - \theta\sum_{k=0}^{N-1} x_k$$

$$\frac{\partial}{\partial\theta}\ln p(x_0, x_1, \cdots, x_{N-1};\theta) = \frac{N}{\theta} - \sum_{k=0}^{N-1} x_k = 0$$

so

$$\hat{\theta} = \frac{N}{\sum_{k=0}^{N-1} x_k} \qquad (5.73)$$

■

Example:

Suppose x_0 through x_{N-1} are i.i.d. with $x_k \sim \mathcal{N}(\theta, \sigma^2)$. The MLE is found as

$$p(x_0, x_1, \cdots, x_{N-1}; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{1}{2\sigma^2} \sum_{k=0}^{N-1} (x_k - \theta)^2\right]$$

$$\ln p(x_0, x_1, \cdots, x_{N-1}) = -\frac{1}{2} \ln 2\pi\sigma^2 + \ln \exp\left[-\frac{1}{2\sigma^2} \sum_{k=0}^{N-1} (x_k - \theta)^2\right]$$

$$= -\frac{1}{2} \ln 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{k=0}^{N-1} (x_k - \theta)^2$$

$$= -\frac{1}{2} \ln 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{k=0}^{N-1} x_i^2 - 2x_k\theta + \theta^2$$

$$\frac{\partial}{\partial \theta} \ln p(x_0, x_1, \cdots, x_{N-1}) = \frac{1}{2\sigma^2} \sum_{k=0}^{N-1} -2x_k\theta + 2\theta = 0$$

so

$$\hat{\theta}_{ML} = \frac{1}{N} \sum_{k=0}^{N-1} x_k \tag{5.74}$$

Now

$$\mathcal{E}\left\{\hat{\theta}_{ML}\right\} = \mathcal{E}\left\{\frac{1}{N} \sum_{k=0}^{N-1} x_k\right\} = \int_{-\infty}^{\infty} \frac{1}{N} \sum_{k=0}^{N-1} x_k p(x_0, x_1, \cdots, x_{N-1} | \theta) \, dx$$

$$= \frac{1}{N} \sum_{k=0}^{N-1} \int_{-\infty}^{\infty} x_k p(x_0, x_1, \cdots, x_{N-1} | \theta) \, dx_k$$

$$= \frac{1}{N} \sum_{k=0}^{N-1} \mathcal{E}\left\{x_k | \theta\right\}$$

Because of the i.i.d. assumption,

$$\sum_{k=0}^{N-1} \mathcal{E}\left\{x_k | \theta\right\} = N\mathcal{E}\left\{x_k | \theta\right\}$$

and, thus,

$$\mathcal{E}\left\{\hat{\theta}_{ML}\right\} = \frac{1}{N} N\mathcal{E}\left\{x_k | \theta\right\} = \theta \tag{5.75}$$

Therefore, this is an unbiased estimator.

■

An unbiased estimator is *minimum variance unbiased* (MVUB) if the error variance σ^2 cannot be made smaller by any other unbiased estimator of θ. $t(x)$ is a *sufficient statistic* for θ if the pdf of x given $t(x)$ is independent of θ. This is true if and only if the pdf of x factors as $a(x)b_\theta(t)$; $a(x)$ cannot depend on θ and $b_\theta(t)$ cannot depend on x except through t. A sufficient statistic is a *minimal sufficient statistic* for θ if it is a function of every other sufficient statistic for θ.

Two characteristics of sufficient statistics are:

- A sufficient statistic summarizes all of the information about θ in x.
- If a sufficient statistic for θ exists, the maximum likelihood estimate $\hat{\theta}$ is a function of the sufficient statistic.

Two characteristics of MVUB estimators are:

- Every efficient estimator is MVUB;
- An estimator may be MVUB but not achieve the CRLB.

Property: Rao-Blackwell

Given that $\hat{g}(x)$ is an unbiased estimator of $g(\theta)$ and that $T(x)$ is a sufficient statistic for θ, define

$$\tilde{g}(x) = \mathcal{E}_\theta\left\{\hat{g}(X)\big|T(X) = T(x)\right\} \qquad (5.76)$$

then

(a) $\tilde{g}(x)$ is also unbiased.
(b) $\mathrm{var}_\theta\left\{\tilde{g}(x)\right\} \leq \mathrm{var}_\theta\left\{\hat{g}(x)\right\}$.
(c) with equality if and only if $P_\theta\left[\tilde{g}(\mathbf{X}) = \hat{g}(\mathbf{X})\right] = 1$.

■

Definition: The family $\{P_\theta;\ \theta \in \Theta\}$ is said to be *complete* if

$$\mathcal{E}_\theta\left\{f(\mathbf{X})\right\} = 0 \ \forall \theta \in \Theta \rightarrow P_\theta\left[f(X) = 0\right] = 1 \qquad (5.77)$$

If the family $\{P_\theta;\ \theta \in \Theta\}$ is complete, then there is no nontrivial sufficient statistic for θ, that is, the x and all sufficient statistics are 1:1.

■

Suppose $T(x)$ is sufficient for $\{P_\theta;\ \theta \in \Theta\}$ and let Q_θ be the distribution for $T(x)$ if $X \sim P_\theta$. If $\{Q_\theta;\ \theta \in \Theta\}$ is a complete family, then $T(x)$ is a complete sufficient statistic. Any unbiased estimator that is a function of a complete sufficient statistic is unique and thus is a MVUB.

A class of distributions is said to be an *exponential family* if there exist the real-valued functions $C, Q_1, ..., Q_m, T_1, ..., T_m,$ and h such that

$$p_\theta(x) = C(\theta)h(x)\exp\left[\sum_{i=1}^{m} Q_i(\theta)T_i(x)\right]\ \forall x \in \mathbf{G},\ \theta \in \Theta\frac{\partial^2 \Omega}{\partial u^2} \qquad (5.78)$$

These T_i are complete sufficient statistics if Θ contains an m-dimensional rectangle.

Example:

Given that x_0 through x_{N-1} are i.i.d. Gaussian with unknown mean θ and known variance σ^2, a sufficient statistic for θ is found as

$$p(x_0, \cdots, x_{N-1}; \theta) = \frac{1}{(2\pi)^{N/2}\sigma^N}\exp\left[-\frac{1}{2\sigma^2}\sum_{k=0}^{N-1}(x_k - \theta)^2\right]$$

$$= \frac{1}{(2\pi)^{N/2}\sigma^N}\exp\left[-\frac{1}{2\sigma^2}\sum_{k=0}^{N-1}x_i^2 - 2x_k\theta + \theta^2\right]$$

$$= \frac{1}{(2\pi)^{N/2}\sigma^N}\exp\left[-\frac{1}{2\sigma^2}\sum_{k=0}^{N-1}x_i^2\right]$$

$$\times \exp\left[-\frac{1}{2\sigma^2}\left(-2\theta\sum_{k=0}^{N-1}x_k + N\theta^2\right)\right]$$

so

$$a(x) = \frac{1}{(2\pi)^{N/2}\sigma^N}\exp\left[-\frac{1}{2\sigma^2}\sum_{k=0}^{N-1}x_i^2\right] \qquad (5.79)$$

and

$$b_\theta(t) = \exp\left[-\frac{1}{2\sigma^2}\left(-2\theta\sum_{k=0}^{N-1}x_k + N\theta^2\right)\right] \qquad (5.80)$$

The sufficient statistic is

$$t = \sum_{k=0}^{N-1}x_k \qquad (5.81)$$

Note that the MLE of θ is a function of this statistic.

■

Example:

Given that x_0 through x_{N-1} are i.i.d. exponential with unknown parameter θ, a sufficient statistic for θ is found as follows:

$$p(x_0,\cdots,x_{N-1};\theta) = \theta^N \exp\left[-\theta\sum_{k=0}^{N-1}x_k\right]$$
$$a(x) = 1 \qquad (5.82)$$
$$b_\theta(t) = \theta^N \exp\left[-\theta\sum_{k=0}^{N-1}x_k\right]$$

The sufficient statistic is

$$t = \exp\left[\sum_{k=0}^{N-1}x_k\right] \qquad (5.83)$$

Note that the *maximum likelihood* (ML) estimator of θ is a function of this statistic.

■

Example:

Given that x_0 through x_{N-1} are i.i.d. binomial with unknown probability θ, a sufficient statistic is found as:

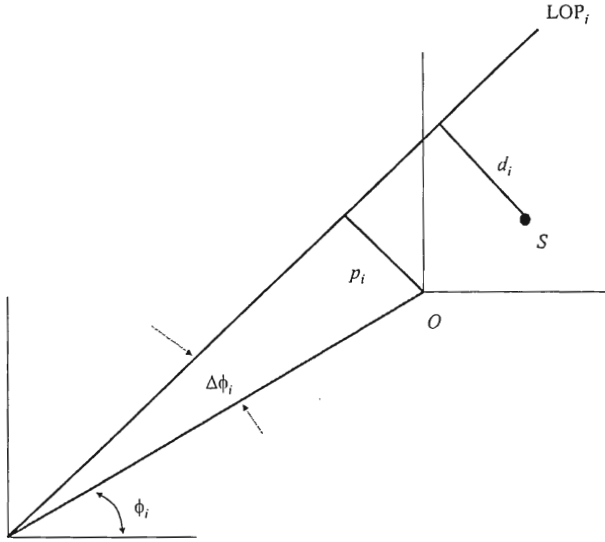$$P(x_k = 0;\theta) = 1-\theta = \theta^{x_k}\left(1-\theta\right)^{1-x_k}$$

Figure 5.6 Stansfield's PF geometry. (*From:* [5], © 1947, IRE. Reprinted with permission.)

$$P(x_k = 1; \theta) = \theta = \theta^{x_k} (1-\theta)^{1-x_k}$$

$$p(x_0, \cdots, x_{N-1}; \theta) = \prod_{k=0}^{N-1} \theta^{x_k} (1-\theta)^{1-x_k}$$

$$= \theta^{\sum_{k=0}^{N-1} x_k} (1-\theta)^{N - \sum_{k=0}^{N-1} x_k} \tag{5.84}$$

∎

Stansfield's Fix Algorithm

Stansfield's original technique for calculating the PF of an emitting target is an example of an EW application of maximum likelihood estimation [5]. Stansfield's approach can be viewed as a small error approximation to the MLE. The geometry involved is illustrated in Figure 5.6. As described in [6, 7], the goal is to minimize the expression for the joint probability of miss given as a function of the miss distances given by

$$P(d_1, d_2, \cdots, d_N) = \frac{1}{(2\pi)^{N/2} \sum_{i=1}^{N} \sigma_{p_i}} \exp\left(-\frac{1}{2} \sum_{i=1}^{N} \frac{d_i^2}{\sigma_{p_i}^2}\right) \qquad (5.85)$$

This expression is minimized when the summation in the argument of exp is maximized. The result is the PF that is the most likely location of the target.

5.4 Concluding Remarks

Some of the more salient aspects of extracting parameters from emitting communication EW targets were presented in this chapter. In general, target recognition proceeds in two steps:

(1) Signal parameter estimation;
(2) Target recognition based on those parameters.

The parameters chosen to estimate depend on the SOIs, and some fairly general ones were discussed. The power level was one of these, and it could be expected to remain much the same from one transmission to the next, as long as the transmitter and receiver are not moving rapidly and the target is stationary in frequency.

MLE was the method presented in the chapter to estimate the parameters. There are other techniques, each with their strong points and weak points. The best technique to estimate parameters depends on the particular application.

References

[1]   Whalen, A. D., *Detection of Signals in Noise,* New York: Academic Press, 1971, Chapter 10.

[2]   Helstrom, C. W., *Statistical Theory of Signal Detection,* New York: Pergamon Press, 1960, Ch. 7 and 5.

[3]   Van Trees, H. L., *Detection, Estimation, and Modulation Theory,* New York: John Wiley & Sons, 1965.

[4]   Ho, K. C., "Modified CRLB on the Modulation Parameters of a PSK Signal," *Proceedings IEEE MILCOM, 99,* Paper 2.2.

[5]   Stansfield, R. G., "Statistical Theory of DF Fixing," *Journal of the Institute of Electrical Engineers,* Vol. 94, Part IIIa, 1947, pp. 762–770.

[6]   Poisel, R. A., *Introduction to Communication Electronic Warfare Systems,* Norwood, MA: Artech House, 2002, pp. 381–383.

[7]   Gething, P. J. D., *Radio Direction Finding and Resolution of Multicomponent Wave-Fields,* London: Peter Peregrinus, Ltd., 1978, pp. 271–275.

Chapter 6

Spectrum Estimation

Detecting targets in communication EW systems consists of determining if energy at a frequency is present and, if it is, subsequently processing the signal to establish whether it is an SOI or not. This is normally accomplished by computing the psd of the signal and finding where the peaks occur. Therefore, even though this process is usually referred to as spectrum estimation, it is not the spectrum that is of primary interest—it is some parameters associated with the spectrum that are of interest for target acquisition.

In this chapter, two basic methods of power spectrum estimation are presented. These are relatively low resolution processes (the meaning of resolution will subsequently become clear). Higher resolution methods, which may be necessary as the spectrum crowding increases, are presented in Chapter 8.

The two methods presented in this chapter are the periodogram and Blackman-Tukey estimation. The former applies when the signal environment is deterministic or almost so, while the latter assumes that the SOIs are stochastic.

6.1 Spectrum Estimation with the Periodogram

Because the FFT of a signal is a representation of the signal in the frequency domain, and the power at each frequency is the information contained therein, searching for the peaks in the FFT can be used for signal detection while simultaneously providing estimates of the location of those signals in the frequency domain [1, 2]. The FFT is a filter bank and is illustrated in Figure 6.1. This is the notion behind the *periodogram*.

The periodogram is the classical nonparametric spectral estimation method. The periodogram assumes that the process being analyzed consists of several harmonically related sinusoids with additive noise. It estimates the psd by forming
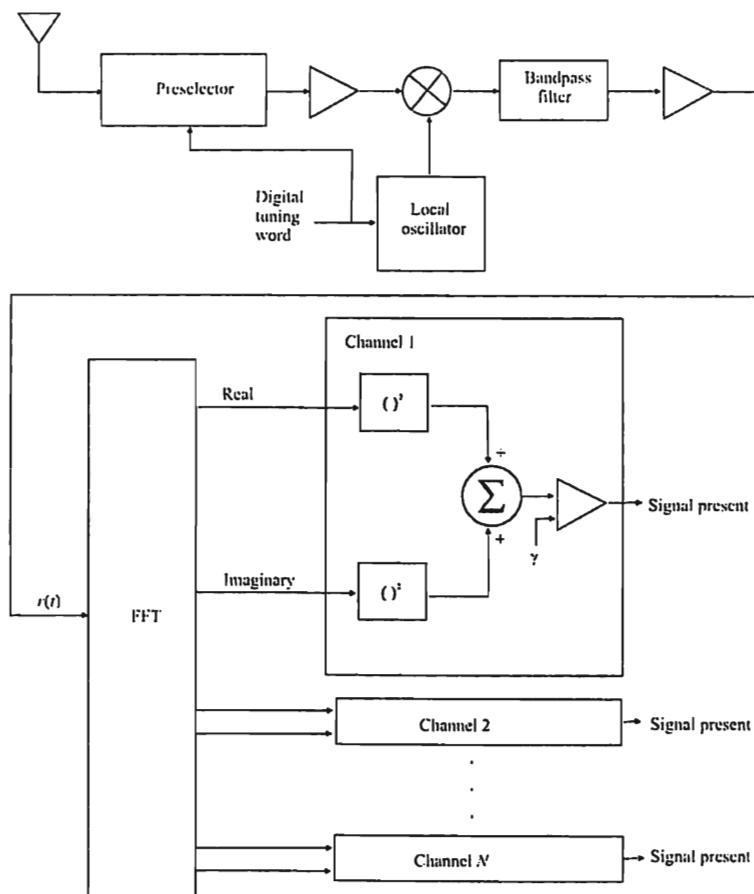
Figure 6.1 Structure for an FFT detector.

a bank of narrowband filters with sin x/x-like response. The periodogram is an unbiased estimator, but the variance does not decrease with an increase in N, which is usually an undesirable characteristic. Even though the periodogram is not a good estimator for all types of problems, it is still widely used. It can be shown that the discrete periodogram spectral estimate computes a least squares fit of a harmonically related set of complex sinusoids to the data [3].

When $x(n)$ is a wss random process with autocorrelation function $\gamma_{xx}(m)$ and sample psd,

$$X(f) = \sum_{m=-\infty}^{\infty} \gamma_{xx,m} e^{-j2\pi fm} \tag{6.1}$$

create the windowed process from $x(n)$ as

$$x_{N,n} = a_n x_n \tag{6.2}$$

where a_n are the time samples of the window. This window has the property that $a_n = 0$ for $n < 0$ and $n \geq N$. The z-transform of x_n is

$$X_N(z) = \sum_{n=-\infty}^{\infty} x_{N,n} z^{-n} = \sum_{n=0}^{N-1} a_n x_n z^{-n} \tag{6.3}$$

Let

$$X_N^*(f) = X_N\left(e^{j2\pi f}\right) \tag{6.4}$$

The sample autocorrelation function for the windowed sequence is

$$\tilde{\gamma}_{N,n} = \frac{1}{N} \sum_{k=-\infty}^{\infty} x_{N,k+n} x_{N,k} = \frac{1}{N} x_{N,n} * x_{N,-n} \tag{6.5}$$

Because of the window, however, $\tilde{\gamma}_{N,n} = 0$ for $|n| \geq N$. The z-transform of $\tilde{\gamma}_{N,n}$ is given by

$$P_N^*(z) = \sum_{n=-\infty}^{\infty} \tilde{\gamma}_N(n) z^{-n} = \frac{1}{N} X_N(z) X_N(z^{-1}) \tag{6.6}$$
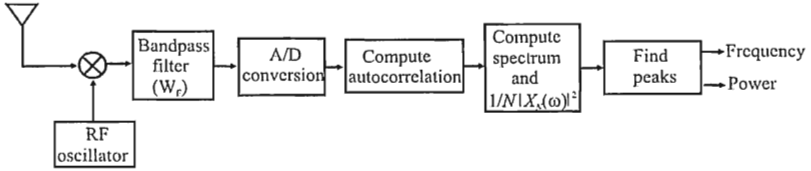
Figure 6.2 Signal flow diagram for signal detection with the periodogram and signal.

Therefore,

$$P_N(f) = \frac{1}{N} X_N(f) X_N(-f) = \frac{1}{N}|X_N(f)|^2 \qquad (6.7)$$

$P_N(f)$ is the periodogram. The signal flow for the periodogram used for signal detection is shown in Figure 6.2. Thus, the intuitive technique for signal detection illustrated in Figure 6.1 is in fact optimal in some cases, as long as the magnitude of the FT is used.

For real signals, the autocorrelation function is real and even, and therefore the psd is real and even. The psd is a measure of the relative probability that the signal contains energy at frequency bin k.

The above described periodogram method for estimating the power spectrum is accomplished by dividing the time series into consecutive blocks, calculating the autocorrelation estimate for each block, taking the DFT of each of these autocorrelation estimates, and averaging. By the correlation theorem, this is the same as averaging the squared-magnitude DFTs of the signal blocks directly, without calculating the autocorrelation functions. That is,

$$P_N(f) = \frac{1}{N}|X_N(f)|^2 \leftrightarrow \frac{1}{N}\left|\sum_{k=0}^{N-1} x_k x_{k+l}\right| \qquad (6.8)$$

where $X_N(f)$ is normally obtained with the FFT.

For a single FFT there is no noise averaging resulting in high false alarm rates. If the signal is available for long enough, then multiple FFTs can be obtained and averaged. Since the signal is at least partially deterministic while the noise is random, it would be expected that with multiple spectrum samples available the noise would tend to cancel when they are averaged. That is indeed what happens, and for the first few averages the gain in SNR is about 2.2 dB/doubling of samples. The variance of the resulting average, however, does not tend to zero [3], but settles asymptotically to 1.6 dB/doubling of FFTs [1].

Windows are frequently used with FFT detection to improve the sidelobe response. With a rectangular window applied to the time series, which occurs

simply by sampling the signal over a finite period, the peak sidelobe response, which occurs in the adjacent channel, is only 13 dB less than the main peak response. With more sophisticated windows applied to the time series, this response can be substantially reduced, but at the expense of increasing the width of the main lobe.

6.1.1 Averaged Periodogram

The periodogram is a random variable itself with a standard deviation on the same order of magnitude as its mean. Any individual periodogram can deviate significantly from the true spectrum estimate. This can be addressed by averaging several sample periodograms, which reduces the variance of the spectral estimate at the expense of lower resolution. Such averaging in the frequency domain corresponds to windowing in the time domain. This is the basis of Bartlett's method for estimating the psd of a time series.

One concern with this technique is the number of periodograms averaged. In order for the periodogram approach to work, the signal being analyzed must be stationary. The more individual periodograms are used for the averaging, the longer the signal must be stationary. In the RF environments typified by EW system employments, the stationary assumption may be short lived.

This method segments the data into K nonoverlapping blocks of length L, and computes the periodogram on each block. The results are then averaged to obtain the psd. This method can reduce the variance in proportion to K if those blocks are uncorrelated. Since the data are usually continuous, they are rarely totally uncorrelated; thus, the variance reduction factor will be less than K.

Given a fixed-length data set, the block length, L, cannot be made arbitrarily short. The rectangular window function generates a bias in the estimation, and when the window length is too short, distortion is introduced. To avoid excessive smearing effects, the block length L must be large enough so that the narrowest peak in the psd can still be resolved. Therefore, when averaging the blocked periodograms, there is a trade-off between the variance and the bias.

In both the periodogram method and the averaged periodogram methods, the higher lags of the estimated correlation produce poorer estimates, since they involve fewer lag products. One way to avoid this problem is to weight the higher lags less using appropriately structured windowing functions.

Definition: Bartlett's Method

Let the lth sample in the kth segment of the nonoverlapped random time series x be given by $x_{k,l}$, where $l = 0, 1, ..., L - 1$ and $k = 0, 1, ..., K - 1$, and K is the number of frames. The periodogram of the kth block is then

$$P_k(f_l) = \frac{1}{L}\left|DFT(x_l)\right|^2 = \frac{1}{L}\left|\sum_{n=0}^{N-1} x_{l,n} e^{-j2\pi nk/N}\right|^2 \tag{6.9}$$

and the *Bartlett estimate* of the psd is

$$\hat{S}_x(f_l) \triangleq \frac{1}{K}\sum_{k=0}^{K-1} P_k(f_l) \tag{6.10}$$

∎

Finding the periodogram via the Bartlett method by averaging K individual periodograms based on K nonoverlapping segments of the time series reduces the variance of the periodogram by a factor of K. However, since the frequency resolution is given by $1/T$, where T is the length of the time window, and each individual sample set has length T/K, the frequency resolution is reduced by the factor of K. A modification to the Bartlett method is given by the Welch method.

Definition: Welch Method

In the Welch method of spectral estimation, the time series segments are overlapped and windowed prior to computation of the DFT. Such overlapping is illustrated in Figure 6.3, where the overlap is 50%. The resulting time series is then $w_k x_{k,l}$ where $l = 0, 1, ..., L-1$ and $k = 0, 1, ..., K-1$, and where w_n is the window function. Such overlapping recovers some of the lost frequency resolution while maintaining the advantage of reducing the variance.

∎

Note that $\left|X_{xx}\right|^2 \leftrightarrow x \bullet x$ is a circular autocorrelation. This can be converted into a linear correlation with zero padding of the time series; that is, $\{x_l\}$ is replaced by $\{x_l, 0, 0, ..., 0\}$. Estimator (6.10) is biased, however, in that the true autocorrelation is weighted by $N - |l|$,6 which can be removed by division. It is common to leave it in, however, since it is equivalent to smoothing the power spectrum with a sinc^2 kernel by multiplying the autocorrelation function with a triangular window, called the *Bartlett window*. This smoothing puts less weight in the large-lag estimates that are less reliable. The Bartlett window is given by
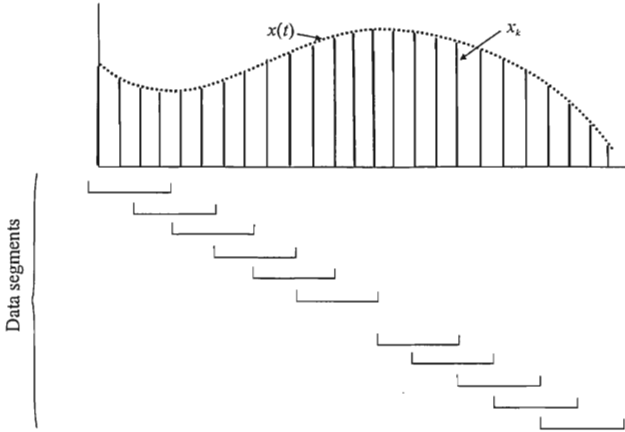
Figure 6.3 Data segment overlapping for the Welch method.

$$w_k = \begin{cases} 1 - \dfrac{|k|}{N}, & |k| \le N \\ 0, & \text{otherwise} \end{cases} \tag{6.11}$$

with spectrum given by

$$S_k = N \operatorname{sinc}^2(kN) \tag{6.12}$$

These functions are shown in Figure 6.4.

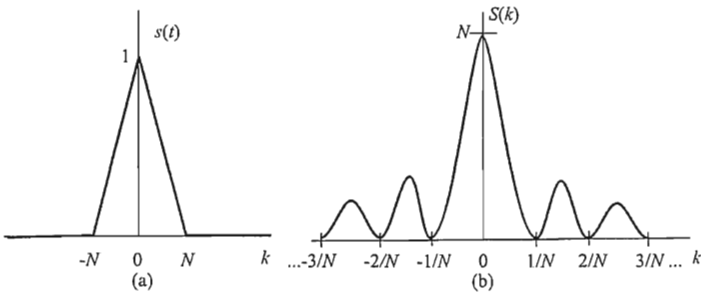A comparison of the three techniques just discussed for detection of a sinusoid

Figure 6.4 Bartlett window: (a) time domain and (b) frequency domain.

and frequency estimation in AWGN was conducted by So et al. [4]. When the frequency was $f_0 = 0.25$ (in other words, precisely at a DFT point), the ROC for the normal periodogram and that computed with the Bartlett method (4 segments) are shown in Figure 6.5. For these curves, $N = 256$, so there are 64 data points per segment. The regular periodogram has better P_d versus P_{fa} performance, as expected since the periodogram is the optimum ML detector. The variance of the lower curves, however, is smaller by a factor of 6.

The CRLB for estimating the frequency with the periodogram is given by [4]

$$\sigma_f^2(N) = \frac{3}{\pi^2 N(N^2 - 1)\nu} \tag{6.13}$$

where ν is the SNR. When SNR is above a threshold, γ, this bound is achieved. If $\nu < \gamma$, however, the frequency estimation performance, as measured by the *mean square frequency error* (MSFE) deviates substantially from this value. Examples are shown in Figure 6.6 when $f_0 = 0.26$. Again, the normal periodogram achieves the CRLB at the lowest SNR. The Welch method with 50% overlap and rectangular window is next for low SNRs but remains high after about $\nu = -5$ dB.

Next is the Bartlett method, which also remains higher for larger SNRs. The worst performance is for the Welch method with no overlap and the Hanning window.

6.2 Blackman-Tukey Spectral Estimation

If the time sequence is stochastic, then, as discussed in Chapter 2, the FT of the time sequence does not exist and some other technique must be used to determine the frequency content.

In 1958, Blackman and Tukey published a short book that presented an implementation of Wiener's autocorrelation to the power spectral estimation method [5]. First, the autocorrelation lags of the time sequence are determined. This sequence is then windowed with an appropriate window that tapers to small values at the edges. The FFT of this sequence is then obtained to obtain the psd estimate [3]. Thus, the autocorrelation function of $s(t)$ is denoted by

$$\gamma_{ss}(\tau) = \mathcal{E}\{s(t + \tau)s^*(t)\} \tag{6.14}$$

The psd is obtained by taking the FT of $\gamma_{ss}(\tau)$ due to the Wiener-Khinchin theorem,
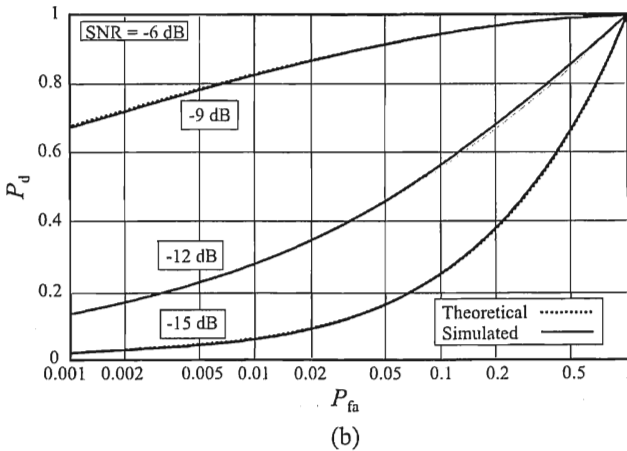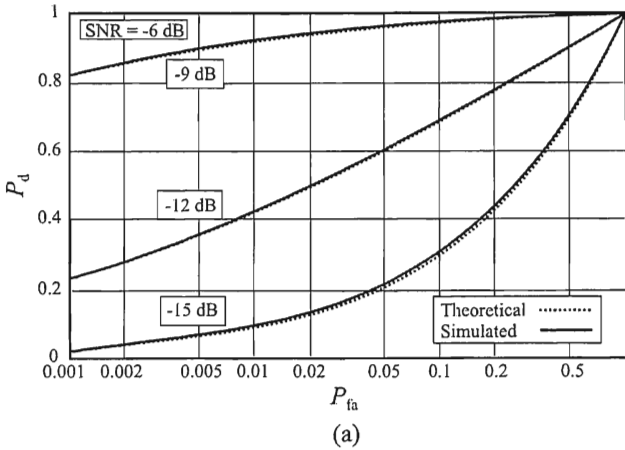
Figure 6.5 Signal detection performance when $f_0 = 0.25$ for (a) normal periodogram and (b) Bartlett method with 4 segments. (*From:* [4], © 1999 IEEE. Reprinted with permission.)
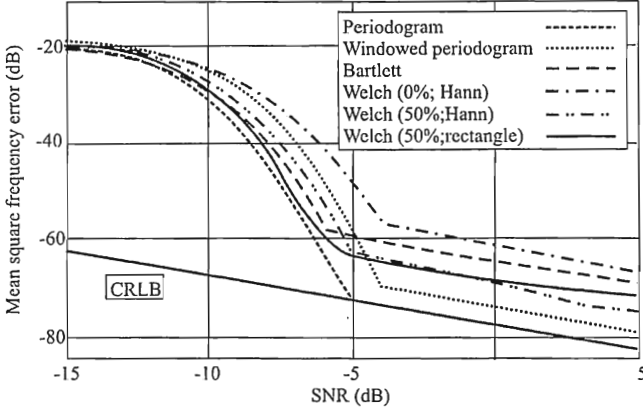
Figure 6.6 MSFE performance comparison for the three techniques to compute the periodogram when $f_0 = 0.26$. (*From:* [4], © 1999 IEEE. Reprinted with permission.)

$$S(f) = \int_{-\infty}^{\infty} \gamma_{ss}(\tau) \, \exp(-j2\pi f\tau) d\tau \qquad (6.15)$$

When $s(t)$ is ergotic, which in practice it almost always is, then the time statistics are equal to the ensemble statistics, and the autocorrelation function can be obtained by averaging over time as opposed to over the ensemble. In that case,

$$\gamma_{ss}(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} s(t+\tau)s^*(t)dt \qquad (6.16)$$

and the psd is obtained as

$$S(f) = \lim_{T \to \infty} \mathcal{E}\left\{ \frac{1}{2T} \left| \int_{-T}^{T} s(t) \, \exp(-j2\pi ft)dt \right|^2 \right\} \qquad (6.17)$$

The psd determined with the BT process can produce negative regions, which leads to complications in interpreting the spectrum function as a psd, since the power cannot be negative in such processes.
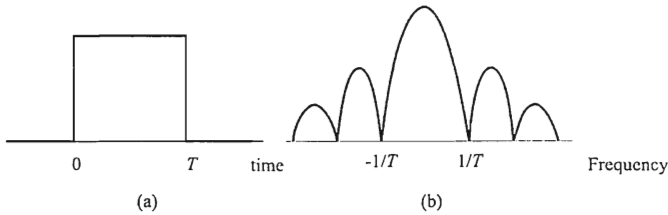
0   *T*   time   -1/*T*   1/*T*   Frequency

(a)   (b)

Figure 6.7 (a) Boxcar time window function and (b) its magnitude spectrum.

6.3 Windows

The periodogram is a poor estimate of the psd of a time series and actually fluctuates more as the record length increases; that is, the variance increases. There are two generally accepted ways of dealing with this problem. The first is to compute the periodogram of several shorter length segments of the series and average the results at each frequency as indicated above. The second method uses a window function with other than simply the unit amplitude and with particular properties to smooth the periodogram. In both these cases, there is a trade-off between spectral resolution and the variance. If high resolution is desired, then a larger variance is necessary and vice versa.

The *resolution* is one of the most important parameters associated with spectrum analysis. It specifies how close two signals can be in the spectrum and still be identified as two signals. The resolution of the periodogram is limited, which is its greatest shortcoming. Irrespective of the characteristics of the signals being analyzed, the resolution of FFT-based spectral analysis is limited to approximately $1/T$, where T is the length of the sample time series. Equivalently, it is the width of the window placed on the time series. High-resolution spectral analysis is discussed in Chapter 8.

Time windowing the data sequence can cause problems with energy leakage into adjacent frequency bins. The amplitude of the spectrum of the boxcar function shown in Figure 6.7(a) has a $|\sin \pi f / \pi f|$ shape as shown in Figure 6.7(b). Multiplication by a boxcar function in (6.2) implies convolution by its spectrum in the frequency domain, and convolution by the spectrum shown in Figure 6.7(b) causes leakage into adjacent frequency bins as well as widening of the main spectral lobe. If the input is a monochromatic sine wave, the output spectrum will have the $\sin \pi f / \pi f$ shape. The null-to-null width of the spectrum is $2/T$, while the 3 dB width is approximately $1/T$. This is the resolution.

6.3.1 Other Windows

Windows other than the boxcar function can also be used that lessen the leakage and main lobe widening; however, applying any time window to the data causes these effects to some degree [6].

6.3.1.1 Welch Window

The Welch window for N points is defined as

$$w_i = 1 - \left(\frac{i - \frac{N}{2}}{\frac{N}{2}} \right)^2 \tag{6.18}$$

where $0 \leq i < N$. This window is illustrated in Figure 6.8 for $N = 1,024$. This window is commonly used as a window for power spectral estimation. The spectrum magnitude is illustrated in Figure 6.9.

6.3.1.2 Cosine Window

The raised cosine window is given by

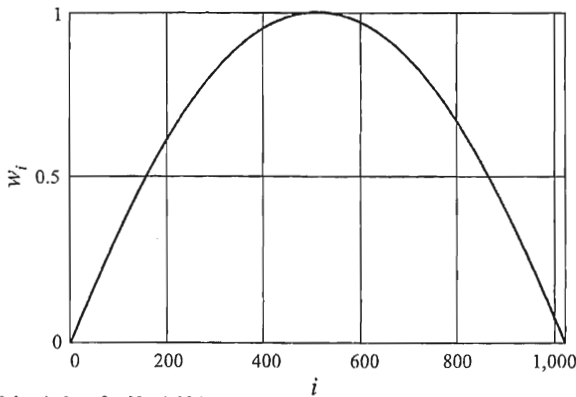$$w_i = \alpha + (1 - \alpha) \cos \left(\frac{2\pi i}{N} \right) \tag{6.19}$$

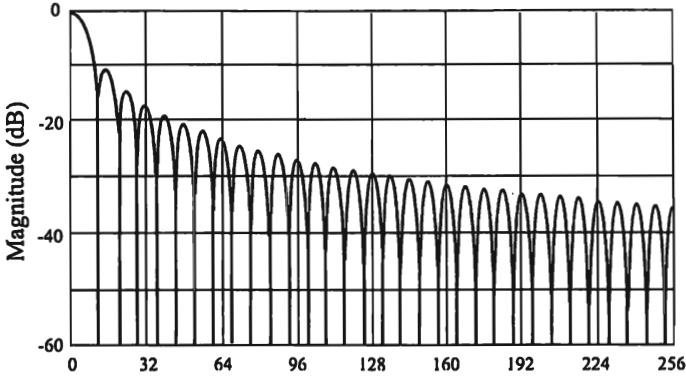**Figure 6.8** Welch window for $N = 1,024$.

Figure 6.9 Magnitude of the Welch spectrum for $N = 256$.

Two specific examples of the raised cosine window are the Hanning and Hamming windows. The Hanning window for N points is defined by $\alpha = 0.5$ as

$$w_i = 0.5 + 0.5\cos\left(\frac{2\pi i}{N}\right) \tag{6.20}$$

where $-N/2 \leq i \leq N/2$. The Hamming window is defined by $\alpha = 0.54$ as

$$w_i = 0.54 + 0.46\cos\left(\frac{2\pi i}{N}\right) \tag{6.21}$$

The Hanning and Hamming windows are illustrated in Figure 6.10. The magnitude of the Hanning spectrum is shown in Figure 6.11.

6.3.1.3 Kaiser-Bessel Window

The Kaiser or Kaiser-Bessel window approximates a time limited function with minimum energy outside a specified frequency band. In the discrete case it is defined as
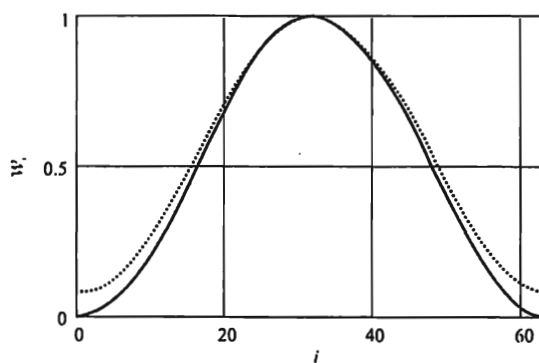
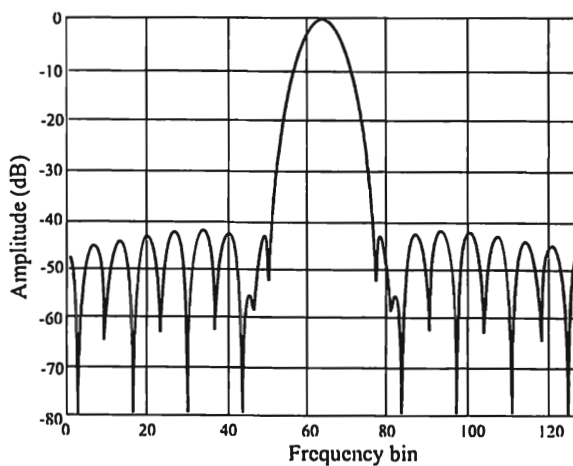**Figure 6.10** Hanning (solid line) and Hamming (dotted line) windows.

Figure 6.11 Magnitude of the spectrum for a Hanning window for $N = 128$.

$$w_{k,i} = \begin{cases} \dfrac{I_0\left\{B\left[1-\left(\dfrac{2i}{N-1}\right)^2\right]^{1/2}\right\}}{I_0(B)}, & -\dfrac{N-1}{2} \le i \le \dfrac{N-1}{2} \\ 0, & \text{otherwise} \end{cases} \qquad (6.22)$$

where B is half the time-bandwidth product. B determines the trade-off between the magnitude of the sidelobes and the energy in the main lobe. It is often specified as a half integer multiple of π. I_0 is the zero-order modified Bessel function of the first kind.

Closed-form expressions for the spectrum are not yet available. It can be shown, however, that the spectrum for the continuous case is proportional to

$$w_k(f) \propto \frac{\sin\left\{B\left[\left(\dfrac{f}{f_b}\right)^2-1\right]^{1/2}\right\}}{\left[\left(\dfrac{f}{f_b}\right)^2-1\right]^{1/2}} \qquad (6.23)$$

where $2\pi f_b$ is the width of the main lobe. Example curves for different values of B are shown in Figure 6.12.

6.3.2 Windows Summary

The definitions for some of the most common windows are given in Table 6.1. The performance of these windows is given in Table 6.2, while the highest sidelobe levels for most of the popular windows are shown in Figure 6.13 [6].

6.4 Frequency Domain Detector Performance

The performance of the frequency domain detector is similar to that in the time domain. The time series in question in this case is a parallel stream of data vectors that correspond to the power in the signal at a particular frequency (frequency bin actually). The probability of false alarm is given by
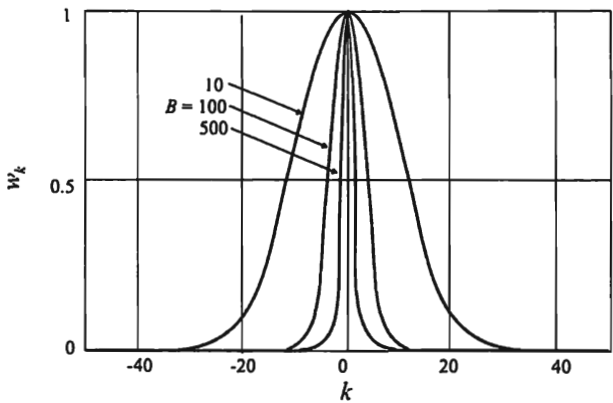
Figure 6.12 Kaiser-Bessel window for $N = 128$.

Table 6.1 Definitions for Some Discrete Windows

Window	Definition
Rectangle	$w_n = \begin{cases} 1, & n = 0,1,\cdots,N-1 \\ 0, & \text{otherwise} \end{cases}$
Triangle	$w_n = \begin{cases} \dfrac{n}{N/2}, & n = 0,1,\cdots,\dfrac{N}{2} \\ w(N-n), & n = \dfrac{N}{2},\cdots,N-1 \\ 0, & \text{otherwise} \end{cases}$
$\cos^\alpha(x)$	$w_n = \begin{cases} \sin^\alpha\left(\dfrac{n}{N}\pi\right), & n = 0,1,\cdots,N-1 \\ 0, & \text{otherwise} \end{cases}$
Hamming	$w_n = \begin{cases} 0.54 + 0.46\cos\left(\dfrac{n}{N}2\pi\right), & n = 0,1,\cdots,N-1 \\ 0, & \text{otherwise} \end{cases}$

Source: [6].

Table 6.1 (Continued)

Tukey	$$w_n = \begin{cases} 1, & 0 \le |n| \le \alpha\dfrac{N}{2} \\[2ex] 0.5\left\{1 + \cos\left[\pi \dfrac{n - \alpha\dfrac{N}{2}}{2(1-\alpha)\dfrac{N}{2}}\right]\right\}, & \alpha\dfrac{N}{2} < |n| \le \dfrac{N}{2} \\[2ex] 0, & \text{otherwise} \end{cases}$$
Bohman	$$w_n = \begin{cases} \left(1 - \dfrac{|n|}{N/2}\right)\cos\left(\pi\dfrac{|n|}{N/2}\right) + \dfrac{1}{\pi}\sin\left(\pi\dfrac{|n|}{N/2}\right), & 0 \le |n| \le \dfrac{N}{2} \\[2ex] 0, & \text{otherwise} \end{cases}$$
Poisson	$$w_n = \begin{cases} \exp\left(-\alpha\dfrac{|n|}{N/2}\right), & 0 \le |n| \le \dfrac{N}{2} \\[2ex] 0, & \text{otherwise} \end{cases}$$
Hanning-Poisson	$$w_n = \begin{cases} 0.5\left[1 + \cos\left(\pi\dfrac{n}{N/2}\right)\right]\exp\left(-\alpha\dfrac{|n|}{N/2}\right), & 0 \le \left|n \le \dfrac{N}{2}\right| \\[2ex] 0, & \text{otherwise} \end{cases}$$
Cauchy	$$w_n = \begin{cases} \dfrac{1}{1 + \left(\alpha\dfrac{n}{N/2}\right)^2}, & 0 \le |n| \le \dfrac{N}{2} \\[2ex] 0, & \text{otherwise} \end{cases}$$
Gaussian	$$w_n = \begin{cases} \exp\left[-\dfrac{1}{2}\left(\alpha\dfrac{n}{N/2}\right)^2\right], & 0 \le |n| \le \dfrac{N}{2} \\[2ex] 0, & \text{otherwise} \end{cases}$$

Table 6.1 (Continued)

Dolph-Chebyshev	$$w_n = \begin{cases} (-1)^n \dfrac{\cos\left\{N\cos^{-1}\left[\beta\cos\left(\pi\dfrac{k}{N}\right)\right]\right\}}{\cosh\left[N\cosh^{-1}(\beta)\right]}, & 0 \leq n \leq N-1 \\ 0, & \text{otherwise} \end{cases}$$ where $$\beta = \cosh\left[\frac{1}{N}\cosh^{-1}(10^\alpha)\right]$$ $$\cos^{-1}(x) = \begin{cases} \dfrac{\pi}{2} - \tan^{-1}\left(\dfrac{x}{\sqrt{1-x^2}}\right), & |x| \leq 1 \\ \ln\left(x + \sqrt{x^2-1}\right), & |x| > 1 \end{cases}$$
Kaiser-Bessel	$$w_n = \begin{cases} \dfrac{I_0\left[\pi\alpha\sqrt{1-\left(\dfrac{n}{N/2}\right)^2}\right]}{I_0(\pi\alpha)}, & 0 \leq |n| \leq \dfrac{N}{2} \\ 0, & \text{otherwise} \end{cases}$$ where $$I_0(x) = \sum_{k=0}^{\infty}\left[\frac{\left(\frac{x}{2}\right)^k}{k!}\right]^2$$
De La Valle Poussin	$$w_n = \begin{cases} 1 - 6\left(\dfrac{n}{N/2}\right)^2\left(1 - \dfrac{|n|}{N/2}\right), & 0 \leq |n| \leq \dfrac{N}{4} \\ 2\left(1 - \dfrac{|n|}{N/2}\right), & \dfrac{N}{4} < |n| \leq \dfrac{N}{2} \\ 0, & \text{otherwise} \end{cases}$$

Table 6.1 (Continued)

Barcilon-Temes	$$w_n = \begin{cases} (-1)^n \dfrac{A\cos[y(n)] + B\left[\dfrac{y(n)}{C}\right]\sin[y(n)]}{(C+AB)\left\{\left[\dfrac{y(n)}{C}\right]^2 + 1\right\}}, & 0 \le |n| \le \dfrac{N}{2} \\ 0, & \text{otherwise} \end{cases}$$ where $$A = \sinh(C) = \sqrt{10^{2\alpha} - 1}$$ $$B = \cosh(C) = 10^{\alpha}$$ $$C = \cosh^{-1}(10^{\alpha})$$ $$\beta = \cosh\left(\dfrac{C}{N}\right)$$ $$y(n) = N\cos^{-1}\left[\beta\cos\left(\pi\dfrac{n}{N}\right)\right]$$
Blackman	$$w_n = \begin{cases} \sum_{m=0}^{N/2} (-1)^m a_m \cos\left(\dfrac{mn}{N} 2\pi\right), & n = 0, 1, \ldots, N/2 \\ 0, & \text{otherwise} \end{cases}$$
Rieman	$$w_n = \begin{cases} \dfrac{\sin\left(\dfrac{n}{N} 2\pi\right)}{\left(\dfrac{n}{N} 2\pi\right)} = \operatorname{sinc}\left(\dfrac{2n}{N}\right), & 0 \le |n| \le \dfrac{N}{2} \\ 0, & \text{otherwise} \end{cases}$$
Reisz	$$w_n = \begin{cases} 1 - \left|\dfrac{n}{N/2}\right|^2, & 0 \le |n| \le \dfrac{N}{2} \\ 0, & \text{otherwise} \end{cases}$$

Table 6.2 Window Performance

Window	High SL Level (dB)	SL Fall-off (dB/oct)	Coh Gain	Equiv Noise BW (bins)	3 dB BW (bins)	Scallop Loss (dB)	Worst Case Process Loss (dB)	6 dB BW (bins)	Overlap Corr (%) 75% OL	50% OL
Rectangle	−13	−6	1.00	1.00	0.89	3.92	3.92	1.21	76.0	50.0
Triangle	−27	−12	0.50	1.33	1.28	1.82	3.0	1.78	71.9	26.0
cos$^\alpha$(x)										
α = 1	−23	−12	0.64	1.23	1.20	2.10	3.01	1.65	76.5	31.8
α = 2 (Hanning)	−32	−18	0.50	1.50	1.44	1.42	3.18	2.00	66.9	16.7
α = 3	−24	−24	0.42	1.73	1.66	1.08	3.47	2.32	56.7	8.5
α = 4	−30	−30	0.38	1.94	1.86	0.86	3.75	2.59	48.6	6.3
Hamming	−43	−6	0.54	1.36	1.30	1.78	3.10	1.81	70.7	23.5
Reisz	−21	−12	0.67	1.20	1.16	2.22	3.01	1.59	76.5	36.4
Rieman	−26	−12	0.59	1.30	1.26	1.89	3.03	1.74	73.4	27.4
De La Valle Poussin	−53	−24	0.38	1.92	1.82	0.90	3.72	2.55	49.3	6.0
Tukey										
α = 0.25	−14	−18	0.88	1.10	1.01	2.96	3.39	1.38	76.1	46.4
α = 0.50	−15	−18	0.75	1.22	1.15	2.24	3.11	1.57	72.7	36.4
α = 0.75	−19	−18	0.63	1.36	1.31	1.73	3.07	1.80	70.5	26.1
Bohman	−46	−24	0.41	1.79	1.71	1.02	3.54	2.38	56.5	7.4
Poisson										
α = 2	−19	−6	0.44	1.30	1.21	2.09	3.23	1.69	69.9	27.8
α = 3	−24	−6	0.32	1.65	1.45	1.46	3.64	2.08	56.8	16.1
α = 4	−31	−6	0.25	2.08	1.75	1.03	6.21	2.58	40.4	7.4
Hanning-Poisson										
α = 0.5	−35	−18	0.43	1.61	1.54	1.26	3.33	2.14	61.3	12.6
α = 1.0	−39	−18	0.38	1.73	1.64	1.11	3.50	2.30	56.0	9.2
α = 2.0	None	−18	0.29	2.02	1.87	0.87	3.94	2.65	46.6	6.7
Cauchy										
α = 3	−31	−6	0.42	1.48	1.34	1.71	3.40	1.90	61.6	20.2
α = 4	−35	−6	0.33	1.76	1.50	1.36	3.83	2.20	48.8	13.2
α = 5	−30	−6	0.28	2.06	1.68	1.13	6.28	2.53	38.3	9.0
Gaussian										
α = 2.5	−42	−6	0.51	1.39	1.33	1.69	3.14	1.86	67.7	20.0
α = 3.0	−55	-6	0.43	1.64	1.55	1.25	3.40	2.18	57.5	10.6
α = 3.5	−69	−6	0.37	1.90	1.79	0.94	3.73	2.52	47.2	6.9
Dolph-Chebyshev										
α = 2.5	−50	0	0.53	1.39	1.33	1.70	3.12	1.85	69.6	22.3
α = 3.0	−60	0	0.48	1.51	1.44	1.44	3.23	2.01	66.7	16.3
α = 3.5	−70	0	0.45	1.62	1.55	1.25	3.35	2.17	60.2	11.9
α = 6.0	−80	0	0.42	1.73	1.65	1.10	3.48	2.31	56.9	8.7
Kaiser-Bessel										
α = 2.5	−46	−6	0.49	1.50	1.43	1.46	3.20	1.99	66.7	16.9
α = 3.0	−57	−6	0.44	1.65	1.57	1.20	3.38	2.20	59.5	11.2
α = 3.5	−69	−6	0.40	1.80	1.71	1.02	3.56	2.39	53.9	7.4
α = 6.0	−82	−6	0.37	1.93	1.83	0.89	3.74	2.57	48.8	6.8
Barcilon-Temes										
α = 3.0	−53	−6	0.47	1.56	1.49	1.34	3.27	2.07	63.0	16.2
α = 3.5	−58	−6	0.43	1.67	1.59	1.18	3.40	2.23	58.6	10.4
α = 6.0	−68	−6	0.41	1.77	1.69	1.05	3.52	2.36	56.4	7.6
Exact Blackman	−51	−6	0.46	1.57	1.52	1.33	3.29	2.13	62.7	16.0
Blackman	−58	−18	0.42	1.73	1.68	1.10	3.47	2.35	56.7	9.0
Minimum 3-Sample Blackman-Harris (B-H)	−67	−6	0.42	1.71	1.66	1.13	3.45	1.81	57.2	9.6
Minimum 4-Sample B-H	−92	−6	0.36	2.00	1.90	0.83	3.85	2.72	46.0	3.8
61 dB 3-Sample B-H	−61	−6	0.45	1.61	1.56	1.27	3.34	2.19	61.0	12.6
74 dB 4-Sample B-H	−74	−6	0.40	1.79	1.74	1.03	3.56	2.44	53.9	7.4
4-Sample α = 3 Kaiser-Beswsel	−69	−6	0.40	1.80	1.74	1.02	3.56	2.44	53.9	7.4

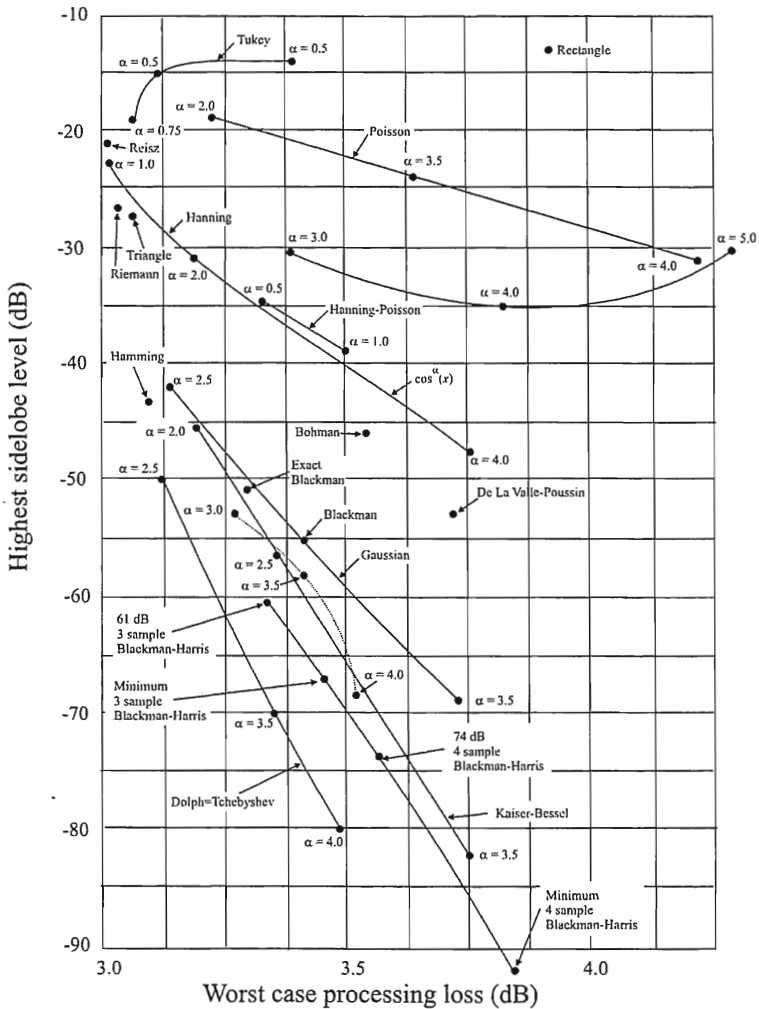SL: Sidelobe, Coh: Coherent, BW: Bandwidth, Corr: Correlation, OL: Overlap
Source: [6].

Figure 6.13 Levels of the highest sidelobes for many window types. (*From:* [6], © 1978 IEEE. Reprinted with permission.)

$$P_{fa} = \int_{\gamma}^{\infty} p(a)da = \int_{\gamma}^{\infty} \frac{a}{N^2\sigma^2} e^{-\frac{a^2}{2N^2\sigma^2}} da = e^{-\frac{\gamma^2}{2N^2\sigma^2}} \qquad (6.24)$$

where σ^2 is the variance of an individual data vector, assumed to be the same, and N represents the number of points used in the detection—the length of the FFT if that is the technique used [7].

The detection performance depends on the frequency of the input signal(s) relative to the placement of the frequency bins in the FFT and the length of the signal sample included in the FFT computation [7]. The variance in the detection performance is also $N\sigma^2$, however.

The amplitude pdf in this case is

$$p(a) = \frac{a}{N\sigma^2} e^{-\frac{a^2+X_i^2}{2N\sigma^2}} I_0\left(\frac{aX_i}{N\sigma^2}\right) \qquad (6.25)$$

The results are summarized in Table 6.3. The first row in Table 6.3 represents the best-case detection performance while that in the fourth row represents the worst-case conditions. For conditions other than those in Table 6.3, the detection performance falls between these two.

6.5 Concluding Remarks

The periodogram assumes that the data values, be they sampled from the original sequence or autocorrelation lags, outside the sample window are zero. This typically is an unrealistic assumption. This leads to distortion of the psd due to smearing and leakage into other frequency bins. Some of the high-resolution spectral estimation techniques discussed in Chapter 8 address this concern.

The advantages to using the FT or periodogram approach for spectrum estimation are:

- They are computationally efficient for small data sets.
- For sinusoidal processes, the peaks of psd are proportional to the power of the signal at that frequency.
- For sinusoids in noise, they provide optimal estimates.

The disadvantages are:

Table 6.3 Detection Probabilities in the Frequency Domain

Condition	Spectral Amplitude $\left\|X\left(k_i\right)\right\| \triangleq X_i$	P_d	Signal-to-Noise Ratio in Time Domain
Signal fills time window and spectral line hit	NA	$1 - \int_0^\gamma p(a)\,da$	$\dfrac{1}{N}\nu$
Signal fills window and input frequency 1/2 way between spectral lines	$\dfrac{2A}{\sqrt{2 - 2\cos\dfrac{\pi}{N}}}$	$1 - \sqrt{1 - P_2}$	$\dfrac{N\left[1 - \cos\left(\dfrac{\pi}{N}\right)\right]}{2}\nu$
Signal fills 1/2 time window and spectral line hit	$\dfrac{NA}{2}$	$1 - \int_0^\gamma p(a)\,da$	$\dfrac{2}{N}\nu$
Signal fills 1/2 window and input frequency 1/2 way between spectral lines	$\dfrac{A}{\sqrt{1 - \cos\left(\dfrac{\pi}{N}\right)}}$	$1 - \sqrt{1 - P_2}$	$N\left[1 - \cos\left(\dfrac{\pi}{N}\right)\right]\nu$

Source: [7].

P_2 represents the probability of detection from two adjacent spectral lines and ν is the SNR in the frequency domain.

- The sidelobes of the main lobe response can hide the main lobe of weaker nearby signals.
- The frequency resolution is limited to the reciprocal of the duration of the sample size. This produces relatively low resolution spectral estimates.
- Distortion of the psd occurs due to the finite sample time.
- Typically, several psd estimates must be averaged for statistical consistency.
- Negative psd regions can occur in some cases.

References

[1]   Roberts, L. R., *Signal Processing Techniques*, Anaheim, CA: Interstate Electronics Corporation, 1981, pp. 7-31–7-33.

[2]   Tsui, J., *Digital Techniques for Wideband Receivers,* Norwood, MA: Artech House, 1995.

[3]   Kay, S. M., and S. L. Marple, "Spectrum Analysis—A Modern Perspective," *Proceedings of the IEEE,* Vol. 69, No. 11, November 1981, pp. 1380–1419.

[4]   So, H. C., et al., "Comparison of Various Periodograms for Sinusoid Detection and Frequency Estimation," *IEEE Transactions on Aerospace and Electronic Systems,* Vol. 35, No. 3, July 1999, pp. 945–952.

[5]   Blackman, R. B., and J. W. Tukey, *The Measurement of Power Spectra from the Point of View of Communications Engineering*, New York: Dover, 1958.

[6]   Harris, F. J., "On the Use of Windows for Harmonic Analysis with the Discrete FT," *Proceedings of the IEEE,* Vol. 66, No. 1, January 1978, pp. 51–83.

[7]   Tsui, J., *Digital Techniques for Wideband Receivers,* Norwood, MA: Artech House, 1995, pp. 293–297.

Chapter 7

Detection of Deterministic Signals

The target set for communication EW systems consists of all types of communication systems of interest to the users of the EW system. It is frequently not known ahead of time what all those signals will be, so the signals arriving at the receiver belonging to the EW system likely have an unknown form until they are further processed [1].

The first step in such processing is to determine the presence of the signal at all. Herein that is known as the *signal detection problem*. To determine if a detected signal is a signal of interest, further processing of the signal is frequently required to determine or estimate parameters associated with the signal. A brief introduction to the theory of signal detection and, in particular, its application to communication EW problems is presented in this chapter.

Typically, in the spectrum of interest to an EW system there will be several signals present. These signals may be close together—indeed they can be in the same frequency channel. This leads to the question of how well can two (or more) signals be separated when they are close together and what are the impacts of signal amplitude on this resolvability. This chapter presents what are more traditional methods of signal detection, couched in terminology of spectral estimation. Spectral estimation, as discussed in Chapter 5, is determining (estimating) the frequency of signals and their amplitudes (or power levels). These techniques typically have low resolution compared to some other techniques. These latter methods are known, appropriately, as high or super resolution techniques for spectral estimation. Such approaches are presented in Chapter 8.

Detecting the presence of a signal is a different problem from estimating the value of some parameter. Detecting presence is a binary problem, or at least a problem of limited dimension—detecting the presence of one of m types of signals, for example. The variate takes on only certain values and there are a finite number of them. In parameter estimation, on the other hand, the variate can, in general, take on any value from a range of values—an infinite number of possibilities.

The communication EW problem is characterized by its noncooperative nature. Two communication nodes, trying to exchange information, know a considerable amount about the other node—that is, they cooperate with one another. The intercept of the signals exchanged between these nodes by a casual interloper is considerably more difficult. Frequently, very little is known by the EW system about the nodes or the signals they pass.

The parameters of signals can be deterministic or random. These parameters are the signal's amplitude, modulation, time of arrival, and phase of the carrier. The frequency may be known because it is either assumed to be the center of the channel being searched, or it can be measured as the position of the centroid of the spectrum or the highest peak of the spectrum.

Therefore, for the signal detection problem in communication EW systems, there are two cases of most interest. The signals are either:

- Deterministic except for one or more unknown parameters. That is, modulated deterministic sinusoidal with unknown frequency, amplitude, time of arrival, and phase.
- Completely random specified only by known (or assumed) statistical distributions.

The noise vector is assumed to be independent of both types of signal vectors.

Coherent detection of signals is possible when the phase of the carrier is known. For coherent detection, the receiver recovers the carrier phase and locks to it, thereby allowing for better detection performance. Coherent detection typically produces 3 dB (SNR) improved performance over incoherent detection. However, the phases of the signals in the pass band of the receiver in an EW system will almost never be known, at least at this stage of signal processing. Therefore, coherent detection is precluded—only incoherent detection will be considered.

Detection of deterministic signals with known as well as unknown parameters is discussed in this chapter. Detection of random signals is presented in Chapter 8.

7.1 Detection of Deterministic Signals with Known Parameters

When all the parameters associated with the signal to be detected are known, except for the delay, and the noise is AWGN, the optimum detector is the matched filter. This filter maximizes the SNR at its output; although it is not necessarily true that maximization of the output SNR is optimum. This will be shown in this section. While this is not the normal problem being addressed by the
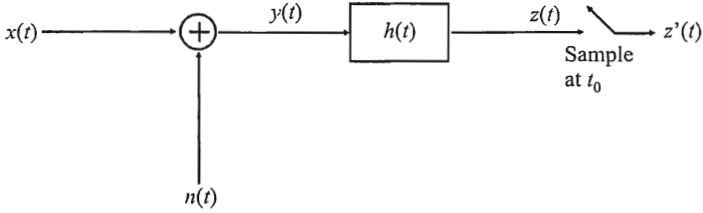
Figure 7.1 Block diagram of a filter.

communication EW system, characteristics of the matched filter are included for completeness and comparison. There could be times when all the parameters of the SOI are known or can be estimated fairly accurately.

7.1.1 Matched Filter Detection

When the signal characteristics are completely known and the signal is embedded in noise of known psd $S_{nn}(t)$, then the optimum detector is the matched filter. The matched filter, while maximizing the output SNR, produces the maximum P_d for a specified value of P_f when the noise is Gaussian [2]. Even when some of the signal parameters are unknown, the matched filter with minor modifications is still the optimum detector [3]. This filter does not necessarily maintain the structure of $x(t)$ at the output, but it does maximize the output SNR. Consider the block diagram shown in Figure 7.1. The filter input signal is

$$y(t) = x(t) + n(t) \tag{7.1}$$

Let the optimizing filter impulse response be given by $h(t)$, which is the function sought. The output $z(t)$ is given by

$$z(t) = \int_{-\infty}^{\infty} H(f)Y(f)e^{j2\pi ft}\, df \tag{7.2}$$

The maximum SNR is produced at time $t = t_0$ when the output is sampled.
    The peak signal at the output is given by

$$\text{peak signal} = \left| \int_{-\infty}^{\infty} H(f)X(f)e^{j2\pi ft_0}\, df \right|^2 \tag{7.3}$$

From Chapter 2, the psd of the noise at the output is $\Phi_{nn}(f)|H(f)|^2$ and the total noise at the output is given by

$$\text{total noise power} = \int_{-\infty}^{\infty} \Phi_{nn}(f)|H(f)|^2 \, df \qquad (7.4)$$

Therefore, the SNR to be maximized is given by

$$\nu = \frac{\left|\int_{-\infty}^{\infty} H(f)X(f)e^{j2\pi f t_0} df\right|^2}{\int_{-\infty}^{\infty} \Phi_{nn}(f)|H(f)|^2 \, df} \qquad (7.5)$$

In a Euclidean vector space, for any two vectors **a** and **b**, the Cauchy-Schwartz inequality states that [4]

$$\|\mathbf{a} \cdot \mathbf{b}\| \le \|\mathbf{a}\| \|\mathbf{b}\| \qquad (7.6)$$

where $\|\mathbf{x}\| = (\mathbf{x}, \mathbf{x})^{1/2}$ is the norm of **x**. The norm of **x** in a Euclidean vector space consisting of vectors α, β, and μ and constant c is any function over that vector space that obeys the following properties [5]:

$$(\alpha, \beta) = (\beta, \alpha) \qquad (7.7)$$
$$(\alpha + \beta, \mu) = (\alpha, \mu) + (\beta, \mu) \qquad (7.8)$$
$$(c\alpha, \beta) = c(\alpha, \beta) \qquad (7.9)$$
$$(\alpha, \alpha) > 0 \text{ unless } \alpha = 0 \qquad (7.10)$$

and **a**·**b** is the dot product of vectors **a** and **b**. Equality holds if and only if the vectors are collinear; that is, $\mathbf{a} = k\mathbf{b}^*$ for some $k \in \mathbb{R}$. In this case, the vectors are functions $U(f)$ and $V(f)$ with inner product given by

$$(\mathbf{a}, \mathbf{b}) = \int_{-\infty}^{\infty} U(f)V(f)df \qquad (7.11)$$

By direct substitution it can be shown that (7.11) satisfies (7.7) through (7.10). Thus,

$$\left| \int_{-\infty}^{\infty} U(f)V(f)df \right|^2 \le \int_{-\infty}^{\infty} |U(f)|^2 df \int_{-\infty}^{\infty} |V(f)|^2 df \qquad (7.12)$$

We can make the following associations in (7.5)

$$U(f) = \sqrt{\Phi_{nn}(f)}H(f) \qquad (7.13)$$

and

$$V(f) = \frac{X(f)e^{j2\pi f t_0}}{\sqrt{\Phi_{nn}(f)}} \qquad (7.14)$$

The Cauchy-Schwartz inequality yields

$$\nu = \frac{\left| \int_{-\infty}^{\infty} H(f)X(f)e^{j2\pi f t_0}df \right|^2}{\int_{-\infty}^{\infty} \Phi_{nn}(f)|H(f)|^2 df}$$

$$\le \frac{\int_{-\infty}^{\infty} \Phi_{nn}(f)H^2(f)df \int_{-\infty}^{\infty} \frac{|X(f)|^2 |e^{j2\pi f t_0}|^2}{\Phi_{nn}(f)}df}{\int_{-\infty}^{\infty} \Phi_{nn}(f)|H(f)|^2 df}$$

$$= \int_{-\infty}^{\infty} \frac{|X(f)|^2}{\Phi_{nn}(f)}df \qquad (7.15)$$

since $\left| e^{j2\pi f t_0} \right| = 1$, and where ν is the SNR. Therefore, the maximum SNR is independent of $H(f)$. The maximum SNR is achieved when equality holds so that

$$\sqrt{\Phi_{nn}(f)}H(f) = k\left[\frac{X(f)e^{j2\pi f t_0}}{\sqrt{\Phi_{nn}(f)}}\right]^*$$ (7.16)

and the optimum filter transfer function that ensues is

$$H(f) = k\frac{X^*(f)e^{-j2\pi f t_0}}{\Phi_{nn}(f)}$$ (7.17)

The filter response is "matched" to the conjugate of the psd of the signal, thus the appellation "matched filter." The resulting maximum SNR is then

$$\nu_{max} = \int_{-\infty}^{\infty}\frac{|X(f)|^2}{\Phi_{nn}(f)}df$$ (7.18)

Calculation of (7.17) requires knowledge of the noise psd, both in the design of the filter and in operation, since the noise environment in which EW systems operate varies with time and location in the frequency spectrum.
    When

$$\Phi_{nn}(f) = \frac{N_0}{2}$$ (7.19)

which corresponds to AWGN, this filter is

$$H(f) = \frac{2k}{N_0}X^*(f)e^{-j2\pi f t_0}$$ (7.20)

and

$$\nu_{max} = \frac{2}{N_0}\int_{-\infty}^{\infty}|X(f)|^2\,df$$ (7.21)

The filter impulse response is

$$h(t) = \int_{-\infty}^{\infty} H(f) e^{j2\pi ft} df$$

$$= \int_{-\infty}^{\infty} \frac{2k}{N_0} X^*(f) e^{-j2\pi ft_0} e^{j2\pi ft} df$$

$$= \frac{2k}{N_0} \int_{-\infty}^{\infty} X^*(f) e^{j2\pi f(t-t_0)} df$$

$$= \frac{2k}{N_0} \left[\int_{-\infty}^{\infty} X(f) e^{j2\pi f(t_0-t)} df \right]^*$$

$$= \frac{2k}{N_0} x^*(t_0 - t) \tag{7.22}$$

Therefore, the filter impulse response is equivalent to the signal conjugated, reversed in time, and delayed by t_0.

· The output signal has the form

$$(x*h)(t) = \frac{2k}{N_0} \left[x(t) * x^*(t_0 - t) \right] = \frac{2k}{N_0} \gamma_{xx}(t - t_0) \tag{7.23}$$

Therefore, matched filter detection when searching for known signals in noise will perform best when the signal has a large and peaked autocorrelation function $\gamma_{xx}(t - t_0)$.

7.1.2 Matched Filter Performance

Given that the signal is given by sample vector **x**, the detection problem is given by the two hypotheses

$$H_0 : x_i = n_i \qquad i = 0, 1, \ldots, N-1 \tag{7.24}$$

$$H_1 : x_i = s_i + n_i \quad i = 1, 2, \ldots, N-1 \tag{7.25}$$

where s_i are samples of the known signal and n_i are samples of the noise. The Neyman-Pearson likelihood ratio test for detecting a known deterministic signal with Gaussian noise of variance σ^2 is given by

$$L(\mathbf{x}) = \frac{p(\mathbf{x}:H_1)}{p(\mathbf{x}:H_0)} > \gamma \tag{7.26}$$

where $\mathbf{x} = [x_0, x_1, \cdots, x_{N-1}]^T$ and

$$p(\mathbf{x}:H_1) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=0}^{N-1} (x_i - s_i)^2 \right] \tag{7.27}$$

$$p(\mathbf{x}:H_0) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=0}^{N-1} x_i^2 \right] \tag{7.28}$$

Thus the test is

$$L(\mathbf{x}) = \exp\left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=0}^{N-1} (x_n - s_n)^2 - \sum_{i=0}^{N-1} x_n^2 \right] \right\} > \gamma \tag{7.29}$$

Taking the natural logarithm of both sides yields

$$\lambda(\mathbf{x}) = \ln[L(\mathbf{x})] = -\frac{1}{2\sigma^2} \left[\sum_{i=0}^{N-1} (x_n - s_n)^2 - \sum_{i=0}^{N-1} x_n^2 \right] > \ln\gamma \tag{7.30}$$

The test statistic is to choose H_1 if

$$\frac{1}{\sigma^2} \sum_{i=0}^{N-1} x_i s_i - \frac{1}{2\sigma^2} \sum_{i=0}^{N-1} s_i^2 > \ln\gamma \tag{7.31}$$

or

$$t(\mathbf{x}) = \sum_{i=0}^{N-1} x_n s_n > \sigma^2 \ln\gamma + \frac{1}{2} \sum_{i=0}^{N-1} s_n^2 = \gamma' \tag{7.32}$$

The second term on the right in (7.32) is the energy in s and is known because it is assumed that s is known. Thus, decide H_1 if

(a)

(b)
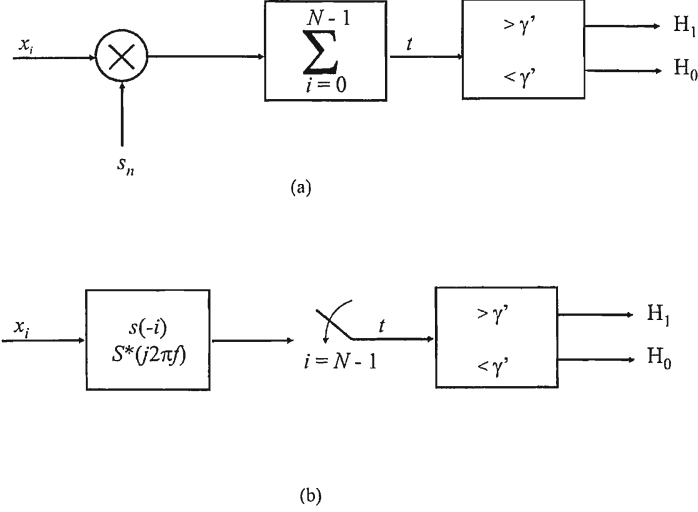
Figure 7.2 Neyman-Pearson detector for a deterministic signal in AWGN (a) correlator and (b) matched filter.

$$t(\mathbf{x}) = \sum_{i=0}^{N-1} x_n s_n > \gamma' \qquad (7.33)$$

Two equivalent forms of the optimum detector employing the Neyman-Pearson criterion for a known signal in AWGN are shown in Figure 7.2. The configuration shown in Figure 7.2(a) is called a *correlator*, while that in Figure 7.2(b) is called a *matched filter*.

Test $t(\mathbf{x})$ is a linear sum of Gaussian random variables and therefore is itself Gaussian. Then

$$\mathcal{E}\{t(\mathbf{x}) : H_0\} = \mathcal{E}\left\{ \sum_{i=0}^{N-1} n_i s_i \right\} = 0 \qquad (7.34)$$

$$\mathcal{E}\{t(\mathbf{x}) : H_1\} = \mathcal{E}\left\{ \sum_{i=0}^{N-1} (s_i + n_i) s_i \right\} = \mathbb{E} \qquad (7.35)$$

where $\mathbb{E}$ is the energy in the signal. In addition
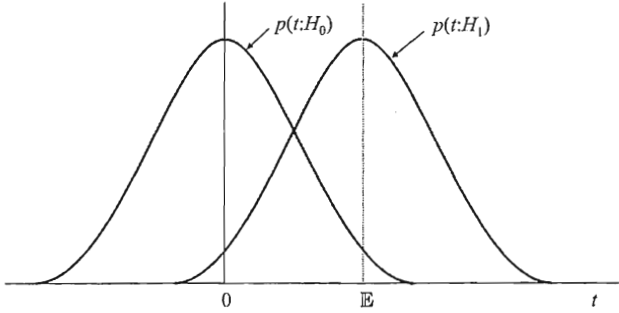
Figure 7.3 Pdfs for the matched filter when the signal is completely known.

$$\mathrm{var}[t(\mathbf{x}) : H_0] = \mathrm{var}\left(\sum_{i=0}^{N-1} w_i s_i\right)$$

$$= \sum_{i=0}^{N-1} \mathrm{var}(w_i) s_i^2$$

$$= \sigma^2 \sum_{i=0}^{N-1} s_i^2 = \sigma^2 \mathbb{E} \qquad (7.36)$$

Likewise, $\mathrm{var}[t(\mathbf{x}) : H_1] = \sigma^2 \mathbb{E}$. Therefore,

$$t \sim \begin{cases} \mathcal{N}(0, \sigma^2 \mathbb{E}) & \text{under } H_0 \\ \mathcal{N}(\mathbb{E}, \sigma^2 \mathbb{E}) & \text{under } H_1 \end{cases} \qquad (7.37)$$

The pdfs corresponding to this test are illustrated in Figure 7.3.

This test statistic is often scaled by $1/\sqrt{\sigma^2 \mathbb{E}}$, which makes the test

$$t \sim \begin{cases} \mathcal{N}(0,1) & \text{under } H_0 \\ \mathcal{N}(\sqrt{\mathbb{E}/\sigma^2}, 1) & \text{under } H_1 \end{cases} \qquad (7.38)$$

As discussed in Kay [6], this test statistic is equivalent to that of a matched filter given by (7.33). Now

$$P_{\mathrm{fa}} = \mathrm{Pr}(t > \gamma' : H_0) \qquad (7.39)$$

$$= Q\left(\frac{\gamma'}{\sqrt{\sigma^2 \mathbb{E}}}\right) \tag{7.40}$$

and

$$P_d = \Pr(t > \gamma' : H_1) \tag{7.41}$$

$$= Q\left(\frac{\gamma' - \mathbb{E}}{\sqrt{\sigma^2 \mathbb{E}}}\right) \tag{7.42}$$

where $Q(\ )$ is the Q-function. Since the Q-function is monotonic in its argument, it has an inverse denoted by $Q^{-1}(\ )$ and

$$\gamma' = \sqrt{\sigma^2 \mathbb{E}} Q^{-1}(P_{fa}) \tag{7.43}$$

Thus,

$$P_d = Q\left(Q^{-1}(P_{fa}) - \sqrt{\frac{\mathbb{E}}{\sigma^2}}\right) \tag{7.44}$$

$d^2 = \mathbb{E}/\sigma^2$ is referred to as the *deflection coefficient*. Equation (7.44) is plotted in Figure 7.4 for several values of the parameters.

7.2 Detection of Deterministic Signals with Unknown Parameters

The detection of deterministic signals where one or more of the parameters is unknown is presented in this section. The parameters in question can be any of the set $\{S, \phi, f_0\}$, where S is the average power, ϕ is the phase angle, $\phi \in [0, 2\pi)$, and f_0 is the carrier frequency.

7.2.1 Quadrature Detector

The quadrature receiver is a noncoherent technique for detecting signals in the time domain. Such a detector is illustrated in Figure 7.5. With no signal present, the Gaussian noises from the lowpass filters have the pdf given by
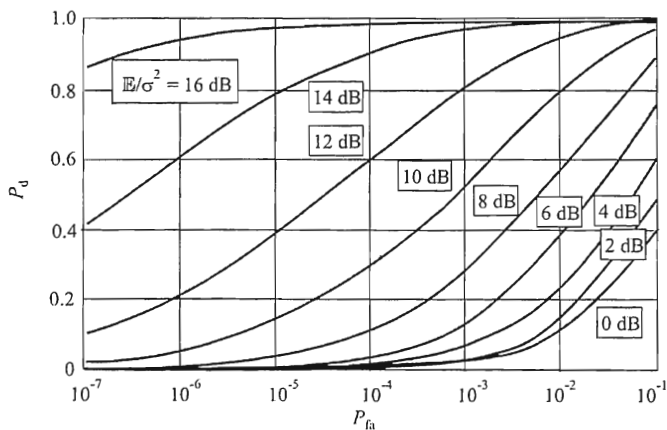
Figure 7.4 Performance of a matched filter.
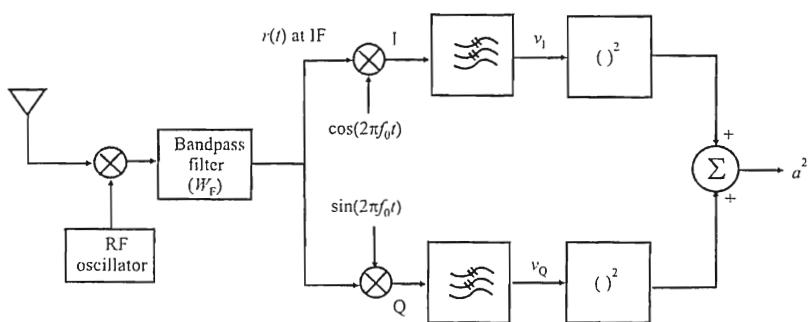
Figure 7.5 Quadrature detector.

$$p(v_1) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{v_1^2}{2\sigma^2}} \qquad (7.45)$$

and

$$p(v_Q) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{v_Q^2}{2\sigma^2}} \qquad (7.46)$$

where σ^2 is the variance of the stochastic noise process, assumed to be the same for both channels. It is also the average power of the noise. Noise processes v_I and v_Q are assumed to be i.i.d., so therefore their joint pdf is given by

$$p(v_1, v_Q) = p(v_1)p(v_Q) \qquad (7.47)$$

The pdf of a is therefore given by

$$p(a) = \int_0^{2\pi} ap(v_1) p(v_Q) d\theta = \frac{1}{2\pi\sigma^2} \int_0^{2\pi} ae^{\frac{-(v_1^2 + v_Q^2)}{2\sigma^2}} d\theta \qquad (7.48)$$

which is

$$p(a) = \frac{a}{\sigma^2} e^{-\frac{a^2}{2\sigma^2}} \qquad (7.49)$$

where $a = \sqrt{v_1^2 + v_Q^2}$ and $\theta = \tan^{-1}(v_Q/v_1)$. Expression (7.49) can be identified as the Rayleigh pdf.

A false alarm occurs when there is no signal present but the detector indicates that there is. The probability of this occurring is given by

$$P_{fa} = \int_\gamma^\infty p(a) da = \int_\gamma^\infty \frac{a}{\sigma^2} e^{-\frac{a^2}{2\sigma^2}} da = e^{-\frac{\gamma^2}{2\sigma^2}} \qquad (7.50)$$

Constant false alarm rate detection is accomplished by setting the probability of false alarm, $P_{fa} = P_1(\Gamma_0)$, to the maximum that is tolerable and making an estimate of the noise power present as given by σ^2. The threshold γ is then adjusted to

maintain the false alarm rate. Note that if the noise level changes, new estimates must be calculated. Changing noise levels are common in EW system operation.

When there is a signal present then

$$p(v_I) = \frac{1}{\sqrt{2\sigma^2}} e^{-\frac{(v_I - \bar{v}_I)^2}{2\sigma^2}} \tag{7.51}$$

and

$$p(v_Q) = \frac{1}{\sqrt{2\sigma^2}} e^{-\frac{(v_Q - \bar{v}_Q)^2}{2\sigma^2}} \tag{7.52}$$

where $\bar{v}_I$ and $\bar{v}_Q$ are the average values of the voltages coming from the lowpass filters. These values are given by

$$\bar{v}_I = A\cos\phi \tag{7.53}$$

and

$$\bar{v}_Q = A\sin\phi \tag{7.54}$$

where ϕ is the initial phase of the input signal. Likewise,

$$v_I = a\cos\theta \tag{7.55}$$

and

$$v_Q = a\sin\theta \tag{7.56}$$

The joint amplitude and phase density function is

$$p(r,\theta|\phi) = ap(v_I)p(v_Q)$$
$$= \frac{a}{2\pi\sigma^2} e^{-\frac{\left[a^2 + A^2 - 2aA(\cos\phi\cos\theta + \sin\phi\sin\theta)\right]}{2\sigma^2}}$$

$$= \frac{a}{2\pi\sigma^2} e^{-\frac{\left[a^2+A^2-2aA\cos(\phi-\theta)\right]}{2\sigma^2}} \tag{7.57}$$

The dependence on the phase can be removed by averaging over $(0, 2\pi)$ as

$$p\left(a|\phi\right) = \int_0^{2\pi} p\left(a,\theta|\phi\right)d\theta$$

$$= \frac{a}{\sigma^2} e^{-\frac{a^2+A^2}{2\sigma^2}} I_0\left(\frac{aA}{\sigma^2}\right) \tag{7.58}$$

which is recognized as the Ricean pdf. Notice that the expression on the right is independent of ϕ.

Note that (7.50) and (7.58) are based on making a detection decision on a single sample. Higher reliability in these decisions can be obtained by using more than one sample. If at least M out of the N samples must indicate the corresponding decision, then

$$P_{\text{fa,multiple}} = \sum_{k=\left\lfloor\frac{M}{2}\right\rfloor+1}^{N} \binom{N}{k} P_{\text{fa}}^k \left(1-P_{\text{fa}}\right)^{N-k} \tag{7.59}$$

and

$$P_{\text{d,multiple}} = \sum_{k=\left\lfloor\frac{M}{2}\right\rfloor+1}^{N} \binom{N}{k} P_{\text{d}}^k \left(1-P_{\text{d}}\right)^{N-k} \tag{7.60}$$

where $\lfloor x \rfloor$ indicates the integer part of x.

The block diagram of a near-optimum receiver where k noncoherent samples are to be used for detection is shown in Figure 7.6. The square root function provides for detection of the envelope of $r(t)$. Samples of this envelope are integrated (summed) in the postdetection integrator, the output of which is sampled after k samples are collected.
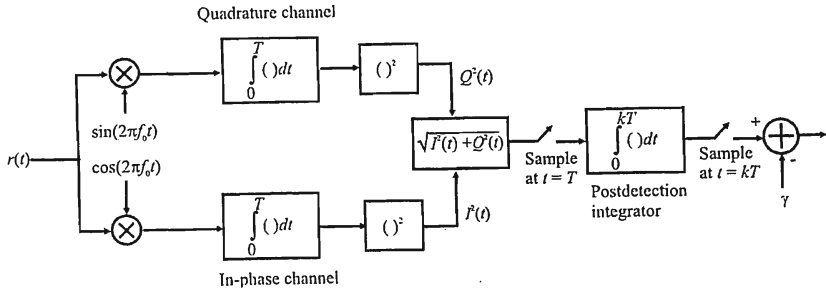
Figure 7.6 Quadrature envelope detector for detection of k noncoherent samples.

There is a noncoherent detection loss associated with the detector in Figure 7.6 compared with coherent detection (when the phase of $r(t)$ is known). This loss is tabulated in the chart shown in Figure 7.7 [7]. Thus, for example, when $v = 10$ dB and $k = 200$, there is an additional approximately 8-dB loss due to noncoherently adding the samples.

Until now it has been assumed that the local oscillator that generates the sin and cos signals is perfectly tuned to the frequency of $r(t)$. This is likely to be rare, and there is an associated loss due to this offset. The loss can be expressed as
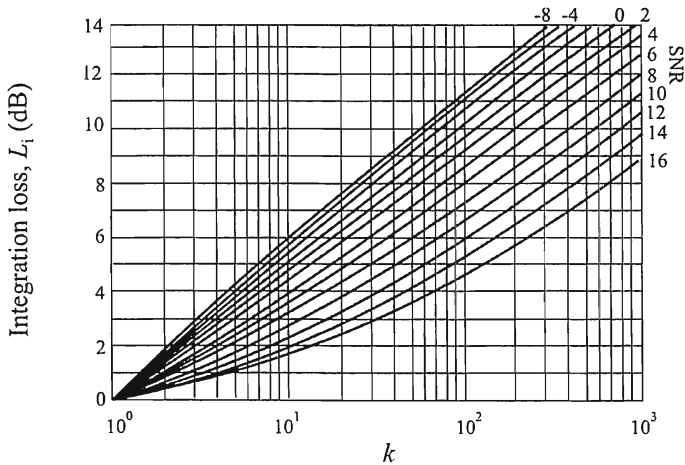

Figure 7.7 Noncoherent integration loss. The SNR in this chart is the postdetection signal power-to-noise power ratio. (*From:* [7]. © 1988 Computer Science Press. Reprinted with permission.)
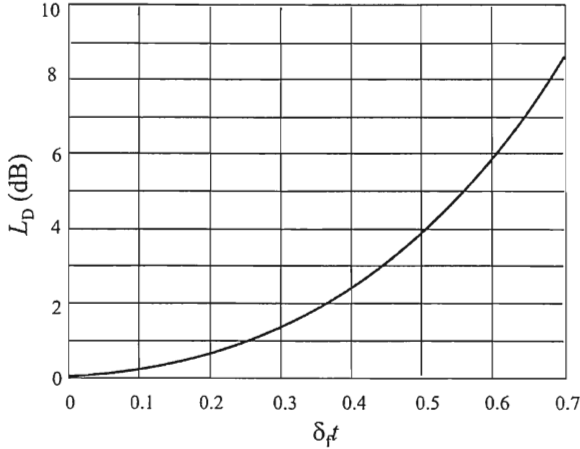
Figure 7.8 Frequency offset loss for quadratic detectors.

$$L_D = \left[\frac{\sin(\pi \delta_f t)}{\pi \delta_f t} \right]^2 \tag{7.61}$$

where δ_f is the frequency offset. This is plotted in Figure 7.8.

7.2.2 GLRT Detection

The development in this section follows that in Kay [6] fairly closely. Assume that the observation samples are measured over a time window given by N samples. The two hypotheses for this problem are given by

$$
\begin{aligned}
H_0 &: x_i = n_i, & i = 0,1,\dots,N-1 \\
H_1 &: x_i = s_i + n_i, & i = 0,1,\dots,N-1
\end{aligned}
\tag{7.62}
$$

where

$$s_n = A\cos(2\pi f_0 n + \phi) \tag{7.63}$$

and $A = \sqrt{2S}$ when S is the average power in the signal. In this case, any or all of the set $\{A, f_0, \phi\}$ are unknown. The noise is represented by samples $n_i \sim \mathcal{N}(0, \sigma^2)$.

Figure 7.6 Quadrature envelope detector for detection of k noncoherent samples.

There is a noncoherent detection loss associated with the detector in Figure 7.6 compared with coherent detection (when the phase of $r(t)$ is known). This loss is tabulated in the chart shown in Figure 7.7 [7]. Thus, for example, when $v = 10$ dB and $k = 200$, there is an additional approximately 8-dB loss due to noncoherently adding the samples.

Until now it has been assumed that the local oscillator that generates the sin and cos signals is perfectly tuned to the frequency of $r(t)$. This is likely to be rare, and there is an associated loss due to this offset. The loss can be expressed as

Figure 7.7 Noncoherent integration loss. The SNR in this chart is the postdetection signal power-to-noise power ratio. (*From:* [7]. © 1988 Computer Science Press. Reprinted with permission.)

Figure 7.8 Frequency offset loss for quadratic detectors.

$$L_{\mathrm{D}} = \left[\frac{\sin\left(\pi\delta_{\mathrm{f}}t\right)}{\pi\delta_{\mathrm{f}}t} \right]^2 \tag{7.61}$$

where δ_{f} is the frequency offset. This is plotted in Figure 7.8.

7.2.2 GLRT Detection

The development in this section follows that in Kay [6] fairly closely. Assume that the observation samples are measured over a time window given by N samples. The two hypotheses for this problem are given by

$$\begin{aligned} H_0 &: x_i = n_i, & i = 0, 1, \ldots, N-1 \\ H_1 &: x_i = s_i + n_i, & i = 0, 1, \ldots, N-1 \end{aligned} \tag{7.62}$$

where

$$s_n = A\cos(2\pi f_0 n + \phi) \tag{7.63}$$

and $A = \sqrt{2S}$ when S is the average power in the signal. In this case, any or all of the set $\{A, f_0, \phi\}$ are unknown. The noise is represented by samples $n_i, \sim \mathcal{N}(0, \sigma^2)$.

The optimal detector, that is, a UMP test, does not exist when any of the parameters is unknown. The GLRT discussed in Chapter 3 produces a suboptimal detector, but in most cases an acceptable one. The GLRT will decide H_1 if

$$\frac{p(\mathbf{x}:\hat{A},\hat{\phi},\hat{f}_0,H_1)}{p(\mathbf{x}:H_0)} > \gamma \tag{7.64}$$

where $\hat{A},\hat{\phi},$ and $\hat{f}_0$ are the MLE estimates of A, ϕ, and f_0, respectively, which are given by [6]

$$\hat{A} = \sqrt{\hat{\alpha}_1^2 + \hat{\alpha}_2^2} \tag{7.65}$$

$$\hat{\phi} = \tan^{-1}\left(-\frac{\hat{\alpha}_2}{\hat{\alpha}_1}\right) \tag{7.66}$$

where

$$\hat{\alpha}_1 = \frac{2}{N}\sum_{i=0}^{N-1} x_i \cos(2\pi\hat{f}_0 i) \tag{7.67}$$

$$\hat{\alpha}_2 = \frac{2}{N}\sum_{i=0}^{N-1} x_i \sin(2\pi\hat{f}_0 i) \tag{7.68}$$

and $\hat{f}_0$ is determined by computing the periodogram and finding the frequency where the largest peak occurs. Denoting the periodogram evaluated at f_0 as $I(f_0)$

$$I(f_0) = \frac{1}{N}\left|\sum_{i=0}^{N-1} x_i\, e^{-j2\pi f_0 i}\right|^2 \tag{7.69}$$

and using (7.64),

$$\ln\frac{p(\mathbf{x}:\hat{A},\hat{\phi},f_0,H_1)}{p(\mathbf{x}:H_0)} = \frac{1}{\sigma^2} I(f_0) > \ln\gamma \tag{7.70}$$

Thus: for a deterministic signal when the amplitude, phase, and frequency are all unknown, decide H_1 if

$$I(f_0) > \sigma^2 \ln\gamma \tag{7.71}$$
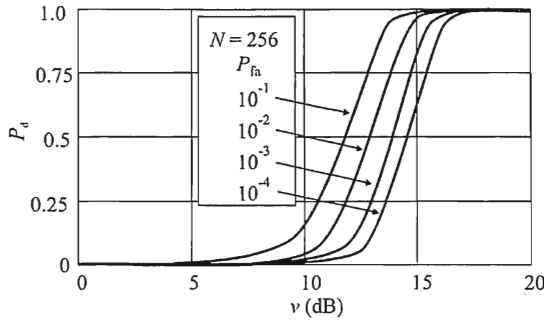
Figure 7.9 Typical performance curves for the GLRT decision logic when the signal is deterministic with unknown amplitude, phase, and frequency.

Therefore, the decision that a signal is present is made if the maximum value of the periodogram exceeds a threshold. When it does, that peak is at the MLE of the frequency.

The probability of detection is related to the other parameters through the expression

$$P_d = Q_{\chi_2'^2\left(\frac{NA^2}{2\sigma^2}\right)}\left(2\ln\frac{N/2-1}{P_{fa}}\right) \tag{7.72}$$

where there are N observations and

$$
Q_{\chi_2'^2\left(\frac{NA^2}{2\sigma^2}\right)}(x) =
\begin{cases}
\exp\left(-\dfrac{1}{2}x\right)\displaystyle\sum_{k=0}^{\frac{\nu}{2}-1}\dfrac{\left(\dfrac{x}{2}\right)^k}{k!}, & \nu > 1, \nu \text{ even} \\[3ex]
2Q(\sqrt{x}), & \nu = 1 \\[2ex]
2Q(\sqrt{x}) + \dfrac{\exp\left(-\dfrac{1}{2}x\right)}{\sqrt{\pi}}\displaystyle\sum_{k=1}^{\frac{\nu-1}{2}}\dfrac{(k-1)!(2x)^{k-\frac{1}{2}}}{(2k-1)!}, & \nu > 2, \nu \text{ odd}
\end{cases}
\tag{7.73}
$$

is the right-tail probability for a χ-square rv. The noncentral χ-square rv with noncentrality parameter ξ is denoted by χ_ξ^2. Typical performance curves for this GLRT detector are illustrated in Figure 7.9.
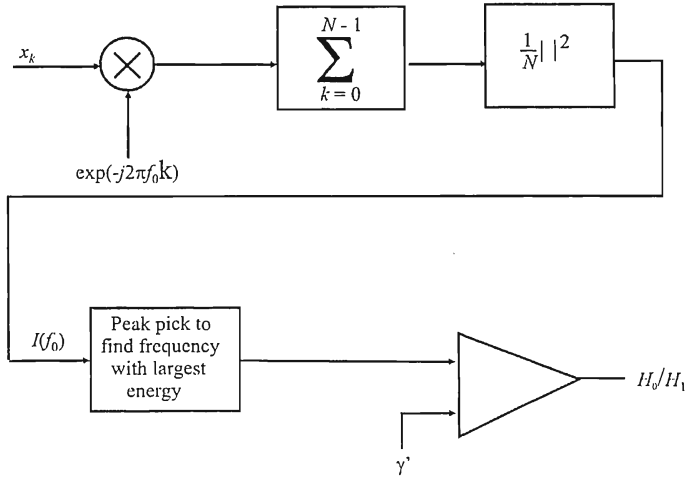
Figure 7.10 GLRT detector when the signal is deterministic with unknown amplitude, phase, and frequency.

The block diagram for the structure that implements this decision logic is shown in Figure 7.10.

Urkowitz followed a similar line of reasoning to develop an approximation to the optimal receiver for deterministic signals with unknown parameters [8]. He based his development on the approximation that a signal uniformly sampled at a rate of $2W$, where W is the highest frequency component present, for T seconds can be approximately recovered from these samples. The rationale for this approximation is presented in the appendix of [8]. Thus, for any signal $s(t)$,

$$s(t) \approx \sum_{i=1}^{2TW} s_i \operatorname{sinc}(2Wt - i) \tag{7.74}$$

where

$$s_i = s_i / 2W \tag{7.75}$$

This decomposition is based on the orthogonal property of the sinc function

$$\int_{-\infty}^{\infty} \operatorname{sinc}(2Wt - i)\operatorname{sinc}(2Wt - k)dt = \begin{cases} 1/2W, & i = k \\ 0, & i \neq k \end{cases} \tag{7.76}$$

Assuming that the signal is represented as a bandpass process, then the noise can be expressed as

$$n(t) = \mathrm{Re}[\tilde{n}(t)e^{j2\pi f_0 t}]$$

where

$$\tilde{n}(t) = n_I(t) + jn_Q(t) \tag{7.77}$$

is known as the complex envelope while $n_I(t)$ and $n_Q(t)$ are the real and imaginary components. Thus,

$$
\begin{aligned}
n(t) &= \mathrm{Re}\{[n_I(t) + jn_Q(t)][\cos(2\pi f_0 t) + j\sin(2\pi f_0 t)]\} \\
&= \mathrm{Re}[n_I(t)\cos(2\pi f_0 t) + jn_I(t)\sin(2\pi f_0 t) \\
&\qquad + jn_Q(t)\cos(2\pi f_0 t) - n_Q(t)\sin(2\pi f_0 t)] \\
&= \mathrm{Re}\{[n_I(t)\cos(2\pi f_0 t) - n_Q(t)\sin(2\pi f_0 t)] \\
&\qquad + j[n_I(t)\sin(2\pi f_0 t) + n_Q(t)\cos(2\pi f_0 t)]\} \\
&= n_I(t)\cos(2\pi f_0 t) - n_Q(t)\sin(2\pi f_0 t) \tag{7.78}
\end{aligned}
$$

where f_0 is the frequency in the center of the pass band, $n_I(t)$ is the in-phase component of the noise, and $n_Q(t)$ is the quadrature component. The noise energy is

$$
\begin{aligned}
\int_0^T n_I^2(t)\,dt &= \frac{1}{W}\sum_{i=1}^{TW} a_{Ii}^2 \\
\int_0^T n_Q^2(t)\,dt &= \frac{1}{W}\sum_{i=1}^{TW} a_{Qi}^2
\end{aligned}
\tag{7.79}
$$

where

$$
\begin{aligned}
a_{Ii} &= n_{Ii}/W \\
a_{Qi} &= n_{Qi}/W
\end{aligned}
\tag{7.80}
$$

Define measures of SNR as

$$b_{Ii} = \frac{a_{Ii}}{\sqrt{2WN_0}} \qquad b_{Qi} = \frac{a_{Qi}}{\sqrt{2WN_0}} \tag{7.81}$$

where N_0 is the band-limited (to width W) noise psd. WN_0 is the total noise power in either of the quadrature channels when N_0 is the two-sided noise psd. Under H_0, the test statistic is

$$V' = \frac{1}{N_0} \int_0^T n^2(t)dt$$
$$= \sum_{n=1}^{TW} (b_{Ii}^2 + b_{Qi}^2) \tag{7.82}$$

The variance of all of the b_{Ii} and b_{Qi} is unity, so the sum in the second line in (7.82) has a χ-square distribution with TW degrees of freedom.

Just as the noise can be represented in quadrature as in (7.78), the signal can be represented in quadrature as well:

$$s(t) = s_I(t) \cos(2\pi f_0 t) - s_Q(t) \sin(2\pi f_0 t) \tag{7.83}$$

where $s_I(t)$ and $s_Q(t)$ are lowpass functions band-limited to $|f| \leq W/2$. These components can be represented as

$$s_I(t) = \sum_{i=1}^{TW} \alpha_{Ii} \, \text{sinc}(Wt - i)$$
$$s_Q(t) = \sum_{i=1}^{TW} \alpha_{Qi} \, \text{sinc}(Wt - i) \tag{7.84}$$

where

$$\alpha_{Ii} = s_{Ii/W}$$
$$\alpha_{Qi} = s_{Qi/W} \tag{7.85}$$

The observed signal plus noise is thus given by

$$x(t) = [s_I(t) + n_I(t)] \cos(2\pi f_0 t) - [s_Q(t) + n_Q(t)] \sin(2\pi f_0 t)$$
$$= x_I(t) \cos(2\pi f_0 t) - x_Q(t) \sin(2\pi f_0 t) \tag{7.86}$$

These components, in turn, can be represented as

$$x_I(t) = s_I(t) + n_I(t) = \frac{1}{W} \sum_{i=1}^{TW} (\alpha_{Ii} + a_{Ii}) \operatorname{sinc}(Wt - i)$$

$$x_Q(t) = s_Q(t) + n_Q(t) = \frac{1}{W} \sum_{i=1}^{TW} (\alpha_{Qi} + a_{Qi}) \operatorname{sinc}(Wt - i) \tag{7.87}$$

Again, define SNR measures

$$\beta_{Ii} = \frac{s_{Ii/W}}{\sqrt{2WN_0}}$$

$$\beta_{Qi} = \frac{s_{Qi/W}}{\sqrt{2WN_0}} \tag{7.88}$$

The SNR is

$$\frac{1}{N_0} \int_0^T s^2(t) dt = \sum_{i=1}^{TW} (\beta_{Ii}^2 + \beta_{Qi}^2) = \frac{E_s}{N_0} \tag{7.89}$$

where E_s is the signal energy.

Under hypothesis H_1, where it is assumed that the signal is present, the total energy in the observation is

$$\int_0^T x^2(t) dt = \frac{1}{2} \int_0^T [x_I^2(t) + x_Q^2] dt$$

$$= \frac{1}{2} \int_0^T [s_I^2(t) + n_I^2] dt + \frac{1}{2} \int_0^T [s_Q^2(t) + n_Q^2] dt \tag{7.90}$$

which yields the test statistic

$$V' = \sum_{i=0}^{TW} [(b_{Ii} + \beta_{Ii})^2 + (b_{Qi} + \beta_{Qi})^2] \tag{7.91}$$

This statistic has a noncentral χ-square distribution with $2TW$ degrees of freedom and noncentrality parameter

$$\xi = \frac{E_s}{N_0} \qquad (7.92)$$

When the threshold is given by γ, the probability of false alarm is given by

$$P_{fa} = \Pr(V' > \gamma | H_0) = \Pr\{\chi^2_{2TW} > \gamma\} \qquad (7.93)$$

and the probability of detection is given by

$$P_d = \Pr(V' > \gamma | H_1) = \Pr[\chi'^2_{2TW}(\xi) > \gamma] \qquad (7.94)$$

where $\chi'^2_{2TW}(\xi)$ represents a noncentral χ-square variate with noncentrality parameter ξ, given by (7.92).

Urkowitz used an approximation to the noncentral χ-square distribution based on the central χ-square distribution with modified threshold and degrees of freedom. In particular, the modified threshold is given by

$$\gamma' = \frac{\gamma}{\dfrac{2TW + 2\xi}{2TW + \xi}} \qquad (7.95)$$

and the modified degrees of freedom is given by

$$\delta = \frac{(2TW + \xi)^2}{2TW + 2\xi} \qquad (7.96)$$

In that case, the probability of detection is given approximately by

$$P_d = \Pr[\chi'^2_{2TW}(\xi) > \gamma] = \Pr(\chi^2_\delta > \gamma') \qquad (7.97)$$

The ROC curves in Figure 7.11 were computed using this approximation to the noncentral χ-square distribution. (Actually, Figure 7.11(a) was computed using a different method, which is explained in [8].) As these charts indicate, as 2TW gets larger for a given P_{fa}, P_d decreases.
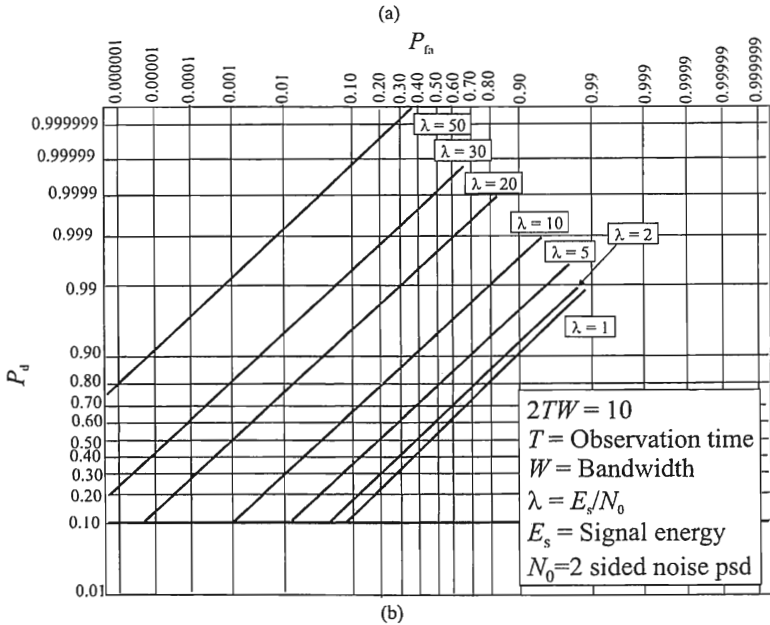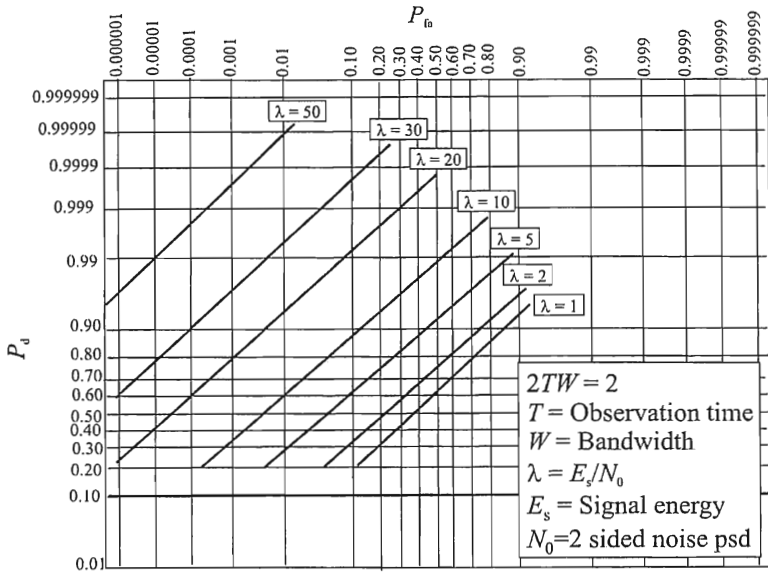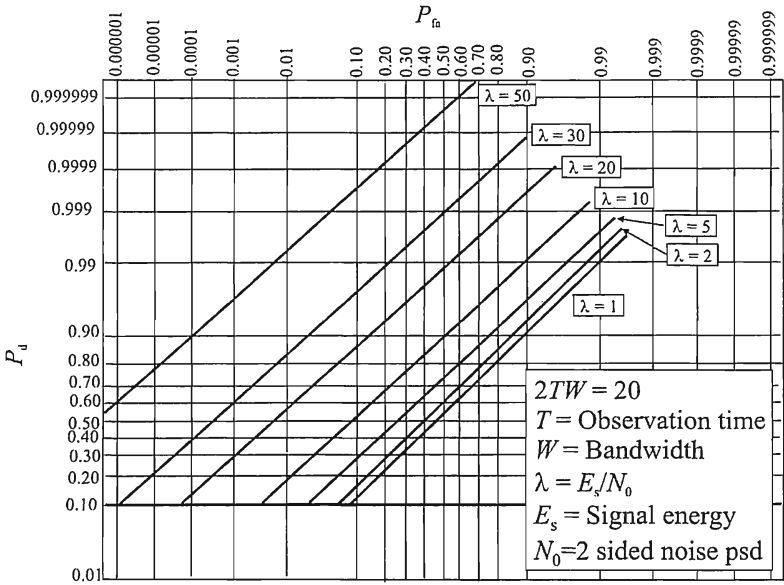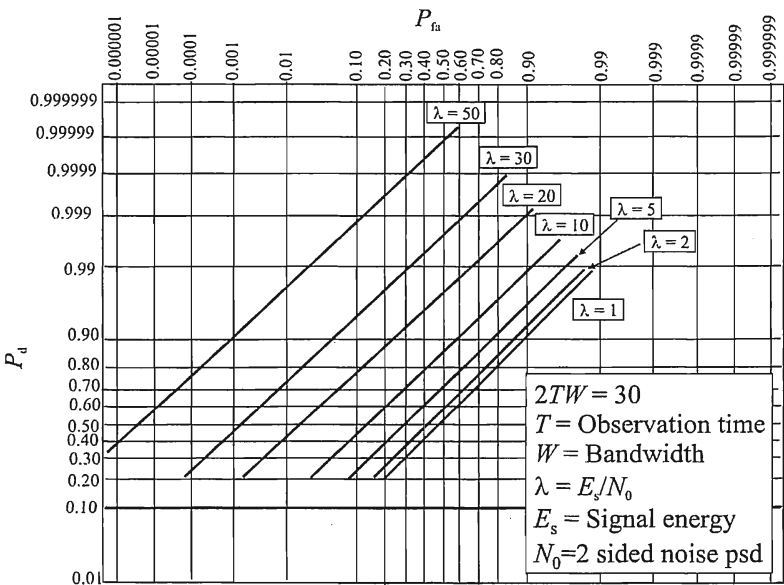
Figure 7.11 ROC curve when (a) $2TW$ = 2; (b) $2TW$ = 10; (c) 2TW = 20; (d) 2TW = 30; amd (e) 2TW = 50. (*From:* [8]. © 1967 IEEE. Reprinted with permission.)

(c)

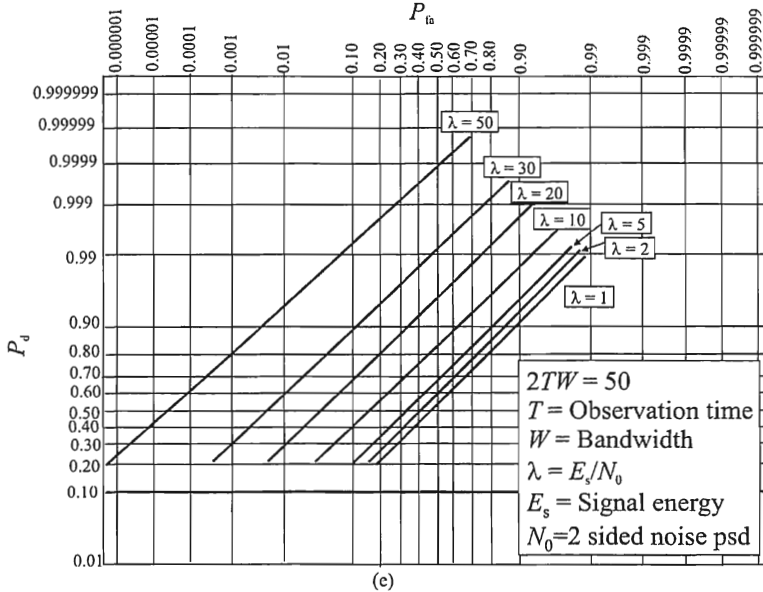
(d)

Figure 7.11 (Continued)

Figure 7.11 (Continued)

Figure 7.12 shows examples of the required SNR as a function of the degrees of freedom. As the number of samples of the signal increases, the SNR required for given levels of P_d and P_{fa} also increases. This is due to the incoherent combining loss that occurs with more samples.

Another curve of interest from [8] is shown in Figure 7.13. This curve shows the increase in SNR required by the energy detection approach as compared to the matched filter, when the decision statistic in the matched filter is based on the envelope of the output of the matched filter. One curve applies to all of the P_d and P_{fa} combinations because the curves in Figure 7.12 are essentially parallel. Recall that a matched filter, having its impulse response matched to the signal to be detected, has knowledge of the signal structure. The energy detector does not have this knowledge. Therefore, Figure 7.13 represents the penalty paid for this ignorance.

7.2.3 Detection of Sinusoidal Carriers with Unknown Parameters
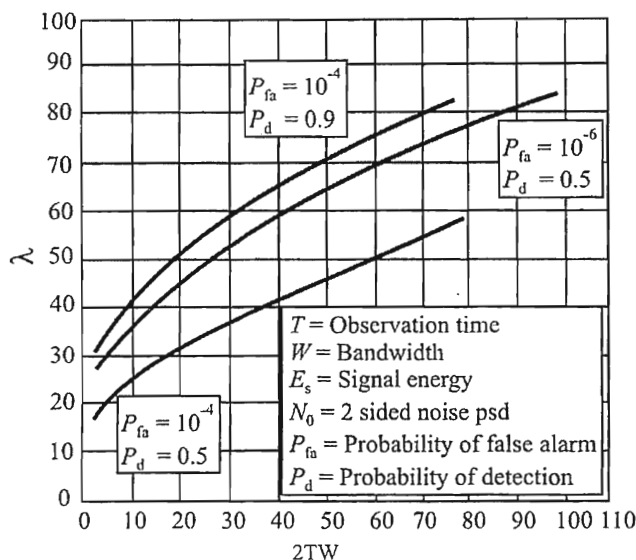
The problem considered is expressed as follows:

Figure 7.12 SNR required for the specified level of performance as a function of the degrees of freedom. (*From:* [8], © 1967 IEEE. Reprinted with permission.)
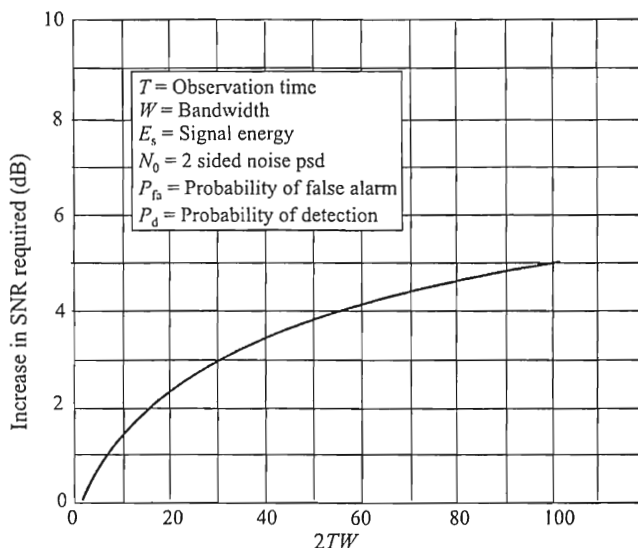

Figure 7.13 Required SNR comparison of energy detector to the matched filter. (*From:* [8], © IEEE 1967. Reprinted with permission.)

$$H_0 : y_k = s_{0k} + n_k, \qquad k = 0,1,\cdots,N-1 \qquad (7.98)$$
$$H_1 : y_k = s_{1k} + n_k, \qquad k = 0,1,\cdots,N-1 \qquad (7.99)$$

where $\mathbf{y} = (y_0, y_1, \ldots, y_{N-1})^T \in \Re^N$ is the vector of observation samples, $\mathbf{n} = (n_0, n_1; \ldots, n_{N-1})^T$ is a vector of noise samples and $\mathbf{s}_0 = (s_{00}, s_{01}, \ldots, s_{0N-1})^T \in \Re^N$ and $\mathbf{s}_1 = (s_{10}, s_{11}, \ldots, s_{1N-1})^T \in \Re^N$ are vectors of samples of two signals.

7.2.3.1 Noncoherent Detection of a Single Modulated Sinusoidal Carrier with Known Amplitude and Unknown Phase

If the signal amplitude is known, or can be estimated, then the hypotheses in this case consist of

$$H_0 : \mathbf{y} = \mathbf{n} \qquad (7.100)$$
$$H_1 : \mathbf{y} = \mathbf{s} + \mathbf{n} \qquad (7.101)$$

where $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. The elements of $\mathbf{s}$ are

$$s_k(\phi_k) = A_k \sin\big[(k-1)2\pi f_0 T_s + \phi_k\big], \qquad k = 1,2,\cdots,N \qquad (7.102)$$

where $\mathbf{A} = (A_1, A_2, \ldots, A_N)^T$ is the vector of the known amplitudes and each ϕ_k is uniformly distributed over $[0, 2\pi]$. The sample time is given by T_s and f_0 is the signal carrier frequency; these two satisfy

$$N f_0 T_s = m \qquad (7.103)$$

for some integer m.

This model is called the *telegraph signal*, where the signal is either present or absent, corresponding to key down and key up events. When the key is down, the sinusoid carrier with known amplitude and random phase is transmitted. To this is added the random thermal noise. For the other case, only noise is present.

The likelihood function for this model is given by [9]

$$\lambda(\mathbf{y}) = e^{-\frac{N\bar{A}^2}{4\sigma^2}} I_0\left(\frac{r}{\sigma^2}\right) \qquad (7.104)$$
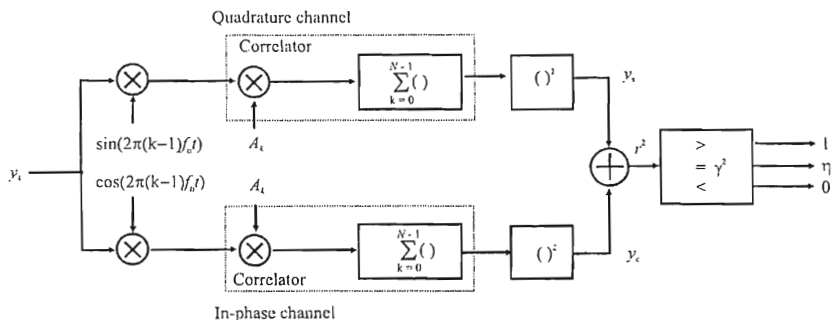
Figure 7.14 Quadrature envelope detector for noncoherent detection of a single sinusoidal carrier with random phase.

where

$$y_c = \sum_{k=1}^{N} y_k A_k \cos\left[2\pi(k-1)f_0 T_s\right] \qquad (7.105)$$

$$y_s = \sum_{k=1}^{N} y_k A_k \sin\left[2\pi(k-1)f_0 T_s\right] \qquad (7.106)$$

$$\overline{A}^2 = \frac{1}{N}\sum_{k=1}^{N} A_k^2 \qquad (7.107)$$

and

$$r = \sqrt{y_c^2 + y_s^2} \qquad (7.108)$$

and $I_0(\ )$ is the modified Bessel function of the first kind and zeroth order.

T_s in these expressions is the sample interval for the telegraph signal. For the EW detection problem of concern here, it is the duration of the sample time as well.

Since $I_0(x)$ is monotonically increasing the decision rule for the optimum detector can be written as

$$\tilde{\delta}_0 = \begin{cases} 0, & r < \gamma \\ \eta, & r = \gamma \\ 1, & r > \gamma \end{cases} \qquad (7.109)$$

with

$$\gamma = \sigma^2 I_0^{-1}\left(\frac{N\bar{A}^2}{\tau e^{4\sigma^2}}\right) \qquad (7.110)$$

where threshold τ depends on the definition of optimality for the test. For example, the threshold for a P_{fa}-level Neyman-Pearson test is given by

$$\tau = \sigma \Phi^{-1}(1 - P_{fa}) + \mu_0 \qquad (7.111)$$

where Φ is the cumulative distribution function for the standard normal $\mathcal{N}(0, 1)$ random variable. The detector for this test is the envelope detector illustrated in Figure 7.14.

7.2.3.2 Noncoherent Detection of Multiple Sinusoidal Signals with Known Amplitude and Unknown Phase

The above can be extended to more that one signal present. Presented here is when there are two signals possible. Extension to more than two signals is straightforward. In this case the two hypotheses are

$$H_0 : \mathbf{y} = \mathbf{s}_0 + \mathbf{w} \qquad (7.112)$$

versus

$$H_1 : \mathbf{y} = \mathbf{s}_1 + \mathbf{w} \qquad (7.113)$$

where

$$s_{jk}(\phi_k) = A_{jk} \sin[2\pi(k-1)f_0 T_s + \phi_k] \qquad (7.114)$$

for $j = 0, 1$. The likelihood ratio is found by assuming a third hypothesis which is the noise-only hypothesis and comparing it to each of (7.112) and (7.113). Thus,

$$\lambda(\mathbf{y}) = \frac{e^{-\frac{N\overline{A_1}}{4\sigma^2}} I_0\left(\frac{r_1}{\sigma^2}\right)}{e^{-\frac{N\overline{A_0}}{4\sigma^2}} I_0\left(\frac{r_0}{\sigma^2}\right)} \tag{7.115}$$

where

$$\overline{A_j^2} = \frac{1}{N}\sum_{k=1}^{N} A_{jk}^2 \tag{7.116}$$

$$r_j = \sqrt{y_{cj}^2 + y_{sj}^2} \tag{7.117}$$

$$y_{cj} = \sum_{k=1}^{N} A_{jk} y_k \cos\left[2\pi(k-1)f_0 T_s\right] \tag{7.118}$$

$$y_{sj} = \sum_{k=1}^{N} A_{jk} y_k \sin\left[2\pi(k-1)f_0 T_s\right] \tag{7.119}$$

Again, T_s is the sample time for detection of the presence of the signals. If $\overline{A_0^2} = \overline{A_1^2}$, with uniform costs and equal priors, then the optimal Bayes detector is given by

$$\delta_B(\mathbf{y}) = \begin{cases} 0, & r_1 < r_0 \\ 0 \text{ or } 1, & r_1 = r_0 \\ 1, & r_1 > r_0 \end{cases} \tag{7.120}$$

If it is further assumed that

$$\sum_{k=1}^{N} A_{jk} A_{lk} = 0, \quad j \neq l \tag{7.121}$$

and that

$$\sum_{k=1}^{N} A_{jk} A_{lk} \sin\left[2\pi(k-1)f_0 T_s + \phi_k\right] = 0 \tag{7.122}$$

for all ϕ_k, then the error probability for this detector is

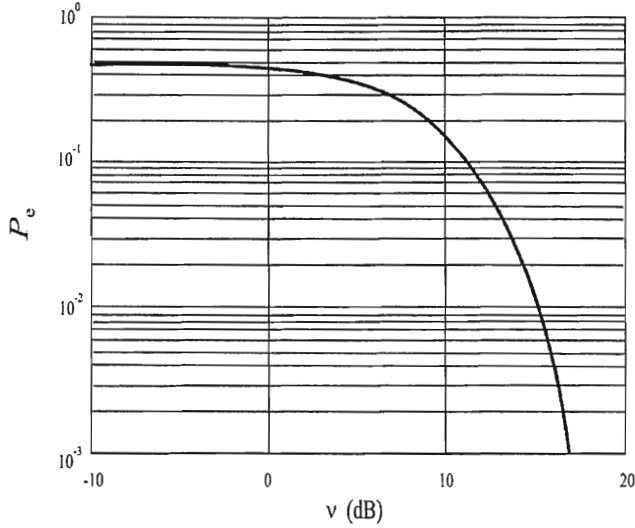$$P_e = \frac{1}{2}e^{-\frac{N\overline{A_0^2}}{8\sigma^2}} \tag{7.123}$$

Figure 7.15 Probability of error for multiple signals optimal Bayes detection.

Recognizing that $\nu = N\overline{A}_0^2 \,/\, \sigma^2$, this function is plotted in Figure 7.15.

7.2.4 Locally Optimum Test for Weak Signal Detection

Hoh [10] reported on the analysis of a *locally optimum detector* (LOD) for weak deterministic signals with unknown amplitude, random phase over $[0, 2\pi)$ (not necessarily uniform), in non-Gaussian noise and interference environments. The hypotheses of the problem are

$$\begin{aligned}
H_0 &: x_k = n_k + i_k, & k &= 0,1,...,N-1 \\
H_1 &: x_k = s_k + n_k + i_k, & k &= 0,1,...,N-1
\end{aligned} \tag{7.124}$$

where n_k are i.i.d. samples from a noise process independent of s_k and i_k, s_k are samples of the signal to be detected independent of n_k and i_k, and i_k represents samples from an interference process such as man-made noise, atmospheric noise, or intentional jamming. The observed process X consists of complex samples

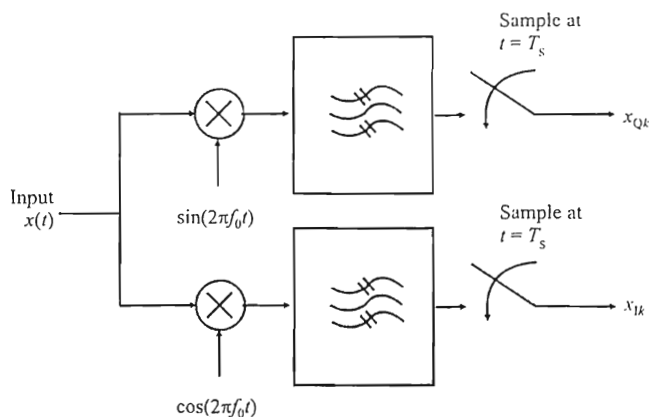$$x_k = x_{Ik} + jx_{Qk} \tag{7.125}$$

Figure 7.16 Quadrature channels.

The real and imaginary components of x_k are obtained by sampling the output of the in-phase and quadrature channels as illustrated in Figure 7.16.

Expression (7.125) can be put into polar form as

$$x_k = a_{x_k} e^{j\phi_{x_k}} \tag{7.126}$$

where

$$a_{x_k} = \sqrt{x_{Ik}^2 + x_{Qk}^2} \tag{7.127}$$

and

$$\phi_{x_k} = \tan^{-1} \frac{x_{Qk}}{x_{Ik}} \tag{7.128}$$

with the obvious associations

$$\begin{aligned} x_{Ik} &= a_{x_k} \cos\phi_{x_k} \\ x_{Qk} &= a_{x_k} \sin\phi_{x_k} \end{aligned} \tag{7.129}$$

Likewise, i_k and n_k can be expressed in rectangular and polar form with the components expressed as $i_{Ik}, i_{Qk}, a_{i_k}, \phi_{i_k}$ and $n_{Ik}, n_{Qk}, a_{n_k}, \phi_{n_k}$, respectively.

The known signal with unknown amplitude is expressed as

$$s_k = A\zeta_k, \qquad k = 1, 2, \ldots, N-1 \qquad (7.130)$$

where A is the unknown amplitude and ζ_n is the signal that has been normalized to have an energy of one:

$$\sum_{k=0}^{N-1} \zeta_k \zeta_k^* = |\zeta_n|^2 = 1 \qquad (7.131)$$

The components of ζ_k are similarly defined in rectangular and polar coordinates as ζ_{Ik}, ζ_{Qk}, $|\zeta_k|$, and ϕ_{ζ_n}.

Let $z_k = n_k + i_k$, and $Z = \{z_k\}$. Under H_0, the pdf for X is given by

$$p_X(X|H_0) = p_Z(X)$$
$$= \prod_{k=0}^{N-1} p_{ZI,ZQ}(x_{Ik}, x_{Qk}) \qquad (7.132)$$

where $p_{ZI,ZQ}(\ )$ is the joint pdf of the in-phase and quadrature phase total interference components. This expression is true because each of the samples x_k is i.i.d. because the underlying noise process n_k is i.i.d. Under H_1, the pdf is

$$p_X(X|H_1) = p_Z(X - S)$$
$$= \prod_{k=0}^{N-1} p_{ZI,ZQ}(x_{Ik} - s_{Ik}, x_{Qk} - s_{Qk}) \qquad (7.133)$$

The test statistic used by the optimum receiver is given by the likelihood ratio

$$\frac{p_X(X|H_1)}{p_X(X|H_0)} = \frac{\displaystyle\prod_{k=0}^{N-1} p_{ZI,ZQ}(x_{Ik} - s_{Ik}, x_{Qk} - s_{Qk})}{\displaystyle\prod_{k=0}^{N-1} p_{ZI,ZQ}(x_{Ik}, x_{Qk})} \qquad (7.134)$$

or its logarithmic equivalent, yielding

$$t(X:A) = \sum_{k=0}^{N-1} \ln \frac{p_{ZI,ZQ}(x_{Ik} - s_{Ik}, x_{Qk} - s_{Qk})}{p_{ZI,ZQ}(x_{Ik}, x_{Qk})} \tag{7.135}$$

and H_0 is chosen if $t(X: A) < \gamma$, and H_1 is chosen in $t(X: A) > \gamma$.

With so little known about the pdf, a UMP cannot be determined. An alternative is to replace A with its MLE and the resulting detector is the well-known GLRT detector. However, this technique performs poorly for small A and small sample size. In addition, GLRT detectors are frequently complex to implement. Another alternative determines locally optimum (LO) detectors for which A can be unknown. In addition to simple implementation, LO detectors are optimum for the case of vanishingly small SNR. For the weak signal LOD being considered here, the appropriate structure maximizes the slope of the likelihood ratio at $A = 0$ while keeping a fixed P_{fa} [see (4.19)]. According to the generalized NP lemma and under mild regularity conditions [11], the LO test is

$$\frac{\left.\dfrac{\partial p(X|H_1)}{\partial A}\right|_{A=0}}{p(X|H_0)} \quad \begin{array}{l} > \gamma \Rightarrow H_1 \\ < \gamma \Rightarrow H_0 \end{array} \tag{7.136}$$

It can be shown that this is the same as the threshold test with the test statistic changed to [12]

$$t'(X:A) = \left.\frac{\partial t(X:A)}{\partial A}\right|_{A=0} \tag{7.137}$$

The Taylor series expansion of $f(x)$ around a point a is given by

$$f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!}f''(a) + \cdots \frac{(x-a)^n}{n!}f''(a) + \cdots \tag{7.138}$$

where the notation $f''(a)$ refers to the nth derivative of $f(x)$ evaluated at point a. In two dimensions this becomes

$$f(a+h,b+k) = f(a,b) + \left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y} \right) f(x,y)\big|_{\substack{x=a\\y=b}} + \cdots$$
$$+ \frac{1}{n!}\left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y} \right)^{n} f(x,y)\big|_{\substack{x=a\\y=b}} + \cdots \qquad (7.139)$$

where the bar and subscripts mean that after differentiation, x is replaced with a and y is replaced with b. Also, in this notation,

$$\left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y} \right) f(x,y) = \left(h\frac{\partial f(x,y)}{\partial x} + k\frac{\partial f(x,y)}{\partial y} \right),$$
$$\left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y} \right)^{2} f(x,y) = h^{2}\frac{\partial^{2} f(x,y)}{\partial x} + 2hk\frac{\partial^{2} f(x,y)}{\partial x \partial y} + k^{2}\frac{\partial f(x,y)}{\partial y}$$

$$\cdots$$

Under the assumption that A is small, $A \approx 0$, the logarithm of the pdf can be approximated by expanding it in a Taylor series yielding (ignoring other than the first two terms)

$$\ln p_{ZI,ZQ}(x_{Ik} - A\zeta_{Ik}, x_{Qk} - A\zeta_{Qk}) = \ln[p_{ZI,ZQ}(x_{Ik}, x_{Qk})]$$
$$-A\zeta_{Ik}\frac{\partial \ln[p_{ZI,ZQ}(x_{Ik}, x_{Qk})]}{\partial x_{Ik}} \qquad (7.140)$$
$$-A\zeta_{Qk}\frac{\partial \ln[p_{ZI,ZQ}(x_{Ik}, x_{Qk})]}{\partial x_{Qk}}$$

Using (7.140), (7.137) becomes

$$t'(X:A) = \frac{\partial}{\partial A}\sum_{k=0}^{N-1} \ln \frac{p_{ZI,ZQ}(x_{Ik} - A\zeta_{Ik}, x_{Qk} - A\zeta_{Qk})}{p_{ZI,ZQ}(x_{Ik}, x_{Qk})}\bigg|_{A=0}$$
$$= \sum_{k=0}^{N-1}\frac{\partial}{\partial A} \ln \frac{p_{ZI,ZQ}(x_{Ik} - A\zeta_{Ik}, x_{Qk} - A\zeta_{Qk})}{p_{ZI,ZQ}(x_{Ik}, x_{Qk})}\bigg|_{A=0}$$
$$= \sum_{k=0}^{N-1}\frac{\partial}{\partial A}\Big[\ln p_{ZI,ZQ}(x_{Ik} - A\zeta_{Ik}, x_{Qk} - A\zeta_{Qk}) - \ln p_{ZI,ZQ}(x_{Ik}, x_{Qk})\Big]\bigg|_{A=0}$$

$$
\begin{aligned}
&\approx \sum_{k=0}^{N-1} \frac{\partial}{\partial A} \left[\begin{array}{l} \ln[p_{ZI,ZQ}(x_{Ik},x_{Qk})] - A\zeta_{Ik} \dfrac{\partial \ln[p_{ZI,ZQ}(x_{Ik},x_{Qk})]}{\partial x_{Ik}} \\[3mm] \quad - A\zeta_{Qk} \dfrac{\partial \ln[p_{ZI,ZQ}(x_{Ik},x_{Qk})]}{\partial x_{Qk}} - \ln p_{ZI,ZQ}(x_{Ik},x_{Qk}) \end{array} \right]_{A=0} \\[4mm]
&= \sum_{k=0}^{N-1} \frac{\partial}{\partial A} \left[-A\zeta_{Ik} \frac{\partial \ln[p_{ZI,ZQ}(x_{Ik},x_{Qk})]}{\partial x_{Ik}} - A\zeta_{Qk} \frac{\partial \ln[p_{ZI,ZQ}(x_{Ik},x_{Qk})]}{\partial x_{Qk}} \right]_{A=0} \\[4mm]
&= \sum_{k=0}^{N-1} \left[-\zeta_{Ik} \frac{\partial \ln[p_{ZI,ZQ}(x_{Ik},x_{Qk})]}{\partial x_{Ik}} - \zeta_{Qk} \frac{\partial \ln[p_{ZI,ZQ}(x_{Ik},x_{Qk})]}{\partial x_{Qk}} \right]_{A=0}
\end{aligned}
$$

so

$$
t'(X:A) = \sum_{k=0}^{N-1} T_k(X) \tag{7.141}
$$

where

$$
T_k(X) = - \left\{ \zeta_{Ik} \frac{\partial \ln[p_{ZI,ZQ}(x_{Ik},x_{Qk})]}{\partial x_{Ik}} + \zeta_{Qk} \frac{\partial \ln[p_{ZI,ZQ}(x_{Ik},x_{Qk})]}{\partial x_{Qk}} \right\} \tag{7.142}
$$

The pdf expressed in terms of a_{x_k} and ϕ_{x_k} can be determined from the pdf expressed in terms of x_{Ik} and x_{Qk} using the *pdf transformation theorem* [13]. This theorem states that

$$
p_Z(a_{x_k}, \phi_{x_k}) = \frac{1}{|J(x_{Ik}, x_{Qk})|} p_Z(x_{Ik}, x_{Qk}) \tag{7.143}
$$

where $|J(x_{Ik}, x_{Qk})|$ is the absolute value of the Jacobian of the transformation given by

$$
J(x,y) = \begin{vmatrix} \dfrac{\partial g(x,y)}{\partial x} & \dfrac{\partial g(x,y)}{\partial y} \\[3mm] \dfrac{\partial h(x,y)}{\partial x} & \dfrac{\partial h(x,y)}{\partial y} \end{vmatrix} \tag{7.144}
$$

where $g(\ )$ and $h(\ )$ are the equations of the transformation from one set of variables to another. In this case,

$$g(x, y) = g(x_{Ik}, x_{Qk}) = \sqrt{x_{Ik}^2 + x_{Qk}^2} \tag{7.145}$$

and

$$h(x, y) = h(x_{Ik}, x_{Qk}) = \tan^{-1}\left(\frac{x_{Qk}}{x_{Ik}}\right) \tag{7.146}$$

Now, from (7.127)

$$\frac{\partial g(x, y)}{\partial x} = \frac{\partial a_{x_k}}{\partial x_{Ik}} = \frac{x_{Ik}}{\sqrt{x_{Ik}^2 + x_{Qk}^2}}$$

and substituting (7.129) into this

$$\frac{\partial a_{x_k}}{\partial x_{Ik}} = \frac{a_{x_k} \cos\phi_{x_k}}{\sqrt{a_{x_k}^2 \cos^2\phi_{x_k} + a_{x_n}^2 \sin\phi_{x_k}}}$$

so

$$\frac{\partial a_{x_k}}{\partial x_{Ik}} = \cos\phi_{x_k} \tag{7.147}$$

Likewise,

$$\frac{\partial g(x, y)}{\partial y} = \frac{\partial a_{x_k}}{\partial x_{Qk}} = \frac{x_{Qk}}{\sqrt{x_{Ik}^2 + x_{Qk}^2}}$$

yielding

$$\frac{\partial a_{x_k}}{\partial x_{Qk}} = \sin\phi_{x_k} \tag{7.148}$$

As for the second function in the Jacobian,

$$
\frac{\partial h(x,y)}{\partial x} = \frac{\partial \phi_{x_k}}{\partial x_{1k}} = \frac{\partial \tan^{-1}\left(\dfrac{x_{Qk}}{x_{1k}}\right)}{\partial x_{1k}}
$$

$$
= \frac{1}{1+\left(\dfrac{x_{Qk}}{x_{1k}}\right)^2} \frac{\partial \dfrac{x_{Qk}}{x_{1k}}}{\partial x_{1k}}
$$

$$
= \frac{-1}{1+\left(\dfrac{x_{Qk}}{x_{1k}}\right)^2} \frac{x_{Qk}}{x_{1k}^2}
$$

$$
= \frac{-\sin \phi_{x_k}}{\left(1+\dfrac{\sin^2 \phi_{x_k}}{\cos^2 \phi_{x_k}}\right) a_{x_k} \cos^2 \phi_{x_k}}
$$

$$
= \frac{-\sin \phi_{x_k}}{a_{x_k}} \tag{7.149}
$$

Likewise,

$$
\frac{\partial h(x,y)}{\partial y} = \frac{\partial \phi_{x_k}}{\partial x_{Qk}} = \frac{\partial \tan^{-1}\left(\dfrac{x_{Qk}}{x_{1k}}\right)}{\partial x_{Qk}}
$$

$$
= \frac{\cos \phi_{x_k}}{a_{x_k}} \tag{7.150}
$$

Therefore,

$$
J(x,y) = \begin{vmatrix} \sin \phi_{x_k} & \sin \phi_{x_k} \\ \dfrac{-\sin \phi_{x_k}}{a_{x_k}} & \dfrac{\cos \phi_{x_k}}{a_{x_k}} \end{vmatrix}
$$

$$= \frac{\cos^2 \phi_{x_k}}{a_{x_k}} + \frac{\sin^2 \phi_{x_k}}{a_{x_k}}$$

$$= \frac{1}{a_{x_k}} \tag{7.151}$$

and, from (7.143),

$$p_Z(a_{x_k}, \phi_{x_k}) = a_{x_k} p_Z(x_{1k}, x_{Qk}) \tag{7.152}$$

$$= a_{x_k} f(a_{x_k}, \phi_{x_k}) \tag{7.153}$$

where

$$f(a_{x_k}, \phi_{x_k}) = p_{Z1,ZQ}(x_{1k}, x_{Qk}) \tag{7.154}$$

Using the total differential for the logarithm of the pdf and (7.147) through (7.150) the terms in (7.142) can be determined. The total differential is

$$\partial \ln p_Z(X_{1k}, X_{Qk}) = \frac{\partial \ln p_Z(X_{1k}, X_{Qk})}{\partial a_{x_k}} \partial a_{x_k} + \frac{\partial \ln p_Z(X_{1k}, X_{Qk})}{\partial \phi_{x_k}} \partial \phi_{x_n}$$

but, from (7.147),

$$\partial a_{x_k} = \cos \phi_{x_k} \partial x_{1k}$$

$$= \frac{a_{x_k} \cos \phi_{x_k}}{a_{x_k}}$$

$$= \frac{x_{1k}}{a_{x_k}} \partial x_{1k} \tag{7.155}$$

Likewise, from (7.149),

$$\partial \phi_{x_k} = -\frac{\sin \phi_{x_k}}{a_{x_k}} \partial \phi_{1k}$$

$$= -\frac{a_{x_k} \sin \phi_{x_k}}{a_{x_k}^2} \partial x_{1k}$$

$$= -\frac{x_{Qk}}{a_{x_k}^2} \partial x_{1k} \qquad (7.156)$$

Therefore,

$$\partial \ln p_Z(X_{1k}, X_{Qk}) = \frac{\partial \ln p_Z(X_{1k}, X_{Qk})}{\partial a_{x_k}} \frac{x_{1k}}{a_{x_k}} \partial x_{1k}$$
$$- \frac{\partial \ln p_Z(X_{1k}, X_{Qk})}{\partial \phi_{x_k}} \frac{a_{x_k} \sin \phi_{x_k}}{a_{x_k}^2} \partial x_{1k}$$

so

$$\frac{\partial \ln p_Z(X_{1k}, X_{Qk})}{\partial x_{1k}} = \frac{x_{1k}}{a_{x_n}} \frac{\partial \ln p_Z(X_{1k}, X_{Qk})}{\partial a_{x_k}}$$
$$- \frac{a_{x_k} \sin \phi_{x_k}}{a_{x_k}^2} \frac{\partial \ln p_Z(X_{1k}, X_{Qk})}{\partial \phi_{x_k}} \qquad (7.157)$$

In a similar manner it can be established that

$$\frac{\partial \ln p_{ZI,ZQ}(x_{1k}, x_{Qk})}{\partial x_{Qk}} = \frac{x_{Qk}}{a_{x_n}} \frac{\partial \ln f(a_{x_k}, \phi_{x_k})}{\partial a_{x_k}} + \frac{x_{1k}}{a_{x_k}^2} \frac{\partial \ln f(a_{x_k}, \phi_{x_k})}{\partial \phi_{x_k}} \qquad (7.158)$$

Putting (7.157) and (7.158) into (7.141) yields

$$t'(X) = \sum_{k=0}^{N-1} [\mathrm{Re}(x_k \zeta_k^*) g_a(a_{x_k}, \phi_{x_k}) + \mathrm{Im}(x_k \zeta_k^*) g_\phi(a_{x_k}, \phi_{x_k})]$$
$$= \sum_{k=0}^{N-1} \mathrm{Re}[(x_k g_k) \zeta_k^*] \qquad (7.159)$$

where

$$g_k = g_a(a_{x_k}, \phi_{x_k}) - jg_\phi(a_{x_k}, \phi_{x_k}) \tag{7.160}$$

with

$$g_a(a_{x_k}, \phi_{x_k}) = -\frac{1}{a_{x_k}} \frac{\partial \ln f(a_{x_k}, \phi_{x_k})}{\partial a_{x_k}} \tag{7.161}$$

$$g_\phi(a_{x_k}, \phi_{x_k}) = -\frac{1}{a_{x_k}^2} \frac{\partial \ln f(a_{x_k}, \phi_{x_k})}{\partial \phi_{x_k}} \tag{7.162}$$

The LOD structure corresponding to this analysis is shown in Figure 7.17.

A performance measure can be defined to determine the gain in detection performance achieved by including the nonlinearity in the processing path. This measure is the *processing gain* (PG), defined by the ratio of the signal-to-interference ratio (*SZR*) at the output of the detector with the nonlinearity present (*SZR*$_{NL}$) to that without it, when the receiver is linear (*SZR*$_L$). Thus,

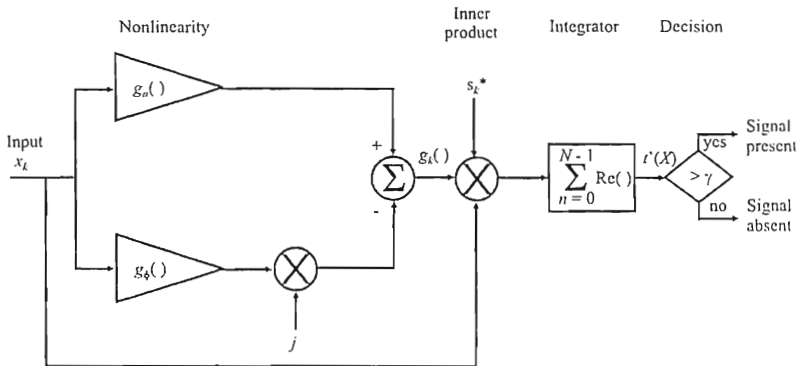$$PG = \frac{SZR_{NL}}{SZR_L} \tag{7.163}$$

Figure 7.17 LOD structure for known weak signal in arbitrary noise. (*From:* [10], © 1990 IEEE. Reprinted with permission.)

As determined by a similar analysis of Ingram and Houlse [14], with the nonlinearity present,

$$SZR_{NL} = \frac{\left[\sum_{k=0}^{N-1} \mathcal{E}\{T_k(X)\}\right]^2}{\sum_{k=0}^{N-1} \sigma_{T_k}^2} \tag{7.164}$$

where

$$T_k(X) = \mathrm{Re}[(x_k g_k)\zeta_k^*] \tag{7.165}$$

Without the nonlinearity

$$SZR_L = \frac{\sum_{k=0}^{N-1} s_k s_k^*}{\sigma^2} \tag{7.166}$$

where σ^2 is the variance of x, which is also the power in x.

Using (7.140) and (7.142),

$$\mathcal{E}\{T_k(X)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{ZI,ZQ}(x_{Ik} - A\zeta_{Ik}, x_{Qk} - A\zeta_{Qk}) dx_{Qk} dx_{Ik}$$

$$= A \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ \zeta_{Ik} \frac{\partial \ln[p_{ZI,ZQ}(x_{Ik}, x_{Qk})]}{\partial x_{Ik}} + \zeta_{Qk} \frac{\partial \ln[p_{ZI,ZQ}(x_{Ik}, x_{Qk})]}{\partial x_{Qk}} \right\}^2 dx_{Qk} dx_{Ik} \tag{7.167}$$

while the variance of $T_k(X)$ is

$$\sigma_{T_k}^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [T_k(X) - \mathcal{E}\{T_k(X)\}]^2 p_{ZI,ZQ}(x_{Ik} - A\zeta_{Ik}, x_{Qk} - A\zeta_{Qk}) dx_{Qk} dx_{Ik} \tag{7.168}$$

Using a Taylor series expansion of the argument in the integrals in (7.168), and ignoring terms higher than the first two, this expression reduces to

$$\sigma_{T_k}^2 = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \left\{ s_{In}\frac{\partial \ln[p_{ZI,ZQ}(x_{Ik},x_{Qk})]}{\partial x_{Ik}} + s_{Qk}\frac{\partial \ln[p_{ZI,ZQ}(x_{Ik},x_{Qk})]}{\partial x_{Qk}} \right\}^2 p_{ZI,ZQ}(x_{Ik},x_{Qk})dx_{Qk}dx_{Ik} \quad (7.169)$$

Therefore,

$$PG = \frac{\displaystyle\sum_{k=0}^{N-1}\mathcal{E}^2\{T_k(X)\}\sigma^2}{\displaystyle\sum_{k=0}^{N-1}\sigma_{T_k}^2 A^2 \sum_{k=0}^{N-1}\zeta_k\zeta_k^*}$$

$$= \sigma^2\sum_{k=0}^{N-1}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \left\{ s_{Ik}\frac{\partial \ln[p_{ZI,ZQ}(x_{Ik},x_{Qk})]}{\partial x_{Ik}} + s_{Qk}\frac{\partial \ln[p_{ZI,ZQ}(x_{Ik},x_{Qk})]}{\partial x_{Qk}} \right\}^2 p_{ZI,ZQ}(x_{Ik},x_{Qk})dx_{Qk}dx_{Ik}$$

$$= \sigma^2\sum_{k=0}^{N-1}|\zeta_k|^2\int\limits_{0}^{\infty}\int\limits_{0}^{2\pi} \left\{ -\frac{\partial \ln f(a_{x_k},\phi_{x_k})}{\partial a_{x_k}}\cos(\phi_{x_k}-\phi_{\zeta_k}) + \frac{1}{a_{x_k}}\frac{\partial \ln f(a_{x_k},\phi_{x_k})}{\partial \phi_{x_k}}\sin(\phi_{x_k}-\phi_{\zeta_k}) \right\}^2 \times f(a_{x_k},\phi_{x_k})a_{x_k}d\phi_k da_k \quad (7.170)$$

Averaging ϕ_{s_k} over $[0, 2\pi)$ in (7.170) yields

$$PG = \frac{\sigma^2}{2}\sum_{k=0}^{N-1}|\zeta_k|^2\int\limits_{0}^{\infty}\int\limits_{0}^{2\pi} \left\{ \left[-\frac{\partial \ln f(a_{x_k},\phi_{x_k})}{\partial a_{x_k}}\right]^2 + \left[\frac{1}{a_{x_k}}\frac{\partial \ln f(a_{x_k},\phi_{x_k})}{\partial \phi_{x_k}}\right]^2 \right\} f(a_{x_k},\phi_{x_k})a_{x_k}d\phi_k da_k \quad (7.171)$$

Inside the integrals the k subscript can be dropped because the variables are not subject to the summation. Using (7.131) and (7.153), the performance gain can be written

$$PG = \frac{\sigma^2}{2} \int\limits_0^\infty \int\limits_0^{2\pi} \left\{ \left[\frac{\partial \ln \frac{p_z(a_x, \phi_x)}{a_x}}{\partial a_x} \right]^2 + \left| \frac{1}{a_x} \frac{\partial \ln \frac{p_z(a_x, \phi_x)}{a_x}}{\partial \phi_x} \right|^2 \right\} p_z(a_x, \phi_x) d\phi_x \, da_x \qquad (7.172)$$

The processing gain due to the amplitude processing is given by the first part of the integrand in (7.172)

$$PG_a = \frac{\sigma^2}{2} \int\limits_0^\infty \int\limits_0^{2\pi} \left[\frac{\partial \ln \frac{p_z(a_x, \phi_x)}{a_x}}{\partial a_x} \right]^2 p_z(a_x, \phi_x) d\phi_x \, da_x \qquad (7.173)$$

while that due to the phase processing is given by the second part of the integrand

$$PG_\varphi = \frac{\sigma^2}{2} \int\limits_0^\infty \int\limits_0^{2\pi} \left[\frac{1}{a_x} \frac{\partial \ln \frac{p_z(a_x, \phi_x)}{a_x}}{\partial \phi_x} \right]^2 p_z(a_x, \phi_x) d\phi_x \, da_x \qquad (7.174)$$

and, from (7.172) the total processing gain is given by their sum

$$PG = PG_a + PG_\varphi \qquad (7.175)$$

As an example of a LOD detector for a weak signal, assume that the noise is white Gaussian and that the interference is a *continuous wave* (CW) signal. Then

$$p_{ZI,ZQ}(x_{Ik}, x_{Qk}) = p_{nI,nQ}(x_{Ik} - i_{Ik}, x_{Qk} - i_{Qk})$$
$$= p_{nI}(x_{Ik} - i_{Ik}) p_{nQ}(x_{Ik}, x_{Qk}) \qquad (7.176)$$

since the two quadrature components of white noise are independent. Here,

$$p_{nI}(x_{Ik} - i_{Ik}) = \frac{1}{\sigma_n \sqrt{2\pi}} \exp\left[\frac{-(x_{Ik} - a_{i_k} \cos\phi_{i_k})^2}{2\sigma_n^2}\right] \tag{7.177}$$

and

$$p_{nQ}(x_{Qk} - i_{Qk}) = \frac{1}{\sigma_n \sqrt{2\pi}} \exp\left[\frac{-(x_{Qk} - a_{i_k} \sin\phi_{i_k})^2}{2\sigma_n^2}\right] \tag{7.178}$$

where σ_n^2 is the variance (power) of the noise. Combining (7.153) and (7.176) yields

$$p_Z(a_X, \phi_X) = \frac{a_X}{2\pi\sigma_n^2} \exp\left\{-\frac{[a_X^2 + a_i^2 - 2a_X a_i \cos(\phi_X - \phi_i)]}{2\sigma_n^2}\right\} \tag{7.179}$$

Substituting (7.179) into (7.173) yields the processing gain due to the amplitude processing as

$$PG_a = C\int_0^\infty \int_0^{2\pi} \left[\begin{array}{c} a_X[a_X^2 - 2a_X a_i \cos(\phi_X - \phi_i) + a_i^2 \cos^2(\phi_X - \phi_i)] \\ \times \exp\left[\frac{a_X a_i}{\sigma_n^2} \cos(\phi_X - \phi_i)\right] \end{array}\right] d\phi_X da_X \tag{7.180}$$

$$= 2\pi C \int_0^\infty a_X[a_X I_0(u) - 2a_X a_i I_0'(u) + a_i^2 I_0''(u)] \exp\left[-\frac{a_X^2}{2\sigma_n^2}\right] da_X \tag{7.181}$$

where

$$C = \frac{\sigma^2}{4\pi\sigma_n^6} \exp\left(-\frac{a_i^2}{2\sigma_n^2}\right) \tag{7.182}$$

$$u = \frac{a_X a_i}{\sigma_n^2} \tag{7.183}$$

$$I_0(u) = \text{modified Bessel function of zero order} \tag{7.184}$$

$$I_0'(u) = \frac{dI_0(u)}{du} \tag{7.185}$$

$$I_0''(u) = \frac{d^2 I_0(u)}{du^2} \tag{7.186}$$

Expression (7.180) reduces to

$$PG_a = \frac{\sigma^2}{2\sigma_n^2}\left[1 + \exp\left(-\frac{a_i^2}{2\sigma_n^2}\right)\right] \tag{7.187}$$

The interference power, denoted by P_i, is given by

$$P_i = \frac{a_i^2}{2} \tag{7.188}$$

and the noise power denoted by P_n is given by σ_n^2. The total nonsignal power is given by

$$\sigma^2 = P_i + P_n \tag{7.189}$$

The processing gain due to amplitude processing can be expressed in terms of these powers as

$$PG_a = \frac{1}{2}\left(1 + \frac{P_i}{P_n}\right)\left[1 + \exp\left(-\frac{P_i}{P_n}\right)\right] \tag{7.190}$$

The processing gain due to phase is derived as

$$PG_\varphi = a_i^2 C \int_0^\infty \int_0^{2\pi}\left\{a_X \exp\left(-\frac{a_X^2}{2\sigma_n^2}\right)\sin^2(\phi_X - \phi_i)\exp\left[\left(\frac{a_X a_i}{\sigma_n^2}\right)\cos(\phi_X - \phi_i)\right]\right\}d\phi_X da_X$$

$$= 2\pi a_i^2 C \int_0^\infty a_X[I_0(u) - I_0''(u)]\exp\left(-\frac{a_X^2}{2\sigma_n^2}\right)da_X$$

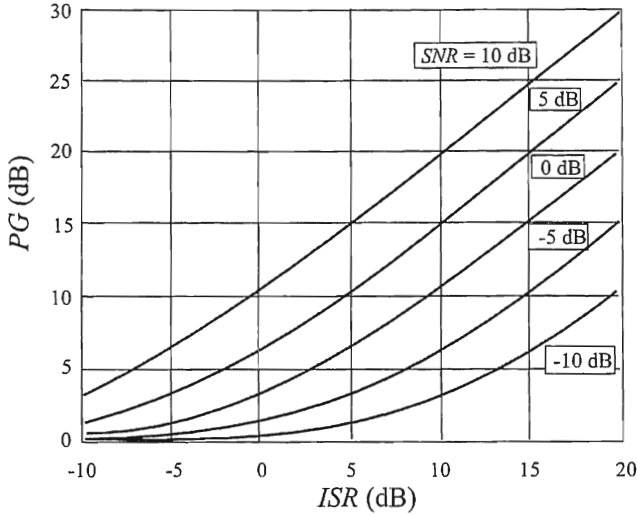$$= \frac{\sigma^2}{2\sigma_n^2}\left[1 - \exp\left(-\frac{a_i^2}{2\sigma_n^2}\right)\right]$$

Figure 7.18 Processing gain for Gaussian noise and a CW interferer.

$$= \frac{1}{2}\left(1 + \frac{P_i}{P_n}\right)\left[1 - \exp\left(-\frac{P_i}{\sigma_n}\right)\right] \tag{7.191}$$

where C, u, and the other factors are the same as for the amplitude processing.
    The total processing gain, then, is given by

$$
\begin{aligned}
PG &= PG_a + PG_\varphi \\
&= \frac{1}{2}\left(1 + \frac{P_i}{P_n}\right)\left[1 - \exp\left(-\frac{P_i}{\sigma_n}\right)\right] + \frac{1}{2}\left(1 + \frac{P_i}{P_n}\right)\left[1 + \exp\left(-\frac{P_i}{P_n}\right)\right] \\
&= \frac{1}{2}\left(1 + \frac{P_i}{P_n}\right)\left[1 - \exp\left(-\frac{P_i}{\sigma_n}\right) + 1 + \exp\left(-\frac{P_i}{P_n}\right)\right] \\
&= \left(1 + \frac{P_i}{P_n}\right) \tag{7.192}
\end{aligned}
$$

Let P_s denote the power in the signal. Then

$$PG = 1 + \frac{P_i}{P_n}$$

$$= \frac{P_s \Big/ P_n}{P_s \Big/ (P_i + P_n)} \tag{7.193}$$

The denominator in (7.193) is the output SZR of a linear detector with total interference given by $i + n$ of power $P_i + P_n$. The numerator is the SZR of the output of a nonlinear detector that can totally suppress the interference i. This function is plotted in Figure 7.18 for some representative values of SNR.

For a given SNR, the larger the level of the CW interference, the larger the processing gain.

7.2.5 Bayes Linear Model

In the Bayes linear model for detecting deterministic signals with unknown parameters, under H_1 it is assumed that [15]

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{n} \tag{7.194}$$

where $\mathbf{x}$ is the vector of observations, $\mathbf{H}$ is a known $N \times p$ $(N > p)$ observation matrix of rank p, $\boldsymbol{\theta}$ is a $p \times 1$ vector of model parameters (some of which are unknown), and $\mathbf{n}$ is an $N \times 1$ random noise vector $\mathbf{n} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$, and it is assumed that the noise variance σ^2 is known. The linear model asserts that either $\mathbf{A}\boldsymbol{\theta} = \mathbf{b}$, where A is a known $r \times p$ $(r \leq p)$ matrix of rank r and $\mathbf{b}$ is a known $r \times 1$ vector, or not. Thus the hypotheses are

$$\begin{aligned} H_0 &: \mathbf{A}\boldsymbol{\theta} = \mathbf{b} \\ H_1 &: \mathbf{A}\boldsymbol{\theta} \neq \mathbf{b} \end{aligned} \tag{7.195}$$

The GLRT for this model is given by the following property of the model.

Property: GLRT for Bayes Linear Model

With the GLRT for the hypothesis problem given by (7.195), then decide H_1 if

$$\begin{aligned} t(\mathbf{x}) &= 2\ln \lambda_G(\mathbf{x}) \\ &= \frac{1}{\sigma^2}(\mathbf{A}\hat{\boldsymbol{\theta}}_1 - \mathbf{b})^T [\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}(\mathbf{A}\hat{\boldsymbol{\theta}}_1 - \mathbf{b}) > \gamma' \end{aligned} \tag{7.196}$$

where

$$\hat{\boldsymbol{\theta}}_1 = (\mathbf{H}^\mathrm{T}\mathbf{H})^{-1}\mathbf{H}^\mathrm{T}\mathbf{x} \tag{7.197}$$

is the MLE of θ under H_1. The detection performance is given by

$$P_{\mathrm{fa}} = Q_{\chi_r^2}(\gamma') \tag{7.198}$$

$$P_{\mathrm{d}} = Q_{\chi_r'^2}(\gamma') \tag{7.199}$$

where the noncentrality parameter is

$$\xi = \frac{1}{\sigma^2}(\mathbf{A}\boldsymbol{\theta}_1 - \mathbf{b})^\mathrm{T}[\mathbf{A}(\mathbf{H}^\mathrm{T}\mathbf{H})^{-1}\mathbf{A}^\mathrm{T}]^{-1}(\mathbf{A}\boldsymbol{\theta}_1 - \mathbf{b}) \tag{7.200}$$

∎

7.2.5.1 Detection of a Deterministic Signal with Unknown Amplitude in Unknown Interference

This problem occurs when there is more than one signal in the pass band of the receiver filter. For processing the signal digitally, the interference is assumed to be a tone at one of the DFT points. The interference could be intentional as with a jammer, or it could be unintentional, as with cochannel interference. Cochannel interference frequently occurs in dense RF environments because all sides share the RF spectrum. If the receiver can hear two or more signals, then cochannel interference will occur. This interference could be from friendly or adversarial transmitters.

We denote the interfering signal with [16]

$$i_k = B\cos(2\pi f_i k + \phi) + n_k, \qquad k = 0, 1, \ldots, N-1 \tag{7.201}$$

The hypotheses are

$$H_0 : x_k = B\cos(2\pi f_i k + \phi) + n_k, \qquad k = 0, 1, \ldots, N-1 \tag{7.202}$$

$$H_1 : x_k = As_k + B\cos(2\pi f_i k + \phi) + n_k, \qquad k = 0, 1, \ldots, N-1 \tag{7.203}$$

where $n_k \sim \mathcal{N}(0, \sigma^2)$. It is assumed that B and ϕ are unknown but the frequency is known. Then, expanding the cos ( ) term in (7.201) using standard trigonometric methods,

$$B\cos(2\pi f_i k + \phi) + n_k = \alpha_1 \cos(2\pi f_i k) + \alpha_2 \sin(2\pi f_i k) \tag{7.204}$$

then

$$\mathbf{x} = \underbrace{\begin{bmatrix} s_0 & 1 & 0 \\ s_1 & \cos(2\pi f_i) & \sin(2\pi f_i) \\ \vdots & \vdots & \vdots \\ s_{N-1} & \cos[2\pi f_i(N-1)] & \sin[2\pi f_i(N-1)] \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} A \\ \alpha_1 \\ \alpha_2 \end{bmatrix}}_{\theta} + \mathbf{n} \tag{7.205}$$

The hypotheses are now

$$\begin{aligned} H_0 &: A = 0 \\ H_1 &: A \neq 0 \end{aligned} \tag{7.206}$$

or (letting $\mathbf{b} = [0\ 0\ 0]^T$)

$$\begin{aligned} H_0 &: \mathbf{A}\theta = \mathbf{0} \\ H_1 &: \mathbf{A}\theta \neq \mathbf{0} \end{aligned} \tag{7.207}$$

where $\mathbf{A} = [\ 1\ 0\ 0]$. Expression (7.196) then becomes

$$\frac{1}{\sigma^2} \hat{\theta}_1^T \mathbf{A}^T [\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1} \mathbf{A}\hat{\theta}_1 > \gamma' \tag{7.208}$$

where $\hat{\theta}_1$ is given by (7.197). If f_i is near the edge of the filter, then amplitude and/or phase distortion will change the characteristics of the interfering signal. Therefore, assuming f_i is not near the edges of the pass band,

$$\mathbf{H}^T\mathbf{H} \approx \begin{bmatrix} \sum_{k=0}^{N-1} s_k^2 & \sum_{k=0}^{N-1} s_k \cos(2\pi f_i k) & \sum_{k=0}^{N-1} s_k \sin(2\pi f_i k) \\ \sum_{k=0}^{N-1} s_k \cos(2\pi f_i k) & \dfrac{N}{2} & 0 \\ \sum_{k=0}^{N-1} s_k \sin(2\pi f_i k) & 0 & \dfrac{N}{2} \end{bmatrix} \tag{7.209}$$

Let $S_c = \sum_{k=0}^{N-1} s_k \cos(2\pi f_i k)$ and $S_s = \sum_{k=0}^{N-1} s_k \sin(2\pi f_i k)$, then

$$(\mathbf{H}^T\mathbf{H})^{-1} = \frac{\begin{bmatrix} -\dfrac{N}{2} & S_c & S_s \\[2ex] S_c & -\sum_{k=0}^{N-1} s_k^2 + \dfrac{2}{N} S_k^2 & -\dfrac{2}{N} S_c S_s \\[2ex] S_s & -\dfrac{2}{N} S_c S_s & -\sum_{k=0}^{N-1} s_k^2 + \dfrac{2}{N} S_c^2 \end{bmatrix}}{S_c^2 + S_s^2 - \dfrac{N}{2}\sum_{k=0}^{N-1} s_k^2} \tag{7.210}$$

and

$$\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1} = \frac{1}{S_c^2 + S_s^2 - \dfrac{N}{2}\sum_{k=0}^{N-1} s_k^2} \begin{bmatrix} -\dfrac{N}{2} & S_c & S_s \end{bmatrix} \tag{7.211}$$

$$\mathbf{H}^T\mathbf{x} = \begin{bmatrix} \sum_{k=0}^{N-1} x_k s_k \\[2ex] \sum_{k=0}^{N-1} x_k \cos(2\pi f_i k) \\[2ex] \sum_{k=0}^{N-1} x_k \sin(2\pi f_i k) \end{bmatrix} \tag{7.212}$$

Therefore,

$$\mathbf{A}\hat{\boldsymbol{\theta}}_1 = \mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$$

$$= \frac{-\dfrac{N}{2}\sum_{k=0}^{N-1} x_k s_k + S_c \sum_{k=0}^{N-1} x_k \cos(2\pi f_i k) + S_s \sum_{k=0}^{N-1} x_k \sin(2\pi f_i k)}{S_c^2 + S_s^2 - \dfrac{N}{2}\sum_{k=0}^{N-1} s_k^2} \tag{7.213}$$

and

$$\mathbf{A}(\mathbf{H}^{\mathsf{T}}\mathbf{H})^{-1}\mathbf{A}^{\mathsf{T}} = \frac{-\dfrac{N}{2}}{S_c^2 + S_s^2 - \dfrac{N}{2}\displaystyle\sum_{k=0}^{N-1}s_k^2} \tag{7.214}$$

Define the FTs

$$S(f_i) = \sum_{k=0}^{N-1}s_k\exp(-j2\pi f_i k) = S_c - jS_s \tag{7.215}$$

$$X(f_i) = \sum_{k=0}^{N-1}s_k\exp(-j2\pi f_i k) \tag{7.216}$$

Then

$$\mathbf{A}\hat{\theta}_1 = \frac{\displaystyle\sum_{k=0}^{N-1}x_k s_k - \dfrac{2}{N}\,\mathrm{Re}[X(f_i)S^*(f_i)]}{-\dfrac{2}{N}|S(f_i)|^2 + \displaystyle\sum_{k=0}^{N-1}s_k^2} \tag{7.217}$$

and

$$\mathbf{A}(\mathbf{H}^{\mathsf{T}}\mathbf{H})^{-1}\mathbf{A}^{\mathsf{T}} = \frac{-\dfrac{N}{2}}{|S(f_i)|^2 - \dfrac{N}{2}\displaystyle\sum_{k=0}^{N-1}s_k^2} \tag{7.218}$$

Then

$$t(\mathbf{x}) = \frac{\left[\displaystyle\sum_{k=0}^{N-1}x_k s_k - \dfrac{2}{N}\,\mathrm{Re}[X(f_i)S^*(f_i)]\right]^2}{\sigma^2\left[\displaystyle\sum_{k=0}^{N-1}s_k^2 - \dfrac{2}{N}|S(f_i)|^2\right]} \tag{7.219}$$

The test statistic can also be expressed in terms of the DFT coefficients. Since

$$\sum_{k=0}^{N-1} x_k s_k = \frac{1}{N} \sum_{k=0}^{N-1} X_k S_k^* \tag{7.220}$$

where X_k and S_k are the DFT coefficients. Assume that the interference is at frequency $f_i = l/N$. Then

$$t(\mathbf{x}) = \frac{\left(\frac{1}{N} \sum_{k=0}^{N-1} X_k S_k^* - \frac{1}{N} X_l S_l^* - \frac{1}{N} X_{N-l} S_{N-l}^* \right)^2}{\sigma^2 \left(\frac{1}{N} \sum_{k=0}^{N-1} |S_k|^2 - \frac{1}{N} |S_l|^2 - \frac{1}{N} |S_{N-l}|^2 \right)} \tag{7.221}$$

From this,

$$t(\mathbf{x}) = \frac{\left(\frac{1}{N} \sum_{\substack{k=0 \\ k \neq l, N-l}}^{N-1} X_k S_k^* \right)^2}{\sigma^2 \frac{1}{N} \sum_{\substack{k=0 \\ k \neq l, N-l}}^{N-1} |S_k|^2} > \gamma' \tag{7.222}$$

Therefore, decide H_1 if

$$\left(\frac{1}{N} \sum_{\substack{k=0 \\ k \neq l, N-l}}^{N-1} X_k S_k^* \right)^2 > \frac{\sigma^2}{N} \gamma' \sum_{\substack{k=0 \\ k \neq l, N-l}}^{N-1} |S_k|^2 = \gamma'' \tag{7.223}$$

The detection logic for this detector is illustrated in Figure 7.19. This is the same detector for detection of a deterministic signal with unknown amplitude, except that the FFT bins that contain the interference are zeroed.

7.2.6 MLE of the Unknown Parameters of Sinusoids in AWGN

The detection of deterministic sinusoidal signals with unknown parameters is discussed in this section. The sinusoids are deterministic since knowing the value of the signals at any point in time allows the value of the signal to be predicted at any future time. The parameters of interest in this case are the frequency of the signal, its phase, and its amplitude. It is assumed that the number of signals
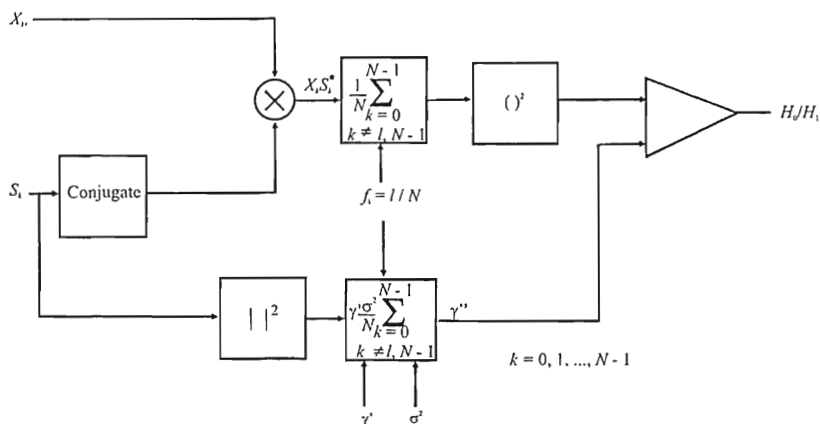
Figure 7.19 GLRT detector with interference. The sums are over all frequencies with the interference frequency deleted.

present, denoted by q, is known or accurately estimated by the techniques presented in Chapter 8.

The samples of the signal consisting of q complex sinusoids in complex AWGN are given by

$$x_i = \sum_{k=1}^{q} A_i e^{j2\pi f_k i + \phi_k} + n_i \qquad (7.224)$$

where A_k, ϕ_k, and f_k are the amplitude, phase, and frequency of the kth sinusoid, respectively. Denoting S_k as the average power in signal k, then $S_k = A_k^2 / 2$.

We should note that the results presented here apply to unmodulated sinusoids and only approximate the case when the signals have modulation applied. Applying these results to narrowband (but not zero bandwidth) signals may lead to significant errors.

7.2.6.1 MLE of the Unknown Parameters of a Single Sinusoid in AWGN

Kay has shown [17] that estimation of the parameters of a single sinusoid is possible. The frequency of a complex sinusoid in complex AWGN is given by the peak of the periodogram of the signal. This yields the MLE of the frequency, $\hat{f}_0$. The same result applies for a real sinusoid in real AWGN. If the estimate of the frequency is known, the MLE of the amplitude and phase is given by

$$\hat{A} = \left| \frac{1}{N} \sum_{i=0}^{N-1} x_i e^{-j2\pi f_0 i} \right| \tag{7.225}$$

and

$$\hat{\phi} = \tan^{-1} \left[\frac{\mathrm{Im}\left(\sum_{i=0}^{N-1} x_i e^{-j2\pi f_0 i} \right)}{\mathrm{Re}\left(\sum_{i=0}^{N-1} x_i e^{-j2\pi f_0 i} \right)} \right] \tag{7.226}$$

It can be shown that the CRLB for the frequency estimate is given by [17]

$$\mathrm{var}(\hat{f}_1) \geq \frac{6\sigma_n^2}{A_1^2 N(N^2 - 1)(2\pi)^2} = \frac{3}{\nu_1 N(N^2 - 1)(2\pi)^2} \tag{7.227}$$

where $\nu_1 = S_1 / \sigma_n^2$ is the SNR for the signal. Similarly,

$$\mathrm{var}(\hat{A}_1) \geq \frac{\sigma_n^2}{2N} \tag{7.228}$$

and

$$\mathrm{var}(\hat{\phi}_1) \geq \frac{\sigma_n^2 (2N - 1)}{A_1^2 N(N + 1)} = \frac{(2N - 1)}{2\nu_1 N(N + 1)} \tag{7.229}$$

Expressions (7.227) and (7.229) are plotted in Figure 7.20 (log scale) versus the SNR for a few typical values of N.

7.2.6.2 MLE of the Unknown Parameters of Multiple Sinusoids in AWGN

Determining the MLEs for the sinusoidal parameters when there is more than one constituent sinusoid present is much more difficult. An example of the surface to
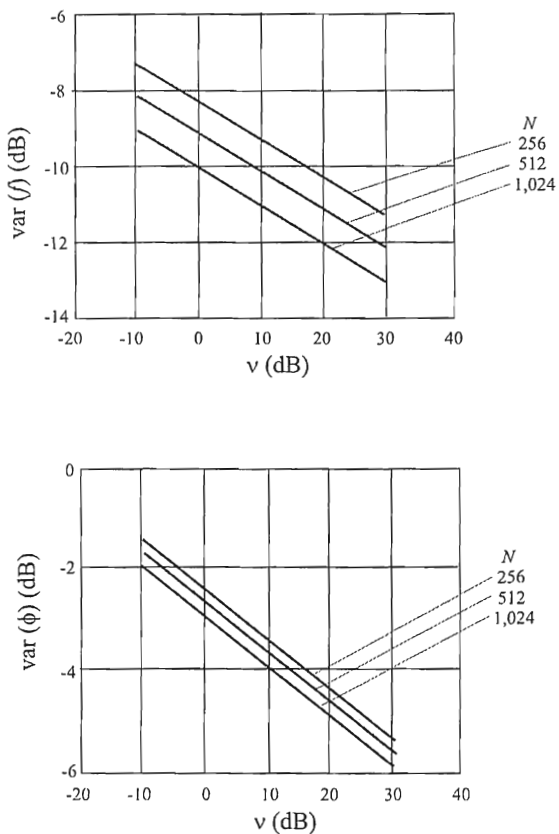
Figure 7.20 CRLBs on the variance of frequency and phase MLEs as a function of SNR. Note that the ordinate is plotted on a log scale.
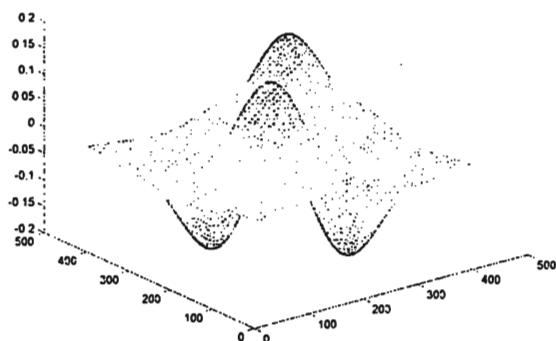
Figure 7.21 Surface to be searched for maximum.

be searched to determine the maximum is shown in Figure 7.21. The local maxima and minima are clearly discernable, and if one of these local maximums is chosen, then erroneous results ensue.

There are two cases, however, when two signals can be discerned. These are:

1. $f_1 = k/N$ and $f_2 = l/N$ where k and l are different integers over the ranges

$$\left[-\frac{N}{2}, \frac{N}{2}-1\right] \text{ for } N \text{ even}$$

and

$$\left[-\left(\frac{N-1}{2}\right), \left(\frac{N-1}{2}\right)\right] \text{ for } N \text{ odd.}$$

2. $|f_1 - f_2| \gg 1/N$ \hfill (7.230)

The MLE estimates of the frequencies are determined from the two highest peaks in the periodogram. This requires a one-dimensional search. The second case is when the periodogram peaks are discernable, which requires wide separation in frequency. It was this difficulty in estimating the parameters of two or more sinusoids in AWGN that was largely responsible for motivating the development of the high-resolution methods presented in Chapter 9.

The CRLBs for multiple signals is difficult to calculate in closed form and therefore must be determined numerically. As indicated in Kay [17], define

$$\boldsymbol{\theta} = [f_1 \quad A_1 \quad \phi_1 \quad f_2 \quad A_2 \quad \phi_2 \quad \cdots \quad f_q \quad A_q \quad \phi_q] \tag{7.231}$$

The CR bounds are then

$$\mathrm{var}(\hat{\theta}_i) \geq \frac{\sigma_n^2 [\mathbf{M}^{-1}]_{ii}}{2B_i^2} \tag{7.232}$$

where $B_i = A_i$ if θ_i corresponds to a phase parameter, $B_i = 1$ if θ_i corresponds to an amplitude parameter, and $B_i = 2\pi A_i$ if θ_i corresponds to a frequency parameter. $\mathbf{M}$ is a $3q$ x $3q$ matrix and $[\mathbf{M}^{-1}]_{ii}$ corresponds to the iith element of its inverse. The matrix is defined as

$$\mathbf{M}_{ij} = \begin{bmatrix} \sum_{k=0}^{N-1} k^2 \cos \Delta_k[i,j] & -\sum_{k=0}^{N-1} k \sin \Delta_k[i,j] & \sum_{k=0}^{N-1} k \cos \Delta_k[i,j] \\ \sum_{k=0}^{N-1} k \sin \Delta_k[i,j] & \sum_{k=0}^{N-1} \cos \Delta_k[i,j] & \sum_{k=0}^{N-1} \sin \Delta_k[i,j] \\ \sum_{k=0}^{N-1} k \cos \Delta_k[i,j] & -\sum_{k=0}^{N-1} \sin \Delta_k[i,j] & \sum_{k=0}^{N-1} \cos \Delta_k[i,j] \end{bmatrix} \tag{7.233}$$

and $\Delta_k[i,j] = 2\pi(f_i - f_j)k + (\phi_i - \phi_j)$. $\mathbf{M}_{ij} \approx 0$, $i \neq j$, if all adjacent signal pairs are such that $|f_i - f_j| \gg 1/N$. This condition is the same as (7.230) and implies that the peaks of the periodogram are discernable for all frequency pairs. In such a situation, $\mathbf{M}$ is block diagonal and the CRLBs are the same as that for a single sinusoid given by (7.227) through (7.229).

7.2.7 Optimum Detection of Deterministic Signals with Unknown Parameters in Impulsive Noise

Tsihrintzis and Nikias examined the performance of signal detection when the noise is modeled as impulsive [18]. The pdf for *bivariate isotropic symmetric, alpha-stable* (BISαS) impulsive noise is defined by the inverse FT given by

$$p_{\alpha,\gamma,\delta_1,\delta_2}(x_1,x_2) = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \begin{array}{l} \{\exp[j2\pi(f_1\delta_1 + f_2\delta_2) \\ -\gamma(2\pi)^2(f_1^2 + f_2^2)^{\alpha/2}] \\ \times\exp[-j2\pi(x_1 f_1 + x_2 f_2)]\} \end{array} \quad df_1 df_2 \qquad (7.234)$$

Parameters α and γ are known as the *characteristic exponent* and *dispersion*, respectively. The other two parameters δ_1 and δ_2 are location parameters and are set equal to zero for this discussion.

Closed-form pdfs of (7.234) only exist for two cases: $\alpha = 1$ and $\alpha = 2$. The former is known as the Cauchy pdf and the latter is the familiar Gaussian pdf. These expressions are

$$p_{\alpha,\gamma}(x_1,x_2) = \begin{cases} \dfrac{\gamma}{2\pi(\rho^2 + \gamma^2)^{3/2}}, & \alpha = 1 \\[4mm] \dfrac{1}{4\pi\gamma}\exp\left(-\dfrac{\rho^2}{4\gamma}\right), & \alpha = 2 \end{cases} \qquad (7.235)$$

These functions are shown in Figure 7.22.

The hypotheses model considered is given by
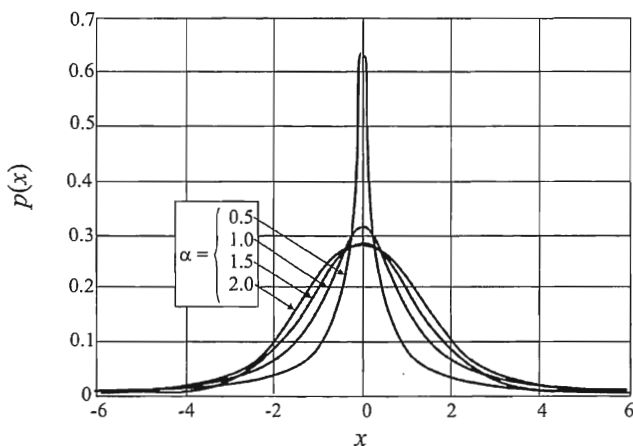
Figure 7.22 Comparison of the Gaussian pdf ($\alpha = 2$) with the Cauchy ($\alpha = 1$) and other pdfs ($\alpha = 0.5$, 1.5), the latter three of which are used to model impulsive noise. For $\alpha = 0.5$ and 1.5, the curves were obtained with a series expansion.

$$H_0 : x_k = n_k, \qquad\qquad k = 0,1,\ldots,N-1$$
$$H_1 : x_k = e^{j\phi}s_k + n_k, \qquad k = 0,1,\ldots,N-1 \qquad (7.236)$$

where s_k is the signal of interest, ϕ is a random phase angle on $[0, 2\pi)$, and n_k are samples from BISαS noise with dispersion γ. It is assumed that either hypothesis is equally likely, which is a marginal assumption for the signal search problem. There is no reason to believe that the signal being sought will occur half the time and not the other half. It does, however, provide a useful bound and serves to illustrate the effects of impulsive noise considerations on detection.

The optimum receiver for this problem implements the test statistic given by the likelihood ratio

$$\lambda_{opt} = \frac{1}{2\pi N}\sum_{k=0}^{N-1}\int_0^{2\pi}\ln\frac{p_{1,\gamma}(x_k - e^{j\phi}s_k)}{p_{1,\gamma}(x_k)}d\phi \qquad (7.237)$$

which, by integrating over the range for ϕ, removes its random effects. This expression is difficult to solve in closed form, so a suboptimum statistic for the Cauchy receiver is used, which is given by

$$\lambda_c = \frac{1}{N}\sum_{k=0}^{N-1}\ln\frac{\dfrac{1}{2\pi}\displaystyle\int_0^{2\pi}\dfrac{1}{(a^2 + b\cos\phi + c\sin\phi)^{3/2}}d\phi}{\dfrac{1}{(|x_k|^2 + \gamma^2)^{3/2}}} \qquad (7.238)$$

where

$$a^2 = |x_k|^2 + |s_k|^2 + \gamma^2$$
$$b = -\operatorname{Re}(x_k s_k)$$
$$c = -2\operatorname{Im}(x_k s_k^*)$$

On the other hand, the test statistic for the optimum Gaussian receiver is

$$\lambda_G = \frac{1}{N}\left|\sum_{k=0}^{N-1}x_k s_k^*\right| \qquad (7.239)$$

Because only two closed-form expressions exist for the pdf, a Monte-Carlo simulation was performed to provide detection performance for when $\alpha \neq 1, 2$. Comparison of the two receivers in the form of ROC curves for a few values of α in BISαS ($\gamma = 1$) noise are shown in Figure 7.23. The Cauchy receiver is shown as a solid line while the Gaussian results are shown as dotted lines. For all but $\alpha = 2$, where the Gaussian receiver is optimum, the Cauchy receiver has significantly better performance. The Gaussian receiver is best when $\alpha = 2$ because then the pdf is normal and the Gaussian receiver is optimal in that case.

Performance of the individual receivers in the same type of noise are shown in Figure 7.24. Again, the optimality of the Gaussian receiver when $\alpha = 2$ is evident. However, when $\alpha \neq 2$, the Cauchy receiver outperforms the Gaussian receiver.

The conclusion of the analysis is that when the noise is known or suspected to be other than Gaussian, the Cauchy receiver implementing the statistic given by (7.238) will provide significantly better detection performance than the Gaussian receiver.

7.3 Concluding Remarks

The fundamentals of detection of deterministic signals with known and unknown parameters were presented in this chapter. Deterministic signals with unknown and random parameters are perhaps the most prolific signal types of interest to communication EW system designers. The signals are deterministic because they are generated by oscillators in transmitters that produce sinusoidal signals. These signals are used as high-frequency carriers for the information-bearing modulations.

The matched filter is the optimum detector when everything is known about a signal. This forms a comparison for other types of detection. When there are nuisance parameters present, and in most real circumstances there are, then the UMP test is not available and some less optimal test is required for signal detection. Described in this chapter were two of these nonoptimal tests. The first was the GLRT test where the variates are replaced with their MLE estimates. The second was a LOD test for weak signals in AWGN.

Other forms of testing for deterministic signals were also presented. These include the Bayes linear model and the quadrature detector.

Finally, the effects on optimal detection in other than AWGN were discussed. In particular, the effects of impulsive noise, which occurs frequently in practice, on optimum detection was discussed.
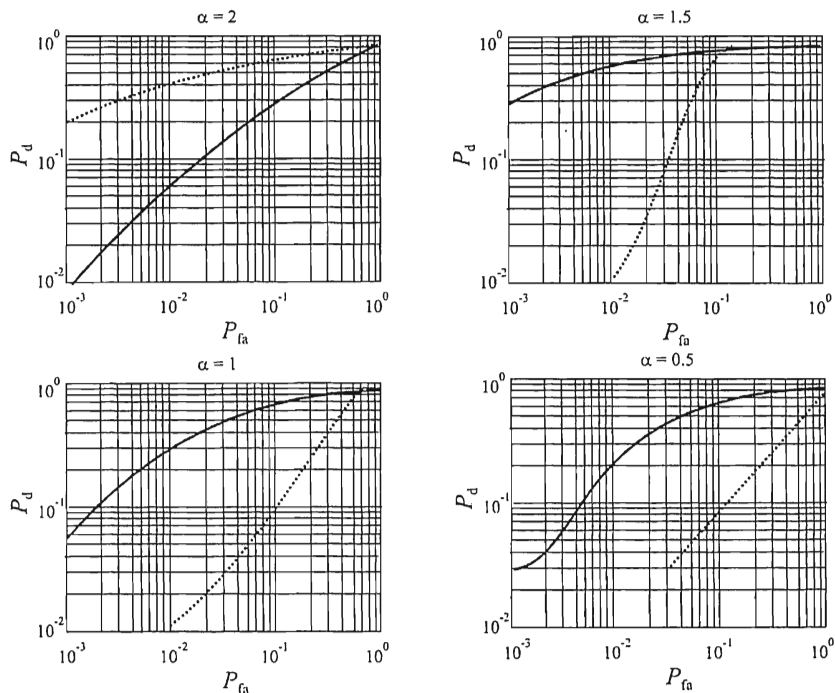
Figure 7.23 Performance of a Cauchy receiver compared to Gaussian receiver in unit dispersion BISαS noise. (*From:* [18], © 1995 IEEE. Reprinted with permission.)
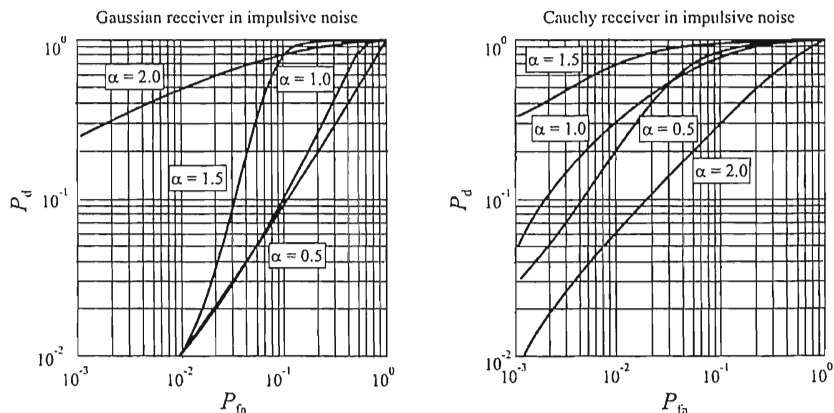
Figure 7.24 Comparison of Cauchy receiver with the Gaussian receiver in impulsive noise. (*From:* [18], © 1995 IEEE. Reprinted with permission.)

References

[1]   McNicol, D., *A Primer of Signal Detection Theory,* Sydney: George Australian Publishing Company, 1972.

[2]   Kay, S. M., *Fundamentals of Statistical Signal Processing Detection Theory,* Upper Saddle River, NJ: Prentice Hall, 1998, p. 65.

[3]   Turin, G. L., "An Introduction to Matched Filters," *IRE Transactions on Information Theory,* June 1960, pp. 311–329.

[4]   Indritz, J., *Methods in Analysis,* New York: Macmillan, 1964, p. 7.

[5]   Indritz, J., *Methods in Analysis,* New York: Macmillan, 1964, p. 5.

[6]   Kay, S. M., *Fundamentals of Statistical Signal Processing Detection Theory,* Upper Saddle River, NJ: Prentice Hall, 1998, pp. 101–105.

[7]   Nicholson, D. L., *Spread Spectrum Signal Design LPE and AJ Systems,* Rockville, MD: Computer Science Press, 1988, p.144.

[8]   Urkowitz, H., "Energy Detection of Unknown Deterministic Signals," *Proceedings of the IEEE,* Vol. 55, No., 4, April 1967, pp. 523–531.

[9]   Whalen, A. D., *Detection of Signals in Noise,* New York: Academic Press, 1971, p. 199.

[10]   Hoh, Y.-S., "Optimum Detection of Known Signals in a Non-Gaussian Interference Environment," *Proceedings of IEEE MILCOM 1990,* pp. 642–657.

[11]   Kassam, S. A., *Signal Detection in Non-Gaussian Noise,* New York: Springer-Verlag, 1988.

[12]   Modestino, J., and A. Ningo, "Detection of Weak Signals in Narrowband Non-Gaussian Noise," *IEEE Transactions on Communications,* Vol. COM-25, September 1977, pp. 592–600.

[13]   Papoulis, A., *Probability, Random Variables, and Stochastic Processes,* New York: McGraw-Hill, 1965, p. 201.

[14]   Ingram, R., and R. Houlse, *Performance of the Optimum and Several Suboptimum Receivers for Threshold Detection of Known Signals in Additive, White, Non-Gaussian Noise,* Naval Underwater Systems Center, New London, CT, NUSC Tech. Report 6339, November 1980.

[15]   Kay, S. M., *Fundamentals of Statistical Signal Processing Detection Theory,* Upper Saddle River, NJ: Prentice Hall, 1998, pp. 272–275.

[16]   Kay, S. M., *Fundamentals of Statistical Signal Processing Detection Theory,* Upper Saddle River, NJ: Prentice Hall, 1998, pp. 284–289.

[17]   Kay, S. M., *Modern Spectral Estimation Theory and Applications,* Upper Saddle River, NJ: Prentice Hall, 1988, Chapter 13.

[18]   Tsihrintzis, G. A., and C. L. Nikias, "Incoherent Receivers in Alpha-Stable Impulse Noise," *IEEE Transactions on Signal Processing,* Vol. 43, No. 9, September 1995, pp. 2225–2229.

Chapter 8

Detection of Stochastic Signals

As indicated at the beginning of Chapter 6, there are two types of signals of primary interest to communication EW systems. The signals are either deterministic with one or more unknown parameters or they are completely random. Detection of deterministic signals was discussed in Chapter 7. Detection of random signals is presented in this chapter.

8.1 Detection of Random Signals with Unknown Parameters

When the medium through which the RF signal is propagating is scattering, a sine wave may appear as though it is noise. A modulated sine wave communication signal in this medium may look like noise with modulation on it. Therefore it is necessary to analyze the detection performance of noise-like signals.

8.1.1 GLRT Detection of Stochastic Signals

When nothing is known about the signal being analyzed, it can be modeled as completely random with unknown and random parameters. Let s_k denote the kth sample of such a Gaussian random signal with zero mean and covariance matrix $\mathbf{C}_s$. Then, with observations x_k, the hypotheses are

$$
\begin{aligned}
H_0 &: x_k = n_k, & k = 0,1,\ldots,N-1 \\
H_1 &: x_k = s_k + n_k, & k = 0,1,\ldots,N-1
\end{aligned}
\tag{8.1}
$$

where $n_k \sim \mathcal{N}(0, \sigma^2)$. The covariance of s_k is given by

$$\gamma_{ss,m} = \mathcal{E}\{s_k s^*_{k+m}\} \tag{8.2}$$

The covariance matrix for s_n is given by

$$C_s = \begin{bmatrix} \gamma_{ss,0} & \gamma_{ss,1} & \cdots & \gamma_{ss,N-1} \\ \gamma_{ss,1} & \gamma_{ss,0} & \cdots & \gamma_{ss,N-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{ss,N-1} & \gamma_{ss,N-2} & \cdots & \gamma_{ss,0} \end{bmatrix} \tag{8.3}$$

Let the covariance matrix for n_k be denoted by $\mathbf{C_n}$. Since n_k are i.i.d. and $\sim \mathcal{N}(0, \sigma^2)$, then $\mathbf{C_n} = \sigma^2 \mathbf{I}$. When the signal is present, the covariance matrix for x_k is given by $\mathbf{C_s} + \mathbf{C_n}$. The pdf under H_1 is thus given by

$$p[\mathbf{x} : \gamma_{ss,0}, H_1] = \frac{1}{2\pi\sqrt{\det(\mathbf{C_s} + \sigma^2 \mathbf{I})}} \exp\left[-\frac{1}{2}\mathbf{x}^T(\mathbf{C_s} + \sigma^2 \mathbf{I})^{-1}\mathbf{x}\right] \tag{8.4}$$

This expression must be maximized over $\gamma_{ss,0}$ in order to find the MLE to use in the GLRT. By making the appropriate substitutions, the log likelihood ratio becomes

$$\ln L(\mathbf{x}) = \ln \frac{\sqrt{|\mathbf{C_n}|}}{\sqrt{|\mathbf{C_n} + \mathbf{C_s}|}} - \frac{1}{2}\mathbf{s}[(\mathbf{C_n} + \mathbf{C_s})^{-1} - \mathbf{C_n^{-1}}]\mathbf{s}^T \tag{8.5}$$

The likelihood test is based on the second term in (8.5), which is of the form $\mathbf{sJs}^T$ where, $-\mathbf{J} = (\mathbf{C_n} + \mathbf{C_s})^{-1} - \mathbf{C_n^{-1}}$ (ignoring the ½ factor has no impact on the result). Thus, H_1 should be chosen when

$$\mathbf{s}(\mathbf{C_n} + \mathbf{C_s})^{-1}\mathbf{C_s}\mathbf{C_n^{-1}}\mathbf{s}^T \geq \gamma' \tag{8.6}$$

Given that there are two random signals and the hypotheses are

$$H_0 : \mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \tag{8.7}$$

$$H_1 : \mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \tag{8.8}$$

where μ_0 and μ_1 are the vectors of means while Σ_0 and Σ_1 are the noise covariance matrices. Σ_0 and Σ_1 are invertible, symmetric, and positive definite. Therefore, all their eigenvalues $\{\lambda_k\}_1^N$ are positive real numbers and the associated eigenvectors $\{v_k\}_1^N$ can form an orthonormal basis for $\mathfrak{R}^N$. The *spectral decomposition* of these matrices is given by

$$\mathbf{S}_0 = \sum_{k=1}^N \lambda_k \mathbf{v}_k \mathbf{v}_k^{\mathrm{T}} \qquad (8.9)$$

and similarly for Σ_1.

The log-likelihood ratio is given by

$$\log L(\mathbf{y}) = \frac{1}{2}\mathbf{y}^{\mathrm{T}}\left(\Sigma_0^{-1} - \Sigma_1^{-1}\right)\mathbf{y} + \left(\mu_1^{\mathrm{T}}\Sigma_1^{-1} - \mu_0^{\mathrm{T}}\Sigma_0^{-1}\right)\mathbf{y}$$
$$+ \frac{1}{2}\left[\log\frac{|\Sigma_0|}{|\Sigma_1|}\right] + \mu_0^{T}\Sigma_0^{-1}\mu_0 - \mu_1^{T}\Sigma_1^{-1}\mu_1 \qquad (8.10)$$

If $\Sigma_0 = \Sigma_1 = \Sigma$, the quadratic term disappears and the optimal test statistic is then

$$t(\mathbf{y}) = (\mu_1 - \mu_0)^{\mathrm{T}}\Sigma^{-1}\mathbf{y} \qquad (8.11)$$

which is a *linear detector* (matched filter). Alternatively, if $\mu_0 = \mu_1 = \mu$, then the optimal test is

$$t(\mathbf{y}) = (\mathbf{y} - \mu)^{\mathrm{T}}\left(\Sigma_0^{-1} - \Sigma_1^{-1}\right)(\mathbf{y} - \mu) \qquad (8.12)$$

When $\mu = \mathbf{0}$, then this statistic becomes

$$t(\mathbf{y}) = \mathbf{y}^{\mathrm{T}}\left(\Sigma_0^{-1} - \Sigma_1^{-1}\right)\mathbf{y} \qquad (8.13)$$

Expression (8.13) corresponds to detecting zero-mean Gaussian signals in Gaussian noise, according to the hypotheses

$$H_0 : \mathbf{y} = \mathbf{n} \qquad (8.14)$$

versus

$$H_1 : \mathbf{y} = \mathbf{s} + \mathbf{n} \tag{8.15}$$

where $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$, $\mathbf{s} \sim \mathcal{N}(\mathbf{0}, \Sigma_S)$, and $\mathbf{n}$ and $\mathbf{s}$ are independent.

When there are two Gaussian signals present, then the covariance matrices are changed to

$$\Sigma_0 = \sigma^2 \mathbf{I} \tag{8.16}$$

and

$$\Sigma_1 = \sigma^2 \mathbf{I} + \Sigma_s \tag{8.17}$$

The corresponding optimal detection statistic is given by

$$\tilde{\delta}_0(\mathbf{y}) = \begin{cases} 0, & \mathbf{y}^T \mathbf{Q} \mathbf{y} < \tau' = 2(\log \tau - C) \\ \gamma, & \mathbf{y}^T \mathbf{Q} \mathbf{y} = \tau' = 2(\log \tau - C) \\ 1, & \mathbf{y}^T \mathbf{Q} \mathbf{y} > \tau' = 2(\log \tau - C) \end{cases} \tag{8.18}$$

where

$$C = \frac{1}{2} \left[\log \frac{|\Sigma_0|}{|\Sigma_1|} \right] + \mu_0^T \Sigma_0^{-1} \mu_0 - \mu_1^T \Sigma_1^{-1} \mu_1 \tag{8.19}$$

and

$$\mathbf{Q} = \sigma^{-2} \mathbf{I} - \left(\sigma^2 \mathbf{I} + \Sigma_s \right)^{-1} = \sigma^{-2} S_s \left(\sigma^2 \mathbf{I} + \Sigma_s \right)^{-1} \tag{8.20}$$

Expressions (8.18) throught (8.20) correspond to a *quadratic detector*.

If $\mathbf{s} = (s_1, s_2, ..., s_N)^T$ and the samples $\{s_k\}_{k=1}^N$ are i.i.d. and $\sim \mathcal{N}\{0, \sigma_S^2\}$, then $\lambda_1 = \lambda_2 = ... = \lambda_N = \sigma_S^2$ and

$$\sigma_0^2 = \frac{\sigma_S^2}{\sigma^2 + \sigma_S^2} \tag{8.21}$$

and

$$\sigma_1^2 = \frac{\sigma_S^2}{\sigma^2} \tag{8.22}$$

The test statistic $t(\mathbf{y})$ is the sum of squares of N i.i.d. $\mathcal{N}\left(0,\sigma_j^2\right)$ random variables and the pdf of $t(\mathbf{y})$ is

$$p_t\left(T|H_j\right) = \begin{cases} \dfrac{1}{\left(2\sigma_j^2\right)^{N/2}\Gamma\left(\dfrac{N}{2}\right)} T^{\frac{N}{2}-1} e^{-\frac{T}{2\sigma_j^2}}, & T > 0 \\ 0, & T \le 0 \end{cases} \tag{8.23}$$

where

$$\Gamma(x) = \int_0^\infty e^{-t}t^{x-1}dt, x > 0 \tag{8.24}$$

is the *Gamma function*, which is $\Gamma(n + 1) = n!$ for $n \ge 0$, integer. Expression (8.23) is illustrated in Figure 8.1.
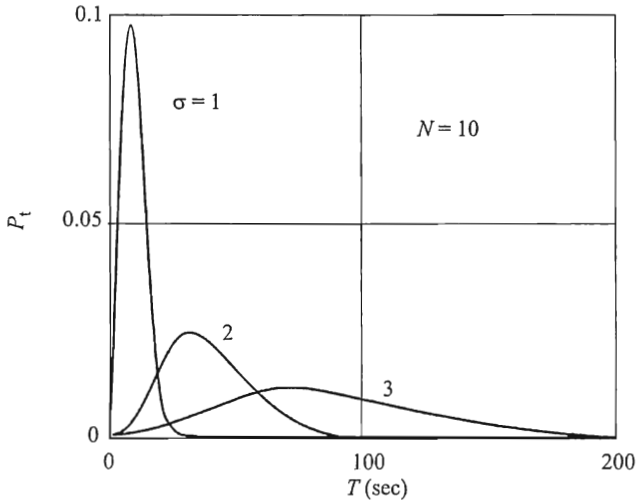
Figure 8.1 Quadratic detector pdf.

The associated cumulative distribution function is given by

$$P_j\left[t(y) \le T\right] = \Gamma\left(\frac{N}{2}, \frac{T}{2\sigma_j^2}\right) \tag{8.25}$$

where

$$\Gamma(x,t) = \frac{1}{\Gamma(x)} \int_0^t e^{-y} y^{x-1} dy \tag{8.26}$$

and if $\sigma_j^2 = 1$, this pdf is the chi-square density with N degrees of freedom, discussed in Chapter 7.

For the P_{fa}-level NP test, the threshold τ' is

$$\tau' = 2\sigma_0^2 \Gamma^{-1}\left(\frac{N}{2}, 1 - P_{fa}\right) \tag{8.27}$$

and

$$P_d(\delta) = 1 - \Gamma\left[\frac{N}{2}, \frac{\sigma_0^2}{\sigma_1^2} \Gamma^{-1}\left(\frac{N}{2}, 1 - P_{fa}\right)\right] \tag{8.28}$$

The hypotheses for detection of random signals in Gaussian noise are

$$x(t) = \begin{cases} s(t) + n(t), & \text{signal present,} \\ n(t), & \text{signal absent,} \end{cases} \quad |t| \le T/2 \tag{8.29}$$

with the detection statistic given by

$$y = \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} k(u,v) x(u) x(v) du dv \tag{8.30}$$

where the weighting function, $k(u,v)$, defines the characteristics of the quadratic detector. The quantity to be maximized used by Gardner [1] for random signal detection was the deflection, defined as

$$d \triangleq \frac{\left| \mathcal{E}\{y|s(t) \text{ present}\} - \mathcal{E}\{y|s(t) \text{ absent}\} \right|}{\left(\text{var}\{y|s(t) \text{ absent}\} \right)^{\frac{1}{2}}} \tag{8.31}$$

The resulting test statistic is

$$y = \frac{1}{N_0^2} \int\limits_{-\infty}^{\infty} S_s(f) P_T(f) df \tag{8.32}$$

where N_0 is the one-sided noise spectral density, $S_s(f)$ is the psd of the random signal to be detected, and $P_T(f)$ is the periodogram of $x(t)$ (discussed in Chapter 5) given by

$$P_T(f) \triangleq \frac{1}{T} |X_T(f)|^2 \tag{8.33}$$

where $X_T(f)$ is a truncated version of $X(t)$:

$$X_T(f) \triangleq \int\limits_{-\frac{T}{2}}^{\frac{T}{2}} x(t) e^{-j2\pi f t} dt \tag{8.34}$$

It is a common when conducting GS in communication EW systems to assume that the signals have a flat spectrum given by S_0 W/Hz, with a finite bandwidth specified by the frequency interval $(-B/2, B/2)$. In that case,

$$y = \frac{S_0}{N_0^2} \int\limits_{-\frac{B}{2}}^{\frac{B}{2}} P_T(f) df \tag{8.35}$$

Note that this detector requires estimating the noise level N_0.

8.1.2 Locally Optimum Detection of Stochastic Signals

Suppose the hypotheses are given by

$$H_0 : \mathbf{y} = \mathbf{n} \tag{8.36}$$

$$H_1 : \mathbf{y} = \theta^{\frac{1}{2}} \mathbf{s} \tag{8.37}$$

where $\theta > 0$ is an unknown constant multiplier, $\mathbf{n} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$, $\mathbf{s} \sim \mathcal{N}(0, \Sigma_S)$, and $\mathbf{n}$ and $\mathbf{s}$ are independent. The optimal test statistic for these hypotheses is

$$t(\mathbf{y}) = \theta \mathbf{y}^\mathsf{T} \left(\mathbf{I} + \theta \Sigma_s \right)^{-1} \mathbf{y} \tag{8.38}$$

The second θ cannot be separated out, so no UMP test exists. An LMP test can be found by differentiating $t(\mathbf{y})$ wrt θ, which ultimately yields

$$t(\mathbf{y}) = \mathbf{y}^\mathsf{T} \Sigma_s \mathbf{y} \tag{8.39}$$

The statistic does not lose its optimality if it is divided by N, so

$$\tilde{t}(\mathbf{y}) = \frac{1}{N} \mathbf{y}^\mathsf{T} \Sigma_s \mathbf{y} \tag{8.40}$$

A zero mean signal is wss if the $(k - l)$th element of Σ_S, denoted by $\rho_{k,l}$, depends only on the difference $(k - l)$. Thus,

$$\rho_{k,l} = \rho_{k-l,0} = \rho_{k-l} \tag{8.41}$$

Therefore,

$$t(\mathbf{y}) = \frac{1}{N} \mathbf{y}^\mathsf{T} \Sigma_s \mathbf{y} \tag{8.42}$$

$$= \frac{1}{N} \begin{bmatrix} y_1 & y_2 & \cdots & y_N \end{bmatrix} \begin{bmatrix} \rho_{1,1} & \rho_{1,2} & \cdots & \rho_{1,N} \\ \rho_{2,1} & \rho_{2,2} & \cdots & \rho_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{N,1} & \rho_{N,2} & \cdots & \rho_{N,N} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

$$= \frac{1}{N} \begin{bmatrix} y_1 & y_2 & \cdots & y_N \end{bmatrix} \begin{bmatrix} \rho_0 & \rho_1 & \cdots & \rho_{N-1} \\ \rho_1 & \rho_0 & \cdots & \rho_{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{N-1} & \rho_{N-2} & \cdots & \rho_0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

Carrying out the matrix math and rearranging terms yield

$$t(\mathbf{y}) = \frac{1}{N} \sum_{k=1}^{N} \sum_{l=1}^{N} y_k y_l \rho_{k-l} = \rho_0 \hat{\rho}_0 + 2 \sum_{k=1}^{N-1} \rho_k \hat{\rho}_k \tag{8.43}$$

where

$$\hat{\rho}_k = \frac{1}{N} \sum_{l=1}^{N-k} y_l y_{l+k} = \frac{N-k}{N} \frac{1}{N-k} \sum_{l=1}^{N-k} y_l y_{l+k} \tag{8.44}$$

$\{\rho_k\}$ can be regarded as a weighted estimate of the covariance $\mathcal{E}\{y_l y_{l+k}\}$ of the observations for $k = 0, 1, ..., N-1$. Hence, the test statistic can be regarded as a weighted estimate of the covariance structure of the observations correlated with the known covariance structure of the signal. Since

$$\mathcal{E}_0 \{y_l y_{l+k}\} = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases} \tag{8.45}$$

and

$$\mathcal{E}_1 \{y_l y_{l+k}\} = \begin{cases} 1 + \theta \rho_k, & k = 0 \\ \theta \rho_k, & k \neq 0 \end{cases} \tag{8.46}$$

it follows that, for reasonably accurate estimates $\{\hat{\rho}_k\}$,

$$\tilde{t}(\mathbf{Y}) \approx \begin{cases} \rho_0, & \text{under } H_0 \\ \rho_0 + \theta \left(\rho_0^2 + 2 \sum_{k=1}^{N-1} \rho_k^2 \right), & \text{under } H_1 \end{cases} \tag{8.47}$$

Hence, the statistic $\tilde{t}(\mathbf{y})$ is an intuitively reasonable way of detecting the signal, especially if the signal is highly correlated; that is, $\sum_{k=0}^{N-1}\rho_k^2$ is large.

A similar interpretation can be given in terms of the psd of the signal $S_{ss}(\omega)$, which satisfies

$$S_{ss}(\omega) = \sum_{k=-\infty}^{\infty} \rho_k e^{-j\omega k} \tag{8.48}$$

Hence,

$$
\begin{aligned}
\tilde{t}(\mathbf{y}) &= \frac{1}{N}\sum_{k=1}^{N}\sum_{l=1}^{N} y_k y_l \rho_{k-l} \\
&= \frac{1}{N}\sum_{k=1}^{N}\sum_{l=1}^{N} y_k y_l \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} S_{ss}(\omega)e^{j(k-l)\omega}\right] \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{ss}(\omega)\left[\frac{1}{N}\left(\sum_{k=1}^{N} y_k e^{jk\omega}\right)\left(\sum_{l=1}^{N} y_l e^{-jl\omega}\right)\right]d\omega \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{ss}(\omega)\left[\frac{1}{N}\left|\left(\sum_{l=1}^{N} y_l e^{-jl\omega}\right)\right|^2\right]d\omega \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{ss}(\omega)\hat{\phi}_{yy}(\omega)d\omega
\end{aligned}
\tag{8.49}
$$

where

$$\hat{\phi}_{yy}(\omega) = \frac{1}{N}\left|\sum_{l=1}^{N} y_k e^{-jl\omega}\right|^2 \tag{8.50}$$

which is the periodogram estimate of the psd of the observed process.

8.2 Radiometer

A radiometer is a device to measure RF energy. It was adopted from the field of radio astronomy where it was originally designed to detect energy from

astronomical objects. As such, its purpose was to measure minute amounts of energy. It was adapted to the communication field for the detection of weak signals—that is, signals where $P_s/N_0W_F = v < 0$, or when the signal parameters are unknown—deterministic or random. One example of a weak signal is a low probability of intercept signal. In that case, the instantaneous bandwidth of the radiometer is large—equal to the total hopping bandwidth in the case of frequency hopping targets or the total spread bandwidth in the case of direct sequence spread spectrum.

8.2.1 Radiometer Detector

The radiometer is an optimum detector in the maximum likelihood sense when the only known parameters of the signal are its bandwidth W_F and duration T [2].

If the noise is Gaussian with one-sided spectral density given by N_0, the normalized energy of a signal plus noise received over the duration T is given by

$$z(\tau) = \frac{2}{N_0} \int_{t-T}^{t} r^2(\tau) d\tau \qquad (8.51)$$

Time interval T could be the entire duration of the signal as it is received or it could be over a shorter interval, such as the symbol duration of a digital signal. For the latter case, the detection of the presence of a symbol is desired after T seconds, after which the detector begins detection of the next symbol. In that case, the filter is dumped after the detection interval. In the former case, the detector is not dumped after the signal has disappeared.

A block diagram of the radiometer detector is shown in Figure 8.2. The incoming signal is first appropriately amplified, filtered, and down-converted to some convenient frequency. It is assumed that such filtering, amplification, and frequency conversion do not change the characteristics of the signal to be detected. That is, the bandwidth of such filtering is much wider than the bandwidth of the
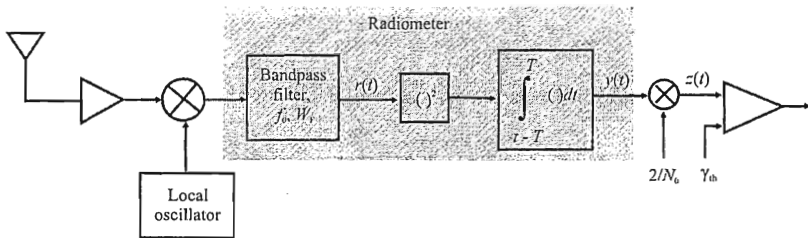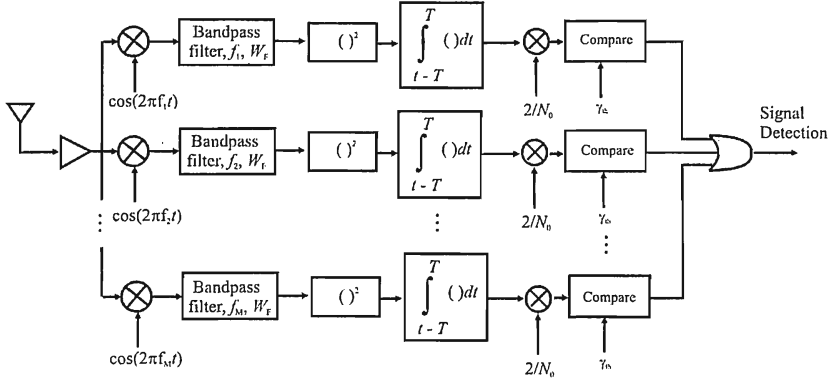

Figure 8.2 Radiometer.

Figure 8.3 Filter bank combiner.

signal of interest. We assume that the filter center frequency f_0 and bandwidth W_F are matched to the signal. The center frequency is the IF frequency of the receiver, while W_F is assumed to be known to match the signal. It is further assumed that W_F is rectangular, with amplitude one. Thus, the noise that passes through W_F is flat of bandwidth W_F.

Note that the normalized energy contains a multiplicative term that contains the reciprocal of the noise spectral density. An estimate of this density is therefore required for the radiometer to function properly. The performance of the radiometer is relatively sensitive to this value, however, as discussed in Section 8.2.2.

It is also possible to implement channelized radiometers, as illustrated in Figure 8.3. The bandwidth of each channel is matched to the instantaneous bandwidth of the target channel, and there are enough channels to cover all or a significant portion of the total bandwidth of interest. This is called the *filter bank combiner* (FBC) [3, 4]. It is not necessary to cover the entire bandwidth, in which case it is called a *partial band FBC* [5].

8.2.2 Radiometer Performance

When the radiometer declares a signal present when there is only noise, a false alarm has occurred. The probability of false alarm is therefore

$$P_{fa} = \Pr\left\{z > \gamma_{th} \,\middle|\, \text{no signal present}\right\} \tag{8.52}$$

When the radiometer declares that there is a signal present when the signal is there, detection has occurred. The probability of detection is thus

$$P_d = \{z \geq \gamma_{th} | \text{signal present}\} \tag{8.53}$$

It is well known that under the assumptions above, test statistic $z(t)$ has a central chi-square distribution with $2TW_F$ degrees of freedom with noise only and noncentral chi-square distribution with $2TW_F$ degrees of freedom and noncentrality parameter $\lambda = 2E_s/N_0$, where E_s is the energy in the signal [6, 7]. The probability of false alarm is given by [7]

$$P_{fa} = \int_{\gamma_{th}/2}^{\infty} \frac{x^{TW_F-1}e^{-x}}{\Gamma(TW_F)} dx \tag{8.54}$$

where $\Gamma(\ )$ is the gamma function. The probability of detection is given by [7]

$$P_d = \int_{\gamma_{th}/2}^{\infty} \left(\frac{2x}{\lambda}\right)^{\frac{TW_F-1}{2}} \exp\left(-x+\frac{\lambda}{2}\right) I_{TW_F-1}\left(2\sqrt{\frac{x\lambda}{2}}\right) dx \tag{8.55}$$

where $I_a(\ )$ is the modified Bessel function of the first kind and order a. Expression (8.55) reduces to the generalized Q-function [8] as

$$P_d = Q_{TW_F}\left(\sqrt{\lambda}, \sqrt{\gamma_{th}}\right) \tag{8.56}$$

Expressions (8.54) and (8.56) are plotted in Figure 8.4 against each other as TW_F is varied between 2 and 100. For values of γ_{th} in the range of 10 to 50, these curves do not vary much.

8.2.3 Radiometer Models

Mills and Prescott [9] compared five different mathematical models of radiometers. The first was developed by Edell, and is described in Simon et al. [2], where the Gaussian approximation to the chi-square distribution was used. The probability of false alarm is
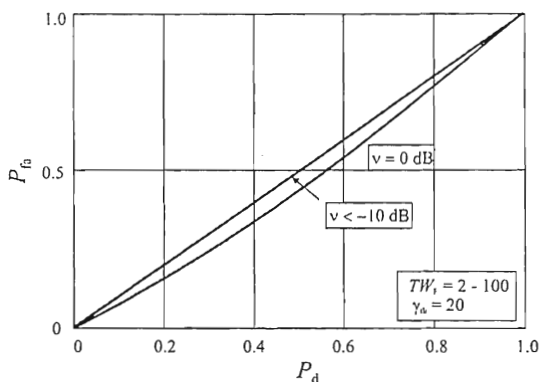
Figure 8.4 Radiometer performance.

$$P_{fa} = Q\left(\frac{\gamma_{th} - \mu_n}{\sigma_n}\right) \tag{8.57}$$

and the probability of detection is

$$P_d = Q\left(\frac{\gamma_{th} - \mu_{sn}}{\sigma_{sn}}\right) \tag{8.58}$$

where μ_n and σ_n are the mean and standard deviation when there is no signal present given by

$$\begin{aligned} \mu_n &= 2TW_F \\ \sigma_n^2 &= 4TW_F \end{aligned} \tag{8.59}$$

and μ_{sn} and σ_{sn} are the corresponding quantities with the signal present given by

$$\begin{aligned} \mu_{sn} &= 2TW_F + 2\frac{E_s}{N_0} \\ \sigma_{sn}^2 &= 4TW_F + 8\frac{E_s}{N_0} \end{aligned} \tag{8.60}$$

Since $Q(\ )$ is a monotonically increasing function of its arguments, it is invertible. Assuming that P_s/N_0W_F, which is the true signal power to noise power SNR, is low—low enough so that it makes no difference whether the signal is present or not—then $\sigma_n = \sigma_{sn}$, and a factor d can be determined by solving (8.57) and (8.58) for γ_{th} and equating, yielding

$$d = \frac{u_{sn} - u_n}{\sigma_n} = \frac{1}{\sqrt{TW_F}}\left(\frac{E_s}{N_0}\right)_{req} = \sqrt{\frac{T}{W_F}}\left(\frac{P_s}{N_0}\right)_{req} = Q^{-1}(P_{fa}) - Q^{-1}(P_d) \qquad (8.61)$$

Parameter d^2 is a measure of the detector output SNR. The power in the signal is given by $S = E_s/T$. This yields the required SNR[1] to achieve the specified levels of P_{fa} and P_d given in Table 8.1. This expression does not apply when TW_F is small ($TW_F < 100$), however. In that case, a factor η is included, which is given by [9]

$$\eta = \frac{\left(P_s/N_0\right)_{req} \text{ assuming } \chi^2 \text{ statistics}}{\left(P_s/N_0\right)_{req} \text{ assuming Gaussian statistics}} \qquad (8.62)$$

This factor is also available from the chart shown in Figure 8.5 for $P_d = 0.9$ [4].

The second model considered was one developed by Torrieri [10]. It also used the Gaussian approximation to the χ-square distribution and therefore is only applicable for large TW products. The approximations are

$$P_{fa} = Q\left(\frac{\gamma'_{th} - N_0TW_F}{\sqrt{N_0^2TW_F}}\right) \qquad (8.63)$$

and

$$P_d = Q\left(\frac{\gamma'_{th} - N_0TW_F - E_s}{\sqrt{N_0^2TW_F + 2N_0E_s}}\right) \qquad (8.64)$$

where the modified threshold is

---

[1] Note that this is not the SNR defined previously, but actually P_s/N_0, which has the units of frequency because W/W/Hz = Hz.

Table 8.1 SNR$_{req}$ for Various Radiometer Models

EVGA (equal variance Gaussian assumption) Edell	$\left(\dfrac{P_s}{N_0}\right)_{req} = d\sqrt{\dfrac{W_F}{T}},\ TW_F > 100$ $\left(\dfrac{P_s}{N_0}\right)_{req} = \eta d\sqrt{\dfrac{W_F}{T}}$
Torrieri	$\left(\dfrac{P_s}{N_0}\right)_{req} = \sqrt{\dfrac{2W_F}{T}}\,(\beta-\varsigma)+\dfrac{1}{T}\Psi(\beta,\varsigma,TW_F),\quad TW_F > 100$ $\varsigma = \dfrac{Q^{-1}(P_d)}{\sqrt{2}}$ $\beta = \dfrac{Q^{-1}(P_{fa})}{\sqrt{2}}$ $\Psi(\beta,\varsigma,TW_F) = 2\varsigma^2 - \varsigma\sqrt{2TW_F}\left(\sqrt{1+\dfrac{2\varsigma^2}{TW_F}+\dfrac{8\beta}{\sqrt{2TW_F}}}-1\right)$
Engler Park	$\left(\dfrac{P_s}{N_0}\right)_{req} = \dfrac{d^2+\sqrt{d^4+16TW_Fd^2}}{4T}$ $\left(\dfrac{P_s}{N_0}\right)_{req} = \dfrac{d^2+\sqrt{d^4+18.4TW_Fd^2}}{4T}$
Dillard	$\left(\dfrac{P_s}{N_0}\right)_{req} = \dfrac{4.6W_F}{\sqrt{1+9.2TW_F\dfrac{\rho+2.3}{\rho^2}}-1}$ $\rho = \dfrac{1}{2}\left[\sqrt{-2\ln(P_{fa})}-Q^{-1}(P_d)\right]^2$

$d = Q^{-1}(P_{fa}) - Q^{-1}(P_d)$

η = correction factor obtained from curves
T = integration time (sec)
W_F = receiver bandwidth (Hz)
Source: [9].
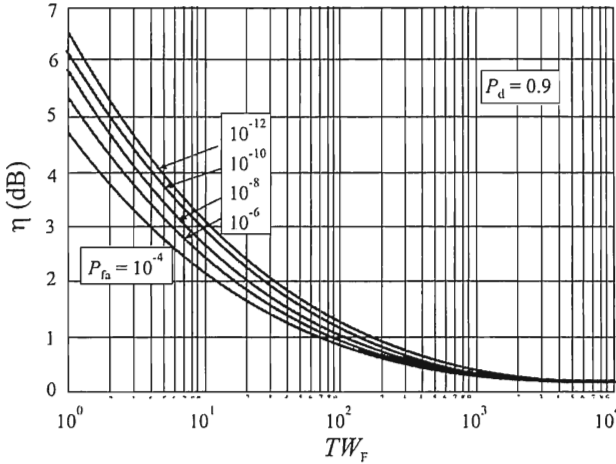
Figure 8.5 Woodring's adjustment factor η for small values of TW when $P_d = 0.9$. (*From:* [4].)

$$\gamma'_{th} = \frac{N_0 \gamma_{th}}{2} \tag{8.65}$$

Equation (8.64) is solved for E_s/N_0, again by inverting the error function. Equation (8.63) is solved for γ'_{th} / N_0, which is substituted into the solution for (8.64), using $E_s = P_s T$. Then letting $TW_F \to \infty$ yields the expression in Table 8.1.

The next model analyzed was due to Engler [11]. This model can be used even for small TW_F, which is sometimes the case for frequency-hopping LPI targets.[2] The derivation of this model was adopted from the theory of detection of radar pulses originally developed by Barton [12]. In this model an equivalent SNR is determined. A conversion is made between the SNR in a noncoherent receiver to an equivalent SNR in a coherent receiver with $TW_F = 1$ that yields the same detection performance. The coherent SNR is given by d^2, where d is given by (8.61). The resulting equation for the SNR required is given in Table 8.1.

Parks model is also usable for all values of TW_F. The basic model is given by

$$\nu_{req} = \varsigma d \sqrt{\frac{W_F}{T}} \tag{8.66}$$

---

[2] For example, if $W_F = 1$ kHz and hop rate = 1 khps ($T = 1$ ms), then $TW_F = 1$.

In this case, however, the multiplicative constant ς (as opposed to η) is given by

$$\varsigma = \frac{1}{4}\sqrt{\frac{d^2}{TW_{\mathrm{F}}}}\left(1 + \sqrt{1 + 18.4\frac{TW}{d^2}}\right) \tag{8.67}$$

The last model analyzed is due to Dillard [13], which was also based on the method in Barton as derived by Urkowitz [14]. In this model the loss incurred for noncoherent integration of n samples is given by

$$L_n = \frac{nD_n}{D_1} \tag{8.68}$$

where D_n is the required SNR for a single sample given the required P_d and P_{fa}. This is also given by

$$L_n = \frac{C_n}{C_1} = \frac{(D_n + 2.3)\big/D_n}{(D_1 + 2.3)\big/D_1} \tag{8.69}$$

where C_n is the detector loss when the SNR at the input is D_n. Equating (8.68) and (8.69), and letting $n = TW_{\mathrm{F}}$, $\rho = D_1$, and $D_{TW_{\mathrm{F}}} = P_s/N_0W_{\mathrm{F}}$, the expression in Table 8.1 ensues. The single sample SNR ρ is available from tables for single sample noncoherent radar detection curves or can be approximated by the expression given in Table 8.1 [15].[3]

Table 8.1 presents the required SNR in order to achieve the specified P_d and P_{fa}. In the first two instances, $TW_{\mathrm{F}} > 100$ is required in order to satisfy the requirement that the chi-square distribution with $2TW_{\mathrm{F}}$ degrees of freedom can be approximated by the Gaussian distribution due to the central limit theorem.

Comparison of the models is shown in Figure 8.6 and 8.7. Figure 8.6 compares the required SNR deviation from the exact SNR required. The exact P_{fa} is given by

$$P_{fa} = \int_{2\gamma_{th}/N_0}^{\infty} p_n(x)\,dx \tag{8.70}$$

---

[3] The values from the reference must be adjusted by 3 dB since the reference uses $2E_s/N_0$.
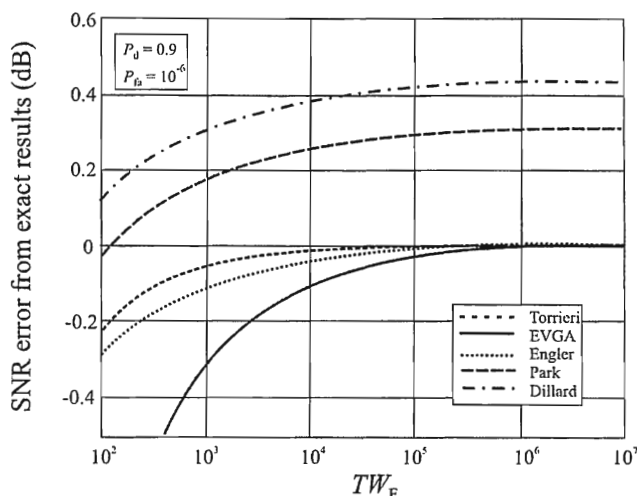
Figure 8.6 Radiometer model comparison. In this case the graphs show the deviation from the exact SNR required to achieve $P_d = 0.9$ and $P_{fa} = 10^{-8}$. (*From:* [9], © 1996 IEEE. Reprinted with permission.)
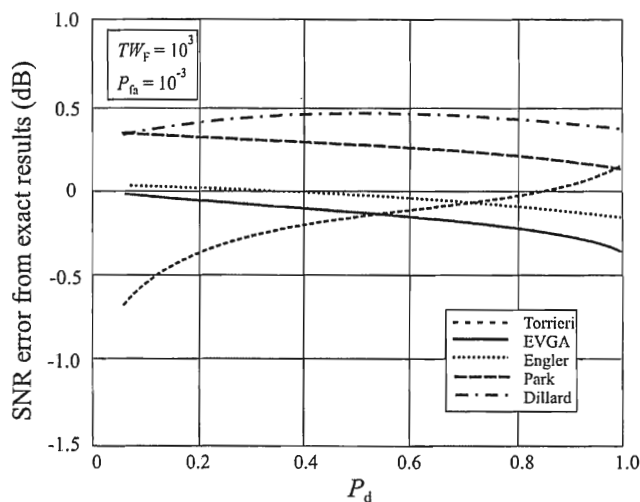

Figure 8.7 Radiometer model comparison. In this case the graphs show the deviation from the exact SNR required to achieve $P_{fa} = 10^{-3}$ with $TW_F = 10^3$. (*From:* [9], © 1996 IEEE. Reprinted with permission.)

where $p_n(x)$ is the probability distribution when there is no signal present given by [10]

$$p_n(x) = \frac{1}{2^{TW_F} \Gamma(TW_F)} x^{TW_F-1} \exp\left(-\frac{x}{2}\right) \qquad (8.71)$$

The exact P_d is given by

$$P_d = \int_{2\gamma_{th}/N_0}^{\infty} p_{sn}(x)dx \qquad (8.72)$$

where $p_{sn}(x)$ is the probability distribution with a signal present given by [10]

$$p_{sn}(x) = \frac{1}{2}\left(\frac{x}{\lambda}\right)^{\frac{TW_F-1}{2}} \exp\left(-\frac{x+\lambda}{2}\right) I_{TW_F-1}\left(\sqrt{x\lambda}\right) \qquad (8.73)$$

where $\lambda = 2E_s/N_0$ and $I_m(\ )$ is the modified Bessel function of the first kind and order m.

As the TW product increases, the Torrieri, EVGA, and Engler models all approach the exact required SNR. On the other hand, the Park and Dillard models remain biased away from the exact value.

As P_d increases, as illustrated in Figure 8.7, the required SNR deviation remains relatively constant. The Engler model varies the most.

8.2.4 Uncertain Noise Power

As mentioned, for a radiometer to function properly, a fairly accurate estimate of the amount of noise is required. Sonnenschein and Fishman conducted an analysis to determine what the impacts are of not having good knowledge of this noise [15]. They used the model developed by Edell discussed above.

The required SNR for a radiometer for some typical values of parameters is shown in Figure 8.8 with P_{fa} and $P_m = 1 - P_d$ as parameters, versus the time bandwidth product. This is the theoretical required SNR assuming perfect knowledge of the noise. This chart is correct assuming that $TW \gg 1$.

For typical signal detection systems in communication EW systems, in the low VHF range, where the channels are 25 kHz wide, the filter bandwidth W_F is normally set at 25 kHz. In the low UHF range it is typically 50 kHz, while in the HF range it is typically 3 kHz. Due to the requirement to detect modern LPI
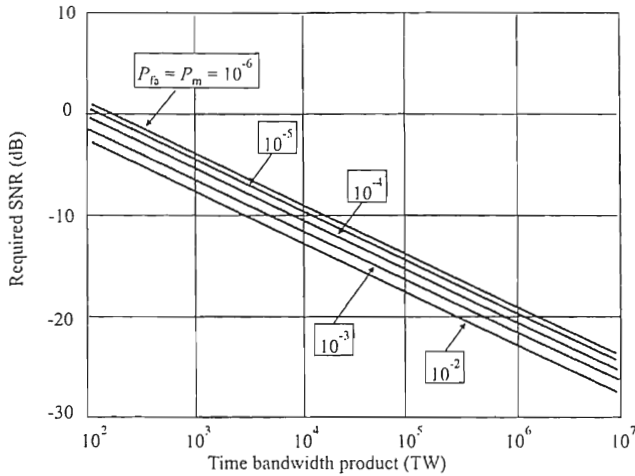
Figure 8.8 The required SNR for proper operation of a radiometer when the noise is perfectly known. (*From:* [15], © 1992 IEEE. Reprinted with permission.)

targets, the sample time T is usually between 1 and 10 ms in the low VHF and low UHF range. LPI targets in the HF range frequency hop at even slower rates. Thus, the value of TW for the usual targets of communication EW systems are as given in Table 8.2. It is best if $TW \geq 100$ or so. It can be argued that for some of these values, TW is not much greater than 1, so care must be used in using these results. In some cases, however, in particular for longer sample windows, this requirement is easily met.

For direct sequence spread spectrum (DSSS) targets, which can occur anywhere in the spectrum, but for practical reasons such as avoiding interference from TVs and radio stations, are only found in specific places, W does not correspond to the channels mentioned above. An example is the cellular international standard IS 95, which is a DSSS signal where W is 1.25 MHz. In the DSSS case, since W is so wide, T can be correspondingly smaller and the results

Table 8.2 Time-Bandwidth Products

Frequency Range	T (msec)	W (Hz)	TW
HF	100	300	30
HF	100	3,000	300
Low VHF	1	25,000	25
Low VHF	10	25,000	250
Low UHF	1	50,000	50
Low UHF	10	50,000	500
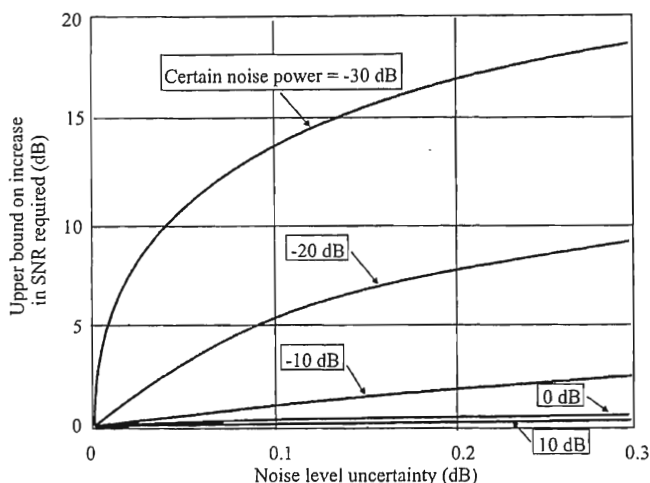
Figure 8.9 Upper bound on the required SNR when the noise level uncertainty is between 0 and 0.3 dB. (*From:* [15], © 1992 IEEE. Reprinted with permission.)

here apply. For 1.25 MHz, for example, T can be 80 μs and TW will still be 100.

The maximum increase in SNR required for specified detection when the uncertainty in the knowledge of the noise is limited to 0.3-dB is shown in Figure 8.9. Clearly, even small amounts of uncertainty in the noise level can cause large increases in the SNR necessary to achieve stated performance. This same information is illustrated in Figure 8.10 for larger values of the noise level uncertainty.

8.2.5 Local Oscillator Offset

Local oscillator frequency offset from the true frequency causes deterioration in the ability of a radiometer to detect energy. This deterioration shows up in an increase in the SNR required to achieve a given level of performance. The loss is given by

$$L = \left[\frac{\sin \pi \delta_f T}{\pi \delta_f T} \right]^2 \tag{8.74}$$

where T is the integration time. The loss is illustrated in Figure 8.11.
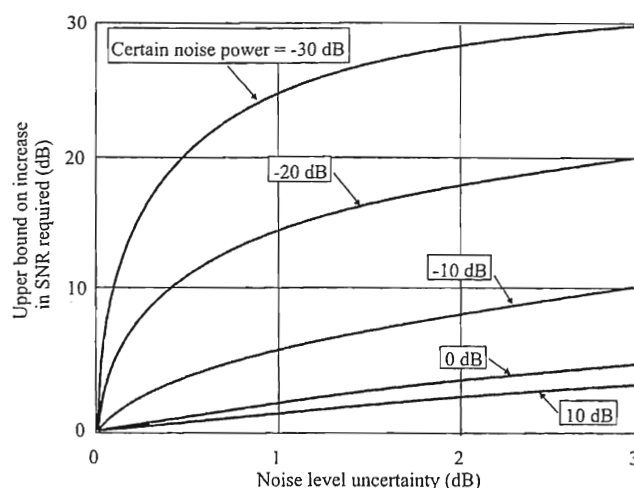
Figure 8.10 Increase in SNR required for optimum performance when the noise level uncertainty is between 0 and 3 dB. (*From:* [15], © 1992 IEEE. Reprinted with permission.)
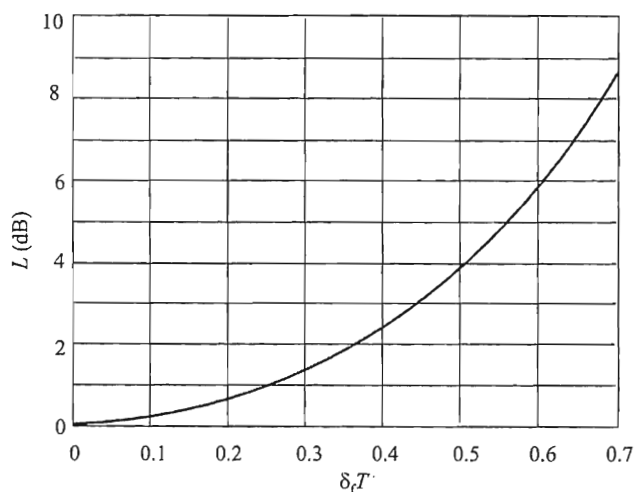
Figure 8.11 Frequency offset loss due to a mismatch in received signal frequency and local oscillator frequency.

8.3 Concluding Remarks

Reliably detecting totally random signals is more difficult than detecting deterministic signals. We must resort to GLRT testing or LOD testing since UMP tests are not defined when any of the parameters are random. A GLRT detector was discussed in this chapter along with its quadratic detector. LOD testing for random signals was also presented.

The radiometer is the device that is normally employed to detect signals when very little is known about their structure. A radiometer is an energy detector that integrates as long as necessary or as long as time available to detect signals. The performance of several models of radiometers was presented, with the results summarized in the form of the required SNR for specified P_{fa} and P_d.

Accurate use of a radiometer requires knowledge of the amount of noise present. If the noise level is not accurately known, then significant degradation in the radiometer performance occurs. Serious degradation also occurs when the local oscillator in the receiver is offset from the true value as well.

References

[1]   Gardner, W. A., *Introduction to Random Processes with Applications to Signals and Systems,* New York: Macmillan, 1986, pp. 274–278.

[2]   Simon, M. K., et al., *Spread Spectrum Communication Handbook*, New York: McGraw-Hill, 1994, Section 4.2.

[3]   Dillard, R. A., and G. M. Dillard, *Detectability of Spread Spectrum Signals,* Norwood, MA: Artech House, 1989, p. 58.

[4]   Woodring, D. G., *Performance of Optimum and Suboptimum Detectors for Spread Spectrum Waveforms,* Naval Research Laboratory, Washington, D.C., Technical Report No. 8432, December 1980.

[5]   Simon, M. K., et al., *Spread Spectrum Communication Handbook,* New York: McGraw-Hill, 1994, pp. 1048–1053.

[6]   Urkowitz, H., "Energy Detection of Unknown Deterministic Signals," *Proceedings of the IEEE,* Vol. 55, No. 4, April 1967, pp. 523–531.

[7]   Dillard, R. A., and G. M. Dillard, *Detectability of Spread Spectrum Signals*, Norwood, MA: Artech House, 1989, p. 58.

[8]   Poisel, R. A., *Modern Communications Jamming Principles and Techniques*, Norwood, MA: Artech House, 2003, Appendix A.

[9]   Mills, R. F., and G. E. Prescott, "A Comparison of Various Radiometer Detection Models," *IEEE Transactions on Aerospace and Electronic Systems,* Vol. 32, No. 1, January 1996, pp. 467–473.

[10]   Torrieri, D. J., *Principles of Secure Communication Systems, 2nd Ed.*, Norwood, MA: Artech House, 1992, pp. 294–302.

[11]   Engler, H. F., and D. H. Howard, *A Compendium of Analytic Models for Coherent and Non-Coherent Receivers,* Technical Report AFWAL-TR-85-1118, Air Force Wright Aeronautical Laboratory, September 1988.

[12]   Barton, D. K., "Simple Procedures for Radar Detection Calculations," *IEEE Transactions on Aerospace and Electronic Systems,* Vol. AES-5, September 1969, pp. 837–846.

[13]   Dillard, R. A., "Detectability of Spread Spectrum Signals," *IEEE Transactions on Aerospace and Electronic Systems,* Vol. AES-15, No. 3, July 1979, pp. 526–538.

[14]   Urkowitz, H., "Closed-Form Expressions for Noncoherent Radar Integration Gain and Collapsing Loss," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-9, September 1973, pp. 781–783.

[15]   Sonnenschein, A., and P. M. Fishman, "Radiometric Detection of Spread-Spectrum Signals in Noise of Uncertain Power," *IEEE Transactions on Aerospace and Electronic Systems,* Vol. 28, No. 3, July 1992, pp. 654–660.

Chapter 9

High-Resolution Spectrum Estimation

The method of spectral estimation using the periodogram was discussed in Chapter 4, which included classical power spectral estimation methods. Estimating the psd via the FT in the form of the periodogram is fast and efficient. There are occasions, however, when there are two or more signals present and they are close together; those techniques will generate a single peak for both signals. High resolution methods have been developed to address this issue.

There is a group of methods that are model-based. These techniques attempt to fit the time series data to models to compute the frequency spectrum directly, without use of the FT [1, 2]. In those techniques, a random process is assumed to be generated by a filter with an input that is noise-like. Characteristics of the process are derived based on determining parameters of the filter. Therefore, these methods are also referred to as parametric methods. The common parametric methods of spectral estimation are described first in this chapter.

There is another group of algorithms for estimating the spectrum of a stochastic process that is based on the eigen-analysis of the autocorrelation matrix of the samples from the process. Perhaps the most well known of these is the MUSIC algorithm discovered by Schmidt [3]. Some of these algorithms are presented next in this chapter.

The last group discussed in this chapter is based on the maximum likelihood technique, as analyzed by Capon [4].

High-resolution spectral estimation is applicable to both GS and DS. Its utility shows up when the spectrum is crowded with many RF signals.

9.1 Autoregressive Moving Average Modeling

Most modern signal processing is done digitally whenever possible. There are several reasons of this. First, digital signal processing is flexible. It can be

changed at any time frequently by simply reprogramming general-purpose digita signal processing computers. Second, the characteristics of digital signa processing are stable as system parameters vary. These parameters consist of time environmental parameters such as temperature and humidity, physical parameters such as shock and vibration, and so forth. This is frequently not true for analog electronics.

If the predetected signal from a frequency translator (receiver) is digitized, the resultant samples form a representation of the received signal, probably corrupted by noise and interference. As long as the sample rate is more than twice the highest frequency component of the received signal, and the dynamic range of the digital processing is not exceeded, the original signal can be recovered perfectly by converting it back into analog form. This is known as the Nyquist criterion This data stream represents samples of random processes. At any sampling instant the sample can take on any value as dictated by the pdf associated with the random process. If this density function is $\mathcal{N}(0, \sigma^2)$, for example, then the average value of the sample is zero, and with 62.4% assurance the sample will be within 1σ of the mean value.

Random data streams can be modeled as autoregressive processes. Consider the system shown in Figure 9.1, which has the transfer function $H(z)$, processes an incoming digital data stream represented by x_n, and produces an output y_n, which is an estimate of a data stream represented by d_n.

The output is related to the input and the impulse response of the system by

$$y_n = \sum_{i=0}^{N-1} h_i x_{n-i}$$

$$= \mathbf{h}^T \mathbf{x}$$

(9.1)

where $\mathbf{h} = [h_0 \ h_1 \ \dots \ h_{N-1}]^T$ is a vector of samples of the unit pulse response and $\mathbf{x}$ is a vector consisting of the last $N-1$ inputs, $\mathbf{x} = [x_n \ x_{n-1} \ \dots \ x_{n-N+1}]^T$.

The power in the output signal is given by
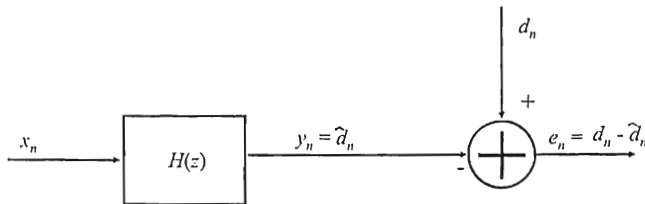
Figure 9.1 Autoregressive process model.

$$P_y = \mathcal{E}\left\{y_n^2\right\} \tag{9.2}$$

Now

$$y_n^2 = \mathbf{y}\mathbf{y}^\dagger = \left(\mathbf{h}^\mathsf{T}\mathbf{x}\right)\left(\mathbf{h}^\mathsf{T}\mathbf{x}\right)^\dagger = \mathbf{h}^\mathsf{T}\mathbf{x}\mathbf{x}^\dagger\mathbf{h} \tag{9.3}$$

when the h_i are real, where $(\ )^\dagger$ denotes conjugate transposition. Then

$$P_y = \mathcal{E}\left\{y_n^2\right\} = \mathcal{E}\left\{\mathbf{h}^\mathsf{T}\mathbf{x}\mathbf{x}^\dagger\mathbf{h}\right\} = \mathbf{h}^\mathsf{T}\mathcal{E}\left\{\mathbf{x}\mathbf{x}^\dagger\right\}\mathbf{h}$$

$$= \mathbf{h}^\mathsf{T}\mathcal{E}\left\{\begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{bmatrix} \begin{bmatrix} x_0^* & x_1^* & \cdots & x_{N-1}^* \end{bmatrix}\right\}\mathbf{h}$$

$$= \mathbf{h}^\mathsf{T}\mathcal{E}\left\{\begin{bmatrix} x_0 x_0^* & x_0 x_1^* & \cdots & x_0 x_{N-1}^* \\ x_1 x_0^* & x_1 x_1^* & \cdots & x_1 x_{N-1}^* \\ \vdots & \vdots & \ddots & \vdots \\ x_{N-1} x_0^* & x_{N-1} x_1^* & \cdots & x_{N-1} x_{N-1}^* \end{bmatrix}\right\}\mathbf{h}$$

$$= \mathbf{h}^\mathsf{T}\begin{bmatrix} \mathcal{E}\left\{x_0 x_0^*\right\} & \mathcal{E}\left\{x_0 x_1^*\right\} & \cdots & \mathcal{E}\left\{x_0 x_{N-1}^*\right\} \\ \mathcal{E}\left\{x_1 x_0^*\right\} & \mathcal{E}\left\{x_1 x_1^*\right\} & \cdots & \mathcal{E}\left\{x_1 x_{N-1}^*\right\} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{E}\left\{x_{N-1} x_0^*\right\} & \mathcal{E}\left\{x_{N-1} x_1^*\right\} & \cdots & \mathcal{E}\left\{x_{N-1} x_{N-1}^*\right\} \end{bmatrix}\mathbf{h} \tag{9.4}$$

$$= \mathbf{h}^\mathsf{T}\mathbf{\Gamma}_{xx}\mathbf{h} \tag{9.5}$$

where $\mathbf{\Gamma}_{xx}$ is the autocorrelation matrix of $\mathbf{x}$ with entries given by

$$\gamma_{ij} = \mathcal{E}\left\{x_i x_j\right\} \tag{9.6}$$

Suppose

$$1 = \mathbf{h}^\mathsf{T}\mathbf{h} = \sum_{k=0}^{N-1}\left|h_k\right|^2 \tag{9.7}$$

Lagrange multipliers with (9.5) and (9.7) produce

$$L = \frac{1}{2}\mathbf{h}^{\mathrm{T}}\mathbf{\Gamma}_{xx}\mathbf{h} + \lambda\left(1 - \mathbf{h}^{\mathrm{T}}\mathbf{h}\right) \tag{9.8}$$

Now take the gradient with respect to $\mathbf{h}$ of L, $\nabla_{\mathbf{h}}L$, and set it to zero to determine the Lagrange multipliers. It can be shown that if $\mathbf{a}$ and $\mathbf{b}$ are two $n \times 1$ nonzero vectors and $\mathbf{A}$ is an $n \times n$ symmetric matrix, then

$$\frac{d}{d\mathbf{b}}\mathbf{a}^{\mathrm{T}}\mathbf{b} = \mathbf{a} \tag{9.9}$$

and

$$\frac{d}{d\mathbf{a}}\mathbf{a}^{\mathrm{T}}\mathbf{A}\mathbf{a} = 2\mathbf{a} \tag{9.10}$$

Applying (9.9) and (9.10) to the gradient yields

$$\nabla_{\mathbf{h}}L = \mathbf{\Gamma}_{xx}\mathbf{h} - \lambda\mathbf{h} = 0 \tag{9.11}$$

Thus,

$$\mathbf{\Gamma}_{xx}\mathbf{h} = \lambda\mathbf{h} \tag{9.12}$$

which is the eigenvalue decomposition of $\mathbf{\Gamma}_{xx}$. Using (9.7),

$$\mathbf{h}^{\mathrm{T}}\mathbf{\Gamma}_{xx}\mathbf{h} = \lambda\mathbf{h}^{\mathrm{T}}\mathbf{h} = \lambda \tag{9.13}$$

The unconstrained optimization of (9.8) leads to $\lambda = \lambda_{min}$ by minimizing (9.13), while $\lambda = \lambda_{max}$ is obtained by maximizing (9.13).

In general, an ARMA model assumes that a time series can be modeled as the output of a pole and zero filter excited by AWGN as

$$y_n = \sum_{k=1}^{P} a_k y_{n-k} + \sum_{k=0}^{Q} b_k x_{n-k} \tag{9.14}$$

where $x_i \sim \mathcal{N}(0, \sigma^2)$. The transfer function for this filter is given by

$$H(z) = \frac{\sum\limits_{k=0}^{Q} b_k z^{-k}}{1 + \sum\limits_{k=1}^{P} a_k z^{-k}} \triangleq \frac{B(z)}{A(z)} \tag{9.15}$$

9.1.1 Moving Average Modeling

An MA process of order Q is defined by $A(z) = 1$, so that

$$H(z) = B(z) \tag{9.16}$$

and the output is a function of the last Q inputs,

$$y_n = b_0 x_n + b_1 x_{n-1} + \cdots b_Q x_{n-Q} \tag{9.17}$$

By the Weiner-Khinchin theorem (2.94), the psd of y_n is given by

$$\Phi_{yy}\left(e^{j\omega}\right) = \left|Y\left(e^{j\omega}\right)\right|^2 = \sum_{m=-\infty}^{\infty} \phi_{yy}(m) e^{-jm\omega}, \qquad |\omega| < \pi \tag{9.18}$$

Therefore, the psd for an MA process is given by

$$\Phi_{yy}\left(e^{j\omega}\right) = \delta_t \left|\sum_{k=0}^{Q} b_k e^{-jk\omega}\right|^2 \tag{9.19}$$

which has $2Q$ zeros. δ_t is the sampling interval.

Substituting $z = \exp(j\omega)$ into (9.19) yields

$$\Phi_{yy}(z) = \delta_t B(z) B(z^{-1}) \tag{9.20}$$

If $b_i = 1/Q \ \forall i$, then y_n is the average of the last Q input values, hence the name moving average. A block diagram of the technique for generating an MA process is shown in Figure 9.2.
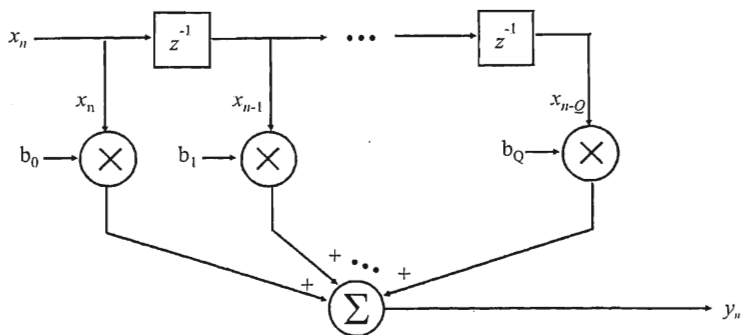
Figure 9.2 Moving average generator block diagram.

9.1.2 Autoregressive Modeling

An AR process is defined by $B(z) = 1$. For an AR process the current output is a function of the current input and last P outputs,

$$y_n = x_n - a_1 y_{n-1} - a_2 y_{n-2} - \cdots - a_P y_{n-P} \qquad (9.21)$$

The psd of the output is given by

$$\Phi_{yy}\left(e^{j\omega}\right) = \delta_t \left| \frac{1}{1 - a_1 e^{-j\omega} - a_2 e^{-j2\omega} - \cdots - a_P e^{-jP\omega}} \right|^2 \qquad (9.22)$$

while

$$\Phi_{yy}\left(z\right) = \delta_t \frac{1}{A(z)A(z^{-1})} \qquad (9.23)$$

which has $2P$ poles.

When noise is included in the process input, (9.21) becomes

$$y_n = x_n + \sum_{k=1}^{P} a_k y_{n-k} + \varepsilon_n \qquad (9.24)$$

The ε_n are the uncorrelated error terms due to noise associated with each sample and are referred to as the *innovations* of the process. These innovations are

assumed independent of prior data samples as well as prior innovations. They are normally assumed to be i.i.d. and Gaussian $\varepsilon_n \sim \mathcal{N}(0, \sigma^2)$. The psd then becomes

$$\Phi_{yy}\left(e^{j\omega}\right) = \delta_t \sigma^2 \left| \frac{1}{1 - a_1 e^{-j\omega} - a_2 e^{-j2\omega} - \cdots - a_p e^{-jP\omega}} \right|^2 \tag{9.25}$$

For most cases of practical interest, the autocorrelation lags will not be known exactly and therefore must be estimated. One estimate for these is given by

$$\hat{\gamma}_{xx,m} = \frac{1}{N-m} \sum_{i=0}^{N-m-1} x_{m+i} x_i^* \tag{9.26}$$

which is an unbiased estimator. Sometimes the biased estimator

$$\hat{\gamma}_{xx,m} = \frac{1}{N} \sum_{i=0}^{N-m-1} x_{m+i} x_i^* \tag{9.27}$$

is used because it frequently has less mean squared error than (9.26).

An autoregressive model assumes that the current output is determined by (possibly) the current input and P past values of the output. The model is therefore sometimes called the linear prediction model because the present value is a prediction based on prior outputs. A block diagram of an AR process generator is shown in Figure 9.3.

9.1.3 ARMA Modeling

The psd of an ARMA process has $2Q$ zeros as well as $2P$ poles. The current output is a function of the current and last Q inputs and last P outputs, so that (9.14)
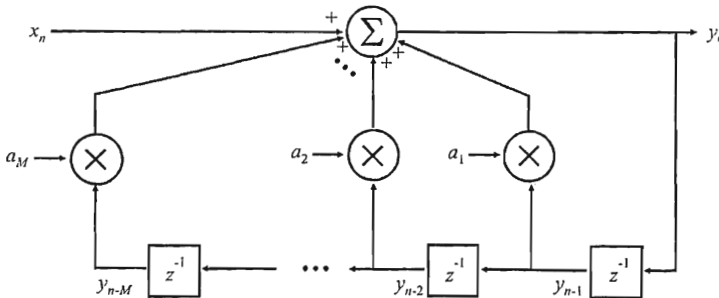
Figure 9.3 Autoregressive process generator.

holds. The psd in an ARMA process is given by

$$\Phi_{yy}\left(e^{j\omega}\right) = \sigma^2\delta_t \left|\frac{b_0 + b_1 e^{-j\omega} + b_2 e^{-j2\omega} + \cdots b_Q e^{-jQ\omega}}{1 - a_1 e^{-j\omega} - a_2 e^{-j2\omega} - \cdots - a_P e^{-jP\omega}}\right|^2 \qquad (9.28)$$

and

$$\Phi_{yy}(z) = \sigma^2\delta_t \frac{B(z)B\left(z^{-1}\right)}{A(z)A\left(z^{-1}\right)} \qquad (9.29)$$

where δ_t is the sample time interval.

For the ARMA(P, Q) data model,

$$\sum_{k=0}^{P} a_k y_{n-k} = \sum_{k=0}^{Q} b_k x_{n-k}, \qquad a_0 = 1 \qquad (9.30)$$

and in vector form is

$$\mathbf{a}^T\mathbf{y} = \mathbf{b}^T\mathbf{x} \qquad (9.31)$$

A block diagram of an ARMA process generator is shown in Figure 9.4.

Multiplying both sides of (9.30) by y_{n-k} and computing the expected value yields

$$\gamma_{xx,k} + a_1\gamma_{xx,k-1} + \cdots a_P\gamma_{xx,k-P} =$$
$$b_0\gamma_{xy,k} + b_1\gamma_{xy,k-1} + \cdots b_Q\gamma_{xy,k-Q} \qquad (9.32)$$

The system transfer function is

$$H(z) = \sum_{k=0}^{\infty} h_k z^{-k} \qquad (9.33)$$

and in the time domain, the output is the convolution of the unit pulse response with the input:
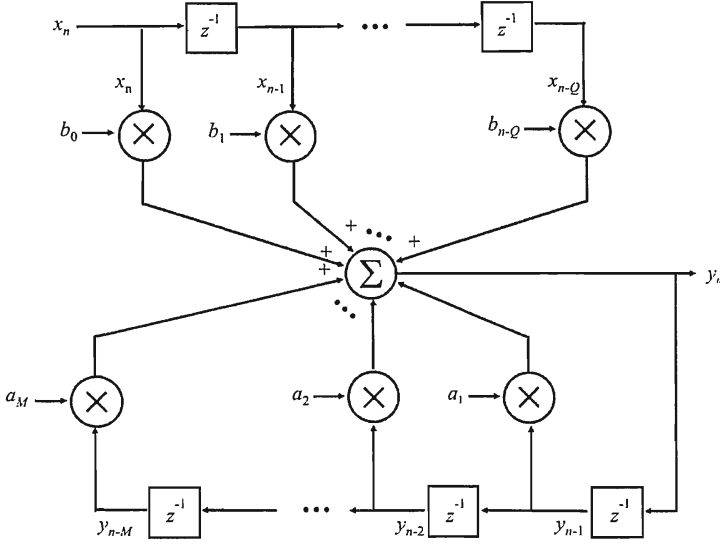
Figure 9.4 Autoregressive moving average process generator.

$$y_n = \sum_{k=0}^{N-1} h_k x_{n-k} = h_n * x_n \tag{9.34}$$

Multiplying (9.34) on both sides by x_{n-k} and calculating the expectations yields

$$\gamma_{yx,k} = \sum_{i=0}^{N-1} h_i \gamma_{xx,k-i} = h_k * \gamma_{xx,k} = \sigma_{xx}^2 h_k \tag{9.35}$$

When $\sigma_{xx}^2 = 1$,

$$\gamma_{xx,k} + a_1 \gamma_{xx,k-1} + \cdots + a_P \gamma_{xx,k-P}$$
$$= b_0 h_{-k}^* + b_1 h_{1-k}^* + \cdots + b_Q h_{Q-k}^* \tag{9.36}$$

Expressions (9.36) are known as the *Yule-Walker equations*. Put into matrix form, these equations become

$$\begin{bmatrix} \gamma_{xx} & \gamma_{xx,-1} & \cdots & \gamma_{xx,-P} \\ \gamma_{xx,1} & \gamma_{xx} & \cdots & \gamma_{xx,1-P} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{xx,Q} & \gamma_{xx,Q-1} & \cdots & \gamma_{xx,Q-P} \\ \gamma_{xx,Q+1} & \gamma_{xx,Q} & \cdots & \gamma_{xx,Q-P+1} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{xx,Q+P} & \gamma_{xx,Q+P-1} & \cdots & \gamma_{xx,Q} \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_P \end{bmatrix} = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_Q \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{9.37}$$

where

$$c_k = b_k * h^*_{-k} \tag{9.38}$$

The right side of (9.37) has zeros at the bottom because for a causal system, $c_k = 0$ for $k > Q$. In block matrix form, (9.37) can be written as

$$\begin{bmatrix} \Gamma_B \\ \Gamma_A \end{bmatrix} \mathbf{a} = \begin{bmatrix} \mathbf{c} \\ \mathbf{0} \end{bmatrix} \tag{9.39}$$

corresponding to the partitions

$$\begin{bmatrix} \gamma_{xx} & \gamma_{xx,-1} & \cdots & \gamma_{xx,-P} \\ \gamma_{xx,1} & \gamma_{xx} & \cdots & \gamma_{xx,1-P} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{xx,Q} & \gamma_{xx,Q-1} & \cdots & \gamma_{xx,Q-P} \\ \hline \gamma_{xx,Q+1} & \gamma_{xx,Q} & \cdots & \gamma_{xx,Q-P+1} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{xx,Q+P} & \gamma_{xx,Q+P-1} & \cdots & \gamma_{xx,Q} \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_P \end{bmatrix} = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_Q \\ \hline 0 \\ \vdots \\ 0 \end{bmatrix} \tag{9.40}$$

Note that for an AR model (when $Q = 0$), (9.37) becomes

$$\begin{bmatrix} \gamma_{xx} & \gamma_{xx,-1} & \cdots & \gamma_{xx,-P} \\ \gamma_{xx,1} & \gamma_{xx} & \cdots & \gamma_{xx,-P+1} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{xx,P} & \gamma_{xx,P-1} & \cdots & \gamma_{xx} \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_P \end{bmatrix} = \begin{bmatrix} |b_0|^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{9.41}$$

This equation for an MA model ($P = 0$) is

$$\begin{bmatrix} \gamma_{xx} \\ \gamma_{xx,1} \\ \vdots \\ \gamma_{xx,Q} \end{bmatrix} = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_Q \end{bmatrix} \tag{9.42}$$

where

$$c_k = b_k * b_{-k}^* \tag{9.43}$$

The $*$ symbol between the factors in (9.43) indicates convolution, which means that the coefficients with offsets of k are multiplied and summed.

The solution to these equations is found as follows. To find the AR parameters, solve the linear equations from (9.39):

$$\mathbf{\Gamma}_A \mathbf{a} = \mathbf{0} \tag{9.44}$$

To find the MA parameters, use the other part of (9.39) as

$$\mathbf{\Gamma}_B \mathbf{a} = \mathbf{c} \tag{9.45}$$

with

$$c_k = b_k * h_{-k}^* \tag{9.46}$$

Thus,

$$a_k * \gamma_{xx,k} = c_k = b_k * h_{-k}^* \tag{9.47}$$

Now take the z-transform
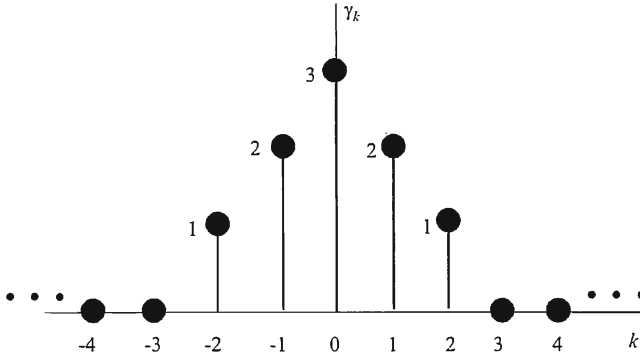
Figure 9.5 Autocorrelation function for the MA and ARMA examples.

$$A(z)Sx(z) = C(z) = B(z)H^*(1/z^*) = B(z)\frac{B^*(1/z^*)}{A^*(1/z^*)} \tag{9.48}$$

To solve this equation the following must be factored

$$S_y(z) = B(z)B^*(1/z^*) \tag{9.49}$$

where

$$S_y(z) \triangleq A(z)S_x(z)A^*(1/z^*) = \underbrace{C(z)A^*(1/z^*)}_{\text{known}} \tag{9.50}$$

Example: (MA)

Suppose that a real random process with the autocorrelation function shown in Figure 9.5 is to be fitted to a second-order ($Q = 2$) MA model. The Yule-Walker equations are of the form $\gamma_{xx,k} = b_k * b^*_{-k}$ so

$$\begin{aligned}
\gamma_0 &= b_0^2 + b_1^2 + b_2^2 = 3 \\
\gamma_1 &= b_1 b_0 + b_2 b_1 = 2 \\
\gamma_2 &= b_2 b_0 = 1
\end{aligned} \tag{9.51}$$

In general, these equations are difficult to solve because they are nonlinear. By inspection, however, $b_0 = b_1 = b_2 = 1$ is one solution.

The spectral factorization approach is one way of solving (9.51). First the complex spectral density function is computed as

$$S_x(z) = \sum_{k=-Q}^{Q} \gamma_{xx,k} z^{-k} = z^2 + 2z + 3 + 2z^{-1} + z^{-2} \tag{9.52}$$

which can be factored as

$$S_x(z) = (1 + z^{-1} + z^{-2}) \underbrace{(1 + z + z^2)}_{B(z)} \tag{9.53}$$

which shows again that $b_0 = b_1 = b_2 = 1$. Normally such a factorization is not as obvious as in this case and the roots must be numerically determined.

For this example there are double roots at each root location on the unit circle so that

$$S_x(z) = (z^{-1} - e^{j2\pi/3})(z^{-1} - e^{-j2\pi/3})(z - e^{j2\pi/3})(z - e^{-j2\pi/3}) \tag{9.54}$$

$B(z)$ is given by the product of the second two factors. The roots are shown in Figure 9.6.

∎

Example: (ARMA)

Suppose the ARMA model is first order with

$$H(z) = \frac{b_0 + b_1 z^{-1}}{1 + a_1 z^{-1}} \tag{9.55}$$

This model is to be fit to the correlation function shown in Figure 9.5. The Yule-Walker equations are of the form

$$\begin{bmatrix} 3 & 2 \\ 2 & 3 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \end{bmatrix} = \begin{bmatrix} c_0 \\ c_1 \\ 0 \end{bmatrix} \tag{9.56}$$
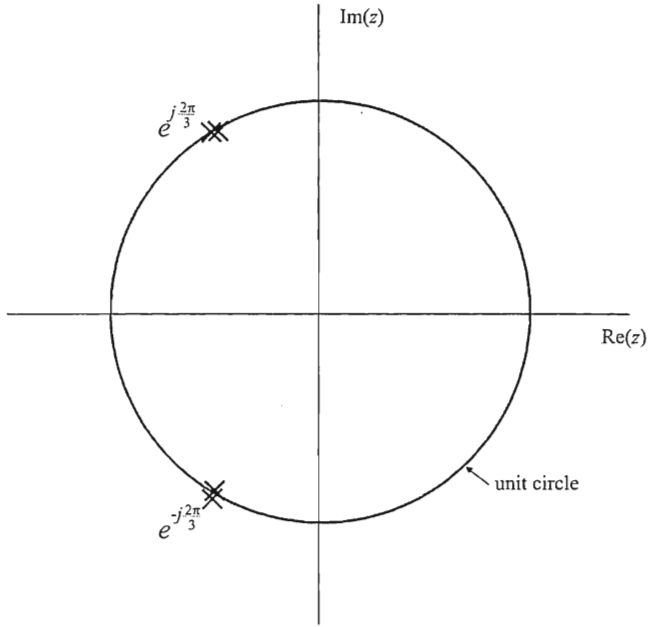
Figure 9.6 Roots for $S_x(z)$ for the MA example. There are two roots at each of $\exp(j2\pi/3)$ and $\exp(-j2\pi/3)$.

The lower part of (9.56) yields a single equation

$$1 + 2a_1 = 0 \qquad (9.57)$$

so that $a_1 = 1/2$. Substituting this into the upper part of (9.56) yields the two equations

$$\begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ -\dfrac{1}{2} \end{bmatrix} = \begin{bmatrix} 2 \\ \dfrac{1}{2} \end{bmatrix} \qquad (9.58)$$

The negative power terms in $S_y(z)$ are obtained from

$$S_y(z) = C(z)A^*(1/z^*) = \left(2 + \frac{1}{2}z^{-1}\right)\left(1 - \frac{1}{2}z\right) \qquad (9.59)$$

$$= \frac{7}{4} + \frac{1}{2}z^{-1} \tag{9.60}$$

Because of symmetry, the entire function must be of the form

$$S_y(z) = \frac{1}{2}z + \frac{7}{4} + \frac{1}{2}z^{-1} \tag{9.61}$$

The factors of (9.61) are

$$S_y(z) = (1.262 + 0.396z^{-1})(1.262 + 0.396z) \tag{9.62}$$

so that $b_0 = 1.262$ and $b_1 = 0.396$. Therefore,

$$H(z) = \frac{1.262 + 0.396z^{-1}}{1 - 0.5z^{-1}} \tag{9.63}$$

■

The Yule-Walker equations can be efficiently calculated recursively [5]. That process is simpler than solving the equations directly.

The Kolmogorov theorem states that any wss process can be modeled as an AR process of infinite order, denoted by AR(∞) [6]. In theory, the resolution of an AR spectral estimator can be improved by increasing the order of the model. In practice, however, as the model order increases, so does the number of peaks, irrespective of the spectral content of the time series.

9.1.4 Maximum Entropy Spectral Analysis

For Gaussian random processes and known (not estimated from the random process) autocorrelation lags, the AR method is equivalent to the one-dimensional *maximum entropy* (ME) technique analyzed by Burg [7]. The assumptions for the lag values outside the time window are different, however. For the AR method they are assumed to be zero, while for the ME technique they are selected such that the entropy, or randomness, of the process is maximized but the first p autocorrelation lags match that of the data. In practice, of course, the autocorrelation lags must be estimated from the data. In that case (practical cases), the maximum entropy is not really obtained due to the estimation of the autocorrelation coefficients.

9.1.4.1 Entropy

The occurrence of a random event provides a certain amount of information about the underlying random process. The occurrence of an improbable event provides much more information than if such an occurrence were highly likely. The amount of information, denoted by I_i, corresponding to the realization of an event $y_i \in Y$, where Y is an rv, should therefore be a decreasing function of its probability $\Pr(y_i)$. In fact, the formal definition of information was provided by Shannon [8] as

$$I_i = \log_2 \frac{1}{\Pr(y_i)} \text{ bits} \tag{9.64}$$

(Of course, different bases for the logarithm can be used that yield different units on I_i. Here we will use base 2.)

Over time interval T, y_i will occur, on average, $T\Pr(y_i)$ times. The total information generated during this interval is then

$$I = \sum_{i=1}^{N} \Pr(y_i) \log_2 \frac{1}{\Pr(y_i)} \tag{9.65}$$

The *entropy*, H, of a discrete rv Y is defined by [9]

$$H(Y) = -\sum_{y \in Y} \Pr(y) \log_2 \Pr(y) \tag{9.66}$$

The entropy rate of stochastic process $Y = \{y_1, y_2, \ldots, y_N\}$ is defined by [10]

$$h(Y) = \lim_{N \to \infty} H(y_1, y_2, \cdots, y_N) \tag{9.67}$$

In the case of a Gaussian process, the entropy rate is proportional to

$$h \propto \int_{-1/2}^{1/2} \log_2 \hat{\Phi}_{ME}(f) df \tag{9.68}$$

where $\hat{\Phi}_{ME}(f)$ is the psd estimated by the maximum entropy method, and f is the frequency.

Maximum entropy is an AR process in that it employs parametric modeling with all poles. The time series being analyzed is assumed to be an AR process, and as such is generated with the structure shown in Figure 9.3, where $x_n = \varepsilon_n$ and $a_i < 0 \; \forall i$.

9.1.4.2 Spectral Estimate

The ME estimate is the psd, $\hat{\Phi}_{ME}(f)$, that maximizes h while being consistent with the known values of the autocorrelation coefficients. That is,

$$\gamma_{Y,k} = \int_{-f_N}^{f_N} \hat{\Phi}_{ME}(f)e^{-j2\pi f \delta_t} df, \qquad -K \le k \le K \qquad (9.69)$$

where $\gamma_{Y,k}$ are the $2K+1$ known autocorrelation lag values and δ_t is the autocorrelation lag interval. The solution can be found by using, for example, the Lagrange multipliers technique described next, with constraints provided by (9.69) that yield

$$\hat{\Phi}_{ME}(f) = \frac{\sigma^2}{\left|1 + \sum_{i=1}^{M} a_i e^{-j2\pi fi\delta_t}\right|^2} \qquad (9.70)$$

and

$$\sigma^2 = \mathcal{E}\{\varepsilon\varepsilon^*\} \qquad (9.71)$$

The psd obtained by using the maximum entropy method is equivalent to the psd of an AR process of order M. Different algorithms are available to estimate the parameters a_i, $i = 1,...,M$. The Yule-Walker technique [11] and the Burg method [7] are two. The Marple method provides a good estimator and is described below in Section 9.1.4.4.

The original maximum entropy technique, as discovered by Burg, has some problems associated with it. First of all, as previously mentioned, the exact results for the statistics for (9.70) are not available. Secondly, just as for the Yule-Walker method for estimating the AR coefficients, the Burg technique causes spectral line splitting, as discovered by Fourgere et al. [12]. Such line splitting is most likely to occur when: (1) the SNR is high; (2) the initial phase of sinusoidal components is an odd multiple of $\pi/4$ radians; (3) the time duration of the data sequence is such

that sinusoidal components have an odd number of quarter cycles; and (4) the number of AR parameters estimated is a large percentage of the number of data values used for the estimation [13]. The utility of the maximum entropy method is limited by its dependence on the SNR. The resolution decreases as the SNR decreases. Lastly, also shared with the Yule-Walker method, is the problem of bias in the estimation of the location of the peaks with respect to the actual location of those peaks. Considerable error is possible in this estimation.

9.1.4.3 The Lagrange Multipliers

The Lagrange multipliers $a_{M,i}$ correspond to the coefficients of an AR time series $Y(t)$ of order M according to

$$Y_{t+1} = \sum_{i=1}^{M} a_{M,i} Y_{t-M+i} + w \qquad (9.72)$$

where w is residual white noise. We saw in Section 9.1 that the knowledge of the $\{a_{M,i}\}$ coefficients directly gives an estimate of the power spectrum of the series. These coefficients can be computed by solving a set of linear equations whose coefficients form a symmetric Toeplitz matrix:

$$\begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_M \\ \gamma_1^* & \gamma_0 & \cdots & \gamma_{M-1} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_M^* & \gamma_{M-1}^* & \cdots & \gamma_0 \end{bmatrix} \begin{bmatrix} 1 \\ a_{M,0} \\ \vdots \\ a_{M,M} \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad (9.73)$$

where

$$\gamma_i = \mathcal{E}\{x_j x_{j+i}\} \qquad (9.74)$$

is the autocorrelation function at lag i and, for a stationary process, $\gamma_i^* = \gamma_{-i}$. Expression (9.73) is known as the Yule-Walker normal equation.

9.1.4.4 The Marple Algorithm

The Marple algorithm is based on a least-squares procedure [13]. The parameters of an AR process of order M can be estimated either with the linear forward or the

linear backwards procedures. These procedures have prediction errors $e_{M,i}^{(f)}$ and $e_{M,i}^{(b)}$ given by

$$e_{M,i}^{(f)} = \sum_{k=0}^{M} a_{M,k} y_{i+M-k} \qquad 1 \le i \le N-M \tag{9.75}$$

and

$$e_{M,i}^{(b)} = \sum_{k=0}^{M} a_{M,k}^{*} y_{i+k} \qquad 1 \le i \le N-M \tag{9.76}$$

where $a_{M,k}$ are the parameters of an AR process defined as

$$y_i = -\sum_{k=1}^{M} a_{M,k} y_{i-k} + \varepsilon_i \tag{9.77}$$

where $\varepsilon_i \sim \mathcal{N}(0, \sigma_M^2)$. Complex-valued data is assumed and $a_{M,0}$ is defined as unity. Since stationarity is assumed, the backward coefficients are simply the conjugate of the forward coefficients [13].

The least-squares estimate of the AR parameters is found by minimizing the sum of the backward and forward prediction error energies e_M defined as

$$e_M = \sum_{k=1}^{N-m} \left| e_{M,k}^{(f)} \right|^2 + \sum_{k=1}^{N-m} \left| e_{M,k}^{(b)} \right|^2 \tag{9.78}$$

The minimum of e_M is determined by setting its derivative with respect to all AR parameters to zero

$$\frac{\partial e_M}{\partial a_{M,j}} = 0 = 2 \sum_{k=0}^{M} a_{M,k} c_M(j,k) \tag{9.79}$$

where

$$c_M(j,k) = \sum_{i=1}^{N-M} (y_{i+M-k} y_{i+M-j}^{*} + y_{i+j}^{*} y_{i+k}) \qquad j,k = 0,1,\dots,M \tag{9.80}$$

The minimum prediction error energy is found to be

$$e_{M,\min} = \sum_{k=0}^{M} a_{M,k} c_M(0,k) \tag{9.81}$$

Expressions (9.79) and (9.81) can be combined in a $(M+1) \times (M+1)$ matrix

$$\mathbf{C}_M \mathbf{A}_M = \mathbf{E}_M \tag{9.82}$$

where

$$\mathbf{C}_M = \begin{bmatrix} c_M(0,0) & \cdots & c_M(0,M) \\ c_M(1,0) & \cdots & c_M(1,M) \\ \vdots & \ddots & \vdots \\ c_M(M,0) & \cdots & c_M(M,M) \end{bmatrix}$$

$$\mathbf{A}_M = \begin{bmatrix} 1 \\ a_{M,1} \\ \vdots \\ a_{M,M} \end{bmatrix}$$

$$\mathbf{E}_M = \begin{bmatrix} e_{M,\min} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$\mathbf{C}_M$ is not a Toeplitz matrix but it can be demonstrated that it is composed of the sum of two Toeplitz data matrix products allowing a Levinson-*like* recursive algorithm as follows.

$$\mathbf{C}_M = \mathbf{Y}'_M \mathbf{Y}_M + (\mathbf{Y}'_M)' \mathbf{Y}'_M \tag{9.83}$$

where

$$\mathbf{Y}_M = \begin{bmatrix} y_{M+1} & y_M & \cdots & y_1 \\ y_{M+2} & y_{M+1} & \cdots & y_2 \\ \vdots & \vdots & \ddots & \vdots \\ y_N & y_{N-1} & \cdots & y_{N-M} \end{bmatrix}$$

is a $(N - M) \times (M + 1)$ Toeplitz data matrix and $\mathbf{Y}_M'$ is the conjugate and reverse version of $\mathbf{Y}_M$ given by

$$\mathbf{Y}_M' = \begin{bmatrix} y_1^* & \cdots & y_M^* & y_{M+1}^* \\ y_2^* & \cdots & y_{M+1}^* & y_{M+2}^* \\ \vdots & \ddots & \vdots & \vdots \\ y_{N-M}^* & \cdots & y_{N-1}^* & y_N^* \end{bmatrix}$$

In order to exploit this *near* Toeplitz form of $\mathbf{C}_M$ and give a Levinson-*like* recursive algorithm for the AR parameters, it is necessary to introduce time-index-shifted variants of (9.78). Define the time-shifted prediction error energies e_M' and e_M'' as

$$e_M' = \sum_{k=1}^{N-m-1} \left| e_{M,k+1}^{(f)} \right|^2 + \sum_{k=1}^{N-m-1} \left| e_{M,k}^{(b)} \right|^2 \tag{9.84}$$

and

$$e_M'' = \sum_{k=1}^{N-m-1} \left| e_{M,k}^{(f)} \right|^2 + \sum_{k=1}^{N-m-1} \left| e_{M,k+1}^{(b)} \right|^2 \tag{9.85}$$

These can be minimized in a manner similar to that used for e_M, yielding

$$\mathbf{C}_M' \mathbf{A}_M' = \mathbf{E}_M' \tag{9.86}$$

and

$$\mathbf{C}_M'' \mathbf{A}_M'' = \mathbf{E}_M'' \tag{9.87}$$

This structure enables a recursive algorithm that must satisfy the AR parameters

$$a_{M,m} = a'_{M-1,m} + a_{M,M}(a'_{M-1,M-m})^*$$ (9.88)

$$e_{M,\min} = e_{M-1} + a_{M,M}\Delta_M^*$$ (9.89)

$$0 = \Delta_M + a_{M,M}e'_{M-1}$$ (9.90)

where

$$\Delta_M = [c_M(M,0) \quad c_M(M,1) \quad \cdots \quad c_M(M,M)]A_M^T$$

The least squares estimates of the AR parameters applied to the problem of spectral estimation leads to better performance than the Burg estimate, especially in the case of small amounts of data. This method does not produce spectral line splitting, presents less sensitivity to initial phase, and reduces bias in the frequency estimate.

9.1.5 Model Order Determination

Selecting the correct model order is critical for proper estimation by these techniques. This is illustrated with an example [14].

Example: (AR Model Order Determination)

Suppose the input to an AR estimator is the sum of three sinusoids given by

$$x_n = \cos(2\pi \times 0.21n + 0.1) + 2\cos(2\pi \times 0.36n) + 1.9\cos(2\pi \times 0.38n) \quad (9.91)$$

for $n = 0, 1, \ldots, 31$. The FT of this signal, after being zero-padded with 4,064 zeros, is shown in Figure 9.7. When an AR estimate of order $p = 14$ is made of (9.91), the chart in Figure 9.8 ensues. The signal at 0.21 is evident, but the signals at 0.36 and 0.38 have been combined into one. When the order is increased to $p = 20$, then the psd in Figure 9.9 results. In that case, all three signals are evident with no spurious peaks that might indicate other signals. Further increasing the order to $p = 30$ results in the psd shown in Figure 9.10. There are several spurious peaks in this case, and the signals at 0.36 and 0.38 are again combined in some fashion. This indicates that the order is set too high.

∎

Thus, the selection of the model order is an important issue. If the order is not large enough, the spectrum may be too smooth. On the other hand, if the order is too high, superfluous peaks can occur.
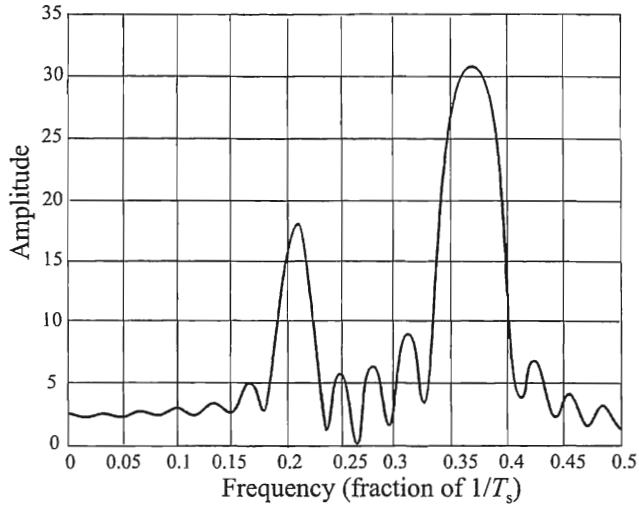
Figure 9.7 FT of expression (9.91). (*From:* [14], © 1995 Horizon House, Inc. Reprinted with permission.)
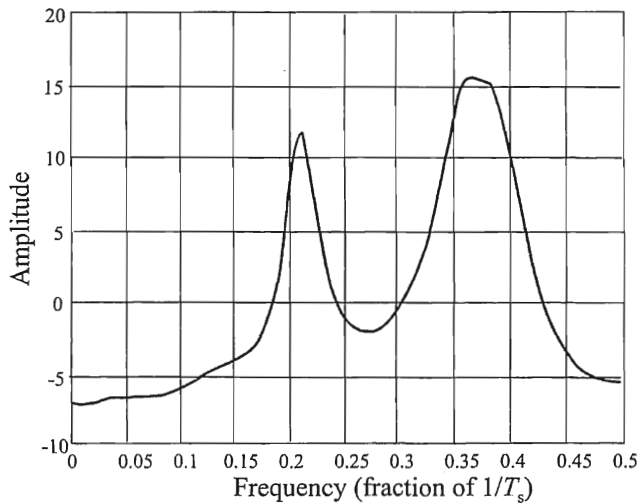
Figure 9.8 AR psd estimate of (9.91) when $p = 14$. (*From:* [14], © 1995 Horizon House, Inc. Reprinted with permission.)
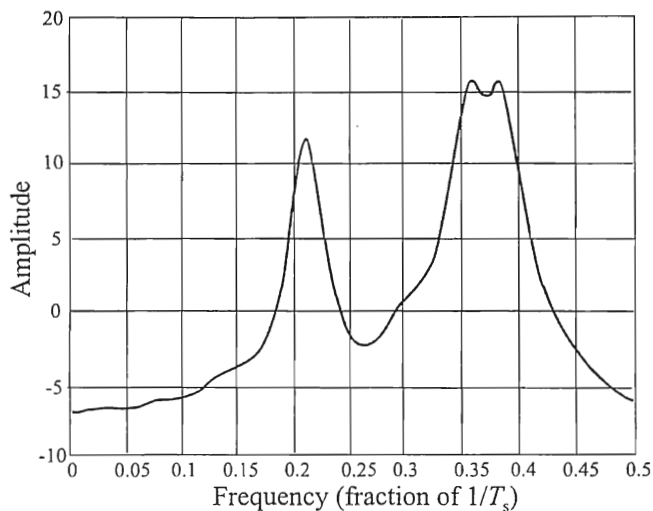
Figure 9.9 AR psd estimate of (9.91) when $p = 20$. (*From:* [14], © 1995 Horizon House, Inc. Reprinted with permission.)
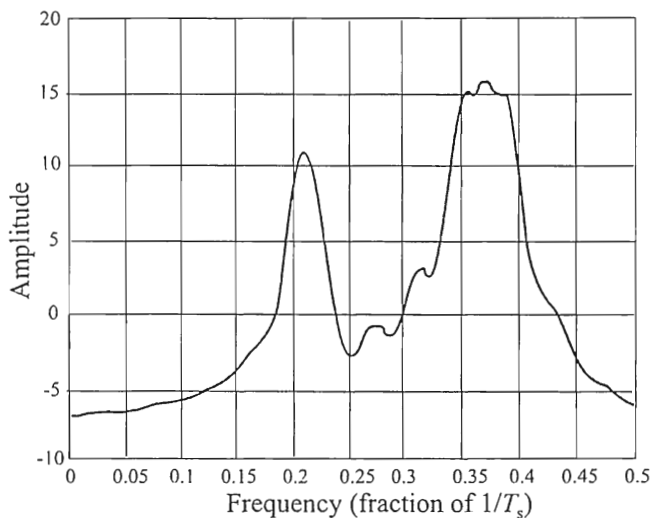
Figure 9.10 AR psd estimate of (9.91) when $p = 30$. (*From:* [14], © 1995 Horizon House, Inc. Reprinted with permission.)

There are several techniques for estimating the order of a process. Two of them will be discussed here, called the *final prediction error* (FPE) criterion and *Akaike information criterion* (AIC). Like the other minimization criteria, both of these information criteria require evaluating the criterion for the various candidate models and selecting the model that minimizes the criterion.

Assume that the sequence $\{x_1, x_2, \ldots\}$, contains samples from an AR process of order m. Then

$$x_k = \sum_{p=1}^{m} a_p^* x_{k-p} + \varepsilon_k \tag{9.92}$$

where $\mathbf{a} = [a_1, a_2, \ldots, a_m]^{\mathrm{T}}$ is the vector of AR coefficients, ε_k is the kth sample from an innovations process $\sim \mathcal{N}(0, \sigma_\varepsilon^2)$, and * denotes conjugation.

9.1.5.1 Final Prediction Error Criterion

Let $\iota_F(p)$ denote the value of the FPE criterion for a particular value of p given by

$$\iota_F(p) = \hat{\sigma}_\varepsilon^2 \frac{N+p+1}{N-p-1} \tag{9.93}$$

The rational term in this expression increases with p and constitutes a penalty term designed to prevent the use of too many terms. For this FPE criterion to be used, it is computed for every potential model of the process and the model with the minimum criterion is selected as the best one.

9.1.5.2 Akaike Information Criterion

Let $\iota_A(p)$ denote the value of the AIC. The second method uses the AIC that determines the order by minimizing

$$\iota_A(p) = \ln \hat{\sigma}_\varepsilon^2 + \frac{2(p+2)}{N-p-3} \tag{9.94}$$

Again, this criterion is computed for each potential model and the one with the minimum criterion is selected. The second term in (9.94) is also a penalty term that imposes a higher cost for larger values of p.
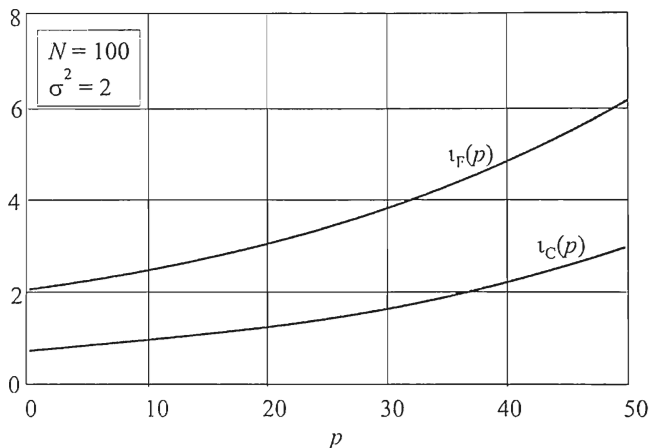
Figure 9.11 Comparison of the two model order determination factors for the example.

The model selected by the AIC criterion frequently has fewer components (smaller p) than the one selected with the FPE criterion. The reason for this is that the penalty term increases faster for the former, so the smaller criterion values typically occur for smaller values of p.

Example:

Suppose $N = 100$ and $\sigma^2 = 2$. These two model order determination values are plotted in Figure 9.11 versus p. Clearly, $\iota_F(p)$ is larger than $\iota_C(p)$ and increases more rapidly.

■

9.1.6 Resolution of AR Spectral Analysis

Zhang [15] developed a theory on the resolvability of the AR technique for spectral analysis. The sequence $\{x_1, x_2, \ldots\}$ is assumed to be samples from an AR process or order m, composed of two sinusoids in Gaussian noise. Then, based on (9.92), the AR spectrum can be expressed as

$$P_{AR}(f) = \frac{1}{\left(\mathbf{w}^\dagger \mathbf{a} - 1\right)^* \left(\mathbf{w}^\dagger \mathbf{a} - 1\right)} \tag{9.95}$$

where † denotes conjugate transpose. In (9.95), vector $\mathbf{w}$ is

$$\mathbf{w} = \begin{bmatrix} e^{-j\omega} \\ e^{-j2\omega} \\ \vdots \\ e^{-jm\omega} \end{bmatrix} \tag{9.96}$$

The null spectrum is given by

$$S(f) = \left(\mathbf{w}^\dagger \mathbf{a} - 1\right)^\cdot \left(\mathbf{w}^\dagger \mathbf{a} - 1\right) \tag{9.97}$$

The null spectrum at frequency k is denoted by S_k.

Two signals are *resolvable* if statistic γ is negative, where

$$\gamma = S_1 + S_2 - 2S_3 \tag{9.98}$$

and where f_3 is halfway between f_1 and f_2,

$$f_3 = \frac{f_1 + f_2}{2} \tag{9.99}$$

If $\gamma > 0$, the two signals are not resolvable.

Since the AR coefficients are estimated from the data, the null spectrum is not known exactly. Denote the estimate of S_k with $\hat{S}_k$. Then define statistic κ as

$$\kappa = \hat{S}_1 + \hat{S}_2 - 2\hat{S}_3 \tag{9.100}$$

The resolution issue can then be expressed as

$$\kappa \begin{cases} < 0, & \text{resolvable} \\ \geq 0, & \text{irresolvable} \end{cases} \tag{9.101}$$

Let the autocorrelation matrix corresponding to the received sequence be denoted by $\mathbf{\Gamma}_{xx}^{(m)}$, that is,

$$\mathbf{\Gamma}_{xx}^{(m)} = \mathcal{E}\left\{\mathbf{x}_{n+m}\mathbf{x}_n^\dagger\right\} \tag{9.102}$$

Note that the variance of the innovations process can be obtained from

$$\sigma_\varepsilon^2 = \frac{\det \Gamma_{xx}^{(m+1)}}{\det \Gamma_{xx}^{(m)}} \tag{9.103}$$

Let matrix Γ_a be defined as

$$\Gamma_a = \sigma_\varepsilon^2 \left[N\Gamma_{xx}^{(m)} \right]^{-1} \tag{9.104}$$

where N is the number of samples in the sequence.
    Define covariance matrix Γ as

$$\Gamma = \begin{bmatrix} \mathbf{w}_1^\dagger \Gamma_a \mathbf{w}_1 & \mathbf{w}_1^\dagger \Gamma_a \mathbf{w}_2 & \mathbf{w}_1^\dagger \Gamma_a \mathbf{w}_3 \\ \mathbf{w}_2^\dagger \Gamma_a \mathbf{w}_1 & \mathbf{w}_2^\dagger \Gamma_a \mathbf{w}_2 & \mathbf{w}_2^\dagger \Gamma_a \mathbf{w}_3 \\ \mathbf{w}_3^\dagger \Gamma_a \mathbf{w}_1 & \mathbf{w}_3^\dagger \Gamma_a \mathbf{w}_2 & \mathbf{w}_3^\dagger \Gamma_a \mathbf{w}_3 \end{bmatrix} \tag{9.105}$$

and a mean vector μ as

$$\mu = \begin{bmatrix} \mathbf{w}_1^\dagger \mathbf{a} - 1 \\ \mathbf{w}_2^\dagger \mathbf{a} - 1 \\ \mathbf{w}_3^\dagger \mathbf{a} - 1 \end{bmatrix} \tag{9.106}$$

The square root of Γ, denoted by $\mathbf{Q}$, is defined as

$$\Gamma = \mathbf{Q}\mathbf{Q}^\dagger \tag{9.107}$$

Let $\Omega = \mathbf{Q}^\dagger \mathbf{D} \mathbf{Q}$, where

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix} \tag{9.108}$$

and let
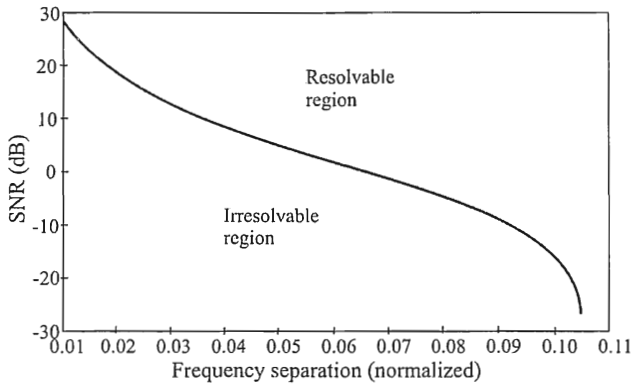
Figure 9.12 Resolvable and irresolvable regions defined by $\gamma = 0$. This is an AR(7) process. (*From:* [15], © 1998 IEEE. Reprinted with permission.)

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \tag{9.109}$$

be the matrix of eigenvalues of Ω.

Finally, using the above formulation, the probability of resolving two narrowband signals corrupted by Gaussian noise is given by

$$P_r = \Pr\{\kappa < 0\} = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\sin \alpha(z)}{z\beta(z)} dz \tag{9.110}$$

where

$$\alpha(z) = \sum_{k=1}^3 \left[\tan^{-1}(\lambda_k z) + \frac{|\mu_k|^2 z^2}{1 + \lambda_k^2 z^2} \right] \tag{9.111}$$

and

$$\beta(z) = \prod_{k=1}^3 \left(1 + \lambda_k^2 z^2\right)^{1/2} \exp\left\{ \frac{|\lambda_k \mu_k|^2 z^2}{1 + \lambda_k^2 z^2} \right\} \tag{9.112}$$

The μ_k are the entries from μ.

The function γ in (9.98) depends on the power in the signals and the noise variance (power) through the SNR. Furthermore, it depends on the frequencies of the two sinusoids through their frequency separation. Setting $\gamma = 0$ separates the parameter space composed of the SNR and frequency separation into two regions. On one side of this curve, the signals are resolvable and on the other, they are not. This is illustrated in Figure 9.12.

Example:

A numerical example is as follows. This example is from [15]. Suppose two narrowband Gaussian signals $s_{1,k}$ and $s_{2,k}$ are generated by the AR processes

$$s_{i,k} = c_i s_{i,k-1} + \varepsilon_{i,k} \tag{9.113}$$

where

$$c_1 = 0.94 e^{j2\pi f_1} \tag{9.114}$$

and

$$c_2 = 0.96 e^{j2\pi f_2} \tag{9.115}$$

The innovations process $\varepsilon_{i,k} \sim \mathcal{CN}(0, \sigma_{\varepsilon_i}^2)$. The correlation matrix of $s_{i,k}$ is given by

$$\mathbf{C}_{s_i} = \begin{bmatrix} 1 & c_i & \cdots & c_i^m \\ c_i^* & 1 & \cdots & c_i^{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ (c_i^*)^m & (c_i^*)^{m-1} & \cdots & 1 \end{bmatrix} \tag{9.116}$$

where $m = 9$.

The observed sequence is given by

$$x_k = \eta s_{1,k} + \eta s_{2,k} + w_k \tag{9.117}$$

with covariance matrix

$$\mathbf{\Gamma}_{xx}^{(m+1)} = \eta \mathbf{C}_{s_1} + \eta \mathbf{C}_{s_2} + \sigma_w^2 \mathbf{I} \tag{9.118}$$

η is used to control the SNR. P_r for this example is plotted in Figure 9.13 for representative values of the SNR versus the normalized frequency separation. The frequency separation is normalized to the mean value of the two frequencies. In this case, $f_1 = 0.15$, $f_2 = 0.24$, so $f_{norm} = 0.195$. As can be seen, a frequency separation of 0.10 or more produces a probability of resolution of 90% or better. P_r is plotted in Figure 9.14 versus SNR.

■

9.2 Line Spectra

In some cases a more appropriate model of the stochastic process is based on a combination of sinusoids in noise. If it is known, for example, that this is the case for the process under investigation, then this model may yield more accurate results than the AR models discussed in Section 9.1.

In this model for stochastic processes the process is assumed to consist of a linear combination of M sinusoidal signals with constant amplitudes and frequencies and random phase θ uniformly distributed on $[0, 2\pi)$. Thus,

$$x_n = \sum_{i=1}^{M} a_i \cos\left(2\pi f_i n + \theta_i\right) \tag{9.119}$$

The autocorrelation function for this random process is

$$\gamma_{xx,k} = \sum_{i=1}^{M} P_i \cos\left(2\pi f_i k\right) \tag{9.120}$$

where

$$P_i = \frac{1}{2} a_i^2 \tag{9.121}$$

The psd of this process is the FT of (9.120) and is given by
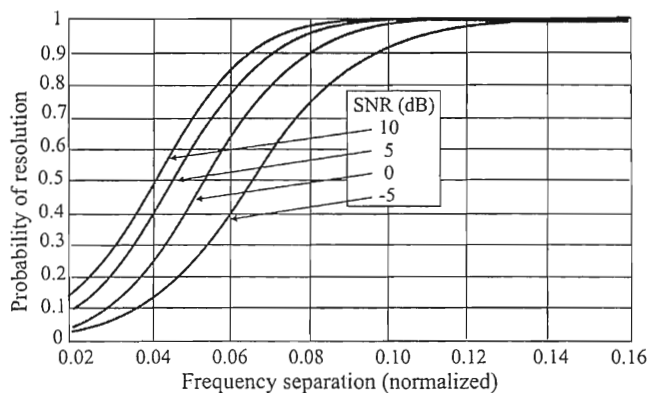
Figure 9.13 Probability of resolution for a sample AR process versus normalized frequency separation. (*From:* [15], © 1998 IEEE. Reprinted with permission.)
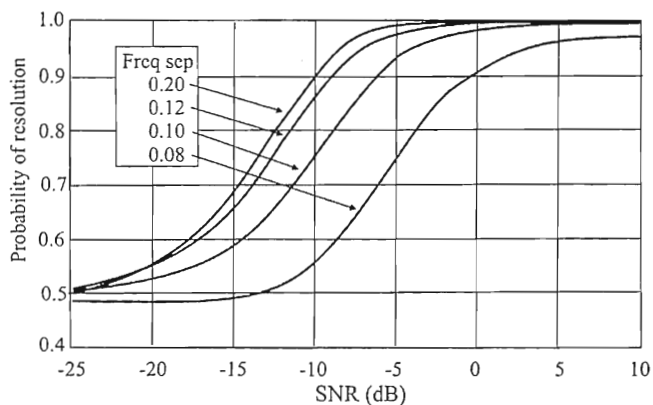
Figure 9.14 Probability of resolution for a sample AR process versus SNR. (*From:* [15], © 1998 IEEE. Reprinted with permission.)

$$P_{xx}(f) = \pi \sum_{i=1}^{M} P_i \left\{ \delta\left[2\pi(f-f_i)\right] + \delta\left[2\pi(f+f_i)\right] \right\} \tag{9.122}$$

For the process described by (9.119),

$$\mathcal{E}\left\{x_n^2\right\} = \sum_{i=1}^{M} P_i \tag{9.123}$$

One approach to modeling the response of systems as the sum of sinusoids was developed by Prony [16]. In that method, the received signal is modeled as a sum of N complex sinusoids given by

$$r(t) = \sum_{i=1}^{N} A_i e^{\lambda_i t} \tag{9.124}$$

where A_i are the residues and $\lambda_i = \alpha_i + j\omega_i$ are the complex poles. The α_i are referred to as the damping constants and ω_i are the frequencies. Once digitized, (9.124) becomes

$$r_k = \sum_{i=1}^{N} A_i e^{\lambda_i k \delta_{\mathrm{T}}}, \qquad k = 0, 1, \cdots, M-1 \tag{9.125}$$

where δ_{T} is the time step size and M is the total number of samples. The set of equations (9.125) is a set of M nonlinear equations in $2N$ unknowns.

This model is motivated by some results from system theory. Many physical systems can be modeled with differential equations of the type

$$\kappa_n \frac{d^n r(t)}{dt^n} + \kappa_{n-1} \frac{d^{n-1} r(t)}{dt^{n-1}} + \cdots + \kappa_1 \frac{dr(t)}{dt} + \kappa_0 r(t) = C, \, t \geq 0 \tag{9.126}$$

which typically have solutions of the form of (9.124) where A_i are the amplitudes and $\mathrm{Re}\{\lambda_i\} = \alpha_i \leq 0$.

9.2.1 Least Squares

In practice, $r(t)$ is not observed exactly since there is always observation noise present. Instead, the observations are $y_i = r_i + e_i$, where the e_i are random observation errors. If the errors can be approximated as Gaussian, then the

unknown parameters in the exponential signal are estimated by minimizing the sum of squared errors given by

$$\sum [y_i - r_i]^2 \tag{9.127}$$

This is a nonlinear least squares problem in the unknown parameters A_j and λ_j.

The Prony estimation approach is to represent the signal $r(t)$ not in terms of the A_j and λ_j but in terms of the coefficients of the differential equation (9.126). Computing the eigenvectors of a suitably calculated covariance matrix identifies the coefficients.

9.2.2 Prony's Method

Prony's method is a technique for extracting the signal parameters by solving a set of linear equations for the coefficients of the recurrence equation that the signals satisfy. It is closely related to Pisarenko's method, discussed in Section 9.3.1, which uses the smallest eigenvalue of an estimated covariance matrix.

In particular the r_k in (9.125) must satisfy

$$\sum_{i=0}^{N} \alpha_i r_{i+n} = 0, \qquad n = 0, 1, \cdots, \eta - 1 \tag{9.128}$$

where $\eta = M - N$. The roots z_i of

$$\sum_{i=1}^{N} \alpha_i z^i = 0 \tag{9.129}$$

define the frequencies as

$$z_i = e^{\lambda \delta_\tau}, \qquad i = 1, 2, \cdots, N \tag{9.130}$$

Setting $\alpha_N = 1$, then the remainder of the damping coefficients can be obtained by solving

$$\sum_{i=0}^{N-1} \alpha_i r_{i+n} = -r_{N+n} \tag{9.131}$$

When exactly $2N$ samples are used, then (9.131) can be solved exactly for the α's. It is normally the case, however, that more than $2N$ samples are available and least-squares estimation of the α's is used.

When the α's are available, $z_i = \exp(\lambda_i \delta_T)$ of (9.129) can be calculated with

$$\lambda_i = \frac{1}{\delta_T} \ln z_i \qquad (9.132)$$

After that, the A's can be calculated using (9.125).

Prony's method works well when there is no noise present. The addition of noise to the model, however, causes significant problems. Noise tends to make the damping parameters too large. Kahn et al. [17] showed that it is actually inconsistent when noise is included. This has motivated the development of modifications to the Prony method.

9.2.3 Modified Prony Methods

A modified Prony algorithm was developed by Osborne [18] that reacts to noise better than Prony's original technique. It was generalized by Smyth [19] and Osborne and Smyth [20] to estimate any function that satisfies a difference equation with coefficients that are linear and homogeneous in the parameters. Osborne and Smyth considered rational function fitting and proved that the algorithm is asymptotically stable in that case.

The algorithm for exponential fitting will estimate, for fixed p, any function $r(t)$ that solves a constant coefficient differential equation (9.126). Perturbed observations, $y_i = r_i + e_i$, are made at equispaced times t_i, $i = 1, ..., n$, where the e_i are independent and $\sim \mathcal{N}(0, \sigma^2)$. The solutions to (9.126) include complex exponentials, damped and undamped sinusoids, and real exponentials, depending on the roots of the polynomial with the κ_i as coefficients. This algorithm has the practical advantage that it will estimate any of these functions according to which best fits the available observations.

9.3 Signal Subspace Techniques

Because the RF channels are assumed to be linear, they admit to the superposition principle, so it can be assumed that the received signal consists of a sum of M complex sinusoids and noise as

$$x_k = \sum_{i=1}^{M} A_i e^{jk\omega_i} + n_k, \qquad k = 0,1,\cdots,N-1 \qquad (9.133)$$

where $n_k \sim \mathcal{N}\{0, \sigma^2\}$ represents a noise sample. The set of noise samples are Gaussian and i.i.d. These complex amplitudes A_i have random phase

$$A_i = |A_i| e^{j\phi_i}, \qquad -\pi \le \phi_i < \pi \qquad (9.134)$$

and ϕ_i are uniformly distributed. The goal is to estimate the unknown frequencies ω_i and amplitudes $|A_i|$, or, equivalently, power $P_i = \mathcal{E}\{|A_i|^2\}$. The signal subspace techniques use the eigen-decomposition of the estimated autocorrelation matrix of the stochastic process. They are based on the notion that the signals $s_{i,k} = A_i e^{jk\omega_i}$ span a subspace of the vector space of observations given by (9.133). Throughout this section it is assumed that

$$\mathcal{E}\{A_i A_k^*\} = \mathcal{E}\{A_i n_k^*\} = 0 \quad \text{when } i \ne k \qquad (9.135)$$

If there is only one signal present, then

$$x_k = As_k + n_k \qquad (9.136)$$

with

$$s_k = e^{jk\omega_0} \qquad (9.137)$$

$$A = |A_k| e^{j\phi} \qquad (9.138)$$

which is illustrated in Figure 9.15.
   In vector form, this is

$$\mathbf{x} = \mathbf{As} + \mathbf{n} \qquad (9.139)$$

where $\mathbf{x} = [x_0 \, x_1 \, \ldots \, x_{N-1}]^T$. The autocorrelation matrix of $\mathbf{x}$ is given by

$$\Gamma_{xx} = \mathcal{E}\{\mathbf{As}(\mathbf{As})^\dagger\} + \mathcal{E}\{\mathbf{nn}^\dagger\} \qquad (9.140)$$

$$= P_0 \mathbf{ss}^\dagger + \sigma_o^2 \mathbf{I} \qquad (9.141)$$
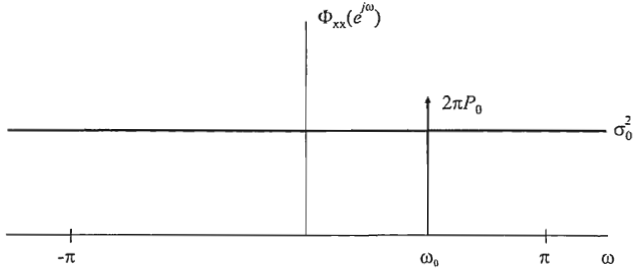
Figure 9.15 Spectrum for a single signal with noise.

where † denotes conjugate transpose and $P_0 = \mathcal{E}\left\{|\mathbf{A}|^2\right\}$. Note that

$$\begin{aligned}
\mathbf{\Gamma}_{xx}\mathbf{s} &= \left(P_0\mathbf{s}\mathbf{s}^\dagger + \sigma_0^2\mathbf{I}\right)\mathbf{s} \\
&= P_0\mathbf{s}\mathbf{s}^\dagger\mathbf{s} + \sigma_0^2\mathbf{s} \\
&= \left(NP_0 + \sigma_0^2\right)\mathbf{s}
\end{aligned} \tag{9.142}$$

so the signal vector is an eigenvector of $\mathbf{\Gamma}_{xx}$ with eigenvalue $\lambda = NP_0 + \sigma_0^2$. Denote the remaining eigenvalues of $\mathbf{\Gamma}_{xx}$ with $\mathbf{e}_i$. Then

$$\begin{aligned}
\mathbf{\Gamma}_{xx}\mathbf{e}_i &= P_0\mathbf{s}\mathbf{s}^\dagger\mathbf{e}_i + \sigma_0^2\mathbf{e}_i \\
&= \sigma_0^2\mathbf{e}_i
\end{aligned} \tag{9.143}$$

since $\mathbf{s}^\dagger\mathbf{e}_i = 0$. Thus, the eigenvalues associated with the remaining eigenvectors are all equal to σ_0^2.

The solution to the single signal situation is therefore as follows:

- Find the autocorrelation matrix and its associated eigenvalues and eigenvectors.
- Find the $N-1$ smallest eigenvalues; they will all be equal to σ_0^2.
- Identify the remaining (largest) eigenvalue which is equal to $NP_0 + \sigma_0^2$; knowing this and σ_0^2 allows computation of P_0.
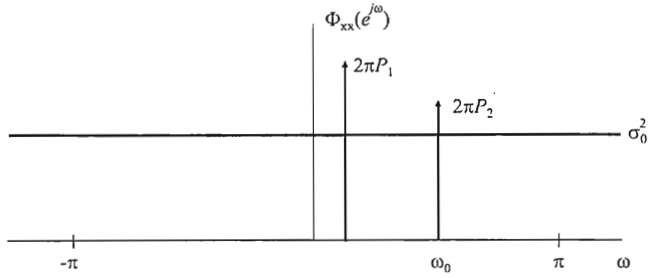
Figure 9.16 Two complex sinusoids with AWGN.

- The eigenvector corresponding to the largest eigenvalue is proportional to $\mathbf{s} = \begin{bmatrix} 1 & e^{j\omega_0} & e^{j2\omega_0} & \cdots & e^{j(N-1)\omega_0} \end{bmatrix}^{\mathrm{T}}$. From this, ω_0 can be found.

Now suppose there are two complex sinusoids present, as illustrated in Figure 9.16. The observed sequence in this case is

$$x_k = A_1 s_{1,k} + A_2 s_{2,k} + n_k \tag{9.144}$$

where

$$s_{i,k} = e^{jk\omega_i} \qquad A_i = |A_i| e^{j\phi_i} \qquad i = 1,2 \tag{9.145}$$

In vector form, this is

$$\mathbf{x} = A_1 \mathbf{s}_1 + A_2 \mathbf{s}_2 + \mathbf{n} \tag{9.146}$$

with autocorrelation matrix

$$\mathbf{\Gamma}_{xx} = P_1 \mathbf{s}_1 \mathbf{s}_1^\dagger + P_2 \mathbf{s}_2 \mathbf{s}_2^\dagger + \sigma_0^2 \mathbf{I} \tag{9.147}$$

where

$$P_i = \mathcal{E}\left\{ |A_i|^2 \right\} \qquad i = 1,2 \tag{9.148}$$

There are $N - 2$ noise eigenvectors orthogonal to both $\mathbf{s}_1$ and $\mathbf{s}_2$, and they all have the eigenvalues equal to σ_0^2 because

$$\begin{aligned}
\mathbf{\Gamma}_{xx}\mathbf{e}_i &= P_1 s_1 \mathbf{s}_1^\dagger \mathbf{e}_i + P_2 s_2 \mathbf{s}_2^\dagger \mathbf{e}_i + \sigma_0^2 \mathbf{e}_i \\
&= \sigma_0^2 \mathbf{e}_i \quad i = 3, 4, \dots, N
\end{aligned} \tag{9.149}$$

because $\mathbf{s}_1^\dagger \mathbf{e}_i = \mathbf{s}_2^\dagger \mathbf{e}_i = 0$. The remaining two largest eigenvectors are in the subspace spanned by $\mathbf{s}_1$ and $\mathbf{s}_2$, as illustrated in Figure 9.17, where for illustrative purposes $N = 3$. The subspace containing the signals is called the *signal subspace* and that containing the noise eigenvectors, orthogonal to the signal subspace, is called the *noise subspace*.

Now consider the general case of M signals in AWGN. The observed sequence is given by

$$x_k = \sum_{i=1}^{M} A_i s_{i,k} + n_k \tag{9.150}$$

or, in matrix form,

$$\mathbf{x} = \sum_{i=1}^{M} A_i \mathbf{s}_i + \mathbf{n} \tag{9.151}$$

The autocorrelation matrix is given by
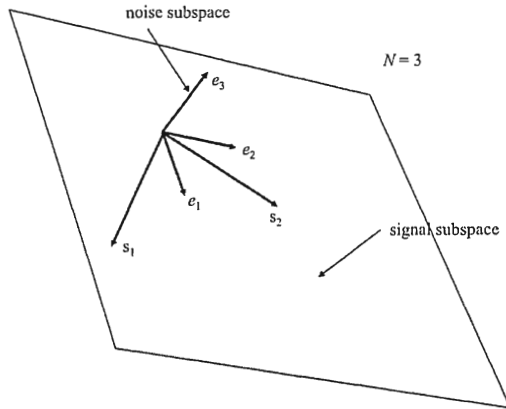
Figure 9.17 Signal and noise subspaces illustrated for $N = 3$.

$$\mathbf{\Gamma}_{xx} = \sum_{i=1}^{M} P_i \mathbf{s}_i \mathbf{s}_i^\dagger + \sigma_0^2 \mathbf{I} \qquad (9.152)$$

which can be written using matrices as

$$\mathbf{\Gamma}_{xx} = \mathbf{S}\mathbf{P}_0\mathbf{S}^\dagger + \sigma_0^2 \mathbf{I} = \mathbf{\Gamma}_p + \sigma_0^2 \mathbf{I} \qquad (9.153)$$

where

$$\mathbf{S} = \begin{bmatrix} \mathbf{s}_1 & \mathbf{s}_2 & \cdots & \mathbf{s}_M \end{bmatrix} \qquad (9.154)$$

is a matrix whose columns are the signal samples. Furthermore,

$$\mathbf{P}_0 = \mathrm{diag}\begin{bmatrix} P_1 & P_2 & \cdots & P_M \end{bmatrix} \qquad (9.155)$$

is a diagonal matrix of the signal powers. $\mathbf{\Gamma}_p$ is a rank p approximation to $\mathbf{\Gamma}_{xx}$.

Thus, the M signal vectors, $\mathbf{s}_1$, $\mathbf{s}_2$, ..., $\mathbf{s}_M$ make up the signal subspace. The first M eigenvectors of the autocorrelation matrix $\mathbf{\Gamma}_{xx}$ (which correspond to the largest eigenvalues) span the signal subspace. These eigenvectors have eigenvalues greater than σ_0^2. The remaining $N - M$ eigenvalues define the noise subspace, which is orthogonal to the signal subspace. These eigenvalues are all equal to σ_0^2. Therefore, the first p eigenvalues of $\mathbf{\Gamma}_{xx}$ will be greater than σ_0^2 and the last $M - p$ eigenvalues will be equal to σ_0^2, so

$$\lambda_i = \begin{cases} \lambda_i^s + \sigma_0^2 & \text{for some } \lambda_i^s \geq 0, \ 1 \leq i \leq p \\ \sigma_0^2 & i = p+1, \cdots, M \end{cases} \qquad (9.156)$$

This decomposes the set of eigenvalues and corresponding eigenvectors into two classes. In the first class are those eigenvectors that span the signal subspace, while the other class consists of those eigenvectors that span the noise subspace. The corresponding eigen-decomposition of $\mathbf{\Gamma}_{xx}$, also known as the *singular value decomposition,* is given by

$$\mathbf{\Gamma}_{xx} = \sum_{i=1}^{p} \overbrace{\left(\lambda_i^s + \sigma_0^2\right)}^{\lambda_i} \mathbf{e}_{s,i}\mathbf{e}_{s,i}^\dagger + \sum_{i=p+1}^{M} \overbrace{\sigma_0^2}^{\lambda_i} \mathbf{e}_{n,i}\mathbf{e}_{n,i}^\dagger \qquad (9.157)$$

$$\overbrace{= \mathbf{E}_s \boldsymbol{\Lambda}_s \mathbf{E}_s^\dagger}^{\substack{\text{signal}\\\text{component}}} + \overbrace{\mathbf{E}_n \boldsymbol{\Lambda}_n \mathbf{E}_n^\dagger}^{\substack{\text{noise}\\\text{component}}} \qquad (9.158)$$

where

$$\mathbf{E}_s = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_p \end{bmatrix} \qquad (9.159)$$

$$\mathbf{E}_n = \begin{bmatrix} \mathbf{e}_{p+1} & \mathbf{e}_{p+2} & \cdots & \mathbf{e}_M \end{bmatrix} \qquad (9.160)$$

$\mathbf{e}_i$ is the eigenvector for the ith eigenvalue of $\boldsymbol{\Gamma}_{xx}$,

$$\boldsymbol{\Lambda}_s = \mathrm{diag}\begin{pmatrix} \lambda_1^s + \sigma_0^2 & \lambda_2^s + \sigma_0^2 & \cdots & \lambda_p^s + \sigma_0^2 \end{pmatrix} \qquad (9.161)$$

and

$$\boldsymbol{\Lambda}_n = \sigma_0^2 \mathbf{I} \qquad (9.162)$$

The $\{\mathbf{e}_i\}$ form an orthonormal basis set.

To estimate the frequency of the component signals, we take advantage of the fact that the signal vectors are orthogonal to the noise subspace. That is,

Property: Orthogonality of Signal Subspace with Noise Subspace

$$\mathbf{s}_i \mathbf{e}_k = 0,\, i = 1, 2, \cdots, M; \quad k = M+1, M+2, \cdots, N \qquad (9.163)$$

Proof:

The $\mathbf{e}_k$ are eigenvectors of $\boldsymbol{\Gamma}_{xx}$ and as such

$$\boldsymbol{\Gamma}_{xx} \mathbf{e}_k = \sigma_0^2 \mathbf{e}_k \qquad (9.164)$$

but (9.153) is also true. Therefore,

$$\begin{pmatrix} \mathbf{SPS}^\dagger + \sigma_0^2 \mathbf{I} \end{pmatrix} \mathbf{e}_k = \sigma_0^2 \mathbf{e}_k \qquad (9.165)$$

and therefore, of necessity,

$$\mathbf{SPS}^{\dagger}\mathbf{e}_k = 0 \tag{9.166}$$

■

The *frequency estimation function* is formed as

$$F(\omega) = \frac{1}{\displaystyle\sum_{i=p+1}^{M} \left| \mathbf{w}^{\dagger}(\omega)\mathbf{e}_i \right|^2} \tag{9.167}$$

where

$$\mathbf{w}(\omega) = \begin{bmatrix} 1 & e^{j\omega} & \cdots & e^{j(M-1)\omega} \end{bmatrix}^{\mathrm{T}} \tag{9.168}$$

The summation in the denominator of (9.167) may be different depending on the method being discussed. $F(\omega)$ will peak at the values of ω corresponding to the frequencies in the component signals ω_1, ω_2, ..., ω_p. In practice, weights are used in the denominator function in (9.167). Also, typically in practice, the inverse of (9.167) is used and the resulting function space is searched for minimums.

The solution of this general problem varies according to the particular algorithm being discussed. The remainder of this section discusses some of these techniques.   The method included here in detail is the Pisarenko technique since all of the subspace approaches are based on the same fundamental principles, and the Pisarenko technique is one of the simplest. The other methods to be briefly discussed are the root Pisarenko method, MUSIC algorithm, minimum norm algorithm, and principle components algorithm. While all these techniques are based on the same fundamental principal, their performance characteristics are distinct.

9.3.1 Pisarenko Method

Assume that there are M signals at unknown frequencies ω_1, ω_2, ..., ω_M for which the frequencies and powers are to be found. Let $N = M + 1$. From the above property, the single eigenvector in the noise subspace $\mathbf{e}_N$ (corresponding to the smallest eigenvector) is orthogonal to each signal vector $\mathbf{s}_i$. Let

$$\mathbf{w} = \begin{bmatrix} 1 & e^{j\omega} & e^{j2\omega} & \cdots & e^{j(N-1)\omega} \end{bmatrix}^{\mathrm{T}} \tag{9.169}$$

while
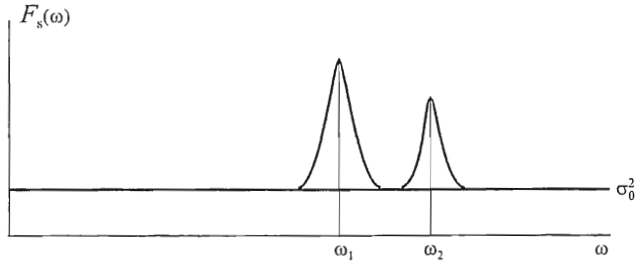
Figure 9.18 Pisarenko frequency estimation function with two signals and noise present.

$$\mathbf{s}_i = \begin{bmatrix} 1 & e^{j\omega_i} & e^{j2\omega_i} & \cdots & e^{j(N-1)\omega_i} \end{bmatrix}^{\mathrm{T}} \tag{9.170}$$

then clearly

$$\mathbf{w}^\dagger \mathbf{e}_N\big|_{\omega=\omega_i} = 0, \quad i = 1, 2, \ldots, M \tag{9.171}$$

Form the frequency estimation function

$$F_{\mathrm{pis}}(\omega) = \frac{1}{\left|\mathbf{w}^\dagger \mathbf{e}_N\right|^2}$$

$$= \frac{1}{\mathbf{w}^\dagger \mathbf{e}_N \mathbf{e}_N^\dagger \mathbf{w}} \tag{9.172}$$

This function will therefore peak at the frequencies of the signals present similar to that shown in Figure 9.18.

The power in the component signals is found as follows. $\boldsymbol{\Gamma}_{xx}$ is related to its eigenvalues λ_i and signal space eigenvectors $\mathbf{e}_i$ as

$$\boldsymbol{\Gamma}_{xx}\mathbf{e}_i = \lambda_i \mathbf{e}_i, \qquad i = 1, 2, \cdots, M \tag{9.173}$$

and since $\{\mathbf{e}_i\}$ forms an orthonormal basis,

$$\mathbf{e}_i^\dagger \mathbf{e}_i = 1 \tag{9.174}$$

Then

$$\lambda_i = e_i^\dagger \Gamma_{xx} e_i$$

$$= e_i^\dagger \left(\sum_{k=1}^{M} P_k s_k s_k^\dagger + \sigma_0^2 I \right) e_i \qquad (9.175)$$

The set of linear equations that results is given by

$$\sum_{k=1}^{M} P_k \left| s_k^\dagger e_i \right|^2 = \lambda_i - \sigma_0^2, \qquad i = 1, 2, \cdots, M \qquad (9.176)$$

For $M = 2$ for example, (9.176) is of the form

$$\begin{bmatrix} |\beta_{11}|^2 & |\beta_{12}|^2 \\ |\beta_{21}|^2 & |\beta_{22}|^2 \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 - \sigma_o^2 \\ \lambda_2 - \sigma_o^2 \end{bmatrix} \qquad (9.177)$$

where

$$\beta_{ik} = e_i^\dagger s_k \qquad (9.178)$$

which can be solved for P_i.

One limitation of the Pisarenko method is that the number of signals, M, must be known or somehow estimated. In addition, it assumes that the noise is AWGN. Whereas Prony's method is inconsistent when noise is considered, the Pisarenko form of the method is consistent but inefficient for estimating sinusoid signals and inconsistent for estimating damped sinusoids or exponential signals.

Example: (Pisarenko)

Suppose there is a single real sinusoidal signal in AWGN present and the frequency and power of this signal are to be determined. The correlation matrix for the data is

$$\Gamma_{xx} = \begin{bmatrix} 3 & 0 & -2 \\ 0 & 3 & 0 \\ -2 & 0 & 3 \end{bmatrix} \qquad (9.179)$$

From Chapter 2, recall that a real sinusoid consists of two complex exponentials, one at the positive frequency and one at the negative frequency. Therefore, $M = 2$ and a 3×3 autocorrelation matrix is required.

The eigenvalues and corresponding eigenvectors for Γ_{xx} are given by the matrices

$$\Lambda = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{9.180}$$

and

$$\mathbf{E} = \begin{bmatrix} -\dfrac{1}{\sqrt{2}} & 0 & -\dfrac{1}{\sqrt{2}} \\ 0 & -1 & 0 \\ \dfrac{1}{\sqrt{2}} & 0 & -\dfrac{1}{\sqrt{2}} \end{bmatrix} \tag{9.181}$$

respectively. Therefore, the noise variance is

$$\sigma_0^2 = \lambda_3 = 1 \tag{9.182}$$

and the corresponding noise eigenvector is

$$\mathbf{e}_3 = \begin{bmatrix} -\dfrac{1}{\sqrt{2}} \\ 0 \\ -\dfrac{1}{\sqrt{2}} \end{bmatrix} \tag{9.183}$$

The frequency estimation function is

$$F(\omega) = \frac{1}{\left| \mathbf{w}^\dagger \mathbf{e}_3 \right|^2} \tag{9.184}$$

This function peaks at $\omega = \pm \pi / 2$, since

$$\mathbf{w} = \begin{bmatrix} 1 \\ e^{\pm j\frac{\pi}{2}} \\ e^{\pm j\pi} \end{bmatrix} = \begin{bmatrix} 1 \\ \pm j \\ -1 \end{bmatrix} \tag{9.185}$$

and

$$\mathbf{w}^\dagger \mathbf{e}_3 = \begin{bmatrix} 1 & \mp j & -1 \end{bmatrix} \begin{bmatrix} -\dfrac{1}{\sqrt{2}} \\ 0 \\ -\dfrac{1}{\sqrt{2}} \end{bmatrix} = 0 \tag{9.186}$$

The power is found as follows. The signal vectors are

$$\mathbf{s}_1 = \begin{bmatrix} 1 \\ e^{j\frac{\pi}{2}} \\ e^{j\pi} \end{bmatrix} = \begin{bmatrix} 1 \\ j \\ -1 \end{bmatrix} \quad \text{and} \quad \mathbf{s}_2 = \begin{bmatrix} 1 \\ e^{j\frac{\pi}{2}} \\ e^{j\pi} \end{bmatrix} = \begin{bmatrix} 1 \\ -j \\ -1 \end{bmatrix} \tag{9.187}$$

Hence,

$$\begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1^\dagger \\ \mathbf{e}_2^\dagger \end{bmatrix} \begin{bmatrix} \mathbf{s}_1 & \mathbf{s}_2 \end{bmatrix} \tag{9.188}$$

$$= \begin{bmatrix} -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ j & -j \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} -\sqrt{2} & -\sqrt{} \\ -j & j \end{bmatrix} \tag{9.189}$$

Then

$$\begin{bmatrix} |\beta_{11}|^2 & |\beta_{12}|^2 \\ |\beta_{21}|^2 & |\beta_{22}|^2 \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 - \sigma_0^2 \\ \lambda_2 - \sigma_0^2 \end{bmatrix}$$
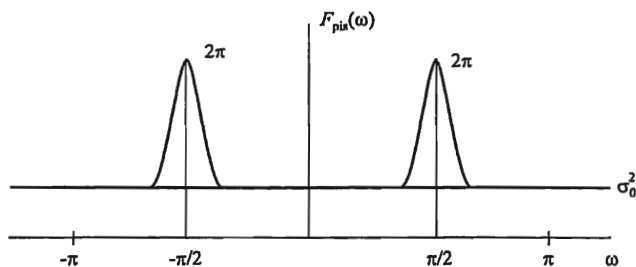
Figure 9.19 Frequency estimation function for the example of the Pisarenko technique.

is

$$\begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$ (9.190)

so

$$P_1 = P_2 = 1$$

Expression (9.184) is plotted in Figure 9.19.   ■

9.3.2 Root Pisarenko

The eigen-filter for the root Pisarenko modification to the technique described in Section 9.3.1 is defined as

$$E_N(z) = e_{N,0} + e_{N,1} z^{-1} + \cdots + e_{N,N-1} z^{-(N-1)}$$ (9.191)

where the noise eigenvector $\mathbf{e}_N$ is given by

$$\mathbf{e}_N = \begin{bmatrix} e_N & e_{N,1} & \cdots & e_{N,N-1} \end{bmatrix}^{\mathrm{T}}$$ (9.192)

Then

$$E_N(e^{j\omega}) = \mathbf{w}^\dagger \mathbf{e}_N$$ (9.193)

which is zero at $\omega = \omega_1, \omega_2, \ldots, \omega_M$. Therefore, the M roots of $E_N(z)$ occurring on the unit circle correspond to the signal frequencies $\omega_1, \omega_2, \ldots, \omega_M$.

9.3.3 MUSIC

The MUSIC algorithm was discovered by Schmidt [3]. It is based on the eigen-decomposition of the autocorrelation matrix and applies for both frequency spectral estimation as well as spatial spectral estimation. The latter of these is useful for determination of the angle of arrival of a signal.

The frequency estimation function for MUSIC is given by

$$F_{\text{MUSIC}}(\omega) = \frac{1}{\sum_{i=p+1}^{M} \left| \mathbf{w}(\omega)\mathbf{e}_i \right|^2} \qquad (9.194)$$

Whereas for the Pisarenko method, the denominator of the frequency estimation function is composed of a single eigenvector [the one corresponding to the smallest eigenvalues (9.172)], the MUSIC function uses all the eigenvectors from the noise subspace. (The signal subspace can also be used—that is the basis of the principle component approach described below.) This essentially averages the results over all the eigenvectors.

In theory, all of the functions in the form of (9.172) would work, with $p + 1 \leq i \leq M$ and with the component frequencies determined by the peaks. The true frequency estimation function, however, is only supposed to exhibit peaks at the p values of the true frequencies present. In practice, the remaining $M - 1 - p$ zeros of $\mathbf{w}^\dagger(\omega)\mathbf{e}_i$ might lie close to the unit circle that could produce spurious peaks. The averaging of the denominator of the MUSIC function tends to ameliorate this concern.

The MUSIC estimation of the psd is shown in Figure 9.20 for (9.91). All three signals are evident at their correct locations, with substantial differences in the power levels, so separation is facilitated.

It turns out that the selection of M is not critical in the MUSIC algorithm [21]. The selection of p is critical, however. If p is too low, the psd will not show the correct peaks, while if it is too large, spurious peaks occur.

Example: (MUSIC)

Let the correlation matrix of the complex exponentials in AWGN be given by
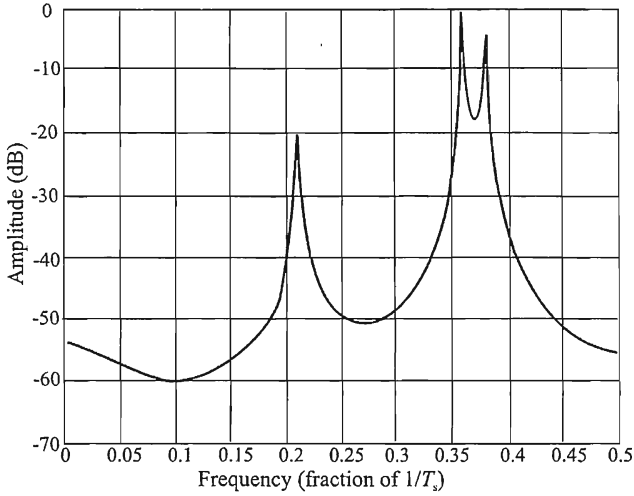
Figure 9.20 MUSIC estimate of the psd of (9.91). (*From:* [21]. © 1995 Horizon House, Inc. Reprinted with permission.)

$$\mathbf{\Gamma}_{xx} = \begin{bmatrix} 2 & -j & -1 \\ j & 2 & -j \\ -1 & j & 2 \end{bmatrix} \tag{9.195}$$

The matrices of eigenvalues and eigenvectors are

$$\mathbf{\Lambda} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \mathbf{E} = \begin{bmatrix} -\dfrac{1}{\sqrt{3}}j & \sqrt{\dfrac{2}{3}} & 0 \\ \dfrac{1}{\sqrt{3}} & -\dfrac{1}{\sqrt{6}}j & \dfrac{1}{\sqrt{2}}j \\ \dfrac{1}{\sqrt{3}}j & \dfrac{1}{\sqrt{6}} & \dfrac{1}{\sqrt{2}} \end{bmatrix} \tag{9.196}$$

There are two smallest eigenvectors and therefore the noise subspace has a dimension of 2 and there is a single complex exponential present. The matrix of the eigenvectors for the noise subspace is

$$\mathbf{E}_{\text{noise}} = \begin{bmatrix} \sqrt{\dfrac{2}{3}} & 0 \\ -\dfrac{1}{\sqrt{6}}j & \dfrac{1}{\sqrt{2}}j \\ \dfrac{1}{\sqrt{6}} & \dfrac{1}{\sqrt{2}} \end{bmatrix} \qquad (9.197)$$

while

$$\mathbf{P}_{\text{noise}} = \mathbf{E}_{\text{noise}}\mathbf{E}_{\text{noise}}^{\dagger} = \begin{bmatrix} \dfrac{1}{3} & \dfrac{1}{3}j & \dfrac{1}{3} \\ -\dfrac{1}{3}j & \dfrac{2}{3} & \dfrac{1}{3}j \\ \dfrac{1}{3} & -\dfrac{1}{3}j & \dfrac{2}{3} \end{bmatrix} \qquad (9.198)$$

The frequency estimation function is

$$F_{\text{MUSIC}} = \frac{1}{\mathbf{w}^{\dagger}P_{\text{noise}}\mathbf{w}} \qquad (9.199)$$

which peaks at

$$\mathbf{w} = \begin{bmatrix} 1 \\ e^{j\frac{\pi}{2}} \\ e^{j\pi} \end{bmatrix} = \begin{bmatrix} 1 \\ j \\ -1 \end{bmatrix} \qquad (9.200)$$

Thus, the signal has frequency $\omega = \pi/2$.

∎

9.3.4 Minimum Norm

The frequency estimation function for the minimum norm technique for frequency estimation is given by

$$F_{\text{MinNorm}}(\omega) = \frac{1}{\left|\mathbf{w}^\dagger(\omega)\mathbf{\psi}\right|^2} \qquad (9.201)$$

where $\mathbf{w}(\omega)$ is given by (9.168) and vector $\mathbf{\psi}$ is chosen according to the following requirements:

1. $\mathbf{\psi}$ is not zero.
2. $\mathbf{\psi}$ lies in the noise subspace.
3. The effects of spurious peaks are minimized.

Criterion 1 is satisfied by making $\psi_1 = 1$. Constraint 2 is enforced by setting $\mathbf{\psi} = \mathbf{Yv}$, where $\mathbf{v}$ is a variable vector and $\mathbf{Y}$ is the projection matrix that projects a vector onto the noise subspace. Thus,

$$\mathbf{Y} = \mathbf{E}_n \mathbf{E}_n^\dagger \qquad (9.202)$$

where

$$\mathbf{E}_n = \begin{bmatrix} \mathbf{e}_{p+1} & \mathbf{e}_{p+2} & \cdots & \mathbf{e}_M \end{bmatrix} \qquad (9.203)$$

contains the noise eigenvectors.

The peaks of $F(\omega)$ are determined by the zeros of the denominator in (9.201), which can be written as

$$\begin{aligned}
d(z) &= \sum_{k=0}^{M-1} \psi_k z^{-k} \\
&= \overbrace{\prod_{k=1}^{p}\left(1 - e^{j\omega}z^{-1}\right)}^{\substack{\text{signal} \\ \text{frequencies}}} \overbrace{\prod_{k=p+1}^{M-1}\left(1 - z_k z^{-1}\right)}^{\substack{\text{spurious} \\ \text{peaks}}}
\end{aligned} \qquad (9.204)$$

The spurious peaks are introduced by the zeros z_k that are close to the unit circle. To minimize the effects of these peaks, it is desirable to choose $\mathbf{\psi}$ such that these zeros are well within the unit circle. This is accomplished by choosing that $\mathbf{\psi}$ with the minimum norm. Therefore, the problem becomes

$$\text{Minimize } \|\boldsymbol{\psi}\|_2 \text{ subject to}$$

$$\boldsymbol{\psi} = \mathbf{Yv}$$

$$\boldsymbol{\psi}^\dagger \mathbf{v}_1 = 1 \tag{9.205}$$

$$\mathbf{v}_1 = [1 \quad 0 \quad \cdots \quad 0]^T$$

This results in the selection of ψ as

$$\boldsymbol{\psi} = \frac{\mathbf{Yv}_1}{\mathbf{v}_1^\dagger \mathbf{Yv}_1} \tag{9.206}$$

An example of the spectrum produced is shown in Figure 9.21. The same signal as for the MUSIC methods is used, $M = 4$ and $p = 20$. All four peaks are difficult to locate; it appears that there are only three peaks present.

9.3.5 Principal Components Spectrum Estimation

The above techniques all use the noise subspace. The signal subspace can also be used. The eigen-decomposition of Γ_{xx} is
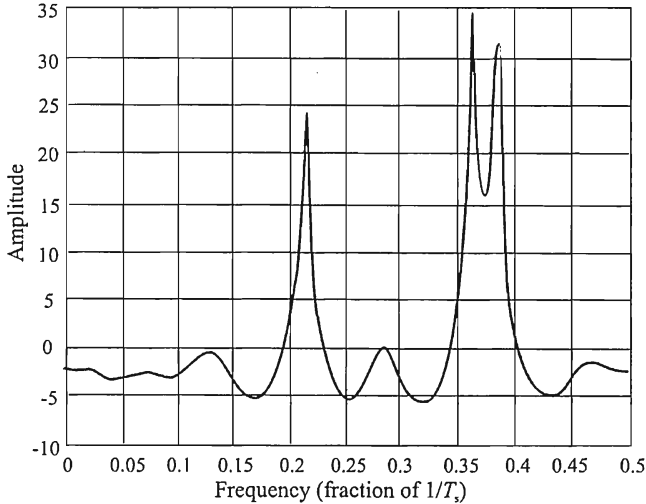
Figure 9.21 Example of the spectrum produced by the min-norm technique. In this case $M = 4$ and $P = 20$. The observation data is the same as for the MUSIC example. (*From:* [21], © 1995 Horizon House, Inc. Reprinted with permission.)

$$\mathbf{\Gamma}_{xx} = \sum_{i=1}^{M} \lambda_i \mathbf{e}_i \mathbf{e}_i^\dagger = \overbrace{\sum_{i=1}^{p} \lambda_i \mathbf{e}_i \mathbf{e}_i^\dagger}^{\substack{\text{signal}\\\text{subspace}}} + \overbrace{\sum_{i=p+1}^{M} \lambda_i \mathbf{e}_i \mathbf{e}_i^\dagger}^{\substack{\text{noise}\\\text{subspace}}} \qquad (9.207)$$

where, without loss of generality, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq \lambda_{p+1} = \lambda_{p+2} = \ldots = \lambda_M$, where it has been assumed that the noise eigenvalues are all equal.

The principal component method uses the rank-p approximation of $\mathbf{\Gamma}_{xx}$ given by $\mathbf{\Gamma}_s$ in (9.153),

$$\mathbf{\Gamma}_{xx} \approx \sum_{i=1}^{p} \lambda_i \mathbf{e}_i \mathbf{e}_i^\dagger = \mathbf{\Gamma}_s \qquad (9.208)$$

Using the signal subspace has the effect of filtering the noise. This estimate for $\mathbf{\Gamma}_{xx}$ is then used in any of the above spectral estimation methods.

9.4 Maximum Likelihood

The ML spectral estimate approach was developed by Capon [4] and is maximum likelihood in name only. The technique here is similar to that for the periodogram discussed in Chapter 4, with the difference being that the filter parameters are adjusted according to the statistics of the stochastic process [22]. Thus, the frequencies of the signals being analyzed need not be (and are normally not) harmonically related. The filters are of the *finite impulse response* (FIR) type shown in Figure 9.22, with tap weights given by

$$\mathbf{\Psi} = [\psi_0 \quad \psi_1 \quad \cdots \quad \psi_{p-1}]^T \qquad (9.209)$$

ML spectral estimates offer higher resolution than the periodogram spectral estimates but less than the AR estimates. The weights are dynamically adjusted so that the frequency response at the frequency under consideration (f_0) is unity and the variance of the output is minimized. The output variance is given by

$$\sigma^2 = \mathbf{\Psi}^\dagger \mathbf{\Gamma}_{xx} \mathbf{\Psi} \qquad (9.210)$$

and the gain constraint can be expressed as

$$\mathbf{W}^\dagger \mathbf{\Psi} = 1 \qquad (9.211)$$
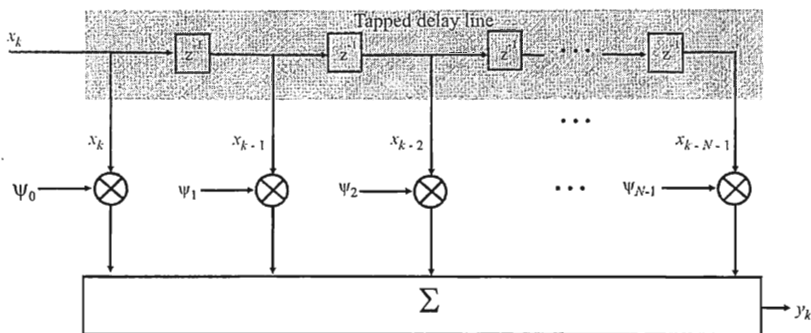
Figure 9.22 Finite impulse response filter.

where Γ_{xx} is the autocorrelation matrix of x_i and $\mathbf{W}$ is given by

$$\mathbf{W} = \begin{bmatrix} 1 & e^{j\omega_0\delta_t} & \cdots & e^{j(p-1)\omega_0\delta_t} \end{bmatrix}^{\mathrm{T}} \tag{9.212}$$

where δ_t is the sampling interval. The optimum weights are given by

$$\mathbf{\Psi}_{opt} = \frac{\Gamma_{xx}^{-1}\mathbf{W}}{\mathbf{W}^{\dagger}\Gamma_{xx}^{-1}\mathbf{W}} \tag{9.213}$$

and the minimum variance is

$$\sigma^2 = \frac{1}{\mathbf{W}^{\dagger}\Gamma_{xx}^{-1}\mathbf{W}} \tag{9.214}$$

The ML frequency estimation function is then given by

$$F_{\mathrm{ML}}(f_0) = \frac{\delta_t}{\mathbf{W}^{\dagger}\Gamma_{xx}^{-1}\mathbf{W}} \tag{9.215}$$

9.5 Resolution Comparison

Computed spectra are illustrated in Figure 9.23 through 9.25 for several of the spectral estimation techniques. The spectra are based on a 64-point real process

that consists of three sinusoids and a colored noise signal that has a broad bandwidth reaching from $f = 0.2/T_s$ to just short of $f = 0.5/T_s$, whose true spectrum is shown in Figure 9.23(a). The sinusoids are located at $f = 0.10/T_s$, $0.20/T_s$, and $0.21/T_s$ with SNRs of 10, 30, and 30 dB, respectively. The noise process is centered at $f = 0.39/T_s$.

The periodogram computed via the Blackman-Tukey approach is shown in Figure 9.23(b). The two sinusoids around $f = 0.2/T_s$ are combined into one and the one at $f = 0.1/T_s$ is not discernable at all due to masking by the sidelobes of the signals at $f = 0.2/T_s$ and $0.21/T_s$. The broad signal centered at $f = 0.35/T_s$ produces a response as well; however, there is no discernable peak. The peaks that are present in the region indicate that there is energy there, however.

The AR psd using the Yule-Walker equations described above is shown in Figure 9.23(c). The signals around $0.2/T_s$ are detected, but they are combined. The signal at $f = 0.1/T_s$ also produces a small peak. The three peaks in the upper frequency range would tend to indicate that there are at least two, if not three, separate signals there rather than the single broadband one.

The psd for the Burg algorithm based on maximizing the information entropy in the AR coefficients is illustrated in Figure 9.23(d) [22]. That approach can
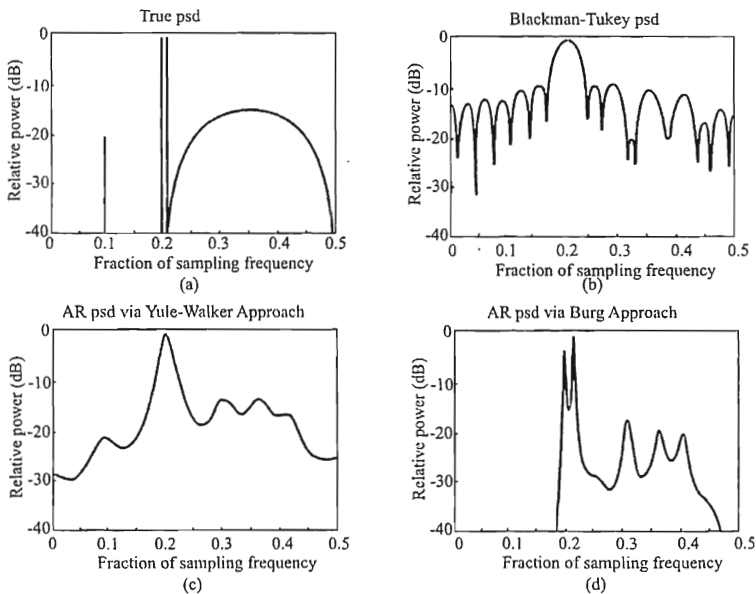
Figure 9.23 The psd estimates for the true psd shown in (a) for (b) periodogram, (c) AR, and (d) AR via the Burg algorithm. (*From:* [22], © 1981 IEEE. Reprinted with permission.)
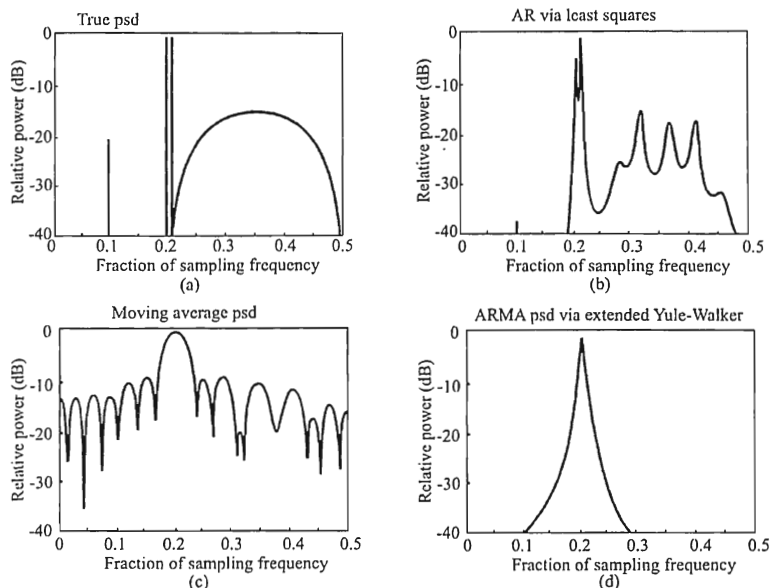
Figure 9.24 The psd estimates for the true psd shown in (a) for (b) AR via least squares, (c) MA, and (d) ARMA via extended Yule-Walker equations. (*From:* [22], © 1981 IEEE. Reprinted with permission.)

produce very sharp peaks at the signals of interest and is therefore useful when signals are closely spaced. The two narrowband signals around $0.2/T_s$ are separately detected with sharp peaks. The spurious peaks where the broadband signal is located are misleading, indicating that three and perhaps more signals are present there. It can be said, however, that the signal was not missed. The signal at $0.01/T_s$ was totally missed in this case, however. Viewing the time series as a sequenced set of data rather than a time series, it is possible also to compute the psd in reverse. Such an approach for the AR estimator yields the curve shown in Figure 9.24(b). In this case, the narrowband signal at $0.1/T_s$ is detected, although it is very weak. The two narrowband signals around $0.2/T_s$ are also detected and resolved. The multiple peaks at the higher frequencies leads to false indications of multiple signals, just as for the Burg algorithm.

The MA psd is shown in Figure 9.24(c). The only signal detected is a combination of the two around $0.2/T_s$. Discerning the other signals from the psd is very difficult, even for a human examining the results.

The Yule-Walker equations can be solved to find the AR coefficients in an ARMA process and then some technique devised to find the MA coefficients
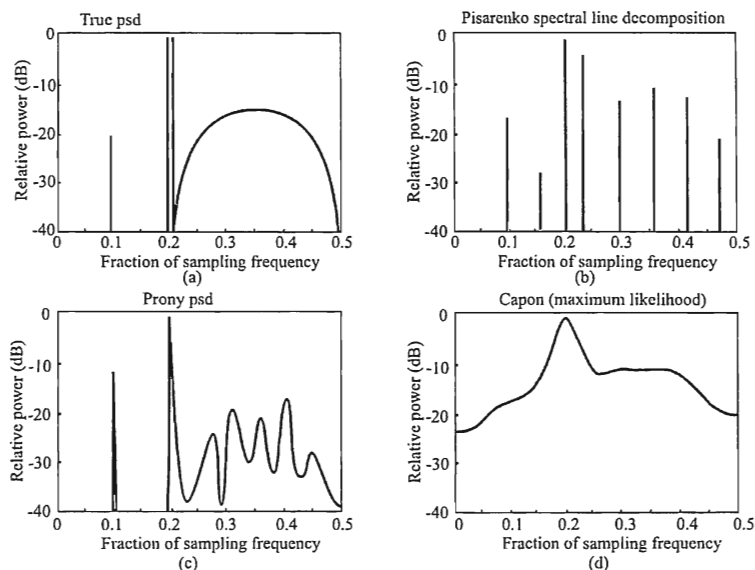
Figure 9.25 The psd estimates for the true psd shown in (a) for (b) Pisarenko spectral line decomposition, (c) Prony method, and (d) Capon method. (*From:* [22], © 1981 IEEE. Reprinted with permission.)

separately. This leads to nonoptimal coefficients, however. Extending the Yule-Walker equations is one such technique. The resulting psd is shown in Figure 9.24(d). The only signals found were the combination of the two narrowband signals around $0.2/T_s$.

The Fourier approach to spectral estimation assumes that the signals present are harmonically related. The Pisarenko harmonic decomposition removes that assumption. It does assume, however, that the signals present are tones in white noise. The technique is a special case of the ARMA approach. It assumes that the autocorrelation lags are perfectly known, even though the frequencies and the order are not a priori known. These results are illustrated in Figure 9.25(b). The narrowband signal at $0.1/T_s$ is found but the two around $0.2/T_s$ are combined. Numerous spurious peaks result as well, including some in the upper frequency ranges where the broadband signal is located. The observed data is clearly not a combination of tones in white noise, however; so one would assume that this model would not work well on this data set.

Prony's technique is based on fitting a linear combination of exponentials to the data. This fit was an exact fit of P exponentials. A more modern version

assumes that an approximate fit of $2P < N$ exponentials is required, which use a least-squares fitting procedure. This latter approach is called the extended Prony method. The resulting psd is shown in Figure 9.25(c). The narrowband signal at $0.1/T_s$ and those around $0.2/T_s$ were found, albeit combined. Several spurious signals emerge where the broadband signal is located, leading to false conclusions about the number of signals present.

Lastly, the Capon method is based on a maximum likelihood approach. It measures the out power in a parallel combination of narrowband filters. This is similar to the periodogram. In the periodogram, however, the width of the filters is fixed, while in the Capon approach the filter widths differ for each frequency and in fact vary as the processing proceeds. Energy not close to the frequency bin of interest is minimized through the filter in question. The output variance in each bin is minimized subject to unity filter amplitude response. The resultant psd is shown in Figure 9.25(d). At best, the two narrowband signals at $0.2/T_s$ were combined and detected. There appears to be somewhat of a correct response to the wideband signals at higher frequencies, however.

Marple performed an analysis of the resolution capability of two high-resolution methods and compared those with the conventional methods of simply taking the DFT of the time series directly as well as the periodogram [23]. As a metric, the normalized resolution given by

$$R = 2\pi M \Delta t \Delta f \qquad (9.216)$$

was used, where M is the number of correlation lags used, Δt is the length of the time series, and Δf is the frequency separation of two signals at the point of just being resolved. Two signals are resolved when

$$S\left(\frac{f_1 + f_2}{2}\right) = \frac{1}{2}[S(f_1) + S(f_2)] \qquad (9.217)$$

That is, two sinusoids are resolved at that $\Delta f = |f_1 - f_2|$ at which the psd evaluated at the center frequency $S(f_c)$, where $f_c = (f_1 + f_2)/2$ is equal to the average of the two separate psds evaluated at the two sinusoidal frequencies. This is illustrated in Figure 9.26.

Resolution of several of the methods of spectrum estimation are illustrated in Figure 9.27. As expected, the resolution limit for the periodogram and straightforward FFT of the data is approximately
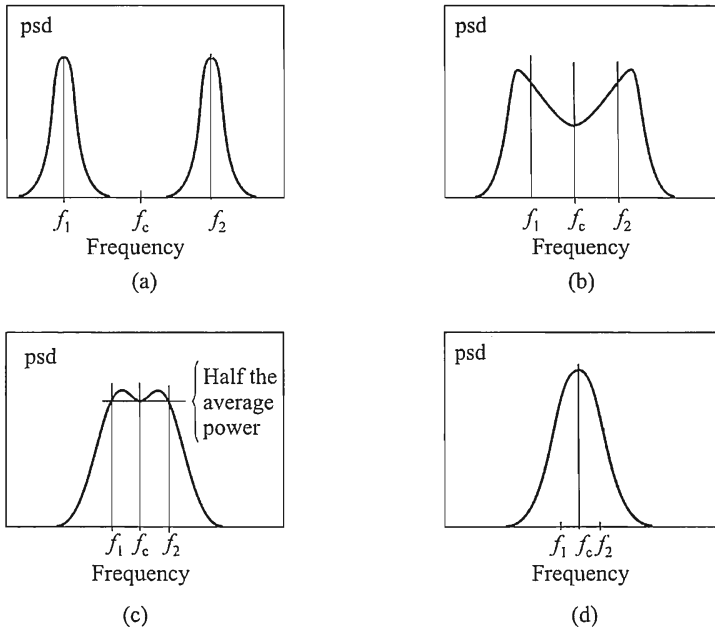
Figure 9.26 Definition of frequency resolution. In (a) the two signals are completely resolved. In (b) they are still resolvable but some of the energy from both signals is mixing. In (c) the two signals are just resolvable, where the psd halfway between them is half their average power (they need not have the same power). This is the definition of resolution. In (d) the signals are not resolvable at all.

$$\Delta f = \frac{1}{\Delta t} \tag{9.218}$$

which means that, for $M = 1$,

$$R \approx 2\pi \tag{9.219}$$

The two high-resolution techniques considered were the AR spectrum as calculated by the Berg method and the Pisarenko decomposition, while the two FT-based techniques are simply transforming the time series directly and the Blackman-Tukey periodogram with a few different windows. The results are illustrated in Figure 9.27 for several combinations of the variables. The limit in (9.219) is also plotted for comparison.
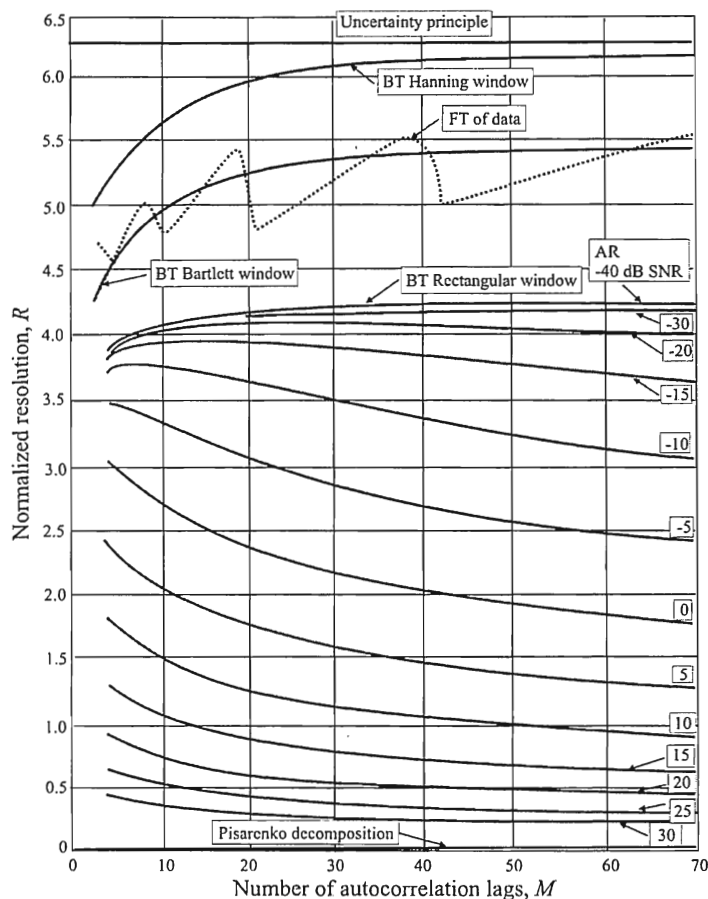
Figure 9.25 Comparisons of the frequency resolution capability of a few high-resolution spectral estimate techniques and FT-based techniques. (*From:* [23].)

The periodogram technique sells itself short when assuming that $\Delta f = 1/\Delta t$ when, in fact, it is a factor of about 1/3 less than that when using a rectangular window. The dotted line represents the mean value of the normalized resolution. The variance of this approach is quite large for all values of M, which in that case, is the number of samples used in the computations.

The superiority of the AR psd is evident over the periodogram for all values of SNR considered. The Pisarenko decomposition is the best of those considered,

however, recall that this technique generates spectral lines—that is, the bandwidth of the peaks detected is zero. Therefore there are no spectra to overlap unless $\Delta f = 0$. The more appropriate question for that technique is how accurate are the peaks located relative to the true position and are spurious peaks generated. From Figures 9.23 through 9.25, such spurs are possible.

9.6 Peak Determination

All of the high-resolution spectrum estimation techniques presented in this chapter produce spectra that have peaks. Two such spectra are illustrated in Figure 9.28 Figure 9.28(a) illustrates a spectrum when the SNR is high and there are two signals present. Figure 9.28(b) illustrates a spectrum when the SNR is low.

EW systems would, more often than not, have to deal with spectra that look like that in Figure 9.28(b). Targets are typically further away than friendly, interfering signals, and therefore subject to considerable noise.

The frequencies corresponding to the peaks in the spectra must be located, and there are several methods to accomplish this. One of the simplest methods is to employ a *central point search* where a few successive points (3 or 5) are typical. The highest point and its corresponding frequency are retained and then the next three points are examined: $k + 1$, $k + 2$, and $k + 3$. This process is repeated until the whole spectrum is examined. This technique will find all the peaks present. If the SNR is high, the peaks likely belong to the signals. If the SNR is low, several of the peaks correspond to noise. Unless some arbitrary rule is invoked, such as allowing no more than K signals or only allowing those peaks above some threshold, the peaks due to noise cannot be discerned from those corresponding to signals.

An algorithm was presented by Judd [24] that attempts to deal with this extraneous peak problem. In that algorithm the spectrum is first smoothed with an N-point *maximum envelope peak* (MEP) search. N adjacent points are examined at a time and only the maximum of the N points is retained. This process is repeated sequentially until all points have been examined. The MEP search is then followed by the central peak search using 3 or 5 points. In many cases the spurious peaks will be smoothed and will not cause false alarms.

9.7 Concluding Remarks

In general, when the characteristics of the data sequence are unknown, an ARMA model is better than simply the AR estimator because of the additional robustness of the ARMA model.

(a)

Frequency

(b)
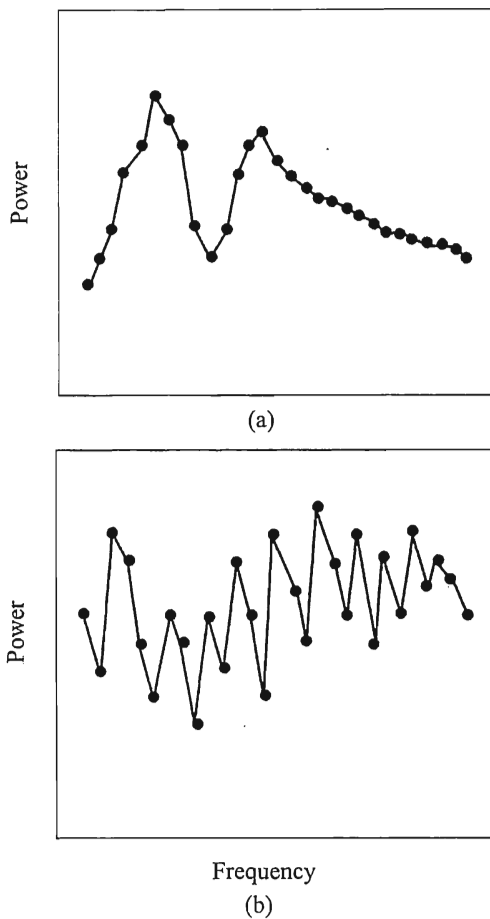
Figure 9.28 Psd spectra for (a) high SNR and (b) low SNR.

It is important to note that while some methods have higher resolution than others, these techniques can have bias associated with them. Therefore, though a high-resolution algorithm has better resolution than some others it may have worse accuracy due to this bias.

Table 9.1 summarizes the various high-resolution spectrum estimation methods discussed in this chapter and compares them to the two low-resolution techniques from Chapter 4.

Table 9.1 Spectral Estimation Techniques

Technique	Computational Complexity	Models	Advantages and Disadvantages	Remarks
Periodogram: FFT Version	$N\log_2 N$	Sum of harmonically related sinusoids	Output directly proportional to power Most computationally efficient Resolution roughly the reciprocal of the observation interval Performance poor for short data records Leakage distorts spectrum and masks weak signals	Harmonic least-squares fit Requires some type of frequency domain statistical averaging to stabilize spectrum Windowing can reduce sidelobes at expense of resolution
Periodogram: Blackman-Tukey	Lag: NM psd*: MS	Identical to MA with windowing of the lags	Most computationally efficient if $M << N$ Resolution roughly the reciprocal of the observation interval Leakage distorts spectrum and masks weak signals	Negative psd values in spectrum may result with some window weightings and autocorrelation estimates
AR Yule-Walker Version	Lag: NM AR Coeff: M^2 psd*: MS	Autoregressive process	Model order must be selected Better resolution than FFT or BT, but not as good as other AR methods Spectral line splitting occurs Implied windowing distorts spectrum No sidelobes	Model applicable to seismic, speech, radar clutter data Minimum-phase linear prediction filter guaranteed if biased log estimates computed AR related to linear prediction analysis and adaptive filtering

Table 9.1 (Continued)

Technique	Computational Complexity	Models	Advantages and Disadvantages	Remarks
AR Burg Version	Lag: $NM + M$ psd*: MS	Autoregressive process	High resolution for low noise levels Good spectral fidelity for short data records Spectral line splitting can occur Bias in the frequency estimates of peaks No implied windowing No sidelobes Must determine order	Stable linear prediction filter guaranteed Adaptive filtering applicable Uses constrained recursive least-squares approach
AR Least squares or forward backward linear prediction version	AR coeff: $MN + M^2$ psd*: MS	Autoregressive process	Sharper response for narrowband processes than other AR estimates No spectral line splitting observed Bias reduced in the frequency estimates Must determine order No sidelobes	Stable linear prediction filter not guaranteed, though stable filter results in most instances Based on exact recursive least squares solution with no constraint
MA	MA coeff: nonlinear simul. Eq. set Lag: NM Psd*: MS	Moving average process	Broad spectral responses (low resolution) Must determine order Has sidelobes	Generalized form of BT technique
ARMA (Yule Walker version)	Lag Ests: NM Coeff: M^3 psd*: MS	ARMA process	Must determine AR & MA orders	Models all rational transfer function processes Requires accurate log estimates to obtain good results

Table 9.1 (Continued)

Technique	Computational Complexity	Models	Advantages and Disadvantages	Remarks
Pisarenko Harmonic Decomposition	Lag: NM Eigen. eq; M^2 to M^3 Poly. rooting: dependent on root alg. Amp. coeff: M^3 psd*: M^3	Special ARMA with equal MA and AR coefficients Sum of nonharmonically undamped sinusoids in additive white noise	Must determine order Does not work well in high noise levels Eigen-equation and rooting are computationally inefficient	Uses least-squares estimates to obtain exponential parameters First step same as AR least-squares estimation
Prony's Method (Extended)	AR coeffs: $M^2 + NM$ Poly. rooting: dependent on root alg. Amp. coeff: M^3 psd*: MS	Sum of nonharmonically related damped sinusoids ARMA with equal MA and AR coefficients and equal orders	Must determine order Output linearly proportional to power Requires a polynomial rooting Resolution as good as AR technique, sometimes better No sidelobes	Uses least-squares estimation
Capon Maximum Likelihood	Lag: NM Matrix Inversion: M^3 psd*: MS	Forms an optimal bandpass filter for each spectral component	Resolution better than BT, not as good as AR	MLSE is related to AR spectra
MUSIC	Lag: NM psd: M^3	Sum of complex exponentials	Psd not very sensitive to model order p but very sensitive to signal count estimate	
ESPRIT	Lag: NM	Sum of complex exponentials in noise	Eigen-decomposition computed twice Search of entire frequency range not necessary, only close to unit circle	
Minimum Norm	Lag: NM psd: M^3	Sum of complex sinusoids		Similar to MUSIC

*FFT could be used to generate $S = 2^v$ values of the psd
N = Number of data samples
S = Number of spectral samples computed
M = Order of model (or number of autocorrelation lags)
After: [22].

References

[1] Hardin, J. C., *Introduction to Time Series Analysis,* NASA Reference Publication 1145, Hampton, VA: NASA, 1986, pp. 136–139.

[2] Gardner, W. A., *Statistical Spectral Analysis: A Nonprobabilistic Theory*, Englewood Cliffs, NJ: Prentice Hall, 1988, pp. 266–329.

[3] Schmidt, R., "Multiple Emitter Location and Signal Parameter Estimation," *IEEE Transactions on Antennas and Propagation,* Vol. AP–34, March 1986, pp. 276–290.

[4] Capon, J., "High Resolution Frequency-Wave Number Spectrum Analysis," *Proceedings of the IEEE,* Vol. 57, No. 57, August 1969, pp. 1408–1418.

[5] Poisel, R. A., *Modern Communications Jamming Principles and Techniques*, Norwood, MA: Artech House, 2003.

[6] Kay, S. M., *Modern Spectral Estimation Theory and Applications,* Upper Saddle River, NJ: Prentice Hall, 1988, p. 174.

[7] Burg, J. P., "Maximum Entropy Spectral Analysis," *Proceedings of 37th Meeting of the Society of Exploration Geophysics,* Oklahoma City, Oklahoma, October 1967.

[8] Shannon, C. E., "The Mathematical Theory of Communication," *Bell System Technical Journal,* Vol. 27, 1948, pp. 379–423 and 623–659. Also available from http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html and in book form from The University of Illinois Press, 1963.

[9] Cover, T. M., and J. A. Thomas, *Elements of Information Theory,* New York: John Wiley & Sons, 1991, p. 13.

[10] Cover, T. M., and J. A. Thomas, *Elements of Information Theory,* New York: John Wiley & Sons, 1991, p. 63.

[11] Ulrych, T. J., and T. N. Bishop, "Maximum Entropy Spectral Analysis and Autoregressive Decomposition," *Rev. Geophys,* Vol. 13, 1975, pp. 183–200.

[12] Fourgere, P. F., E. J. Zawaqlick, and H. R. Radoski, "Spontaneous Line Splitting in Maximum Entropy Power Spectrum Analysis," *Physical Earth and Plane. International,* Vol. 12, August 1976, pp. 201–209.

[13] Marple, S. L., Jr., "A New Autoregresssive Spectrum Analysis Algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* Vol. ASSP28, August 1980, pp. 441–454.

[14] Tsui, J., *Digital Techniques for Wideband Receivers*, Norwood, MA: Artech House, 1995, pp. 386–390.

[15] Zhang, Q. T., "A Statistical Resolution Theory of the AR Method of Spectral Analysis," *IEEE Transactions on Signal Processing,* Vol. 46, No. 10, October 1998, pp. 2757–2769.

[16] Prony, R., "Essai Expérimental et Analytique sur les Lois de la Dilatabilité de Fluides Elastiques et sur Celles de la Force Expansive de la Vapeur de L'Alkool, A Differentes Températures," *J. L'Ecole Polytech, (Paris),* Vol. 1, No. 2, 1795, pp. 24–79.

[17] Kahn, M., et al., "On the Consistency of Prony's Method and Related Algorithms," *Journal of Computational and Graphical Statistics,* Vol. 1, 1992, pp. 329–349.

[18] Osborne, M. R., "Some Special Nonlinear Least Squares Problems," *SIAM Journal of Numerical Analysis,* Vol. 12, 1975, pp. 571–592.

[19] Smyth, G. K., "Coupled and Nested Iterations in Nonlinear Estimation,". Ph.D. thesis, Australian National University, Canberra, 1985.

[20] Osborne, M. R. and G. K. Smyth, "A Modified Prony Algorithm for Fitting Functions Defined by Difference Equations," *SIAM Journal of Scientific and Statistical Computing*, Vol. 12, 1991, pp. 362-382.

[21] Tsui, J. P., *Digital Techniques for Wideband Receivers,* Norwood, MA: Artech House, 1995, pp. 410–412.

[22]   Kay, S. M., and S. L. Marple, "Spectrum Analysis—A Modern Perspective," *Proceedings of the IEEE,* Vol. 69, No. 11, November 1981, pp. 1380–1419.

[23]   Marple, L., "Frequency Resolution of High-Resolution Spectrum Analysis Techniques," *Proceedings of the RADC Spectrum Estimation Workshop,* May 24–26, 1978, pp. 19–39.

[24]   Judd, M. D., "A Simple, Low-Computation Peak Detection Algorithm for the Angle-of Arrival Spectrum for Signal Subspace Methods," *IEEE Transactions on Aerospace and Electronic Systems,* Vol. 28, No. 4, October 1992, pp. 1158–1163.

Chapter 10

Artificial Reasoning for Target Recognition

As pointed out at the beginning of Chapter 9, the process of recognizing targets in communication EW systems consists generally of two steps:

1. Extract parameters (also called features) from the intercepted signal by measurement.
2. Process these parameters by some applicable technique to ascertain whether the signal is one of interest.

This chapter presents two techniques applicable to the second step in this process. They both are from the generic field of artificial intelligence, where computers are programmed to rudimentarily mimic the way humans think. The techniques are evidential reasoning and fuzzy logic.

10.1 Evidential Reasoning

A way to allow for some uncertainty in machine reasoning processes is to use the *evidential reasoning* theory discovered by Dempster and Schafer and subsequently documented by the latter [1]. Input data are allowed to support a proposition, refute a proposition, or have no affect on a proposition.

A *proposition* is either a basic hypothesis, as in the Bayesian case, or a combination of hypotheses. Although hypotheses cannot overlap, propositions can. In fact, a proposition can consist of hypotheses that have conflicting elements.

Let $\Theta = \{A_1, A_2, ..., A_n\}$ represent the complete and mutually exclusive set of possible propositions, called the *elemental propositions*. Θ is called the *frame of*

discernment. There are $2^n - 1$ general propositions that can be developed from Θ, which, in addition to the elemental propositions, consist of the disjunction of the elements of Θ. (Actually the power set usually includes the null subset $\varnothing$, which brings the number of elements to 2^n. That is not included here.) This is called the power set of Θ, and is usually denoted by 2^Θ. Let A^Θ denote the set of all of the possible distinctions of these elementary propositions. That is,

$$A^\Theta = \{A_1 \vee A_2, A_1 \vee A_3, ..., A_1 \vee A_2 \vee ... \vee A_n\} \tag{10.1}$$

and $2^\Theta = \{\Theta, A^\Theta\}$. In this equation, $\vee$ refers to the logical, or Boolean, OR function.

Because the processing involved with evidential reasoning involves the power set of propositions, usually problems that are solved using the technique are limited in size. The power set quickly gets very large.

For each $\theta_i \in 2^\Theta$, a probability mass, $m(\theta_i)$, is assigned that represents the amount of evidence associated with θ_i. This mass is assigned such that

$$m(\theta_i) \leq 1 \tag{10.2}$$

and

$$\sum_i m(\theta_i) = 1 \tag{10.3}$$

From these probability masses, the probability of an elementary proposition A_i is given by

$$P(A_i) = \sum_{\theta_i \subseteq A_i} m(\theta_i) \tag{10.4}$$

Thus, the probability of A_i is given by summing all of the probability masses associated with elements in 2^Θ that are subsets of A_i. For example, if $\Theta = \{a_1, a_2, a_3\}$, $A^\Theta = \{a_1 \vee a_2, a_1 \vee a_3, a_2 \vee a_3, a_1 \vee a_2 \vee a_3\}$, then $P(a_1 \vee a_3) = m(a_1) + m(a_1 \vee a_3) + m(a_3)$.

The last element of 2^Θ, that is, $A_1 \vee A_2 \vee A_3 \vee ... \vee A_n$ is called the *granular proposition* and represents the amount that is unknown, that is, the amount of uncertainty.

Belief in a proposition K is given by an interval whose lower bound is the support for K, denoted by spt(K) and whose upper limit is given by the plausibility

of K, denoted by $\text{pls}(K)$. K could be an elementary proposition or it could be a combination of elementary propositions. These values are given by

$$\text{spt}(K) \;=\; \sum_{\theta_i \subseteq K} m(\theta_i) \tag{10.5}$$

and

$$\text{pls}(K) = 1 - \text{spt}(\neg K) \tag{10.6}$$

where $\neg K$ is not K. The support is calculated by summing all of the probability masses of elements $\theta_i \in 2^\Theta$ that are contained in K, either alone or in combination. The plausibility is the lack of evidence that refutes proposition K. The probability of proposition is bounded by these two quantities, called the *evidential interval*:

$$\text{spt}(K) \leq P(K) \leq \text{pls}(K) \tag{10.7}$$

The net result of these calculations at any given stage in the processing is a set of probability intervals given by

$$[\text{spt}(A_1), \text{pls}(A_1)]$$
$$[\text{spt}(A_2), \text{pls}(A_2)]$$
$$\cdots$$
$$[\text{spt}(A_n), \text{pls}(A_n)]$$
$$[\text{spt}(A_1 \vee A_2), \text{pls}(A_1 \vee A_2)]$$
$$[\text{spt}(A_1 \vee A_3), \text{pls}(A_1 \vee A_3)]$$
$$\cdots$$
$$[\text{spt}(A_1 \vee A_2 \vee \ldots \vee A_n), \text{pls}(A_1 \vee A_2 \vee \ldots \vee A_n)]$$

or, in general,

$$[\text{spt}(K_j), \text{pls}(K_j)] \text{ for all } K_j \in 2^\Theta$$

The following example illustrates these concepts.

Example:

Suppose the frame of discernment consists of the 36 exhaustive and mutually exclusive elementary propositions A_1 through A_{36} shown in Figure 10.1. Suppose
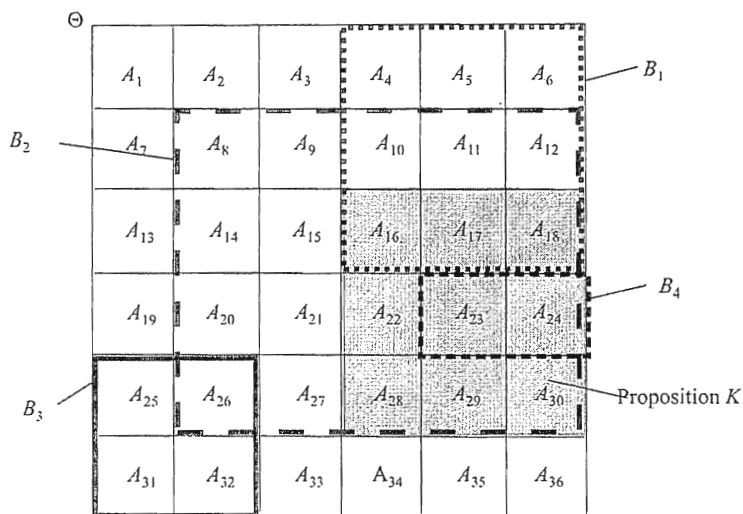
Figure 10.1 Frame of discernment for the example.

that there are four propositions under consideration, B_1 through B_4, which are composed of some of these elementary propositions, and there is a question about proposition K, shown in the figure.

Here,

$$\Theta = \{A_1, A_2, A_3, ..., A_{36}\}$$
$$A^{\Theta} = \{A_1 \vee A_2, A_1 \vee A_3, ..., A_1 \vee A_2 \vee A_3 \vee A_4 \vee ... \vee A_{36}\}$$
$$2^{\Theta} = \{A_1, A_2, A_3, ..., A_{36}, A_1 \vee A_2, A_1 \vee A_3, ..., A_1 \vee A_2 \vee A_3 \vee A_4 \vee ... \vee A_{36}\}$$

A_i represents an elementary proposition about Θ, which are exhaustive (complete) and mutually exclusive. In symbology,

$$K = \{A_{16} \vee A_{17} \vee A_{18} \vee A_{22} \vee A_{23} \vee A_{24} \vee A_{28} \vee A_{29} \vee A_{30}\}$$
$$B_1 = \{A_4 \vee A_5 \vee A_6 \vee A_{10} \vee A_{11} \vee A_{12} \vee A_{16} \vee A_{17} \vee A_{18}\}$$
$$B_2 = \{A_8 \vee A_9 \vee A_{10} \vee A_{11} \vee A_{14} \vee A_{15} \vee A_{16} \vee A_{17} \vee A_{20} \vee A_{21} \vee A_{22} \vee A_{23} \vee A_{26} \vee A_{27} \vee A_{28} \vee A_{29}\}$$
$$B_3 = \{A_{25} \vee A_{26} \vee A_{31} \vee A_{32}\}$$
$$B_4 = \{A_{23} \vee A_{24}\}$$

Suppose that probability masses have been established by some mechanism to be

$m(B_1) = 0.3$
$m(B_2) = 0.2$
$m(B_3) = 0.4$
$m(B_4) = 0.1$
$m(\cdot) = 0$ otherwise.

In this example, $\theta_i \in \{B_1, B_2, B_3, B_4\}$, whereas in general, unless otherwise eliminated from consideration for practical reasons, the B_i's would consist of most, if not all, of 2^Θ. Then

$$\text{spt}(K) = \sum_{\theta_i \subseteq K} m(\theta_i) = m(B_4) = 0.1$$

$$\text{spt}(\neg K) = \sum_{\theta_i \not\subseteq K} m(\theta_i) = m(B_3) = 0.4$$

$$\text{pls}(K) = 1 - \text{spt}(\neg K) = 0.6$$

Thus, the evidential interval for K is $[0.1, 0.6]$.

∎

10.1.1 Rules of Combination

The probability intervals computed as above can be combined in order to compute an overall probability of an event. The combinations must be based on evidence from independent sources over the same frame of discernment Θ in order for the theory to apply, however.

Dempster's rule of combination can be viewed as a generalization of the Bayesian rule of combination—they yield the same results in the same circumstances. Let θ_3 represent the new element of 2^Θ that is represented by combining θ_1 and θ_2.

Property: Dempster's Rule of Combination

For all θ_1, θ_2, $\theta_3 \in 2^\Theta$

$$m_3(\theta_3) = \frac{1}{1-k} \sum_{\theta_1 \cap \theta_2 = \theta_3} m_1(\theta_1) m_2(\theta_2) \tag{10.8}$$

where
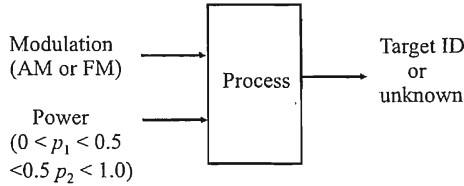
Figure 10.2 Simplified example of target recognition.

$$k = \sum_{\theta_1 \cap \theta_2 = \varnothing} m_1(\theta_1)m_2(\theta_2) \tag{10.9}$$

In these equations, $m_1(\theta_1)$ refers to the probability mass assigned by logic system 1 for $\theta_1 \in 2^{\Theta}$, while $m_2(\theta_1)$ refers to the probability mass assigned by logic system 2 for $\theta_1 \in 2^{\Theta}$. Obviously, Θ must be the same for the two logic systems.

∎

Example:

As a simple example of target recognition using evidential reasoning, consider the block diagram shown in Figure 10.2. The elementary events in this case are given by the four two-tuples as:

A_1: <modulation = AM, power = p_1>
A_2: <modulation = AM, power = p_2>
A_3: <modulation = FM, power = p_1>
A_4: <modulation = FM, power = p_2>

There are $2^4 - 1 = 15$ elements of 2^{Θ}. Suppose that the following mass functions have been assigned by some means.

$\theta_1 = A_1$, $m_1(\theta_1) = 0.10$, $m_2(\theta_1) = 0.09$
$\theta_2 = A_2$, $m_1(\theta_2) = 0.09$, $m_2(\theta_2) = 0.09$
$\theta_3 = A_3$, $m_1(\theta_3) = 0.03$, $m_2(\theta_3) = 0.04$
$\theta_4 = A_4$, $m_1(\theta_4) = 0.01$, $m_2(\theta_4) = 0.01$
$\theta_5 = A_1 \vee A_2$, $m_1(\theta_5) = 0.15$, $m_2(\theta_5) = 0.10$
$\theta_6 = A_1 \vee A_3$, $m_1(\theta_6) = 0.05$, $m_2(\theta_6) = 0.12$
$\theta_7 = A_1 \vee A_4$, $m_1(\theta_7) = 0.04$, $m_2(\theta_7) = 0.02$
$\theta_8 = A_2 \vee A_3$, $m_1(\theta_8) = 0.06$, $m_2(\theta_8) = 0.06$
$\theta_9 = A_2 \vee A_4$, $m_1(\theta_9) = 0.02$, $m_2(\theta_9) = 0.04$
$\theta_{10} = A_3 \vee A_4$, $m_1(\theta_{10}) = 0.04$, $m_2(\theta_{10}) = 0.06$

$\theta_{11} = A_1 \vee A_2 \vee A_3$, $m_1(\theta_{11}) = 0.10$, $m_2(\theta_{11}) = 0.07$
$\theta_{12} = A_1 \vee A_2 \vee A_4$, $m_1(\theta_{12}) = 0.06$, $m_2(\theta_{12}) = 0.08$
$\theta_{13} = A_1 \vee A_3 \vee A_4$, $m_1(\theta_{13}) = 0.02$, $m_2(\theta_{13}) = 0.04$
$\theta_{14} = A_2 \vee A_3 \vee A_4$, $m_1(\theta_{14}) = 0.09$, $m_2(\theta_{14}) = 0.02$
$\theta_{15} = A_1 \vee A_2 \vee A_3 \vee A_4$, $m_1(\theta_{15}) = 0.14$, $m_2(\theta_{15}) = 0.16$

Scholium

Some of these compound sets have meaning as defined by the constituent events. These events are:

$\theta_5 = A_1 \vee A_2$, Is the modulation AM?
$\theta_6 = A_1 \vee A_3$, Is the power p_1?
$\theta_7 = A_1 \vee A_4$, Is there any signal present?
$\theta_8 = A_2 \vee A_3$, Is there any signal present?
$\theta_9 = A_2 \vee A_4$, Is the power p_2?
$\theta_{10} = A_3 \vee A_4$, Is the modulation FM?
$\theta_{15} = A_1 \vee A_2 \vee A_3 \vee A_4$, Is there any signal present?

■

Next, calculate the combined probability masses:

$k = m_1(\theta_1)m_2(\theta_2) + m_1(\theta_1)m_2(\theta_3) + m_1(\theta_1)m_2(\theta_4) + m_1(\theta_1)m_2(\theta_8) + m_1(\theta_1)m_2(\theta_9) + m_1(\theta_1)m_2(\theta_{10}) + m_1(\theta_1)m_2(\theta_{14}) + m_1(\theta_2)m_2(\theta_1) + m_1(\theta_2)m_2(\theta_3) + m_1(\theta_2)m_2(\theta_4) + m_1(\theta_2)m_2(\theta_6) + m_1(\theta_2)m_2(\theta_7) + m_1(\theta_2)m_2(\theta_{10}) + m_1(\theta_2)m_2(\theta_{13}) + m_1(\theta_3)m_2(\theta_1) + m_1(\theta_3)m_2(\theta_2) + m_1(\theta_3)m_2(\theta_4) + m_1(\theta_3)m_2(\theta_5) + m_1(\theta_3)m_2(\theta_7) + m_1(\theta_3)m_2(\theta_9) + m_1(\theta_3)m_2(\theta_{12}) + m_1(\theta_4)m_2(\theta_1) + m_1(\theta_4)m_2(\theta_2) + m_1(\theta_4)m_2(\theta_3) + m_1(\theta_4)m_2(\theta_5) + m_1(\theta_4)m_2(\theta_6) + m_1(\theta_4)m_2(\theta_8) + m_1(\theta_4)m_2(\theta_{11}) + m_1(\theta_5)m_2(\theta_3) + m_1(\theta_5)m_2(\theta_4) + m_1(\theta_5)m_2(\theta_{10}) + m_1(\theta_6)m_2(\theta_2) + m_1(\theta_6)m_2(\theta_4) + m_1(\theta_6)m_2(\theta_9) + m_1(\theta_7)m_2(\theta_2) + m_1(\theta_7)m_2(\theta_3) + m_1(\theta_7)m_2(\theta_8) + m_1(\theta_8)m_2(\theta_1) + m_1(\theta_8)m_2(\theta_4) + m_1(\theta_8)m_2(\theta_7) + m_1(\theta_9)m_2(\theta_1) + m_1(\theta_9)m_2(\theta_3) + m_1(\theta_9)m_2(\theta_6) + m_1(\theta_{10})m_2(\theta_1) + m_1(\theta_{10})m_2(\theta_2) + m_1(\theta_{10})m_2(\theta_5) + m_1(\theta_{11})m_2(\theta_4) + m_1(\theta_{12})m_2(\theta_3) + m_1(\theta_{13})m_2(\theta_2) + m_1(\theta_{14})m_2(\theta_1)$
$= 0.1526$

$m_3(\theta_1) = [m_1(\theta_1)m_2(\theta_1) + m_1(\theta_1)m_2(\theta_5) + m_1(\theta_1)m_2(\theta_6) + m_1(\theta_1)m_2(\theta_7) + m_1(\theta_1)m_2(\theta_{11}) + m_1(\theta_1)m_2(\theta_{12}) + m_1(\theta_1)m_2(\theta_{13}) + m_1(\theta_1)m_2(\theta_{15})] / (1 - 0.1526) = 0.0802$

$m_3(\theta_2) = [m_1(\theta_2)m_2(\theta_2) + m_1(\theta_2)m_2(\theta_5) + m_1(\theta_2)m_2(\theta_8) + m_1(\theta_2)m_2(\theta_9) + m_1(\theta_2)m_2(\theta_{11}) + m_1(\theta_2)m_2(\theta_{12}) + m_1(\theta_2)m_2(\theta_{14}) + m_1(\theta_2)m_2(\theta_{15})] / 0.8474 = 0.0658$

$$m_3(\theta_3) = [m_1(\theta_3)m_2(\theta_3) + m_1(\theta_3)m_2(\theta_6) + m_1(\theta_3)m_2(\theta_8) + m_1(\theta_3)m_2(\theta_{10}) + m_1(\theta_3)m_2(\theta_{11}) + m_1(\theta_3)m_2(\theta_{13}) + m_1(\theta_3)m_2(\theta_{14}) + m_1(\theta_3)m_2(\theta_{15})] / 0.8474 = 0.0202$$

$$m_3(\theta_4) = [m_1(\theta_4)m_2(\theta_4) + m_1(\theta_4)m_2(\theta_7) + m_1(\theta_4)m_2(\theta_9) + m_1(\theta_4)m_2(\theta_{10}) + m_1(\theta_4)m_2(\theta_{12}) + m_1(\theta_4)m_2(\theta_{13}) + m_1(\theta_4)m_2(\theta_{14}) + m_1(\theta_4)m_2(\theta_{15})] / 0.8474 = 0.0058$$

$$m_3(\theta_5) = [m_1(\theta_5)m_2(\theta_5) + m_1(\theta_5)m_2(\theta_{11}) + m_1(\theta_5)m_2(\theta_{12}) + m_1(\theta_5)m_2(\theta_{15})] / 0.8474 = 0.0549$$

$$m_3(\theta_6) = [m_1(\theta_6)m_2(\theta_6) + m_1(\theta_6)m_2(\theta_{11}) + m_1(\theta_6)m_2(\theta_{13}) + m_1(\theta_6)m_2(\theta_{15})] / 0.8474 = 0.0230$$

$$m_3(\theta_7) = [m_1(\theta_7)m_2(\theta_7) + m_1(\theta_7)m_2(\theta_{12}) + m_1(\theta_7)m_2(\theta_{13}) + m_1(\theta_7)m_2(\theta_{15})] / 0.8474 = 0.0142$$

$$m_3(\theta_8) = [m_1(\theta_8)m_2(\theta_8) + m_1(\theta_8)m_2(\theta_{11}) + m_1(\theta_8)m_2(\theta_{14}) + m_1(\theta_8)m_2(\theta_{15})] / 0.8474 = 0.0219$$

$$m_3(\theta_9) = [m_1(\theta_9)m_2(\theta_9) + m_1(\theta_9)m_2(\theta_{12}) + m_1(\theta_9)m_2(\theta_{14}) + m_1(\theta_9)m_2(\theta_{15})] / 0.8474 = 0.0071$$

$$m_3(\theta_{10}) = [m_1(\theta_{10})m_2(\theta_{10}) + m_1(\theta_{10})m_2(\theta_{13}) + m_1(\theta_{10})m_2(\theta_{14}) + m_1(\theta_{10})m_2(\theta_{15})] / 0.8474 = 0.0132$$

$$m_3(\theta_{11}) = [m_1(\theta_{11})m_2(\theta_{11}) + m_1(\theta_{11})m_2(\theta_{15})] / 0.8474 = 0.0271$$

$$m_3(\theta_{12}) = [m_1(\theta_{12})m_2(\theta_{12}) + m_1(\theta_{12})m_2(\theta_{15})] / 0.8474 = 0.0170$$

$$m_3(\theta_{13}) = [m_1(\theta_{13})m_2(\theta_{13}) + m_1(\theta_{13})m_2(\theta_{15})] / 0.8474 = 0.0047$$

$$m_3(\theta_{14}) = [m_1(\theta_{14})m_2(\theta_{14}) + m_1(\theta_{14})m_2(\theta_{15})] / 0.8474 = 0.0191$$

$$m_3(\theta_{15}) = m_1(\theta_{15})m_2(\theta_{15}) / 0.8474 = 0.0264$$

Now the evidential intervals will be calculated based on the combined mass functions.

$$spt(\theta_1) = m_3(\theta_1) = 0.0802$$

$pls(\theta_1) = 1 - spt(\neg\theta_1) = 1 - [m_3(\theta_2) + m_3(\theta_3) + m_3(\theta_4) + m_3(\theta_8) + m_3(\theta_9) + m_3(\theta_{10})$
$+ m_3(\theta_{14})] = 1 - [0.0658+ 0.0202+ 0.0058+ 0.0219 + 0.0071 + 0.0132 + 0.0191] =$
0.8469

$spt(\theta_2) = m_3(\theta_2) = 0.0658$

$pls(\theta_2) = 1 - spt(\neg\theta_2) = 1 - [m_3(\theta_1) + m_3(\theta_3) + m_3(\theta_4) + m_3(\theta_6) + m_3(\theta_7) + m_3(\theta_{10})$
$+ m_3(\theta_{13})] = 1 - [0.0802 + 0.0202 + 0.0058+ 0.0230 + 0.0142 + 0.0132 + 0.0047]$
$= 0.8387$

$spt(\theta_3) = m_3(\theta_3) = 0.0202$

$pls(\theta_3) = 1 - spt(\neg\theta_3) = 1 - [m_3(\theta_1) + m_3(\theta_2) + m_3(\theta_4) + m_3(\theta_5) + m_3(\theta_7) + m_3(\theta_9) +$
$m_3(\theta_{12})] = 1 - [0.0802 + 0.0658 + 0.0058 + 0.0549 + 0.0142 + 0.0071+ 0.0170] =$
0.7550

$spt(\theta_4) = m_3(\theta_4) = 0.0058$

$pls(\theta_4) = 1 - spt(\neg\theta_4) = 1 - [m_3(\theta_1) + m_3(\theta_2) + m_3(\theta_3) + m_3(\theta_5) + m_3(\theta_6) + m_3(\theta_8) +$
$m_3(\theta_{11})] = 1 - [0.0802 + 0.0658 + 0.0202 + 0.0549 + 0.0230 + 0.0219 + 0.0271] =$
0.7069

$spt(\theta_5) = m_3(\theta_5) + m_3(\theta_1) + m_3(\theta_2) = 0.0549 + 0.0802 + .0658 = 0.2009$

$pls(\theta_5) = 1 - spt(\neg\theta_5) = 1 - [m_3(\theta_3) + m_3(\theta_4) + m_3(\theta_{10})] = 1 - [0.0202 + 0.0058 +$
$0.0132] = 0.961$

$spt(\theta_6) = m_3(\theta_6) + m_3(\theta_1) + m_3(\theta_3) = 0.0230 + 0.0802 + 0.0202 = 0.1234$

$pls(\theta_6) = 1 - spt(\neg\theta_6) = 1 - [m_3(\theta_2) + m_3(\theta_4) + m_3(\theta_9)] = 1 - [0.0658 + 0.0058 +$
$0.0071] = 0.9213$

$spt(\theta_7) = m_3(\theta_7) + m_3(\theta_1) + m_3(\theta_4) + m_3(\theta_8) = 0.0142 + 0.0802 + 0.0058 = 0.1002$

$pls(\theta_7) = 1 - spt(\neg\theta_7) = 1 - [m_3(\theta_2) + m_3(\theta_3) + m_3(\theta_4) + m_3(\theta_8)] = 1 - [0.0658 +$
$0.0202 + 0.0219] = 0.8921$

$spt(\theta_8) = m_3(\theta_8) + m_3(\theta_2) + m_3(\theta_3) = 0.0219 + 0.0658 + 0.0202 = 0.1079$

$pls(\theta_8) = 1 - spt(\neg\theta_8) = 1 - [m_3(\theta_1) + m_3(\theta_4) + m_3(\theta_7)] = 1 - [0.0802 + 0.0058 + 0.0142] = 0.8998$

$spt(\theta_9) = m_3(\theta_9) + m_3(\theta_1) + m_3(\theta_3) = 0.0071 + 0.0802 + 0.0202 = 0.0893$

$pls(\theta_9) = 1 - spt(\neg\theta_9) = 1 - [m_3(\theta_1) + m_3(\theta_3) + m_3(\theta_6)] = 1 - [0.0802 + 0.0202 + 0.0230] = 0.8766$

$spt(\theta_{10}) = m_3(\theta_{10}) + m_3(\theta_3) + m_3(\theta_4) = 0.0132 + 0.0202 + 0.0058 = 0.0392$

$pls(\theta_{10}) = 1 - spt(\neg\theta_{10}) = 1 - [m_3(\theta_1) + m_3(\theta_2) + m_3(\theta_5)] = 1 - [0.0802 + 0.0658 + 0.0549] = 0.7791$

$spt(\theta_{11}) = m_3(\theta_{11}) + m_3(\theta_1) + m_3(\theta_2) + m_3(\theta_3) = 0.0271 + 0.0802 + 0.0658 + 0.0202 = 0.1933$

$pls(\theta_{11}) = 1 - spt(\neg\theta_{11}) = 1 - m_3(\theta_4) = 1 - 0.0058 = 0.9942$

$spt(\theta_{12}) = m_3(\theta_{12}) + m_3(\theta_1) + m_3(\theta_2) + m_3(\theta_4) = 0.0170 + 0.0802 + 0.0658 + 0.0058 = 0.1688$

$pls(\theta_{12}) = 1 - spt(\neg\theta_{12}) = 1 - m_3(\theta_3) = 1 - 0.0202 = 0.9798$

$spt(\theta_{13}) = m_3(\theta_{13}) + m_3(\theta_1) + m_3(\theta_3) + m_3(\theta_4) = 0.0047 + 0.0802 + 0.0202 + 0.0058 = 0.1109$

$pls(\theta_{13}) = 1 - spt(\neg\theta_{13}) = 1 - m_3(\theta_2) = 1 - 0.0658 = 0.9342$

$spt(\theta_{14}) = m_3(\theta_{14}) + m_3(\theta_2) + m_3(\theta_3) + m_3(\theta_4) = 0.0191 + 0.0658 + 0.0202 + 0.0058 = 0.1109$

$pls(\theta_{14}) = 1 - spt(\neg\theta_{14}) = 1 - m_3(\theta_1) = 1 - 0.0802 = 0.9198$

$spt(\theta_{15}) = m_3(\theta_{15}) + m_3(\theta_1) + m_3(\theta_2) + m_3(\theta_3) + m_3(\theta_4) = 0.0264 + 0.0802 + 0.0658 + 0.0202 + 0.0058 = 0.1984$

$pls(\theta_{15}) = 1 - spt(\neg\theta_{15}) = 1$

The results of these calculations are summarized in Table 10.1 for convenience.

Table 10.1 Support and Plausibility Summary

	Support	Plausibility
θ_1	0.0802	0.8469
θ_2	0.0658	0.8387
θ_3	0.0202	0.7550
θ_4	0.0058	0.7069
θ_5	0.2009	0.9610
θ_6	0.1234	0.9213
θ_7	0.1002	0.8921
θ_8	0.1079	0.8998
θ_9	0.0893	0.8766
θ_{10}	0.0392	0.7791
θ_{11}	0.1933	0.9942
θ_{12}	0.1688	0.9798
θ_{13}	0.1109	0.9342
θ_{14}	0.1109	0.9198
θ_{15}	0.1984	1.0000

Note that the support for the combined individual propositions is less than the initial probability masses, whereas support for all but one combination (θ_{10}) is larger. This indicates that in this situation there is considerable uncertainty as to which target is present—an indication that additional data need to be collected before a firm prediction can be made. This is further confirmed by the support for the granular proposition, θ_{15}, which has one of the largest values of support so far. The only one that is larger is $\theta_5 = A_1 \vee A_2$, indicating that there is considerable lack of knowledge if the signal is AM or not. There is also considerable support for $\theta_{11} = A_1 \vee A_2 \vee A_3$, which would tend to indicate that the signal is probably not FM with power p_2.

∎

The computations shown above are both associative and communitative. Therefore, it matters not in which order they are computed. The computations can therefore be optimized for the hardware available.

10.1.2 Limitations of the Dempster-Shafer Method

In general, a probability mass needs to be associated with every element of 2^Θ, whereas for the Bayesian system, probabilities are only necessary for elements of Θ. This adds substantially to the amount of calculations necessary, going from a set on the order of n to one on the order of 2^n. As illustrated in the last example, the number of combinations explodes. For realistic problems, this is a significant concern, typically limiting the algorithm to relatively small problem sets.

When combining probability masses, both bodies of evidence to be combined need to be independent and the errors restricted to measurement errors. This is difficult to ensure.

Determining the mass functions can be a problem and is perhaps one of the more difficult issues in using this approach. In some cases, for example, when raw sensor data is being combined, the mass functions are relatively easy to assign. In others, such as the examples used here, it might be much more difficult.

10.2 Fuzzy Logic

The utility of fuzzy logic arises because of its ability to mimic the way humans reason. A human often does not require all the data necessary to make a decision or to otherwise take an action. Normal computer logic does require all the data, every time, and if false data is provided a computer will come to a false conclusion. As with evidential reasoning, fuzzy logic allows for some degree of tolerance for not knowing everything [2–7].

The material in the first part of this section was derived principally from the excellent tutorial paper by Mendel [8].

10.2.1 Fuzzy and Crisp Sets

To understand fuzzy logic, it is convenient to begin with the definitions of (normal, nonfuzzy, crisp) sets and operations on those sets. These notions will then be extended to fuzzy sets and fuzzy logic.

A *set* is any collection of objects, such as the members of a baseball team or the actors in a play. A *crisp set* is a set whose membership can be distinctly delineated, such as those sets just given, or the numbers between 0 and 100 that are evenly divisible by 31 (= {0, 31, 62, 93}). Crisp set C can be characterized by a *membership function* $\mu_C(x)$ such that

$$\mu_C(x) = \begin{cases} 1 & \text{if } x \in C \\ 0 & \text{if } x \notin C \end{cases} \tag{10.10}$$

where $\in$ means "is a member of" and $\notin$ means "is not a member of." If C is the set just defined, then $\mu_C(0) = 1$, $\mu_C(31) = 1$, $\mu_C(62) = 1$, $\mu_C(93) = 1$, and $\mu_C(x_i) = 0$ for all other $x_i \in U = \{0, 1, 2, 3, ...,100\}$. Thus membership functions for crisp sets consist of the values 1 and 0.

A *universe of discourse U* is the totality of elements at issue for a particular problem. For the example given above, $U = \{0, 1, 2, 3, ..., 100\}$. The *cross product*

of two such universes U and V, denoted by $U \times V$, is the complete set of 2-tuples (u, v), where $u \in U$ and $v \in V$. That is, it's simply the ordered pairing of elements of the universes taking one element at a time from each of the universes. A single "element" of $U \times V$ is (u, v). That is, $(u, v) \in U \times V$.

The operations of union, intersection, and complement are defined in the usual way. For crisp sets C_1 and C_2 consisting of elements x_i selected from the universe of discourse U, their *crisp union* is given by

$$C_1 \cup C_2 = \left\{ x_i \in U \, \middle| \, x_i \in C_1 \text{ or } x_i \in C_2 \text{ or both} \right\} \tag{10.11}$$

For those unfamiliar with this notation, this expression should be read: $C_1 \cup C_2$ is the set consisting of those elements that are in C_1 or are in C_2 or both.

Their *crisp intersection* is given by

$$C_1 \cap C_2 = \left\{ x_i \in U \, \middle| \, x_i \in C_1 \text{ and } x_i \in C_2 \right\} \tag{10.12}$$

The crisp *complement* of C_1 is given by

$$\neg C_1 = \left\{ x_i \in U \, \middle| \, x_i \notin C_1 \right\} \tag{10.13}$$

Membership functions exist for both crisp and fuzzy sets. The membership functions for fuzzy sets, can, and normally do, consist of more than just 0 and 1, however. Let X describe a universe of discourse, which even for fuzzy sets is not fuzzy itself, and let $F \subseteq X$ be a fuzzy set on X. $\mu_F(x)$ denotes its membership function. This function describes the degree of similarity of x to the fuzzy set F and is normally a connected curve, although not necessarily continuous nor smooth, describing this similarity. Its values normally range from 0 to 1, where $\mu_F(x_i) = 0$ means there is no similarity between x_i and the set F and $\mu_F(x_i) = 1$ means there is total similarity between x_i and F. Values in between 0 and 1 for $\mu_F(x)$ are allowed; $\mu_F(x_i) = 0.6$, for example, would indicate that x_i is more similar to F than it is dissimilar.

Therefore, a fuzzy set is totally defined by the variables in the set and the membership function. The notation for a fuzzy set is

$$F = \left\{ x, \mu_F(x) \, \middle| \, x_i \in X \right\} \tag{10.14}$$

As an example of a fuzzy set, suppose F is the set "short American men." Its membership function might be as shown in Figure 10.3. Any American man less
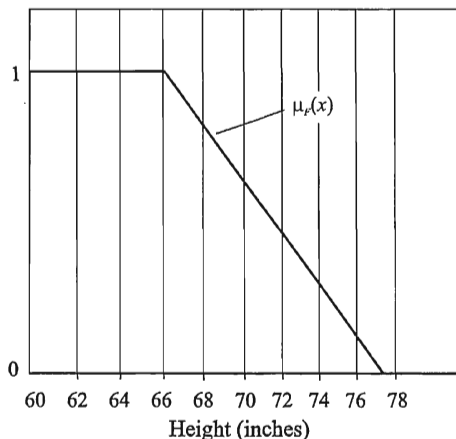
Figure 10.3 Example of a membership function.

than 66 inches is considered short. If a man is taller than 66 inches, then he belongs to a set that is considered short on a linearly decreasing scale until he reaches 77 inches, at which point, if he is larger than 77 inches, he is no longer considered short at all.

Definition: *Support* for fuzzy set F is the crisp set of all points x in U such that $\mu_F(x) > 0$.

In the above example, the support for F is given by

$$Support\ F = \{x_i \mid x_i \le 77 \text{ inches}\}$$

Definition: The element x in U at which $\mu_F(x) = 0.5$ is called a *crossover point*. F, whose support is a single point $x \in U$ with $\mu_F(x) = 1$, is called a *fuzzy singleton*.

Just as there are fundamental operations on crisp sets, there are fundamental operations defined for fuzzy sets. If F and G are two fuzzy sets, then:

Definition: *Fuzzy Union:* For $F \cup G$, the membership function is given by $\mu_{F \cup G}(x) = \max[\mu_F(x),\ \mu_G(x)]$, where "max(arguments)" refers to the maximum arithmetic value of the arguments. This is also defined as the smallest fuzzy set containing both F and G.

Definition: *Fuzzy Intersection*: For $F \cap G$, the membership function is given by $\mu_{F \cap G}(x) = \min[\mu_F(x), \mu_G(x)]$, where "min(arguments)" refers to the smallest arithmetic values of the arguments. It is also defined as the largest fuzzy set that is contained in both F and G.

Definition: *Fuzzy Complement*: The complement of F, denoted by $\neg F$ has a membership function given by $\mu_{\neg F}(x) = 1 - \mu_F(x)$.

In crisp logic there is a law, called the *law of the excluded middle*, which says that

$$F \cup \neg F = U \qquad (10.15)$$

which means that a set and its complement, taken together, defines the entire universe of discourse. Fuzzy logic, on the other hand, does not have this same condition present. That is, normally

$$F \cup \neg F \neq U \qquad (10.16)$$

Likewise there is a law, called the *law of contradiction*, in crisp logic that says that the intersection of a set with its complement must be the null set, the set with no elements in it. In symbols,

$$F \cap \neg F = \varnothing \qquad (10.17)$$

where $\varnothing$ is the null set. In fuzzy logic, this law does not apply; that is, normally

$$F \cap \neg F \neq \varnothing \qquad (10.18)$$

10.2.2 Relationships

Relationships, or functions on sets, describe mappings from one set to another. A relation $y = f(x)$ can be described as sets of tuples, that is, members (often numbers) in pairs given by (x, y).

If $x \in U$ and $y \in V$, then a crisp relation $R(U, V)$ can be characterized by

$$\mu_R(x, y) = \begin{cases} 1, & (x, y) \in R(U, V) \\ 0, & \text{otherwise} \end{cases} \qquad (10.19)$$

where $R(U, V)$ describes the relationship between U and V. This form is useful for relationships between finite sets. It is often inconvenient for infinite sets, including continuous functions.

Fuzzy relations represent a degree of presence or absence of association between the elements of two or more fuzzy sets. An example of a binary fuzzy relationship is "x is much larger than y."

Let U and V be two universes of discourse. A fuzzy relationship $R(U,V)$ is a fuzzy set in $U \times V$.

$$R(U,V) = \left\{ (x,y), \mu_R(x,y) \big| (x,y) \in U \times V \right\} \tag{10.20}$$

Definition: *Containment*: (also called subset): F is contained in G if and only if $\mu_F(x) \leq \mu_G(x)$:

$$F \subseteq G \Leftrightarrow \mu_F(x) \leq \mu_G(x) \tag{10.21}$$

Figure 10.4 shows these operations on the two fuzzy sets $F1$ and $F2$.

Sometimes alternative definitions of AND and OR are used. This is to simplify the computations required. Two of these definitions are

$$\mu_{F \cap G}(x) = \mu_F(x) \mu_G(x) \tag{10.22}$$

that is, the algebraic product, and
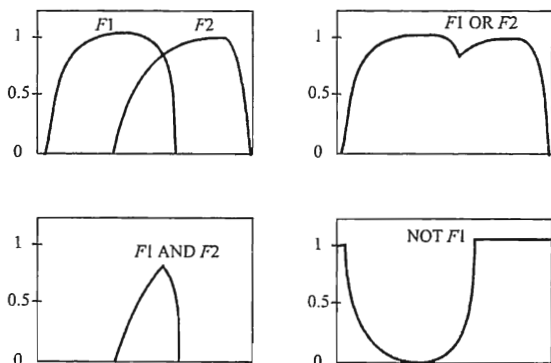
Figure 10.4 Operations on fuzzy sets.
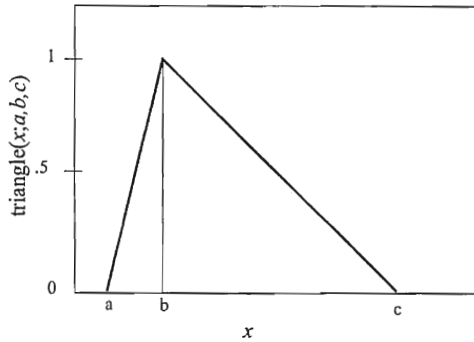
Figure 10.5 Triangular membership function.

$$\mu_{F \cup G}\left(x\right) = \mu_F\left(x\right) + \mu_G\left(x\right) - \mu_F\left(x\right)\mu_G\left(x\right) \tag{10.23}$$

Note that these do not evaluate to the same numbers arithmetically as those definitions given previously, but in some cases effective use can be made, especially in those cases when real-time computations are required. Herein the definitions given first above will be used exclusively.

10.2.3 Common Membership Functions

A common set of mathematical functions is often used in order to simplify the mathematics used in computing fuzzy logic functions [9]. In this section a few of these are presented.

10.2.3.1 Triangular Membership Function

$$\text{triangle}(x; a, b, c) = \max\left(\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right),\, 0\right) \tag{10.24}$$

The parameters a, b, and c specify the three corners of the triangle; see Figure 10.5.

10.2.3.2 Trapezoidal Membership Function

The trapezoid membership function is given by

$$\text{trapezoid}(x; a, b, c, d) = \max\left(\min\left[\frac{x-a}{b-a}, \ 1, \ \frac{d-x}{d-c}\right], \ 0\right) \qquad (10.25)$$

The parameters a, b, c, and d specify the corners of the trapezoid; see Figure 10.6.

10.2.3.3 Gaussian Membership Function

The Gaussian membership function is given by [10]

$$\text{Gaussian}(x; \sigma, \mu) = e^{-\left[\frac{(x-\mu)}{\sigma}\right]^2} \qquad (10.26)$$

where c is the mean of the Gaussian function and σ is the standard deviation. Shown in Figure 10.7 is Gaussian$(x;1,0)$.

10.2.3.4 Generalized Bell Membership Function

The generalized bell membership function is given by [10]

$$\text{bell}(x; a, b, c) = \frac{1}{1 + \left|\frac{x-c}{a}\right|^{2b}}$$

where c is the center of the curve, $c - a$ is the lower point where the curve is 0.5, $c + a$ is the point where the curve is also 0.5, and the slopes at these points are $b/2a$ and $-b/2a$, respectively. Shown in Figure 10.8 is the generalized bell membership function.

10.2.3.5 Logistic Function

The logistic is also a useful continuous membership function [11]. Its equation is given by

$$\text{logistic}(x, k) = \frac{1}{1 - e^{-kx}} \qquad (10.27)$$

This curve is plotted in Figure 10.9 for $k = 1$.
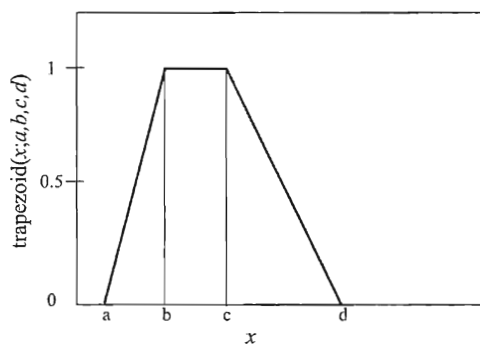
Figure 10.6 Trapezoidal membership function.
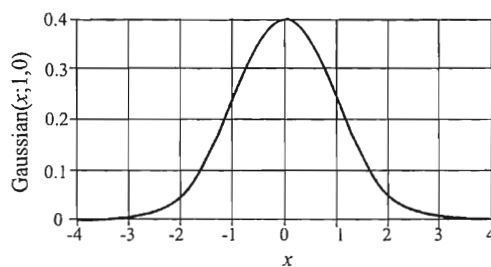
Figure 10.7 Gaussian membership function.
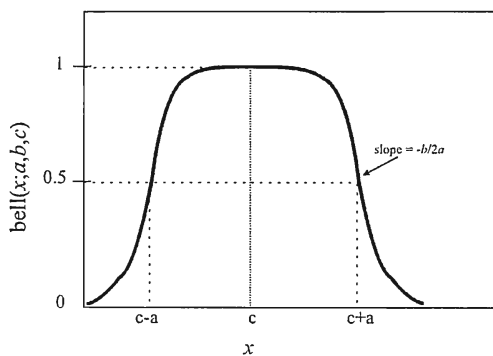
Figure 10.8 Generalized bell membership function.
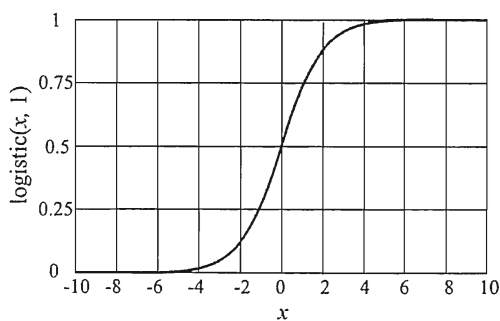
Figure 10.9 The logistic membership function for k = 1.
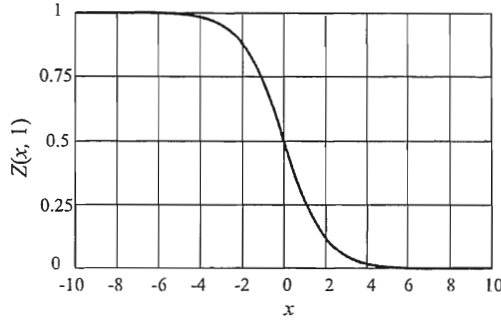
Figure 10.10 Z-function.

10.2.3.6 Z-Function

The Z-function is the logistic function reversed. Its equation is given by

$$Z(x,k) = 1 - \frac{1}{1 + e^{-kx}} \tag{10.28}$$

which is plotted in Figure 10.10 for $k = 1$.

10.2.3.7 Sigmoid Function

The Sigmoid function [12, 13] (or S-function) is similar to the logistic function. It is given by

$$S(x; a, b, c) = \begin{cases} 0, & x \le a \\ 2\left(\dfrac{x-a}{c-a}\right)^2, & a < x \le b \\ 1 - 2\left(\dfrac{x-a}{c-a}\right)^2, & b < x \le c \\ 1, & x > c \end{cases} \tag{10.29}$$

and is illustrated in Figure 10.11.
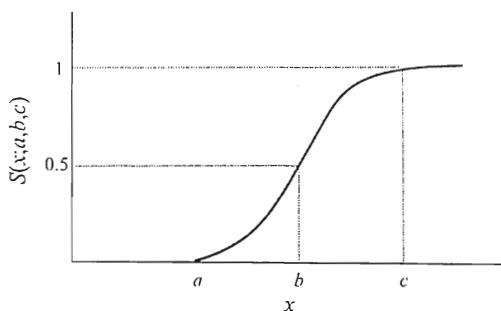
Figure 10.11 Sigmoid function.

10.2.3.8 π-Function

The π-function   [14]   is useful as a continuous function that is bounded at both ends. It is given by

$$\pi(x;a,b) = \begin{cases} S(x;a-b,a), & x \leq a \\ 1-S\left(x;a+\dfrac{b}{2},a+b\right), & x > a \end{cases}$$ (10.30)

This function is illustrated in Figure 10.12.

10.2.4 Fuzzy If-Then Rules

Fuzzy if-then rules, or simply fuzzy rules, are given by
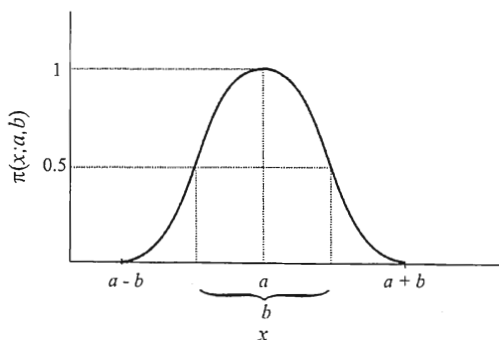
Figure 10.12 π-function.

if x is A then y is B

where A and B are linguistic values defined by fuzzy sets on universes of discourse X and Y, respectively. Examples of linguistic values are "short American men" and "poor troop morale." These values take on ranges, as discussed above, as opposed to specific values, and therefore "x is A" might mean "Joe is a short American man."

In this expression, "x is A" is referred to as the antecedent or premise, and "y is B" is referred to as the consequence or conclusion. Continuing the example with short American men, if "x is A" means "Joe is a short American man" is the antecedent, then "y is B" might mean "the left door is tall enough." Here, y is the left door and B is the set of all doors in the universe of discourse (presumably in the same room with at least a door on the left and another door someplace else).

10.2.5 Fuzzy Reasoning

The fundamental concepts involved with reasoning with fuzzy logic principles are presented in this section.

10.2.5.1 Generalized Modus Ponens

The fundamental law of inference for crisp sets is called *modus ponens* and it says that if A is true and A implies B, then B is true. "A implies B" is referred to as an *implication*. There is an equivalent rule in fuzzy logic that generalizes this concept, called *generalized modus ponens*, which says:

Premise 1 (fact or, in reality, often an assumption): x is A'
Premise 2 (rule):   if x is A then y is B
Consequence (conclusion): y is B'

where A' is *close* to A and B' is *close* to B. The meaning of "close" will become apparent subsequently.

10.2.5.2 Fuzzy Reasoning Based on Max Min Composition

Let A, A', and B be fuzzy sets of X, X, and Y, respectively. Assume that the fuzzy implication $A \rightarrow B$ is expressed as a fuzzy relation R on $X \times Y$. Then the fuzzy set B' induced by "x is A'" and the fuzzy rule "if x is A then y is B" is defined by
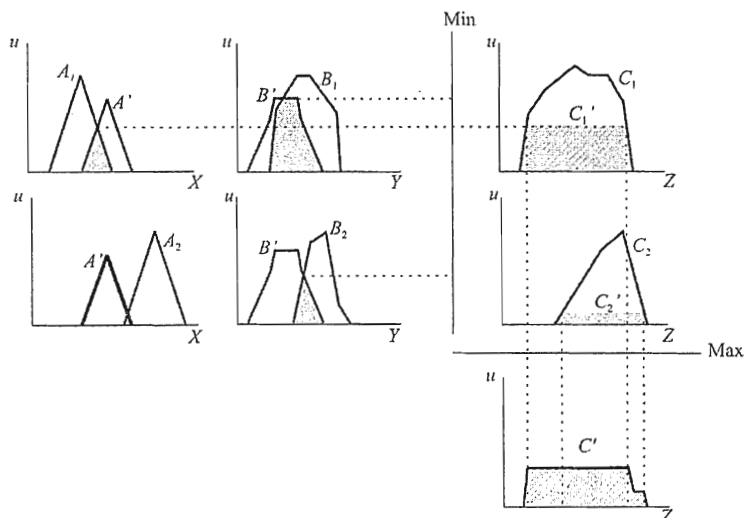
Figure 10.13 Max min composition on fuzzy sets.

$$\mu_{B'}(y) = \max_x \min\left[\mu_{A'}(x), \mu_R(x,y)\right] \tag{10.31}$$

This is called the *max min composition* for fuzzy sets.

For the target recognition application being considered here, the last step (maximization) is not necessary. The target with the largest area under its associated C_i curve would be the identified one, assuming that this area is above some minimum threshold value.

The above described one relationship between fuzzy sets. This can be extended to more than one antecedent or rule as follows. In Figure 10.13,

      Premise 1 (fact or assumption): x is A' and y is B'
      Premise 2 (rule 1):   if x is A_1 and y is B_1 then z is C_1
      Premise 3 (rule 2):   if x is A_2 and y is B_2 then z is C_2
      Consequence (conclusion): z is C'

The shaded areas on the left four charts represent the common area beneath the membership functions, which is the membership function of the intersections. The largest value of these membership functions is projected to the right onto the lines labeled "Min." At that point, the minimum value is selected, which is then projected further to the right onto the Z membership functions C_i. These minimums form a type of weighting for the C_i membership functions. They can be

thought of as the amount of weight given to their respective C_i in computing the max min composition. The minimums of the C_i thus selected are projected downward to the line labeled "Max" where the maximum, or union, function is performed. The resultant membership function is the maximum of each of the individual weighted membership functions.

The question still remains: What is the significance of the last statement "z is C'"? This will be addressed below, after an example is given to illustrate the points just covered.

Example: (Fuzzy Logic)

For a specific example of the use of fuzzy logic to recognize targets, suppose there are six signal parameters necessary given by:

(1) Signal bandwidth expressed as a fraction of the channel width;
(2) Signal power (variance), normalized to a maximum value so the range is (0, 1);
(3) Line of bearing;
(4) Amount of power from an AM demodulator, normalized to a maximum value of 1;
(5) Amount of power from an FM discriminator, normalized to a maximum value of 1;
(6) SNR of the predetected signal.

Membership functions for the bandwidth as a fraction of the channel width is shown in Figure 10.14, where low, moderate, and high spectral occupancy form the categories. There is nothing magic about these curves. They are best estimates of what might be good functions to use to categorize channel occupancy. These functions have been linearized for simplicity in presentation.

These membership functions are given by

$$\mu_{\text{LowBW}}(x) = \text{trapezoid}(x; 0, 0, 0.15, 0.25)$$
$$\mu_{\text{ModerateBW}}(x) = \text{trapezoid}(x; 0.20, 0.40, 0.60, 0.70)$$
$$\mu_{\text{HighBW}}(x) = \text{trapezoid}(x; 0.50, 0.70, 1.0, 1.0)$$

The membership functions for the amount of power in the predetected signal are shown in Figure 10.15. These relationships are given by the following:
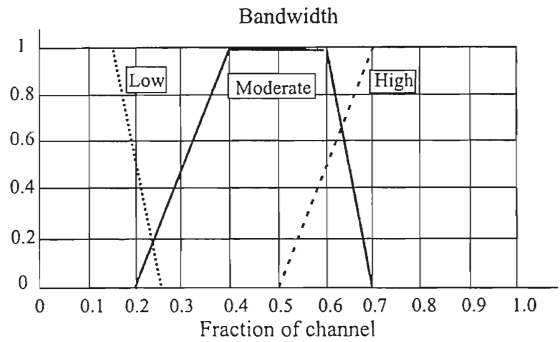
Bandwidth

Figure 10.14 Membership function for signal bandwidth expressed as a fraction of the channel width for the example.

Power (variance)
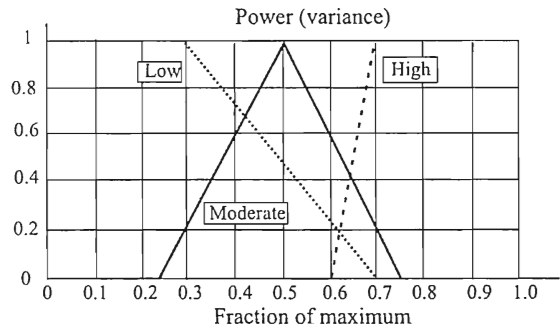
Figure 10.15 Membership function for signal power.
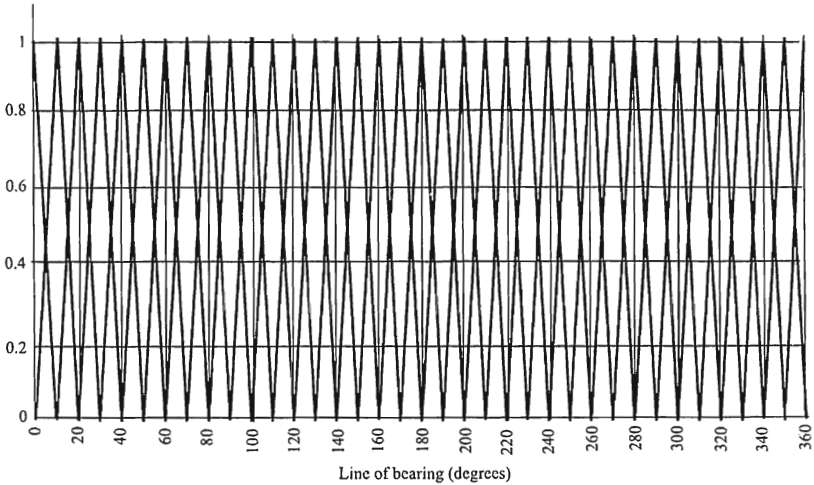
Figure 10.16 Membership functions for the line of bearing.

$$\mu_{\text{LowPower}}(x) = \text{trapezoid}(x; 0, 0, 0.30, 0.70)$$
$$\mu_{\text{ModeratePower}}(x) = \text{triangle}(x; 0.25, 0.50, 0.75)$$
$$\mu_{\text{HighPower}}(x) = \text{trapezoid}(x; 0.60, 0.70, 1.0, 1.0)$$

The membership functions for the line of bearing (LOB) are shown in Figure 10.16. There is a membership function defined for every $10°$ increment in azimuth that is $20°$ wide at the bottom. The ith function is given by

$$\mu_{\text{LOB}_i}(x) = \text{triangle}(x; i-10, i, i+10)$$

The amount of AM on a signal, whether intended or unintended, can be a sorting parameter. Unintentional AM on an FM signal, for example, can be caused by a malfunctioning power supply in the transmitter, which would tend to identify that particular transmitter until the power supply is fixed. The AM modulation membership functions are shown in Figure 10.17. Thus,
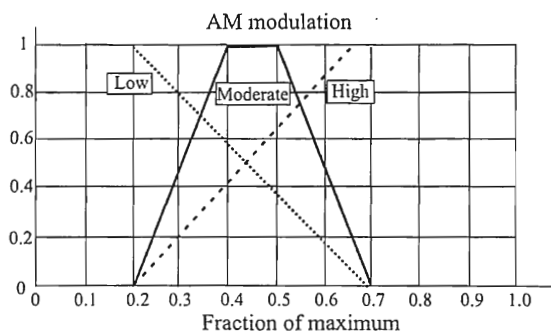
Figure 10.17 Membership function for AM.

$$\mu_{\text{LowAM}}(x) = \text{trapezoid}(x; 0, 0, 0.20, 0.70)$$
$$\mu_{\text{ModerateAM}}(x) = \text{trapezoid}(x; 0.20, 0.40, 0.50, 0.70)$$
$$\mu_{\text{HighAM}}(x) = \text{trapezoid}(x; 0.20, 0.70, 1.0, 1.0)$$

Unintentional FM on an AM or PM signal can also be a sorting parameter. Such FM on an AM signal, for example, could be caused by a local oscillator that is unstable and varying in frequency. This also would tend to stay with that transmitter until it is fixed. Therefore,

$$\mu_{\text{LowFM}}(x) = \text{trapezoid}(x; 0, 0, 0.25, 1.0)$$
$$\mu_{\text{ModerateFM}}(x) = \text{trapezoid}(x; 0.25, 040, 0.70, 1.0)$$
$$\mu_{\text{HighFM}}(x) = \text{triangle}(x; 0.25, 1.0, 1.0)$$

These membership functions are shown in Figure 10.18.

The last membership functions, pertaining to the SNR of the predetected signal, are shown in Figure 10.19. The higher the SNR, in general, the better signal recognition results ensue, and it is important to factor that into the process. A high SNR is also an indication of the proximity of the transmitter to the EW system. The membership functions are given by

$$\mu_{\text{LowSNR}}(x) = \text{trapezoid}(x; 0, 0, 10, 30)$$
$$\mu_{\text{ModerateSNR}}(x) = \text{trapezoid}(x; -2, 10, 15, 30)$$
$$\mu_{\text{HighSNR}}(x) = \text{trapezoid}(x; -2, 27, 45, 45)$$
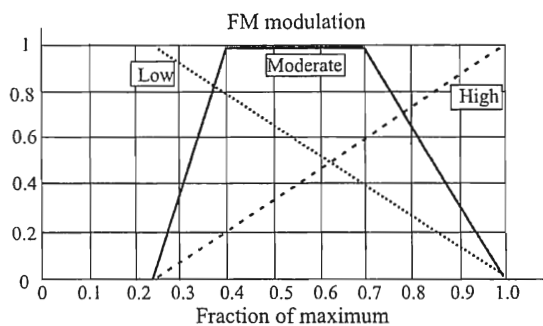
Figure 10.18 Membership function for FM.
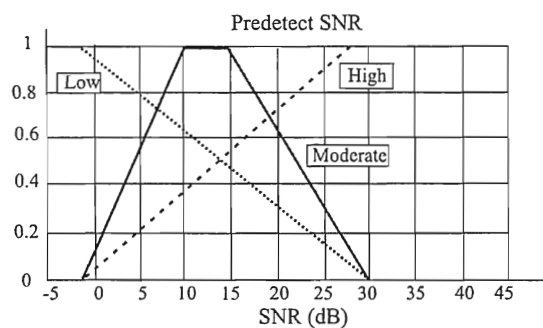
Figure 10.19 Membership function for predetection SNR.
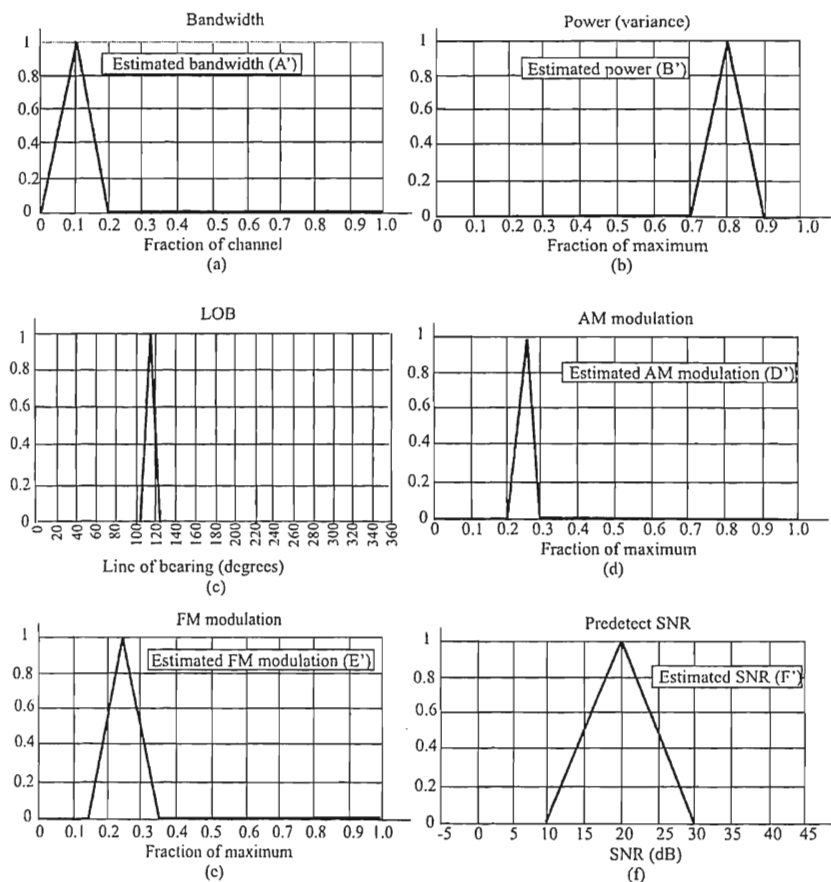
Figure 10.20 Premises for the example.

Given these fuzzy sets of membership functions, we will now proceed to establish the logical inferences of the problem. Premise 1 is given by

$$a \text{ is } A', \ b \text{ is } B', \ c \text{ is } C', \ d \text{ is } D', \ e \text{ is } E' \text{ and } f \text{ is } F'$$

where A', B', C', D', E', and F' are as shown in Figure 10.20. These premises are the estimates of the parameters established with the techniques described in

Chapter 9. The estimation procedure can provide a measure of the confidence in the estimate, and that confidence measure can be used to define the membership functions.

In this example, the estimate of the bandwidth is fairly low—between 0 and 20% of the channel width, with the average at 10%, as seen in Figure 10.20(a). The power in the signal is significant, being close to the maximum as seen in Figure 10.20(b). The estimated LOB, whose membership function is shown in Figure 10.20(c), is 118°, with an associated estimate of the possible error in that value. The membership function for the estimate of the amount of AM modulation in the signal is shown in Figure 10.20(d). There is about 25% of the maximum amount of such modulation in this case. Likewise for the FM modulation, except the membership function is somewhat wider, as shown in Figure 10.20(e). Finally, the predetection SNR membership function, shown in Figure 10.20(e), indicates that the SNR is about 20 dB, with variability of ±10 dB.

The next step is to apply the logical if-then statements, or simply rules. For illustration purposes, the following two will be used:

Rule 1:

If   the fraction of the channel occupied is (low or moderate);
and   the power is (moderate or high);
and   the LOB is (120°);
and   the AM modulation is (> 10%);
and   the FM modulation is (moderate or high);
and   the predetection SNR is (moderate or high);
Then   the target is T_1.

Rule 2:

If   the fraction of the channel occupied is (low);
and   the power is (moderate);
and   the LOB is (110°);
and   the AM modulation is (moderate or high);
and   the FM modulation is (high);
and   the predetection SNR is (moderate or high);
Then   the target is T_2.

In these rules, the parentheses indicate that the logical function contained therein should be evaluated first—the so-called *order of precedence* in mathematics.

The first of these rules with the premise overlaid is shown in Figure 10.21. Notice the graph of where the OR functions have been computed which represent

the max values of each of the individual curves involved. The shaded regions comprise the common areas of the membership functions and the premise—it is the highest point on these membership curves that is used for the max min composition. These points are carried to the right in the figure by the dotted lines. The 45° lines on the right simply reflect the values onto a common line so that the overall minimum can be obtained; they do not change the values. These lines are necessary only to fit all of the parts of the figure on one page. The depiction of the second rule is shown in Figure 10.22, where the same layout applies.

The minimum values from these curves are selected to determine the magnitude of the conclusion sets, as shown in Figure 10.23.

Output

The center of mass of the calculated minimum functions can be calculated with

$$z' = \frac{\int_z z\,\mu(z)dz}{\int_z \mu(z)dz} \tag{10.32}$$

for continuous $\mu(z)$ and

$$z' = \frac{\sum_i z_i\,\mu(z_i)}{\sum_i \mu(z_i)} \tag{10.33}$$

if the membership function is discrete. This is a measure of the center of the membership function, and, if z were a random variable and $\mu(z)$ were its probability density function, this is the equation of the mean. In this example, the target is identified as T_1 since the area under its curve is larger.

10.3 Concluding Remarks

This chapter presented two methods for computer reasoning with the estimates made from measured data in a communication EW system. The estimates are obtained with the methods presented in Chapter 9, and the meanings of the estimates are derived with the methods presented in this chapter.
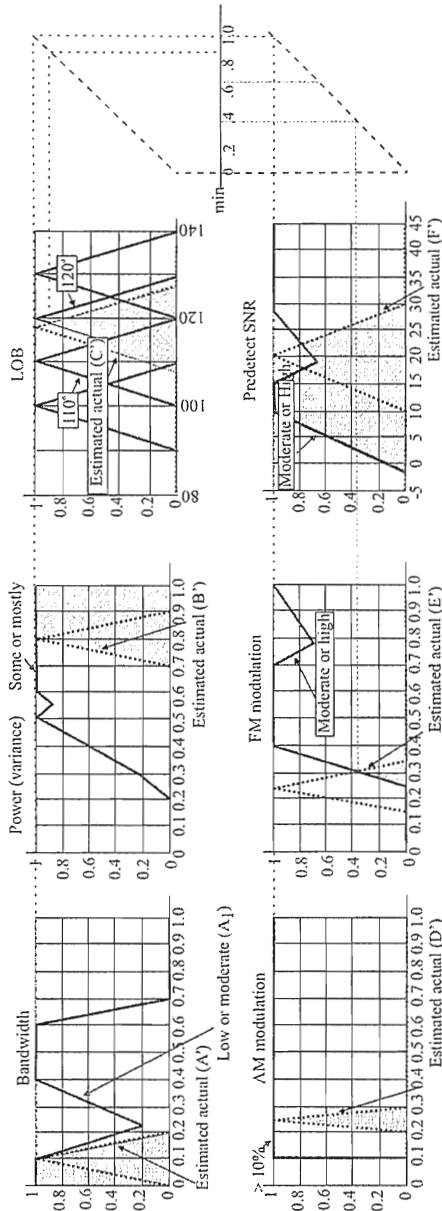
Figure 10.21 Max min diagrams for rule 1.
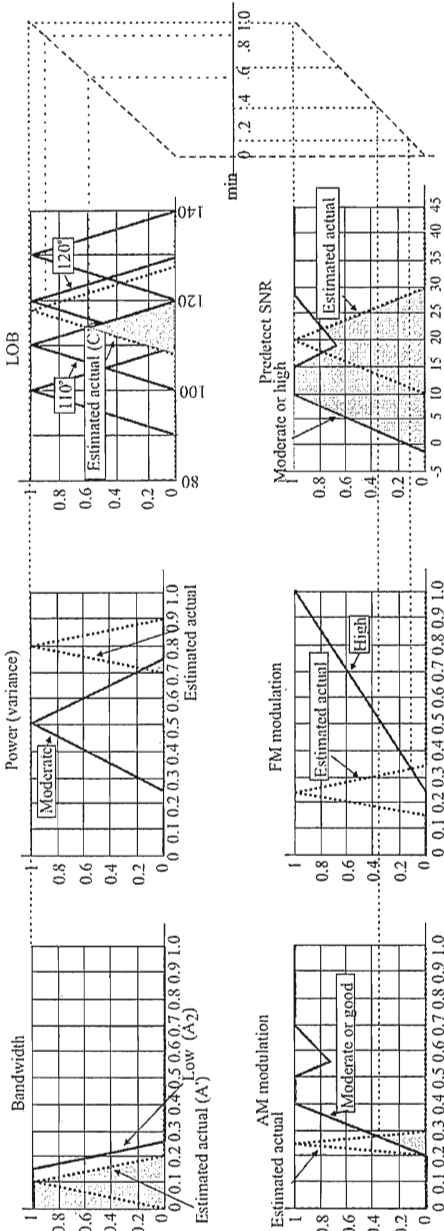
Figure 10.22 Min max diagrams for rule 2.
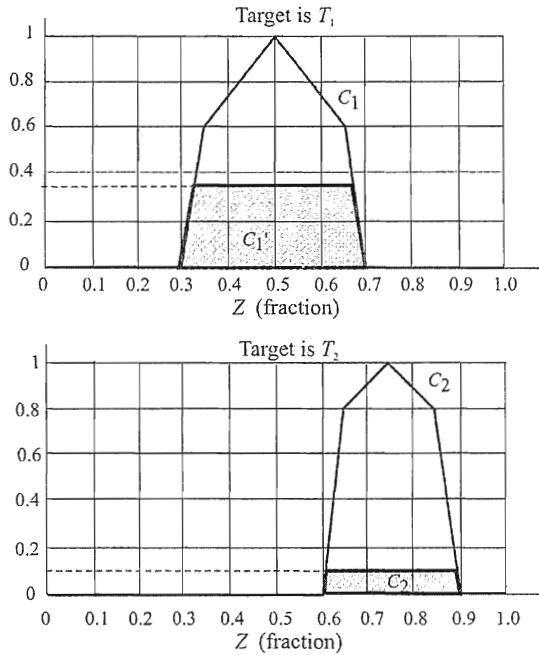
Figure 10.23 Conclusion sets.

The two techniques discussed, evidential reasoning and fuzzy logic, are certainly not the only ones available for computer reasoning. They are, however, relatively straightforward and easy to understand. Both are readily implemented on general purpose computers.

References

[1]   Shafer, G., *A Mathematical Theory of Evidence,* Princeton, NJ: Princeton University Press, 1976.

[2]   Kosko, B., *Fuzzy Engineering,* Upper Saddle River, NJ: Prentice Hall, 1991.

[3]   Kosko, B. *The Fuzzy Future,* New York: Harmony Books, 1999.

[4]   Zadeh, L., K.-S. Fu, K. Tanaka, and M. Shimura, eds., *Fuzzy Sets and Their Application to Cognitive and Decision Processes,* Boston: Academic Press, 1975.

[5]   McNeill, F. M., and R. R. Yager, *Fuzzy Logic: A Practical Approach*, Boston: Academic Press, 1994.

[6]   Dubois, D., and H. Prade, *Fuzzy Sets and Systems: Theory and Applications,* Boston: Academic Press, 1980.

[7]   Cox, E., *The Fuzzy System Handbook: A Practitioner's Guide to Building, Using, and Maintaining Fuzzy Systems,* Boston: Academic Press, 1994.

[8]   Mendel, J. M., "Fuzzy Logic Systems for Engineering: A Tutorial," *Proceedings of the IEEE,* Vol. 83, No. 3, March 1995, pp. 345–3710.

[9]   Jang, J. R. and C. Sun, "Neuro-Fuzzy Modeling and Control," *Proceedings of the IEEE,* Vol. 83, No. 3, March 1995, pp. 378–405.

[10]   Cox, E., *The Fuzzy System Handbook: A Practitioner's Guide to Building, Using, and Maintaining Fuzzy Systems,* Boston: Academic Press, 1994, pp. 705–710.

[11]   McNeill, F. M., and R. R. Yager, *Fuzzy Logic: A Practical Approach,* Boston: Academic Press, p. 34.

[12]   Zadeh, L., et al. (eds.), *Fuzzy Sets and Their Application to Cognitive and Decision Processes,* Boston: Academic Press, 1975, p. 30.

[13]   Cox, E., *The Fuzzy System Handbook: A Practitioner's Guide to Building, Using, and Maintaining Fuzzy Systems,* Boston: Academic Press, 1994, pp. 52–62.

[14]   Cox, E., *The Fuzzy System Handbook: A Practitioner's Guide to Building, Using, and Maintaining Fuzzy Systems,* Boston: Academic Press, 1994, pp. 605–610.

Chapter 11

Resource Allocation

When a signal is detected by some mechanism, in order to determine if it's an SOI, further processing is normally required. Signal processors typically perform this processing in the system, and the number of them is normally limited. Thus, conflicts arise with their utilization. The effects of limiting the number of signal processors, or other system resources, are discussed in this chapter.

11.1 Queues

Queuing theory proves useful for the analysis of such conflicts. This theory was first investigated by A. K. Erlang (1878–1929), a Dutch mathematician, for the study of blocking and waiting times in telephone systems [1]. Allen presented an excellent introduction to queuing theory in designing telephone systems [2]. His particular emphasis was on communication and computing system applications, but the results are applicable to communication EW systems as well.

Queuing theory is the study of waiting lines, or queues [3]. The principal components are an input source (population), the queue, the queue discipline, and one or more servers. A block diagram of a notional queuing system is shown in Figure 11.1.

The input source produces customers. These customers are generated according to a statistical distribution that describes their interarrival time at the queue input. These customers join the queue, waiting for their turn to use a server. The servers will select the next customer to service according to the queue discipline. The head of the queue is the customer that arrived first. The tail of the queue can mean two things: either all of the whole queue except the head, or the last customer in the queue. Which meaning applies depends on the context. The size of the population can be finite or infinite—most of the time it is assumed to be infinite.
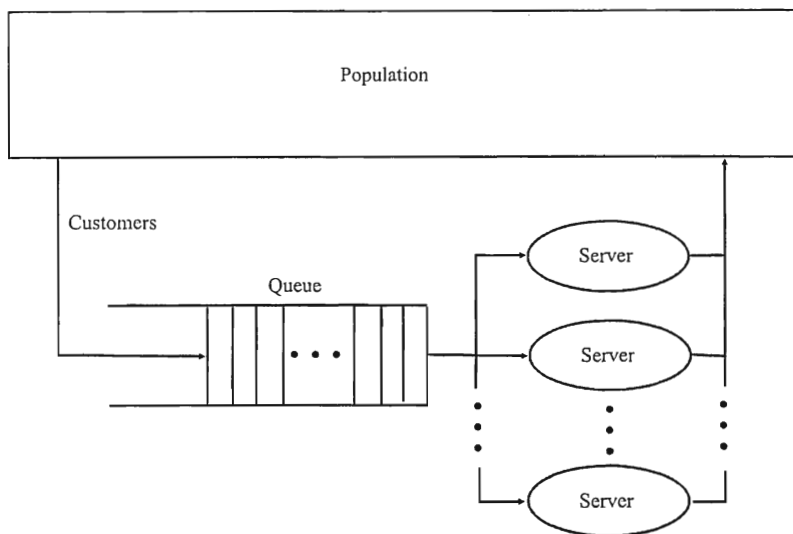
Figure 11.1 Queuing system block diagram.

The size of the queue can also be either infinite or finite. A finite queue can hold only a limited number of customers. If a customer arrives when the queue is full, that customer gets turned away. Most of the time the queue size is assumed to be infinite, however, which simplifies the mathematics describing the process. In any case, if the size of the queue is significantly larger than the likely number of customers then for all intents and purposes the queue is infinite. The amount of time a customer waits in the queue is called the *queuing time*.

The queue discipline is the method by which the server selects the next customer to serve. Normally this is *first-come first-served* (FCFS), also called *first-in first-out* (FIFO), but there are others. In the FIFO queue, the head is selected next. In the *last-in first-out* (LIFO), the last customer to arrive is selected next. There also are priority schemes, where a particular type of customer has a higher priority than the rest of the customers in the queue. If present in the queue, it is selected next. The queue discipline could also dictate that the next customer to be served is selected at random.

The server, or service mechanism, is the method that the customers receive service once they have been selected. The *service time* is the amount of time that a customer requires to be serviced, and it need not be, and normally is not, the same for all customers. The service time is described by a probability distribution. The most normal case is to assume that there is only one server, but that can be generalized to include $N > 1$ servers.

When the system is in *steady state*, it has been in operation for a considerable time. When a queuing system is just started and has not yet achieved steady state, it is in *transition*. Steady state is normally defined as that point at which the state of the system has become independent of the initial state and the total elapsed time.

11.1.1 Statistics for Queuing Theory

There are four generally accepted statistical distributions encountered in queuing theory: the degenerate distribution, the exponential distribution, the Erlang distribution, and the general distribution.

In a degenerate distribution, the variable is a constant value. This would be the case if the variable does not change, so that the distance, velocity, production rate, and so forth are constant. In fact, this distribution is not really a statistical distribution at all—the variable is deterministic.

The time between arrivals T, is a Poisson process which has an exponential distribution with the pdf

$$p(x) = \frac{1}{\alpha} e^{-\frac{x}{\alpha}}, \quad 0 \leq x < \infty, \quad \alpha > 0 \tag{11.1}$$

where $\lambda = 1/\alpha$. Note that

$$\mathcal{E}\{T\} = \frac{1}{\lambda} \tag{11.2}$$

and

$$\text{var}(T) = \frac{1}{\lambda^2} \tag{11.3}$$

An example of this distribution is shown in Figure 11.2.

The number of arrivals in a fixed time interval T_1 is also Poisson distributed with parameter $m = \lambda T_1$. For this distribution the number of customers that arrive in interval T_1 is a discrete rv, denoted by X such that

$$\Pr(X = x) = \frac{m^x}{x!} e^{-\lambda T_1} \tag{11.4}$$

$$\mathcal{E}\{X\} = m \tag{11.5}$$
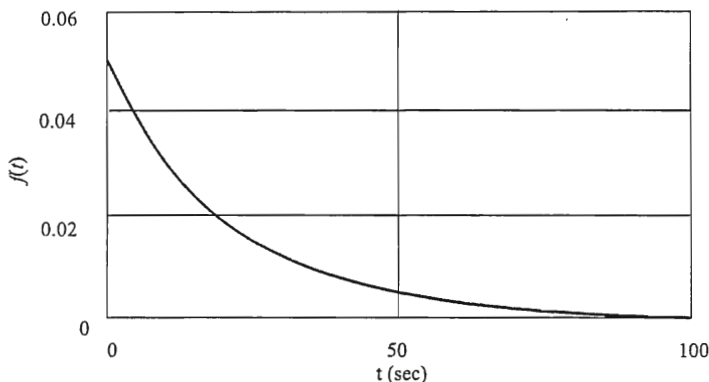
Figure 11.2 Exponential distribution when $\lambda = 1/20$.

and

$$\text{var}(x) = m \qquad (11.6)$$

The exponential distribution is a special case of the *Erlang pdf*, which is given by

$$p(x) = \frac{\beta^{k+1} x^k}{k!} e^{-\beta x}, \qquad 0 \leq x < \infty, \qquad \beta > 0 \qquad (11.7)$$

So the exponential distribution is the Erlang distribution with $k = 1$ and $\alpha = 1/\beta$.

The general distribution actually does not specify a distribution. The results apply to all distributions. When a specific pdf is applied to the problem, further restrictions are applied.

The arrival rate of radar pulses in a dense target environment at an *electronic intelligence* (ELINT) EW system has been analyzed to be approximately exponential [4]. This was verified by analysis and simulation by El-Ayadi et al. [5]. This is basically the same as the arrival of pulses at a communication EW system in a moderate frequency-hopping target environment. The target environment considered in the latter reference was over the range of 30,000 to 60,000 pulses per second. A tactical communication EW system would probably have to deal with about this amount as well. At this pulse rate, if the targets are hopping at 100 hps, there would be about 300 to 500 active emitters at a time. Since, typically, tactical *push-to-talk* (PTT) communication networks use their communication systems at a duty cycle of about 10% on average, that would imply that there are a total of 3,000 to 6,000 total targets that the EW system would have to deal with within its hearing range.

11.1.2 Kendall-Lee Notation

D. G. Kendall (1918 –) proposed a standard notation for queue types [6]. Lee later extended that notation [7]. The standard is given by

$$A / B / S : (d / C / P) \qquad (11.8)$$

where

A = Distribution of interarrival times of customers;
B = Distribution of service times;
S = Number of servers;
d = The queue discipline;
C = Maximum total number of customers that can be accommodated in the system;
P = Calling population size.

The two distributions, A and B, can be any one of the following distributions:

M = Markovian distribution (exponential);
D = Degenerate (or deterministic) distribution;
E_k = Erlang distribution (k = shape parameter);
G = General distribution.

Kendall introduced the first three notations while Lee added the latter three.

As mentioned, S is normally 1 or a variable to be determined. C is normally either infinite or a variable to be determined, as is P. If C and P are infinite they are usually omitted. If C is infinite but P is not, then the infinity symbol is needed for C. When only the first three symbols are used, the last three are assumed to be FIFO/∞/∞.

For example, $E_k/E_l/1$ describes a queuing system that has an Erlang distribution with shape parameter k for the interarrival times of the customers, an Erlang distribution with shape parameter l for the service time distribution, and a single server.

The symbol λ denotes the mean arrival rate when P is infinite and λ_n denotes the mean arrival rate when P is not infinite and there are n customers in the system. Likewise, μ_n denotes the mean service rate when there are n customers in the system and when the mean service rate is constant for all servers, μ denotes the mean service rate. When there are multiple servers, μ and μ_n represent the total

rate of all the servers in the system. In steady state, the arrivals are random and form a Poisson process at a constant average rate.

Some notation in queuing theory is:

P_i = Probability of exactly i customers in the queuing system;
L = Expected number of customers in the queuing system;
L_q = Expected queue length and includes customers currently being served;
$\bar{\omega}$ = Waiting time in the system, including service time, for each customer;
$W = \mathcal{E}\{\bar{\omega}\}$;
$\bar{\omega}_q$ = Waiting time in queue, excluding service time, for each customer;
$W_q = \mathcal{E}\{\bar{\omega}_q\}$.

11.1.3 Queue Relationships

For all queues the following applies:

$$\text{Expected interarrival time} = \frac{1}{\lambda} \qquad (11.9)$$

$$\text{Expected service time} = \frac{1}{\mu} \qquad (11.10)$$

Property: Little's Law

Little's law states that for a steady state system,

$$L = \lambda W \qquad (11.11)$$

as well as

$$L_q = \lambda W_q \qquad (11.12)$$

If the arrival rate of customers is dependent on the size of the queue, then λ in (11.11) and (11.12) must be replaced with the average arrival rate $\bar{\lambda}$. In addition,

$$W = W_q + \frac{1}{\mu} \qquad (11.13)$$

Since λ is the constant average arrival rate and μ is the maximum service rate when the system is busy, when in steady state $\lambda \leq \mu$ or else the queue would

continue to expand. Also, in steady state the average output rate must equal the average input rate.

The system vacillates back and forth between two states: busy and idle. When busy, the system produces an average output rate of μ. In an idle period there are no customers and no output. If P_0 denotes the proportion of time the system is idle, then the utilization of the system, which is defined as the proportion of time the system is busy and denoted by ρ, is given by $1 - P_0$. ρ is also called the *traffic intensity*. Thus, in steady state the arrival rate = departure rate = λ and

$$\lambda = 0P_0 + \mu(1 - P_0)$$

so

$$P_0 = 1 - \frac{\lambda}{\mu}$$

which yields the utilization

$$\rho = \frac{\lambda}{\mu} \tag{11.14}$$

In steady state with a single server, $\rho \leq 1$.

11.1.4 M/M/1 Model

In the M/M/1 model, the customer arrival time is exponential, as is the service time. Assuming the system is in steady state, let $f(t)$ denote the pdf of the customer arrival rate and $g(t)$ the pdf of the service time. Then

$$
\begin{aligned}
f(t) &= \lambda e^{-\lambda t} \\
g(\mu) &= \mu e^{-\mu t}
\end{aligned}
\tag{11.15}
$$

Thus, from (11.14) the server utilization is

$$\rho = \frac{\lambda}{\mu} \tag{11.16}$$
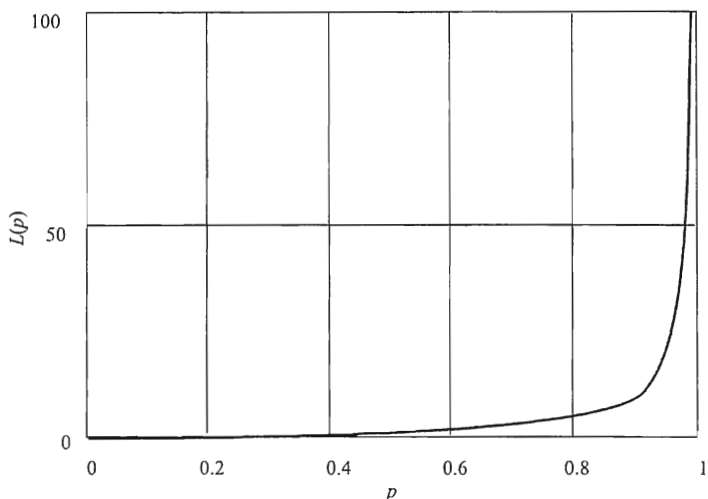
Figure 11.3 Mean number of customers in the system.

The probability that there are n customers in the queue and being serviced at any given time in a steady state system is

$$P_n = (1-\rho)\rho^n \tag{11.17}$$

With this probability, the average number of customers in the system can be found as

$$L = \mathcal{E}(n) = \sum_{n=0}^{\infty} nP_n = \sum_{n=0}^{\infty} n(1-\rho)\rho^n$$

$$= \frac{\rho}{1-\rho} \tag{11.18}$$

This simple expression is shown in Figure 11.3, which illustrates that as the traffic intensity approaches 1, the average number of customers in the system explodes. For numbers below about 0.6 the mean number is reasonably small.

The average time in the system is calculated as follows. By Little's law,

$$L = \lambda W$$

so

$$W = \frac{L}{\lambda} = \frac{\rho}{\lambda(1-\rho)} \tag{11.19}$$

The average time in the queue before being served is given by

$$W_q = W - \frac{1}{\mu}$$

$$= \frac{\rho^2}{\lambda(1-\rho)} \tag{11.20}$$

The average length of the queue is easily calculated.

$$L_q = \mathcal{E}\{n\} - \mathcal{E}\{\text{number in service}\}$$

where

$$\mathcal{E}\{\text{number in service}\} = 0 \times \text{Pr}_0 + 1 \times \text{Pr}(> 0)$$
$$= 0\,\text{Pr}_0 + 1(1 - \text{Pr}_0)$$
$$= 1 - \text{Pr}_0$$
$$= \rho \tag{11.21}$$

Therefore,

$$L_q = \mathcal{E}\{n\} - \rho$$

$$= \frac{\rho}{1-\rho} - \rho$$

$$= \frac{\rho^2}{1-\rho} \tag{11.22}$$

Example: (M/M/1 Queue)

A signal classification subsystem is tasked by the spectrum search subsystem when there is a signal detected at a frequency. The frequency is the only parameter that is passed to the signal classification system. The classification system has its own receiver that is tuned to the tasked frequency, if available, so that the spectrum search system can move on to new frequencies. On average, a signal is

detected every 500 ms and it takes the classification system an average of 200 ms to obtain an answer.

How busy is the classifier?

$$\lambda = \frac{1 \text{ signal}}{500 \text{ ms}} = 2 \text{ per second}$$

The mean service rate is

$$\mu = \frac{1 \text{ signal}}{200 \text{ ms}} = 5 \text{ per second}$$

while the traffic intensity is

$$\rho = \frac{\lambda}{\mu} = \frac{2}{5}$$

The mean number of signals in the system is

$$L = \frac{\rho}{1-\rho} = \frac{2}{5}\frac{5}{3} = \frac{2}{3} \text{ signals}$$

while the mean time in the system is

$$W = \frac{\rho}{\lambda(1-\rho)} = \frac{2}{5}\frac{5}{3}\frac{1}{2} = \frac{1}{3} \text{ second}$$

The average size of the queue is

$$L_q = \frac{\rho^2}{1-\rho} = \frac{4}{25}\frac{5}{3} = \frac{4}{15} \approx 0.25 \text{ signals}$$

If this same processor were used in a denser target environment where the signal detection rate is 4 signals per second how would the performance be impacted?

$$\lambda = \frac{4 \text{ signals}}{1 \text{ second}} = 4 \text{ per second}$$

The traffic intensity becomes

$$\rho = \frac{\lambda}{\mu} = \frac{4}{5}$$

The mean number of signals in the system is

$$L = \frac{\rho}{1-\rho} = \frac{4}{5}\frac{5}{1} = 4 \text{ signals}$$

while the mean time in the system is

$$W = \frac{\rho}{\lambda(1-\rho)} = \frac{4}{5}\frac{5}{3}\frac{1}{2} = \frac{4}{6} \text{ second}$$

The average size of the queue is

$$L_q = \frac{\rho^2}{1-\rho} = \frac{16}{25}\frac{5}{1} = \frac{16}{5} \approx 3 \text{ signals}$$

The size of the queue changed from about 0.25 to 3 signals.

■

11.1.5 Other Queue Types

The performance of the other common types of queues in some cases can be calculated similarly to that above for M/M/1 queues. A summary chart is shown in Table 11.1.

11.2 Concluding Remarks

A brief introduction to queuing theory as it applies to resource allocation in communication EW systems was presented in this chapter. The limited number of resources normally available to these systems dictates that they be shared. This requirement for sharing typically causes scheduling conflicts, and the effects of

Table 11.1 Types of Queues and Their Characteristics

Notation	Server Utilization (ρ)	Average Customer Wait Time (W_q)	Average Length of Queue (L_q)
M/M/1	$\rho = \dfrac{\lambda}{\mu}$	$\dfrac{\rho}{\lambda(1-\rho)}$	$\dfrac{\rho^2}{1-\rho}$
M/M/N	$\dfrac{\rho}{N}$	NA	$P_0 \dfrac{\rho N}{N!} \dfrac{\rho/N}{\left(1-\rho/N\right)^2}$
M/D/1	NA	$\dfrac{L_q}{\lambda}$	$\dfrac{\rho^2}{2(1-\rho)}$
M/D/N		No simple formulas	
M/E$_k$/1		$\dfrac{1+k}{2k}\dfrac{\lambda}{\mu(\mu-\lambda)}$	$\dfrac{1+k}{2k}\dfrac{\lambda^2}{\mu(\mu-\lambda)}$
M/E$_k$/N		No simple formulas	
M/G/1	NA	$\dfrac{L_q}{\lambda}$	$\dfrac{\lambda^2\sigma^2+\rho^2}{2(1-\rho)}$

σ^2 is the variance of the general distribution and $1/\mu$ is its mean.

such sharing can be analyzed with the assistance of queuing theory, as devised by Erlang at the beginning of the twentieth century.

The most common type of queue is the M/M/1 queue, where the population arrival rate is described by a Markov process, as is the server time distribution. The M/M/1 queuing system has one server. Details of the statistical behavior of the M/M/1 queue were presented.

References

[1]   Erlang, A. K., "Solution of Some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges," *Elektrotkeknikeren,* Vol. 13, 1917.
[2]   Allen A. O., "Elements of Queuing Theory for System Design," *Bell System Technical Journal,* No. 2, 1975, pp. 161–187.
[3]   Bunday, D., *Basic Queuing Theory*, London, England: Arnold, 1986.

[4]   Bussgang, J. J., and T. L. Fine, "Interpulse Interval Distribution in the Environment of N Periodic Pulse Radars," *IEEE Transactions on Radio Frequency Interference,* Vol. RFI-5, 1963, pp. 7–10.

[5]   El-Ayadi, M. H., K. El-Barbary, and H. E. Abou-Badr, "Analysis of Queuing Behavior of Automatic ESM Systems," *IEEE Transactions on Aerospace and Electronic Systems,* Vol. 37, No. 3, July 2001, pp. 1010–1021.

[6]   Kendall, D. G., "Stochastic Processes Occurring in the Theory of Queues and Their Analysis by the Method of Imbedded Markov Chains," *Annals of Mathematical Statistics,* Vol. 24, 1953, pp. 339–354.

[7]   Lee, A. M., *Applied Queueing Theory,* London, England: Macmillan, 1966.

Appendix A

Lagrange Multipliers

There are two ways to find the maximum (or minimum) of a function $f(x, y)$ (such maximum or minimum is called a *stationary point*) with two unknowns which is subject to a second, constraint, function $g(x, y) = c$. The first way solves for one variable in the first function in terms of the other variable. This, then, is substituted into the second function to create a single function in one unknown. The derivative of this function is then found and set to zero which is then solved for the variable. This variable is then substituted into the first function to find the second variable. Attempting to solve such problems this way often leads to unwieldy expressions that cannot be simplified.

The second way is based on *Lagrange multipliers*. In this approach, the constraint function is restructured to equal zero, that is,

$$h(x,y) = g(x,y) - c = 0 \qquad (A.1)$$

This new function, equaling zero, is added to the first function after multiplication by a constant λ, called a Lagrange multiplier,

$$L = f(x,y) + \lambda h(x,y) \qquad (A.2)$$

and the maximum is found by setting the partial derivatives of L equal to zero,

$$\frac{\partial L}{\partial x} = 0, \qquad \frac{\partial L}{\partial y} = 0, \qquad \frac{\partial L}{\partial \lambda} = 0 \qquad (A.3)$$

This generates three equations in three unknowns that can be solved for the three variables, although normally the value of λ is only of interest for solving for the x and y values, where $f(x, y)$ is maximum.

Example:

Suppose the maximum and minimum values of

$$f(x, y) = x + 2xy - 5y \tag{A.4}$$

are to be found subject to the constraint

$$-x + y = 4 \tag{A.5}$$

Then

$$h(x, y) = -x + y - 4 \tag{A.6}$$

and

$$L = x + 2xy - 5y + \lambda(-x + y - 4) \tag{A.7}$$

Now

$$\frac{\partial L}{\partial x} = \quad 2y + \lambda + 1 = 0 \tag{A.8}$$

$$\frac{\partial L}{\partial y} = 2x \quad \cdot \quad -\lambda - 5 = 0 \tag{A.9}$$

$$\frac{\partial L}{\partial \lambda} = -x + y \quad -4 = 0 \tag{A.10}$$

Eliminating λ from (A.8) and (A.9) yields

$$2x + 2y \quad -4 = 0 \tag{A.11}$$

Using (A.11) and (A.10) generates

$$4y - 8 = 0$$
$$y = 2 \tag{A.12}$$

Substituting this into (A.8) yields

$$4 + \lambda + 1 = 0$$
$$\lambda = -5 \tag{A.13}$$

Finally, using this in (A.9) yields

$$2x + 5 - 5 = 0$$
$$x = 0 \tag{A.14}$$

So the maximum of $f(x, y)$ occurs at $(0, 2)$.

∎

The Lagrange multiplier technique for finding the stationary points of functions is more general than the substitution technique in that it can be applied in more cases.

Appendix B

Convex Functions

In general, higher order functions have many local minima and maxima. Therefore, it is necessary to determine whether a maximum or minimum found is the global one or one of the local ones. The concept of convex functions is useful for this purpose.

Definition: Suppose K-dimensional vectors $\mathbf{x} = (x_1, x_2, \cdots, x_K)$ and $\mathbf{y} = (y_1, y_2, \cdots, y_K)$ consisting of real numbers x_k and y_k exist in domain R. R is *convex* if $\forall\ x, y \in R$ and for any positive constant $c \leq 1$

$$c\mathbf{x} + (1-c)\mathbf{y} \in R \tag{B.1}$$

■

The meaning of this is evident from Figure B.1, which shows some examples of convex and nonconvex regions. All such lines between any two vectors in R must lie within R for the region to be convex.

Definition: A vector $\mathbf{x} = (x_1, x_2, \cdots, x_K)$ is a *probability vector* if the elements x_i are all nonnegative and

$$\sum_{k=1}^{K} x_k = 1 \tag{B.2}$$

■

The set of probability vectors associated with symbol sources is a convex domain. To see this, suppose $\mathbf{p}_1$ and $\mathbf{p}_2$ are probability vectors. Their components p_k are nonnegative and the sum of p_k in each vector equals unity. The components of
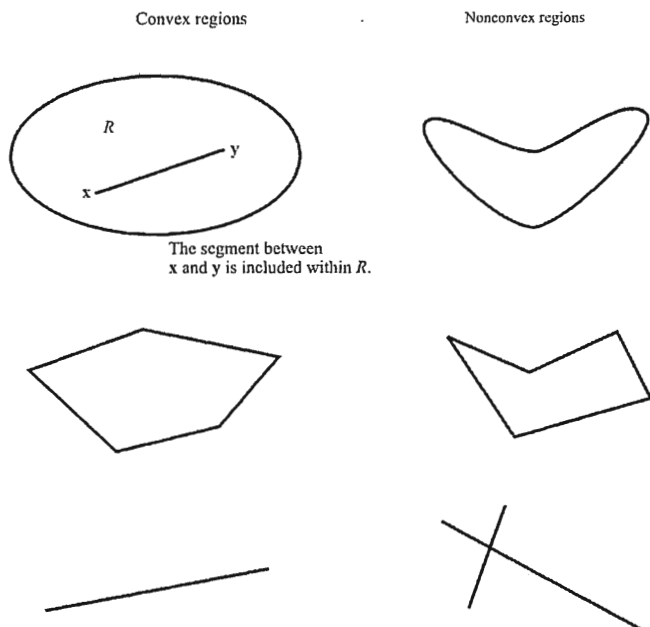
Figure B.1 Example of convex and nonconvex regions R. The domain is convex if all lines, such as the one between x and y, are within R.

$$\mathbf{p} = c\mathbf{p}_1 + (1-c)\mathbf{p}_2, \qquad c \leq 1 \qquad (B.3)$$

are nonnegative and their sum is also unity. Therefore, $\mathbf{p}$ is a probability vector and the domain of probability vectors is convex.

Functions can be convex as well.

Definition: A real valued function $f(\mathbf{x})$ in convex $\cap$ (upward) in a convex domain of a vector space R, if $\forall \mathbf{x} \in R$, $f(\mathbf{x})$ and $f(\mathbf{y})$ satisfies [1]

$$cf(\mathbf{x}) + (1-c)f(\mathbf{y}) \leq f[c\mathbf{x} + (1-c)\mathbf{y}] \qquad (B.4)$$

∎

Geometrically, this requirement is shown in Figure B.2. In this case, the chord expressed in the left of (B.4) must lie below the curve for the function to have a
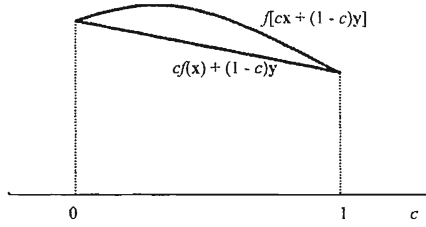
Figure B.2 Convex ∩ function.

maximum between **x** and **y**. Likewise, a real valued function $f(\mathbf{x})$ is convex ∪ (downward, also just convex) under the same conditions when

$$cf(\mathbf{x}) + (1-c)f(\mathbf{y}) \ge f[c\mathbf{x} + (1-c)\mathbf{y}] \tag{B.5}$$

that is, when the chord is above the arc of $f(\mathbf{x})$.

Definition: $\mathbf{x}_0$ is an *extreme point* for convex set ℵ if each line segment that lies completely in ℵ and contains $\mathbf{x}_0$ has $\mathbf{x}_0$ as an end point of the line segment. Mathematically this is expressed as

$$\mathbf{x}_0 \text{ is an extreme point in } ℵ \iff$$
$$\left[\mathbf{x}_0 = c\mathbf{x}_1 + (1-c)\mathbf{x}_2, \mathbf{x}_1, \mathbf{x}_2 \in ℵ, c \in (0,1) \Rightarrow \mathbf{x}_0 = \mathbf{x}_1 = \mathbf{x}_2 \right] \tag{B.6}$$

■

Extreme points for a two-dimensional space, $ℵ^2$, are the ends of lines, and extreme points for $ℵ^3$ are the corners of planes, and so on.

Property: Jensen's Inequality

For a convex ∩ function $f(\mathbf{x})$, Jensen's inequality states that

$$f\left(\sum_{k=1}^{K} w_k x_k \right) \ge \sum_{k=1}^{K} w_k f(x_k) \tag{B.7}$$

where the weights w_k are nonnegative and they sum to unity. Expression (B.7) means that the center of gravity lies below the arc. It also means that

$$\mathcal{E}\{f(\mathbf{x})\} \le f(\mathcal{E}\{\mathbf{x}\}) \tag{B.8}$$

since the w_k define a probability space (defined below). Likewise, for a convex $\cup$ function $f(x)$,

$$f\left(\sum_{k=1}^{K} w_k x_k\right) \leq \sum_{k=1}^{K} w_k f(x_k) \tag{B.9}$$

which says that the center of gravity lies above the arc.

■

Let vector $\mathbf{x} = [x_1, x_2, \cdots, x_K]$ in a convex $\cap$ domain. When $[p_1, p_2, \cdots, p_K]$ are probabilities that sum to unity, then from Jensen's inequality,

$$f\left(\sum_{k=1}^{K} p_k x_k\right) \geq \sum_{k=1}^{K} p_k f'(x_k) \tag{B.10}$$

This means that

$$f(\mathcal{E}\{x\}) \geq \mathcal{E}\{f(x)\} \tag{B.11}$$

where $\mathcal{E}\{.\}$ denotes expectation.

It can be shown that if $f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_K(\mathbf{x})$ are convex $\cap$ functions and $c_1, c_2, \ldots, c_K$ are positive real numbers, then the sum

$$S = \sum_{k=1}^{K} c_k f_k(\mathbf{x}) \tag{B.12}$$

is also convex $\cap$.

For a one-dimensional vector x, $f(x)$ is convex $\cap$ at x whenever

$$\frac{d^2 f(x)}{dx^2} < 0 \tag{B.13}$$

Likewise $f(x)$ is convex $\cup$ at x whenever

$$\frac{d^2 f(x)}{dx^2} > 0 \tag{B.14}$$

If

$$\frac{d^2 f(x)}{dx^2} = 0 \qquad \text{(B.15)}$$

at x, then x is an *inflection point*, changing from convex $\cap$ to convex $\cup$ or vice versa. Expressions (B.13) and (B.14) are simply the requirements learned in elementary calculus for a minimum and maximum, respectively, of a function at a stationary point.

Reference

[1]   Gallager, R. G., *Information Theory and Reliable Communication*, New York: John Wiley & Sons, 1968, p. 84.

List of Acronyms

AIC	Akaike information criteria
AI	artificial intelligence
AM	amplitude modulation
AR	autoregressive
ARMA	autoregressive moving average
AWGN	additive white Gaussian noise
BFSK	binary frequency shift key
BISαS	bivariate isotropic symmetric alpha-stable
BPSK	binary phase shift key
BT	Blackman-Tukey
cdf	cumulative density function
CFAR	constant false alarm rate
CRLB	Cramer-Rao lower bound
CW	continuous wave
dB	decibel
DF	direction finding
DFT	discrete Fourier transform
DS	directed search
DSSS	direct sequence spread spectrum
EA	electronic attack
ELINT	electronic intelligence
EMCON	emission control
EP	electronic protect

ES	electronic support
EVGA	equal variance Gaussian assumption
EW	electronic warfare
FBC	filter bank combiner
FCFS	first-come first-served
FFT	fast Fourier transform
FIFO	first-in first-out
FM	frequency modulation
FPE	final prediction error criterion
FSK	frequency shift key
FT	Fourier transform
GHz	gigahertz
GLRT	generalized likelihood test
GS	general search
HF	high frequency
i.i.d.	independent and identically distributed
IEE	Institute of Electrical Engineers
IEEE	Institute of Electrical and Electronic Engineers
IF	intermediate frequency
IRE	Institute of Radio Engineers
kHz	kilohertz
LIFO	last-in first-out
LMP	locally most powerful
LO	locally optimum
LOB	line of bearing
LOD	locally optimum detector
LPI	low probability of intercept
LRT	likelihood test
LTI	linear time invariant

MA	moving average
MAP	maximum a posteriori
ME	maximum entropy
MEP	maximum envelope peak
MHz	megahertz
MLE	maximum likelihood estimate
ML	maximum likelihood
MMSE	minimum mean square error
MSFE	mean square frequency error
MUSIC	multiple signal classification
MVSE	minimum variance spectral estimation
MVUB	minimum variance unbiased
NP	Neyman-Pearson
PBFBC	partial band filter bank combiner
PCS	personal communication system
pdf	probability density function
PF	position fix
PG	processing gain
PLL	phase lock loop
psd	power spectral density
PSK	phase shift key
PTT	push to talk
QPSK	quaternary phase shift key
RF	radio frequency
ROC	receiver operating characteristic
rv	random variable
SNR	signal-to-noise ratio
SOI	signal of interest
sss	strict-sense stationary
STFT	short-term Fourier transform
SVD	singular value decomposition

SZE   signal-to-interference ratio

TW   time bandwidth product

UAV   unmanned aerial vehicle
UHF   ultra high frequency
UMP   universally most powerful

VHF   very high frequency

wss   wide-sense stationary

About the Author

Richard A Poisel received a B.S. in electrical engineering from the Milwaukee School of Engineering in 1969 and an M.S. in the same discipline from Purdue University in 1971. He spent 3 years in the military service from 1971 to 1973. After his service he attended the University of Wisconsin, where he received a Ph.D. in electrical and computer engineering in 1977. From 1977 to 2004 he was with the same government organization, which has had several different names and is currently known as the U.S. Army Research, Development, and Engineering Command, Intelligence and Information Warfare Laboratory. During the 1993–1994 academic year, Dr. Poisel attended the MIT Sloan School of Management as a Sloan Fellow, receiving an M.B.A. Initially a research engineer, Dr. Poisel eventually rose to the role of the director of the laboratory on an acting basis from 1997 to 1999. He was appointed chief scientist in 1999 and was relocated to the Army's Intelligence Center at Ft. Huachuca, Arizona, where he served as a technical advisor to the command group. He is currently employed by Raytheon Missile Systems as a senior engineering fellow.

Index

The Artech House Information Warfare Library

Electronic Intelligence: The Analysis of Radar Signals, Second Edition, Richard G. Wiley

Electronic Warfare for the Digitized Battlefield, Michael R. Frater and Michael Ryan

Electronic Warfare in the Information Age, D. Curtis Schleher

EW 101: A First Course in Electronic Warfare, David Adamy

Information Warfare Principles and Operations, Edward Waltz

Introduction to Communication Electronic Warfare Systems, Richard A. Poisel

Knowledge Management in the Intelligence Enterprise, Edward Waltz

Mathematical Techniques in Multisensor Data Fusion, Second Edition, David L. Hall and Sonya A. H. McMullen

Modern Communications Jamming Principles and Techniques, Richard A. Poisel

Principles of Data Fusion Automation, Richard T. Antony

Tactical Communications for the Digitized Battlefield, Michael Ryan and Michael R. Frater

Target Acquisition in Communication Electronic Warfare Systems, Richard A. Poisel

For further information on these and other Artech House titles, including previously considered out-of-print books now available through our In-Print-Forever® (IPF®) program, contact:

Artech House   Artech House
685 Canton Street   46 Gillingham Street
Norwood, MA 02062   London SW1V 1AH UK
Phone: 781-769-9750   Phone: +44 (0)20-7596-8750
Fax: 781-769-6334   Fax: +44 (0)20-7630-0166
e-mail: artech@artechhouse.com   e-mail: artech-uk@artechhouse.com

Find us on the World Wide Web at: www.artechhouse.com