PRENTICE HALL SIGNAL PROCESSING SERIES

Alan V. Oppenheim, *Series Editor*

ANDREWS & HUNT *Digital Image Restoration*
BRACEWELL *Two Dimensional Imaging*
BRIGHAM *The Fast Fourier Transform and Its Applications*
BURDIC *Underwater Acoustic System Analysis, 2/E*
CASTLEMAN *Digital Image Processing*
CROCHIERE & RABINER *Multirange Digital Signal Processing*
DUDGEON & MERSEREAU *Multidimensional Digital Signal Processing*
HAYKIN *Advances in Spectrum Analysis and Array Processing. Vols. I, II, & III*
HAYKIN, ED. *Array Signal Processing*
JOHNSON & DUDGEON *Array Signal Processing*
KAY *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*
KAY *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*
KAY *Modern Spectral Estimation*
KINO *Acoustic Waves: Devices, Imaging, and Analog Signal Processing*
LIM *Two-Dimensional Signal and Image Processing*
LIM, ED. *Speech Enhancement*
LIM, & OPPENHEIM, EDS. *Advanced Topics in Signal Processing*
MARPLE *Digital Spectral Analysis with Applications*
MCCLELLAN & RADER *Number Theory in Digital Signal Processing*
MENDEL *Lessons in Estimation Theory for Signal Processing Communications and Control 2/E*
NIKIAS *Higher Order Spectra Analysis*
OPPENHEIM & NAWAB *Symbolic and Knowledge-Based Signal Processing*
OPPENHEIM WILLSKY, WITH YOUNG *Signals and Systems*
OPPENHEIM & SCHAFER *Digital Signal Processing*
OPPENHEIM & SCHAFER *Discrete-Time Signal Processing*
PHILLIPS & NAGLE *Digital Control Systems Analysis and Design, 3/E*
PICINBONO *Random Signals and Systems*
RABINER & GOLD *Theory and Applications of Digital Signal Processing*
RABINER & SCHAFER *Digital Processing of Speech Signals*
RABINER & JUANG *Fundamentals of Speech Recognition*
ROBINSON & TREITEL *Geophysical Signal Analysis*
STEARNS & DAVID *Signal Processing Algorithms in Fortran and C*
TEKLAP *Digital Video Processing*
THERRIEN *Discrete Random Signals and Statistical Signal Processing*
TRIBOLET *Seismic Applications of Homomorphic Signal Processing*
VIADYANATHAN *Multirate Systems and Filter Banks*
WIDROW & STEARNS *Adaptive Signal Processing*

Fundamentals of Statistical Signal Processing

Volume II

Detection Theory

Steven M. Kay

University of Rhode Island

PH PTR

Prentice Hall PTR
Upper Saddle River, New Jersey 07458
http://www.phptr.com

Library of Congress Cataloging-in-Publication Data

Kay, Steven M.
Fundamentals of statistical signal processing : detection theory
/ Steven M. Kay
p. cm. — (PH signal processing series)
Includes bibliographical references and index.
ISBN 0-13-504135-X
1. Signal processing—Statistical methods 2. Detection theory.
I. Title. II. Series: Prentice-Hall signal processing series.
TK5102.5.K379 1993
621.382'2—dc20 92-29495
CIP

Editorial/Production Supervision: *Craig Little*
Acquisitions Editor: *Bernard Goodwin*
Cover Design Director: *Jerry Votta*
Cover Design: *Wanda Lubelska*
Manufacturing Buyer: *Julia Meehan*
Marketing Manager: *Miles Williams*
Composition: *Steven M. Kay, PreTEX, Inc.*

PH PTR

Prentice-Hall, Inc.
Upper Saddle River, New Jersey 07458

Printed in the United States of America

10 9 8 7 6 Sixth Printing

ISBN 0-13-504135-X

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Prentice-Hall Canada Inc., *Toronto*
Prentice-Hall Hispanoamericana, S.A., *Mexico*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*

To my parents
Phyllis and Jack,

to my in-laws
Betty and Walter,

and to my family
Cindy, Lisa, and Ashley

Contents

Preface

This text is the second volume of a series of books addressing statistical signal processing. The first volume, *Fundamentals of Statistical Signal Processing: Estimation Theory*, was published in 1993 by Prentice-Hall, Inc. Henceforth, it will be referred to as [Kay-I 1993]. This second volume, entitled *Fundamentals of Statistical Signal Processing: Detection Theory*, is the application of statistical hypothesis testing to the detection of signals in noise. The series has been written to provide the reader with a broad introduction to the theory and application of statistical signal processing.

Hypothesis testing is a subject that is standard fare in the many books available dealing with statistics. These books range from the highly theoretical expositions written by statisticians to the more practical treatments contributed by the many users of applied statistics. This text is an attempt to strike a balance between these two extremes. The particular audience we have in mind is the community involved in the design and implementation of signal processing algorithms. As such, the primary focus is on obtaining optimal detection algorithms that may be implemented on a digital computer. The data sets are therefore assumed to be samples of a continuous-time waveform or a sequence of data points. The choice of topics reflects what we believe to be the important approaches to obtaining an optimal detector and analyzing its performance. As a consequence, some of the deeper theoretical issues have been omitted with references given instead.

It is the author's opinion that the best way to assimilate the material on detection theory is by exposure to and working with good examples. Consequently, there are numerous examples that illustrate the theory and others that apply the theory to actual detection problems of current interest. We have made extensive use of the MATLAB® scientific programming language (Version 4.2b)[1] for all computer-generated results. In some cases, actual MATLAB programs have been listed where a program was deemed to be of sufficient utility to the reader. Additionally, an abundance of homework problems has been included. They range from simple applications of the theory to extensions of the basic concepts. A solutions manual is available from the author. To aid the reader, summary sections have been provided at the beginning of each chapter. Also, an overview of all the principal detection approaches and the rationale for choosing a particular method can be found in

[1]MATLAB is a registered trademark of The MathWorks, Inc.

Chapter 11. Detection based on simple hypothesis testing is described in Chapters 3–5, while that based on composite hypothesis testing (to accommodate unknown parameters) is the subject of Chapters 6–9. Other chapters address detection in nonGaussian noise (Chapter 10), detection of model changes (Chapter 12), and extensions for complex/vector data useful in array processing (Chapter 13).

This book is an outgrowth of a one-semester graduate level course on detection theory given at the University of Rhode Island. It includes somewhat more material than can actually be covered in one semester. We typically cover most of Chapters 1–10, leaving the subjects of model change detection and complex data/vector data extensions to the student. It is also possible to combine the subjects of estimation and detection into a single semester course by a judicious choice of material from Volumes I and II. The necessary background that has been assumed is an exposure to the basic theory of digital signal processing, probability and random processes, and linear and matrix algebra. This book can also be used for self-study and so should be useful to the practicing engineer as well as the student.

The author would like to acknowledge the contributions of the many people who over the years have provided stimulating discussions of research problems, opportunities to apply the results of that research, and support for conducting research. Thanks are due to my colleagues L. Jackson, R. Kumaresan, L. Pakula, and P. Swaszek of the University of Rhode Island, and L. Scharf of the University of Colorado. Exposure to practical problems, leading to new research directions, has been provided by H. Woodsum of Sonetech, Bedford, New Hampshire, and by D. Mook and S. Lang of Sanders, a Lockheed-Martin Co., Nashua, New Hampshire. The opportunity to apply detection theory to sonar and the research support of J. Kelly of the Naval Undersea Warfare Center, J. Salisbury, formerly of the Naval Undersea Warfare Center, and D. Sheldon of the Naval Undersea Warfare Center, Newport, Rhode Island are also greatly appreciated. Thanks are due to J. Sjogren of the Air Force Office of Scientific Research, whose support has allowed the author to investigate the field of statistical signal processing. A debt of gratitude is owed to all my current and former graduate students. They have contributed to the final manuscript through many hours of pedagogical and research discussions as well as by their specific comments and questions. In particular, P. Djurić of the State University of New York proofread much of the manuscript, and S. Talwalkar of Motorola, Plantation, Florida proofread parts of the manuscript and helped with the finer points of MATLAB.

Steven M. Kay
University of Rhode Island
Kingston, RI 02881
Email: kay@ele.uri.edu

Chapter 1

Introduction

1.1 Detection Theory in Signal Processing

Modern detection theory is fundamental to the design of electronic signal processing systems for decision making and information extraction. These systems include

1. Radar
2. Communications
3. Speech
4. Sonar
5. Image processing
6. Biomedicine
7. Control
8. Seismology,

and all share the common goal of being able to decide when an event of interest occurs and then to determine more information about that event. The latter task, information extraction, is the subject of the first volume [Kay 1993]. The former problem, that of decision making, is the subject of this book and is broadly termed *detection theory*. Other names associated with it are *hypothesis testing* and *decision theory*. To illustrate the problem of detection as applied to signal processing, we briefly describe the first three of these systems.

In radar we are interested in determining the presence or absence of an approaching aircraft [Skolnik 1980]. To accomplish this task we transmit an electromagnetic pulse, which if reflected by a large moving object, will indicate the presence of an aircraft. If an aircraft is present, the received waveform will consist of the reflected pulse (at some time later) and noise due to ambient radiation and the receiver electronics. If an aircraft is not present, then only noise will be present. It is the function of the signal processor to decide whether the received waveform consists of noise only (no aircraft) or an echo in noise (aircraft present). As an example, in Figure 1.1a we have depicted a radar and in Figure 1.1b a typical received waveform for the two possible scenarios. When an echo is present, we see that the character of the received waveform is somewhat different, although possibly not by much. This is because the received echo is attenuated due to propagation loss and possibly distorted due to the interaction of multiple reflections. Of course, if the aircraft is detected, then it is of interest to determine its bearing, range, speed, etc. Hence, detection is the first task of the signal processing system while the second task is information extraction. Estimation theory provides the foundation for the second task and has already been described in Volume I [Kay-I 1993]. The optimal detector for the radar problem is the Neyman-Pearson detector, which is described in Chapter 4. A more practical detector which accommodates signal uncertainties, however, is discussed in Chapter 7.

A second application is in the design of a digital communication system. An example is the binary phase shift keyed (BPSK) system as shown in Figure 1.2a used to communicate the output of a digital data source that emits a "0" or "1" [Proakis 1989]. The data bit is first modulated, then transmitted, and at the receiver, demodulated and then detected. The modulator converts a 0 to the waveform $s_0(t) = \cos 2\pi F_0 t$ and a 1 to $s_1(t) = \cos(2\pi F_0 t + \pi) = -\cos 2\pi F_0 t$ to allow transmission through a bandpass channel whose center frequency is F_0 Hz (such as a microwave link). The phase of the sinusoid indicates whether a 0 or 1 has been sent. In this problem, the function of the detector is to decide between the two possibilities, as in the radar problem, although now, we always have a signal present – the question is *which* signal. Typical received waveforms are shown in Figure 1.2b. Since the sinusoidal carrier has been extracted by the demodulator, all that remains at the detector input is the baseband signal, either a positive or negative pulse. This signal is usually distorted due to limited channel bandwidth and is also corrupted by additive channel noise. The solution to this problem is given in Chapter 4.

Another application is in speech recognition where we wish to determine which word was spoken from among a group of possible words [Rabiner and Juang 1993]. A simple example is to discern among the digits "0", "1", ..., "9". To recognize a spoken digit using a digital computer we would need to *match* the spoken digit with some stored digit. For example, the waveforms for the spoken digits 0 and 1 are shown in Figure 1.3. They have been repeated three times by the same speaker. Note that the waveform changes slightly for each utterance of the same word. We may think of this change as "noise," although it is actually the natural variability of speech. Given an utterance, we wish to decide if it is a 0 or 1. More generally, we would need to decide among the ten possible digits. Such a problem is a generalization of that for radar and for digital communications in which only one of two possible choices need be made. The solution to this problem is discussed in Chapter 4.

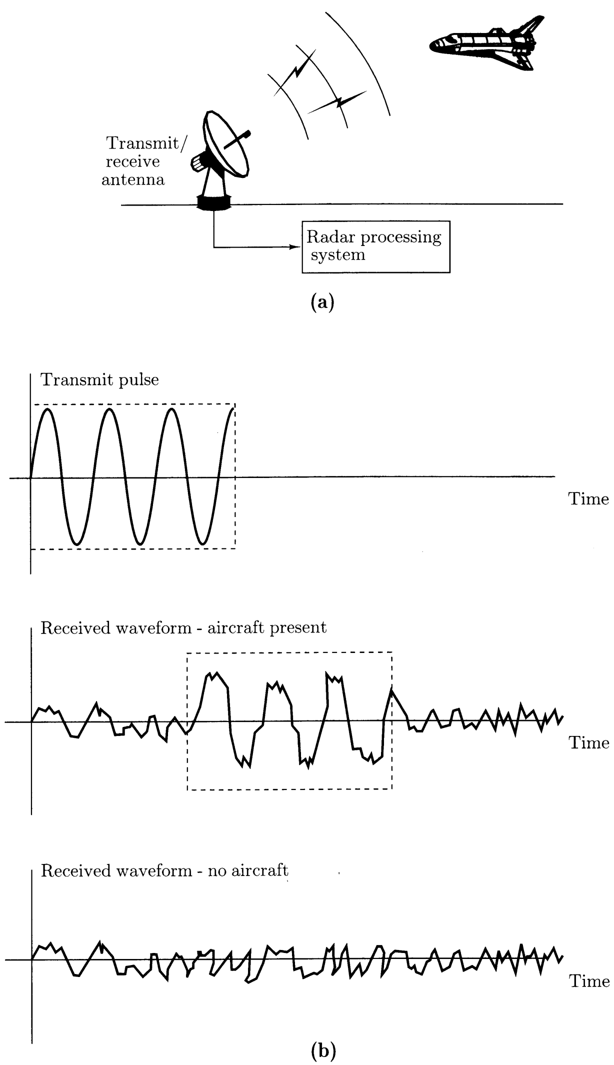


Figure 1.1. Radar system (a) Radar (b) Radar waveforms.

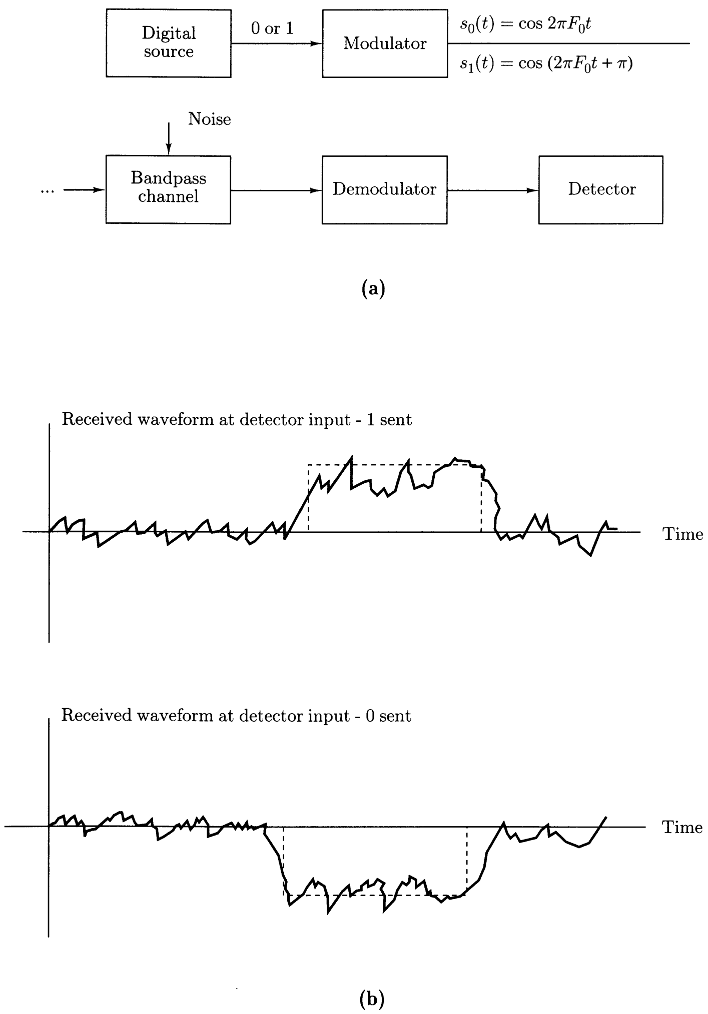


Figure 1.2. Binary phase shift keyed digital communication system (a) Basic system (b) BPSK baseband waveforms.

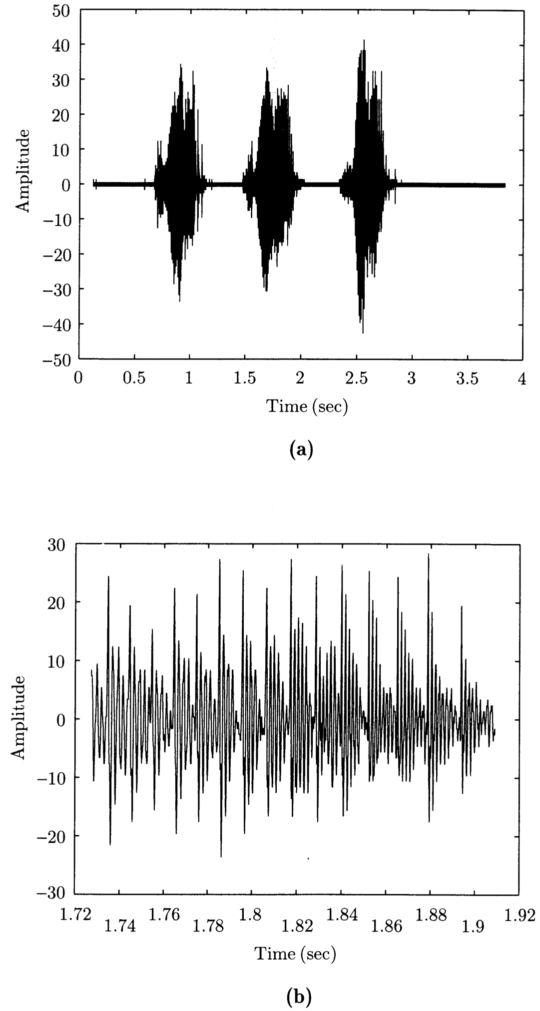
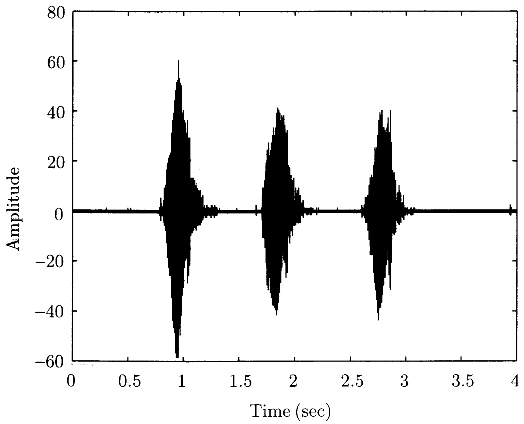
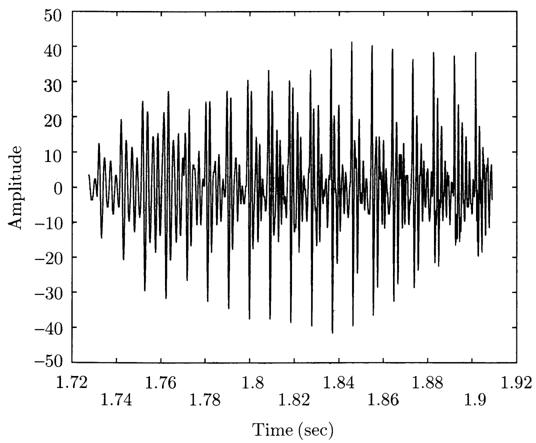


Figure 1.3. Speech waveforms for digits "zero" and "one" (a) "Zero" spoken three times (b) "Zero"–portion of utterance.



(c)



(d)

Figure 1.3. Continued (c) "One" spoken three times (d) "One"–portion of utterance.

In all of these systems, we are faced with the problem of making a decision based on a continuous-time waveform. Modern-day signal processing systems utilize digital computers to sample the continuous-time waveform and store the samples. As a result, we have the equivalent problem of making a decision based on a *discrete-time* waveform or *data set.* Mathematically, we assume the N-point data set $\{x[0], x[1], \ldots, x[N-1]\}$ is available. To arrive at a decision we first form a function of the data or $T(x[0], x[1], \ldots, x[N-1])$ and then make a decision based on its value. Determining the function T and mapping it into a decision is the central problem addressed in *detection theory.* Although electrical engineers at one time designed systems based on analog signals and analog circuits, the future trend is based on discrete-time signals or sequences and digital circuitry. With this transition the detection problem has evolved into one of making a decision based on the observation of a *time series*, which is just a discrete-time process. Therefore, our problem has now evolved into decision-making based on data, which is the subject of *statistical hypothesis testing.* All the theory and techniques developed are now at our disposal [Kendall and Stuart 1976–1979].

Before concluding our discussion of application areas, we complete the previous list.

4. Sonar - detect the presence of an enemy submarine [Knight, Pridham, and Kay 1981, Burdic 1984]

5. Image processing - detect the presence of an aircraft using infrared surveillance [Chan, Langan, and Staver 1990]

6. Biomedicine - detect the presence of a cardiac arrythmia [Gustafson et al. 1978]

7. Control - detect the occurence of an abrupt change in a system to be controlled [Willsky and Jones 1976]

8. Seismology - detect the presence of an underground oil deposit [Justice 1985]

Finally, the multitude of applications stemming from analysis of data from physical phenomena, economics, medical testing, etc., should also be mentioned [Ives 1981, Taylor 1986, Ellenberg et al. 1992].

1.2 The Detection Problem

The simplest detection problem is to decide whether a signal is present, which, as always, is embedded in noise, or if only noise is present. An example of this problem is the detection of an aircraft based on a radar return. Since we wish to decide between two possible hypotheses, signal and noise present versus noise only present, we term this the *binary hypothesis testing problem.* Our goal is to use the received data as efficiently as possible in making our decision and hopefully to be correct most of the time. A somewhat more general form of the binary hypothesis test was encountered in the communication problem. There our interest was in deciding which of two possible signals was transmitted. Our two hypotheses in this case consist of a sinusoid with phase 0° embedded in noise versus a sinusoid with phase 180° embedded in noise.

It also frequently occurs that we wish to decide among more than two hypotheses. In the speech recognition example, our goal was to determine which digit among the ten possible ones was spoken. Such a problem is referred to as the *multiple hypothesis testing problem.* Because we are essentially attempting to determine the speech pattern or to classify the spoken digit as one of a set of possible digits, it is also referred to as the *pattern recognition or classification* problem [Fukunaga 1990].

All these problems are characterized by the need to decide among two or more possible hypotheses based on an observed data set. As always, the data are inherently random in nature, with speech patterns and noise as examples, so that a statistical approach is necessary. In the next section we model the detection problem in a form that allows us to apply the theory of *statistical hypothesis testing* [Lehmann 1959].

1.3 The Mathematical Detection Problem

By way of introduction we consider the detection of a DC level of amplitude $A = 1$ embedded in white Gaussian noise $w[n]$ with variance σ^2. To simplify the discussion we assume that only one sample is available on which to base the decision. Hence, we wish to decide between the hypotheses $x[0] = w[0]$ (noise only) and $x[0] = 1 + w[0]$ (signal in noise). Since the noise is assumed to be zero mean, we might decide that a signal is present if

$$x[0] > \frac{1}{2} \tag{1.1}$$

and noise only is present if

$$x[0] < \frac{1}{2} \tag{1.2}$$

since $E(x[0]) = 0$ if noise only is present and $E(x[0]) = 1$ if a signal is present in noise. (The decision for $x[0] = 1/2$ can be made arbitrarily since the probability of this event is zero. We will henceforth omit this case.) Clearly, we will be in error whenever a signal is present and $w[0] < -1/2$ or whenever noise only is present and $w[0] > 1/2$ (see also Problem 1.1). Hence, we cannot expect to make a correct decision all the time. Hopefully, we will decide correctly most of the time. A better understanding can be obtained by considering what would happen if we repeated the experiment a number of times. This is to say that we observe $x[0]$ for 100 realizations of $w[0]$ when a signal is present and when it is not. Then, some typical results are shown in Figure 1.4a for $\sigma^2 = 0.05$. The "o"'s denote the outcomes when no signal is present and the "x"'s when a signal is present. Clearly, according to (1.1), (1.2) we may make an incorrect decision but only rarely. However, if $\sigma^2 = 0.5$, then our chances of making an error increase dramatically as shown in Figure 1.4b.

Of course, this is due to the increasing spread of the realizations of $w[0]$ as σ^2 increases. Specifically, the probability density function (PDF) of the noise is

$$p(w[0]) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}w^2[0]\right). \tag{1.3}$$

This is illustrated in Figure 1.5 in which histograms of the data shown in Figure 1.4 have been plotted. The dashed plot is for noise only and the solid plot is for a signal in noise. The performance of any detector will depend upon how different the PDFs of $x[0]$ are under each hypothesis. For the same example we plot the PDFs as given by (1.3) in Figure 1.6 for $\sigma^2 = 0.05$ and $\sigma^2 = 0.5$. When noise only is present, they are

$$p(x[0]) = \begin{cases} \frac{1}{\sqrt{0.1\pi}} \exp\left(-10x^2[0]\right) & \sigma^2 = 0.05 \\ \frac{1}{\sqrt{\pi}} \exp\left(-x^2[0]\right) & \sigma^2 = 0.5 \end{cases}$$

and when a signal is embedded in noise, the PDFs are

$$p(x[0]) = \begin{cases} \frac{1}{\sqrt{0.1\pi}} \exp\left(-10(x[0]-1)^2\right) & \sigma^2 = 0.05 \\ \frac{1}{\sqrt{\pi}} \exp\left(-(x[0]-1)^2\right) & \sigma^2 = 0.5. \end{cases}$$

We will see later that the detection performance improves as the "distance" between the PDFs increases or as A^2/σ^2 (the signal-to-noise ratio (SNR)) increases. This example illustrates the basic result that the detection performance depends on the *discrimination* between the two hypotheses or equivalently between the two PDFs (see also Problems 1.2 and 1.3).

More formally, we model the previous detection problem as one of choosing between $\mathcal{H}_0$, which is termed the noise-only hypothesis, and $\mathcal{H}_1$, which is the signal-present hypothesis, or symbolically

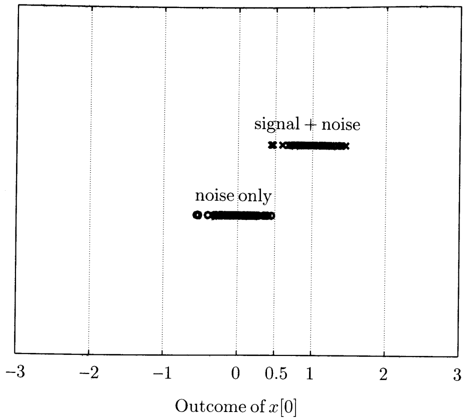
$$\begin{aligned} \mathcal{H}_0 &: x[0] = w[0] \\ \mathcal{H}_1 &: x[0] = 1 + w[0]. \end{aligned} \tag{1.4}$$

The PDFs under each hypothesis are denoted by $p(x[0]; \mathcal{H}_0)$ and $p(x[0]; \mathcal{H}_1)$, which for this example are

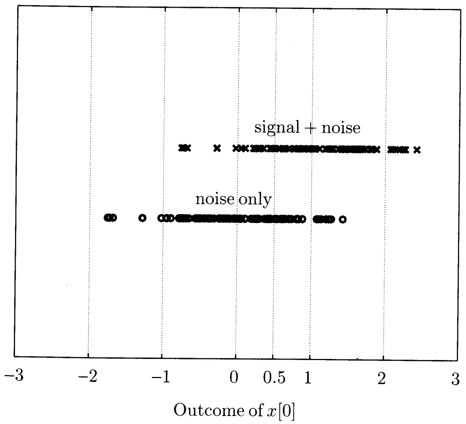
$$\begin{aligned} p(x[0]; \mathcal{H}_0) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}x^2[0]\right) \\ p(x[0]; \mathcal{H}_1) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x[0]-1)^2\right). \end{aligned} \tag{1.5}$$

Note that in deciding between $\mathcal{H}_0$ and $\mathcal{H}_1$, we are essentially asking whether $x[0]$ has been generated according to the PDF $p(x[0]; \mathcal{H}_0)$ or the PDF $p(x[0]; \mathcal{H}_1)$. Alternatively, if we consider the *family* of PDFs

$$p(x[0]; A) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x[0]-A)^2\right) \tag{1.6}$$

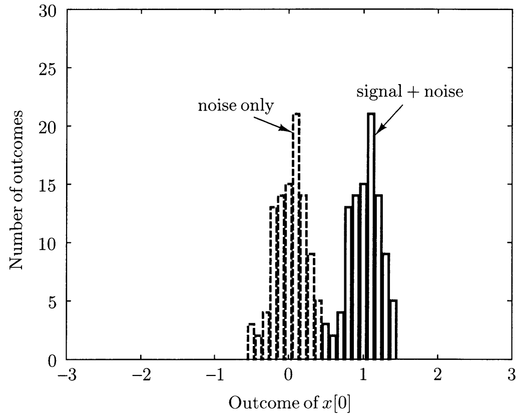


(a)

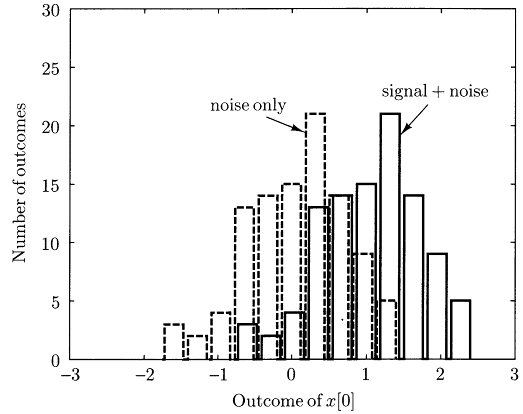


(b)

Figure 1.4. Realizations of $x[0]$ for signal present and signal absent (a) $\sigma^2 = 0.05$ (b) $\sigma^2 = 0.5$.



(a)



(b)

Figure 1.5. Histograms of $x[0]$ for signal present and signal absent (a) $\sigma^2 = 0.05$ (b) $\sigma^2 = 0.5$.

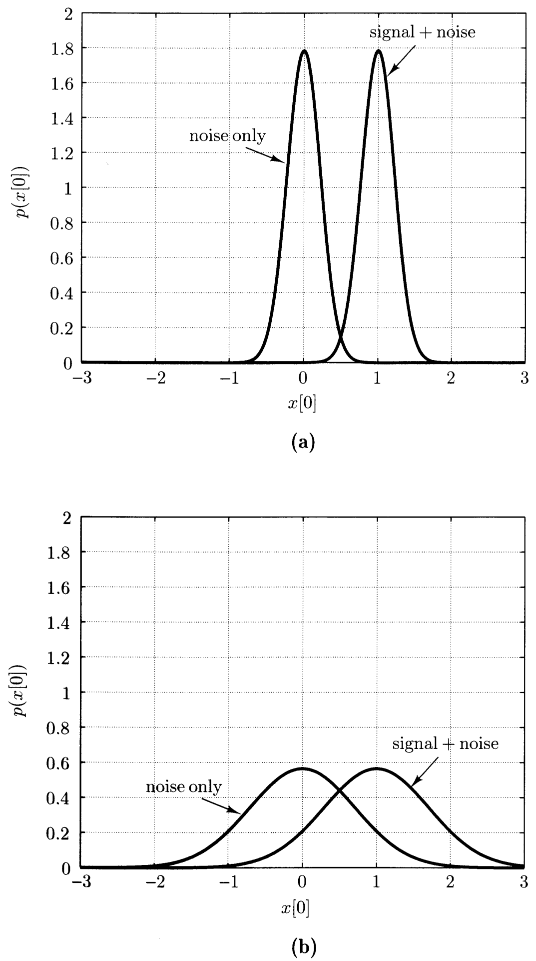


Figure 1.6. PDFs of $x[0]$ for signal present and signal absent (a) $\sigma^2 = 0.05$ (b) $\sigma^2 = 0.5$.

which is *parameterized* by A, then we obtain $p(x[0]; \mathcal{H}_0)$ if $A = 0$, and $p(x[0]; \mathcal{H}_1)$ if $A = 1$. We may, therefore, view the detection problem as a *parameter test*. Given the observation $x[0]$, whose PDF is given by (1.6), we wish to test if $A = 0$ or $A = 1$

or symbolically

$$\begin{aligned} \mathcal{H}_0 &: A = 0 \\ \mathcal{H}_1 &: A = 1. \end{aligned} \tag{1.7}$$

This is termed a parameter test of the PDF, a viewpoint that will be useful later on.

At times it is convenient to assign *prior* probabilities to the possible occurrences of $\mathcal{H}_0$ and $\mathcal{H}_1$. For example, in an on-off keyed (OOK) digital communication system we transmit a "0" by sending no pulse and a "1" by sending a pulse with amplitude $A = 1$. Hence, the corresponding hypothesis test is given by (1.7). In an actual OOK system we will transmit a steady stream of data bits. Since the data bits, 0 or 1, are equally likely to be generated by the source (in the long run), we would expect $\mathcal{H}_0$ to be true half the time and $\mathcal{H}_1$ the other half. It makes sense then to regard the hypotheses as random events with probability 1/2. When we do so, our notation for the PDFs will be $p(x[0]|\mathcal{H}_0)$ and $p(x[0]|\mathcal{H}_1)$, in keeping with the standard notation of a conditional PDF. For this example, we have then that

$$\begin{aligned} p(x[0]|\mathcal{H}_0) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} x^2[0]\right) \\ p(x[0]|\mathcal{H}_1) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (x[0]-1)^2\right) \end{aligned}$$

which should be contrasted with (1.5) (see also Problem 1.4). This distinction is analogous the the classical versus Bayesian approach to parameter estimation [Kay-I 1993].

1.4 Hierarchy of Detection Problems

The detection problems we will address will proceed from the simplest to the more difficult. The degree of difficulty is directly related to our knowledge of the signal and noise characteristics in terms of their PDFs. The ideal case occurs when we have exact knowledge of the PDFs. This is explored in Chapters 4 and 5. Then, at least in theory, we can obtain an optimal detector. When the PDFs are not completely known, the determination of a good (but possibly not optimal) detector is much more difficult. This case is discussed in Chapters 7–9. Another consideration in designing detectors is the mathematical tractability of the PDF. The Gaussian PDF is particularly convenient from a theoretical and practical viewpoint and will be the assumption most often made. In Chapter 10, however, the Gaussian PDF is replaced by the more general nonGaussian PDF. A summary of detection problems and where they are discussed is shown in Table 1.1. Along with the increasing difficulty of determining a detector, we will see that the detection *performance* decreases as we have less specific knowledge of the signal and noise characteristics.

Signal ↓ / Noise →	Gaussian Known PDF	Gaussian Unknown PDF	NonGaussian Known PDF	NonGaussian Unknown PDF
Deterministic Known	4	9	10	*
Deterministic Unknown	7	9	10	*
Random Known PDF	5	9	*	*
Random Unknown PDF	8	*	*	*

* Not discussed (beyond scope of text)

Table 1.1. Hierarchy of detection problems and chapters where discussed.

1.5 Role of Asymptotics

In practice we are principally interested in detecting signals that are *weak* or signals whose SNR is small. If this were not the case, then there would be little need to bother with detection theory in that the signal would not be "buried" in the noise. This is in contrast to the estimation problem in which we usually desire a highly accurate estimate. For accurate estimation we are required to control the SNR so that it is high enough to meet some requirement. It followed then that in estimation problems an asymptotic or high SNR assumption was sometimes useful. In detection problems, however, we are generally faced with a low SNR signal so that our success depends on the data record length. As an illustration, assume we wish to detect the same DC level as before but we will do so by taking multiple measurements. Our data then consist of $x[n] = w[n]$ for $n = 0, 1, \ldots, N-1$ under $\mathcal{H}_0$ and $x[n] = A + w[n]$ for $n = 0, 1, \ldots, N-1$ under $\mathcal{H}_1$, or more formally we have the detection problem

$$
\begin{array}{ll}
\mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\
\mathcal{H}_1 : x[n] = A + w[n] & n = 0, 1, \ldots, N-1
\end{array}
$$

where $w[n]$ is WGN with variance σ^2. A reasonable approach might be to average the samples and compare the value obtained to a threshold γ so that we would decide $\mathcal{H}_1$ if

$$T = \frac{1}{N} \sum_{n=0}^{N-1} x[n] > \gamma.$$

(Note that (1.1) is just a special case when $N = 1$ and $\gamma = 1/2$.) Intuitively, we expect that as N increases, the detection performance should also increase. To

justify our intuition we have plotted a histogram of T for $N = 1$ and $N = 10$ using $\sigma^2 = 0.5$ in Figure 1.7. The experiment has been repeated 100 times so that 100 outcomes of T have been generated. It is seen that, as expected, the overlap between the histograms (which are estimates of the PDFs to within a scale factor) is less as N increases. To quantify this we use a measure that increases with the differences of the means or $E(T; \mathcal{H}_1) - E(T; \mathcal{H}_0)$ and which increases as the variance of each PDF decreases. Noting that $\mathrm{var}(T; \mathcal{H}_0) = \mathrm{var}(T; \mathcal{H}_1)$ (see Problem 1.6) we have the measure, termed the *deflection coefficient*,

$$d^2 = \frac{(E(T; \mathcal{H}_1) - E(T; \mathcal{H}_0))^2}{\mathrm{var}(T; \mathcal{H}_0)}.$$

It can be shown (see Chapter 4) that the detection performance increases with increasing d^2. For the problem at hand it is easily shown that (see Problem 1.6)

$$
\begin{aligned}
E(T; \mathcal{H}_0) &= 0 \\
E(T; \mathcal{H}_1) &= A \\
\mathrm{var}(T; \mathcal{H}_0) &= \sigma^2/N
\end{aligned}
$$

so that

$$d^2 = \frac{A^2}{\sigma^2/N} = \frac{NA^2}{\sigma^2}. \tag{1.8}$$

Hence, as intuited, the detection performance improves as the SNR A^2/σ^2 increases and/or *the data record length N increases.* For weak signals, for which A^2/σ^2 is small, we require N to be large for good detection performance. This has the effect of reducing the noise via averaging since the variance of T, which is due to noise, is σ^2/N. As a result, in detection theory asymptotic analysis (as $N \to \infty$) proves to be appropriate and quite useful. It allows us to derive detectors more easily and also to analyze their performance. For example, if $w[n]$ consisted of independent and identically distributed samples of *nonGaussian* noise, then T would not have a Gaussian PDF. However, as $N \to \infty$, we could invoke the central limit theorem to justify a Gaussian approximation. To determine the detection performance we would need only to obtain the first two moments of T.

1.6 Some Notes to the Reader

Our philosophy in presenting a theory of detection is to provide the user with the main ideas necessary for determining optimal detectors, where possible, and good detectors otherwise. We have included results that we deem to be most useful in practice, omitting some important theoretical issues. The latter can be found in many books on statistical theory, which have been written from a more theoretical viewpoint [Cox and Hinkley 1974, Lehmann 1959, Kendall and Stuart 1976–1979, Rao 1973]. Other books on detection theory that are similar to this one and

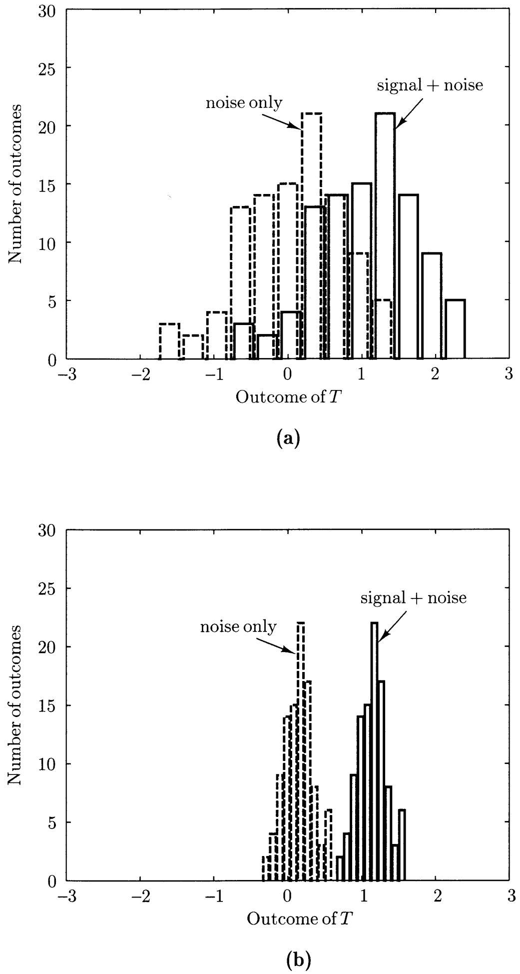


Figure 1.7. Histograms of T for signal present and signal absent (a) $N = 1$ (b) $N = 10$.

which the reader may wish to consult are [Van Trees 1968–1971, Helstrom 1995, McDonough and Whalen 1995]. In Chapter 11 we provide a "road map" for determining a good detector as well as a summary of the various methods and their properties. The reader may wish to read this chapter first to obtain an overview.

This text is part of a two-volume series on statistical signal processing. The first volume, *Fundamentals of Statistical Signal Processing: Estimation Theory*, is referenced as [Kay-I 1993]. For a full appreciation the reader is assumed to be familiar with Volume I, although we have tried to minimize its impact on this text. When this was not possible, references to Volume I by chapter or page number have been made.

We have also tried to maximize insight by including many examples and minimizing long mathematical expositions, although much of the tedious algebra and proofs have been included as appendices. The DC level in noise described earlier will serve as a standard example in introducing almost all of the detection approaches. It is hoped that in doing so the reader will be able to develop his or her own intuition by building upon previously assimilated concepts.

The mathematical notation for all common symbols is summarized in Appendix 2. The distinction between a continuous-time waveform and a discrete-time waveform or sequence is made through the symbolism $x(t)$ for continuous-time and $x[n]$ for discrete-time. Plots of $x[n]$, however, appear continuous in time, the points having been connected by straight lines for easier viewing. All vectors and matrices are **boldface** with all vectors being *column* vectors. All other symbolism is defined within the context of the discussion.

References

Burdic, W.S., *Underwater Acoustic System Analysis*, Prentice-Hall, Englewood Cliffs, N.J., 1984.

Chan, D.S.K, D.A. Langan, S.A. Staver, "Spatial Processing Techniques for the Detection of Small Targets in IR Clutter," *SPIE, Signal and Data Processing of Small Targets*, Vol. 1305, pp. 53–62, 1990.

Cox, D.R., D.V. Hinkley, *Theoretical Statistics*, Chapman and Hall, New York, 1974.

Ellenberg, S.S., D.M. Finklestein, D.A. Schoenfeld, "Statistical Issues Arising in AIDS Clinical Trials," *J. Am. Statist. Asssoc.*, Vol. 87, pp. 562–569, 1992.

Fukunaga, K., *Introduction to Statistical Pattern Recognition*, 2nd ed., Academic Press, New York, 1990.

Gustafson, D.E., A.S. Willsky, J.Y. Wang, M.C. Lancaster, J.H. Triebwasser, "ECG/VCG Rhythm Diagnosis Using Statistical Signal Analysis. Part II: Identification of Transient Rhythms," *IEEE Trans. Biomed. Eng.*, Vol. BME-25, pp. 353–361, 1978.

Helstrom, C.W., *Elements of Signal Detection and Estimation*, Prentice-Hall, Englewood Cliffs, N.J., 1995.

Ives, R.B., "The Applications of Advanced Pattern Recognition Techniques for the Discrimination Between Earthquakes and Nuclear Detonations," *Pattern Recognition*, Vol. 14, pp. 155–161, 1981.

Justice, J.H. ,"Array Processing in Exploration Seismology," in *Array Signal Processing*, S. Haykin, Ed., Prentice-Hall, Englewood Cliffs, N.J., 1985.

Kay, S.M., *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Englewood Cliffs, N.J., 1993.

Kendall, Sir M., A. Stuart, *The Advanced Theory of Statistics*, Vols. 1–3, Macmillan, New York, 1976–1979.

Knight, W.S., R.G. Pridham, S.M. Kay, "Digital Signal Processing for Sonar," *Proc. IEEE*, Vol. 69, pp. 1451–1506, Nov. 1981.

Lehmann, E.L., *Testing Statistical Hypotheses*, J. Wiley, New York, 1959.

McDonough, R.N., A.T. Whalen,*Detection of Signals in Noise*, 2nd Ed., Academic Press, 1995.

Proakis, J.G., *Digital Communications*, 2nd Ed., McGraw-Hill, New York, 1989.

Rabiner, L., B-H Juang, *Fundamentals of Speech Recognition*, Prentice-Hall, Englewood Cliffs, N.J., 1993.

Rao, C.R., *Linear Statistical Inference and its Applications*, J. Wiley, New York, 1973.

Skolnik, M.I., *Introduction to Radar Systems*, McGraw-Hill, New York, 1980.

Taylor, S., *Modeling Financial Time Series*, J. Wiley, New York, 1986.

Van Trees, H.L., *Detection, Estimation, and Modulation Theory*, Vols. I–III, J. Wiley, New York, 1968–1971.

Willsky, A.S., H.L. Jones, "A Generalized Likelihood Ratio Approach to Detection and Estimation of Jumps in Linear Systems," *IEEE Trans. Automat. Contr.*, Vol. AC-21, pp. 108–112, 1976.

Problems

1.1 Consider the detection problem

$$\begin{aligned}\mathcal{H}_0 &: x[0] = w[0] \\ \mathcal{H}_1 &: x[0] = 1 + w[0]\end{aligned}$$

where $w[0]$ is a zero mean Gaussian random variable with variance σ^2. If the detector decides $\mathcal{H}_1$ if $x[0] > 1/2$, find the probability of making a wrong decision when $\mathcal{H}_0$ is true. To do so, determine the probability of deciding $\mathcal{H}_1$ when $\mathcal{H}_0$ is true or $P_0 = \Pr\{x[0] > 1/2; \mathcal{H}_0\}$. For this to be 10^{-3} what must σ^2 be?

1.2 Consider the detection problem

$$\begin{aligned}\mathcal{H}_0 &: x[0] = w[0] \\ \mathcal{H}_1 &: x[0] = 1 + w[0]\end{aligned}$$

where $w[0]$ is a uniformly distributed random variable on the interval $[-a, a]$ for $a > 0$. Discuss the performance of the detector that decides $\mathcal{H}_1$ if $x[0] > 1/2$ as a increases.

1.3 We observe the datum $x[0]$ where $x[0]$ is a Gaussian random variable with mean A and variance $\sigma^2 = 1$. We wish to test if $A = A_0$ or $A = -A_0$, where $A_0 > 0$. Propose a test and discuss its performance as a function of A_0.

1.4 In Problem 1.1 now assume that the probability of $\mathcal{H}_0$ being true is 1/2. If we decide $\mathcal{H}_1$ if $x[0] > 1/2$, find the *total* probability of error or

$$P_e = \Pr\{x[0] > 1/2|\mathcal{H}_0\}\Pr\{\mathcal{H}_0\} + \Pr\{x[0] < 1/2|\mathcal{H}_1\}\Pr\{\mathcal{H}_1\}.$$

Plot P_e versus σ^2 and explain your result.

1.5 In Problem 1.1 now assume that we have two samples on which to base our decision. We decide that a signal is present if

$$T = \frac{1}{2}(x[0] + x[1]) > \frac{1}{2}.$$

Determine all the values of $x[0]$ and $x[1]$ that will result in a decision that a signal is present. Plot these values in a plane. Also, plot the point $[E(x[0]), E(x[1])]^T$ assuming $\mathcal{H}_0$ is true and then assuming $\mathcal{H}_1$ is true. Comment on the results.

1.6 Verify (1.8) by determining the means and variances of T.

1.7 Using (1.8) determine the number of samples required if the deflection coefficient is to be $d^2 = 100$ for adequate detection performance and the SNR is -20 dB.

Chapter 2

Summary of Important PDFs

2.1 Introduction

Evaluation of the performance of a detector depends upon the ability to determine the probability density function of a function of the data samples, either analytically or numerically. When this is not possible, we must resort to Monte Carlo computer simulation techniques. Familiarity with common probability density functions and their properties is essential to the success of the performance evaluation. In this chapter we provide the background material that will be called upon throughout the text. Our discussion is cursory at best due to space limitations. For further details the reader is referred to [Abramowitz and Stegun 1970, Kendall and Stuart 1976–1979, Johnson, Kotz, and Balakrishnan 1995], as well as to the specific references given within the chapter.

2.2 Fundamental Probability Density Functions and Properties

2.2.1 Gaussian (Normal)

The *Gaussian* probability density function (PDF) (also referred to as the *normal* PDF) for a scalar random variable x is defined as

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] \qquad -\infty < x < \infty \tag{2.1}$$

where μ is the mean and σ^2 is the variance of x. It is denoted by $\mathcal{N}(\mu, \sigma^2)$ and we say that $x \sim \mathcal{N}(\mu, \sigma^2)$, where "$\sim$" means "is distributed according to." If $\mu = 0$, its moments are

$$E(x^n) = \begin{cases} 1 \cdot 3 \cdot 5 \cdots (n-1)\sigma^n & n \text{ even} \\ 0 & n \text{ odd.} \end{cases} \tag{2.2}$$

Otherwise, we use

$$E[(x+\mu)^n] = \sum_{k=0}^{n} \binom{n}{k} E(x^k)\mu^{n-k}$$

where $E(x^k)$ is given by (2.2). The cumulative distribution function (CDF) for $\mu = 0$ and $\sigma^2 = 1$, for which the PDF is termed a *standard normal* PDF, is defined as

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt.$$

A more convenient description, which is termed the *right-tail probability* and is the probability of exceeding a given value, is defined as $Q(x) = 1 - \Phi(x)$, where

$$Q(x) = \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt. \tag{2.3}$$

The function $Q(x)$ is also referred to as the *complementary cumulative distribution function*. This cannot be evaluated in closed-form. Its value is shown in Figure 2.1 on linear and log scales. For computational purposes we use the MATLAB program Q.m listed in Appendix 2C. An approximation that is sometimes useful is (see Problem 2.2)

$$Q(x) \approx \frac{1}{\sqrt{2\pi}x} \exp\left(-\frac{1}{2}x^2\right). \tag{2.4}$$

It is shown in Figure 2.2 along with the exact value of $Q(x)$. The approximation is quite accurate for $x > 4$. At times it is important to determine if a random variable has a PDF that is approximately Gaussian. An examination of its right-tail probability reveals whether this is so. By plotting $Q(x)$ on *normal probability paper*, the curve becomes a straight line as shown in Figure 2.3. The normal probability paper can be constructed as described in Appendix 2B. Also contained there is the MATLAB program plotprob.m for plotting the right-tail probability on normal probability paper. As an example, the right-tail probability of the nonGaussian PDF

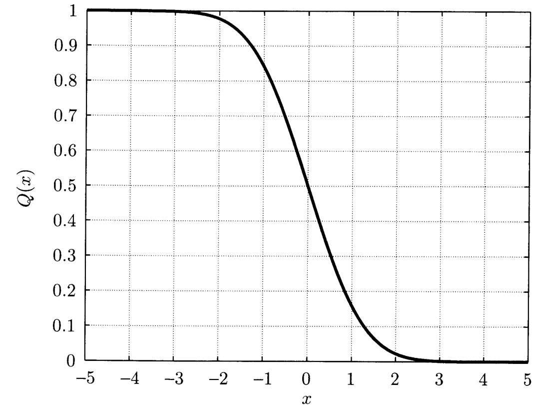
$$p(x) = \frac{1}{2}\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) + \frac{1}{2}\frac{1}{\sqrt{2\pi \cdot 2}} \exp\left(-\frac{1}{2 \cdot 2}x^2\right)$$

which is a Gaussian mixture PDF, is plotted on normal probability paper in Figure 2.4. The functional form for the right-tail probability is easily shown to be $Q(x)/2 + Q(x/\sqrt{2})/2$. The function $Q(x)$ is shown as a dashed straight line for comparison.

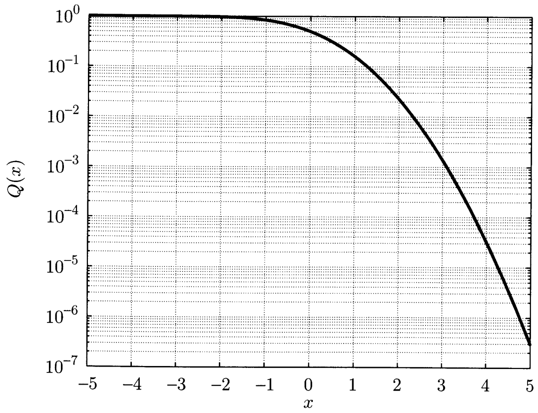
If it is known that a probability is given by $P = Q(\gamma)$, then we can determine γ for a given P. Symbolically, we have $\gamma = Q^{-1}(P)$, where Q^{-1} is the inverse function. The latter must exist since $Q(x)$ is strictly monotonically decreasing. The MATLAB program Qinv.m listed in Appendix 2C computes the function Q^{-1} and can be used to determine γ numerically.

The *multivariate Gaussian* PDF of an $n \times 1$ random vector $\mathbf{x}$ is defined as

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \det^{\frac{1}{2}}(\mathbf{C})} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right] \tag{2.5}$$



(a)



(b)

Figure 2.1. Right-tail probability for standard normal PDF (a) Linear vertical axis (b) Logarithmic vertical axis.

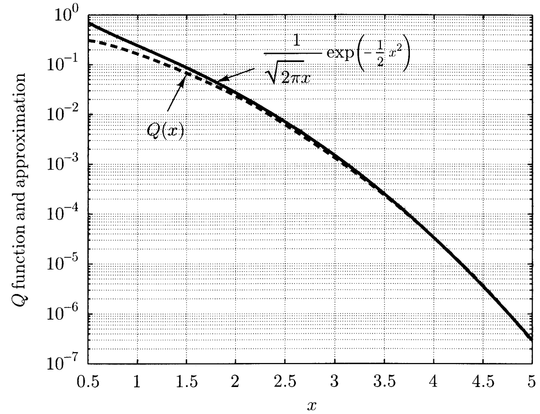


Figure 2.2. Approximation to Q function.

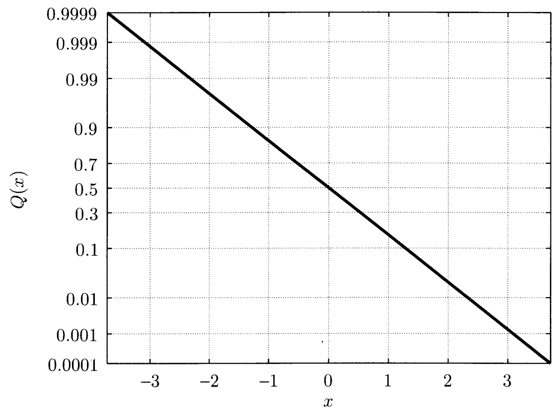


Figure 2.3. Q function plotted on normal probability paper.

where $\boldsymbol{\mu}$ is the mean vector and $\mathbf{C}$ is the covariance matrix and is denoted by $\mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$. It is assumed that $\mathbf{C}$ is positive definite and hence $\mathbf{C}^{-1}$ exists. The mean vector is defined as

$$[\boldsymbol{\mu}]_i = E(x_i) \qquad i = 1, 2, \ldots, n$$

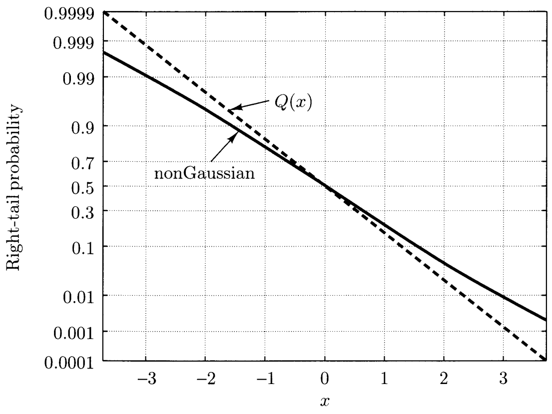


Figure 2.4. Right-tail probability for nongaussian PDF on normal probability paper.

and the covariance matrix as

$$[\mathbf{C}]_{ij} = E[(x_i - E(x_i))(x_j - E(x_j))] \qquad i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, n$$

or in more compact form as

$$\mathbf{C} = E[(\mathbf{x} - E(\mathbf{x}))(\mathbf{x} - E(\mathbf{x}))^T].$$

If $\boldsymbol{\mu} = \mathbf{0}$, then all odd-order joint moments are zero. Even-order moments are found as combinations of second-order moments. In particular, a useful result for $\boldsymbol{\mu} = \mathbf{0}$ is

$$E(x_i x_j x_k x_l) = E(x_i x_j)E(x_k x_l) + E(x_i x_k)E(x_j x_l) + E(x_i x_l)E(x_j x_k). \qquad (2.6)$$

2.2.2 Chi-Squared (Central)

A *chi-squared* PDF with ν degrees of freedom is defined as

$$p(x) = \begin{cases} \dfrac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} x^{\frac{\nu}{2}-1} \exp(-\frac{1}{2}x) & x > 0 \\ 0 & x < 0 \end{cases} \qquad (2.7)$$

and is denoted by χ^2_ν. The degrees of freedom ν is assumed to be an integer with $\nu \geq 1$. The function $\Gamma(u)$ is the Gamma function, which is defined as

$$\Gamma(u) = \int_0^\infty t^{u-1} \exp(-t)dt.$$

The relations $\Gamma(u) = (u-1)\Gamma(u-1)$ for any u, $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, and $\Gamma(n) = (n-1)!$ for n an integer can be used to evaluate it. Some examples of the PDF are given in Figure 2.5. It becomes Gaussian as ν becomes large. Note that for $\nu = 1$ the PDF is infinite at $x = 0$.

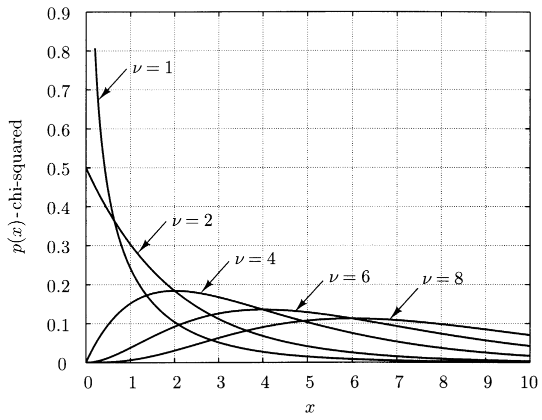


Figure 2.5. PDF for chi-squared random variable.

The chi-squared PDF arises as the PDF of x where $x = \sum_{i=1}^{\nu} x_i^2$ if $x_i \sim \mathcal{N}(0,1)$ and the x_i's are independent and identically distributed (IID). By the latter we mean that each x_i is independent of the others and each x_i has the same PDF (identically distributed). The mean and variance are

$$E(x) = \nu \qquad (2.8)$$

$$\operatorname{var}(x) = 2\nu. \qquad (2.9)$$

A specific case of interest occurs when $\nu = 2$ so that

$$p(x) = \begin{cases} \frac{1}{2}\exp(-\frac{1}{2}x) & x > 0 \\ 0 & x < 0 \end{cases}$$

and is referred to as an *exponential* PDF (see Figure 2.5).

The right-tail probability for a χ^2_ν random variable is defined as

$$Q_{\chi^2_\nu}(x) = \int_x^\infty p(t)dt \qquad x > 0$$

and can be shown to be [Abramowitz and Stegun 1970] for ν even

$$Q_{\chi^2_\nu}(x) = \exp\left(-\frac{1}{2}x\right) \sum_{k=0}^{\frac{\nu}{2}-1} \frac{(\frac{x}{2})^k}{k!} \qquad \nu \geq 2 \qquad (2.10)$$

and for ν odd

$$Q_{\chi^2_\nu}(x) = \begin{cases} 2Q(\sqrt{x}) & \nu = 1 \\ 2Q(\sqrt{x}) + \dfrac{\exp(-\frac{1}{2}x)}{\sqrt{\pi}} \displaystyle\sum_{k=1}^{\frac{\nu-1}{2}} \dfrac{(k-1)!(2x)^{k-\frac{1}{2}}}{(2k-1)!} & \nu \geq 3. \end{cases} \tag{2.11}$$

The MATLAB program Qchipr2.m listed in Appendix 2D can be used to numerically evaluate $Q_{\chi^2_\nu}(x)$.

2.2.3 Chi-Squared (Noncentral)

A generalization of the χ^2_ν PDF arises as a result of summing the squares of IID Gaussian random variables with *nonzero means*. Specifically, if $x = \sum_{i=1}^{\nu} x_i^2$, where the x_i's are independent and $x_i \sim \mathcal{N}(\mu_i, 1)$, then x has a *noncentral chi-squared* PDF with ν degrees of freedom and *noncentrality parameter* $\lambda = \sum_{i=1}^{\nu} \mu_i^2$. The PDF is quite complicated and must be expressed either in integral or infinite series form. As an integral it is

$$p(x) = \begin{cases} \frac{1}{2}\left(\frac{x}{\lambda}\right)^{\frac{\nu-2}{4}} \exp\left[-\frac{1}{2}(x+\lambda)\right] I_{\frac{\nu}{2}-1}(\sqrt{\lambda x}) & x > 0 \\ 0 & x < 0 \end{cases} \tag{2.12}$$

where $I_r(u)$ is the modified Bessel function of the first kind and order r. It is defined as

$$I_r(u) = \frac{(\frac{1}{2}u)^r}{\sqrt{\pi}\,\Gamma(r+\frac{1}{2})} \int_0^{\pi} \exp(u\cos\theta)\sin^{2r}\theta\, d\theta \tag{2.13}$$

and has the series representation

$$I_r(u) = \sum_{k=0}^{\infty} \frac{(\frac{1}{2}u)^{2k+r}}{k!\,\Gamma(r+k+1)}. \tag{2.14}$$

Some examples of the PDF are given in Figure 2.6. Note that the PDF becomes Gaussian as ν becomes large. Using the series expansion of (2.14) the PDF can also be expressed in infinite series form as

$$p(x) = \frac{x^{\frac{\nu}{2}-1}\exp[-\frac{1}{2}(x+\lambda)]}{2^{\frac{\nu}{2}}} \sum_{k=0}^{\infty} \frac{\left(\frac{\lambda x}{4}\right)^k}{k!\,\Gamma(\frac{\nu}{2}+k)}. \tag{2.15}$$

Note that for $\lambda = 0$, the noncentral chi-squared PDF reduces to the chi-squared PDF. The noncentral chi-squared PDF with ν degrees of freedom and noncentrality parameter λ is denoted by $\chi'^2_\nu(\lambda)$.

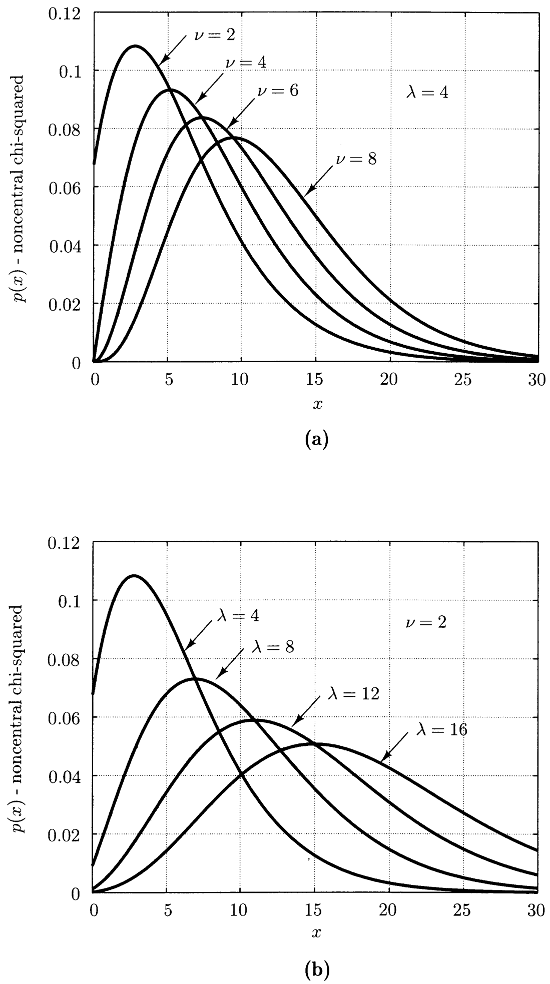


Figure 2.6. PDF for noncentral chi-squared random variable (a) Varying degrees of freedom (b) Varying noncentrality parameter.

The mean and variance are

$$\begin{aligned} E(x) &= \nu + \lambda \\ \mathrm{var}(x) &= 2\nu + 4\lambda. \end{aligned} \tag{2.16}$$

We will denote the right-tail probability as

$$Q_{\chi'^2_\nu(\lambda)}(x) = \int_x^\infty p(t)dt \qquad x > 0$$

Its value can be numerically determined by the MATLAB program Qchipr2.m given in Appendix 2D.

2.2.4 F (Central)

The F PDF arises as the ratio of two independent χ^2 random variables. Specifically, if

$$x = \frac{x_1/\nu_1}{x_2/\nu_2}$$

where $x_1 \sim \chi^2_{\nu_1}$, $x_2 \sim \chi^2_{\nu_2}$, and x_1 and x_2 are independent, then x has the F PDF. It is given by

$$p(x) = \begin{cases} \dfrac{\left(\dfrac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}}}{B\left(\dfrac{\nu_1}{2}, \dfrac{\nu_2}{2}\right)} \dfrac{x^{\frac{\nu_1}{2}-1}}{\left(1 + \dfrac{\nu_1}{\nu_2}x\right)^{\frac{\nu_1+\nu_2}{2}}} & x > 0 \\ 0 & x < 0 \end{cases} \tag{2.17}$$

where $B(u, v)$ is the Beta function, which can be related to the Gamma function as

$$B(u, v) = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)}.$$

The PDF is denoted by F_{ν_1,ν_2} as an F PDF with ν_1 numerator degrees of freedom and ν_2 denominator degrees of freedom. Some examples of the PDF are given in Figure 2.7. The right-tail probability is denoted by $Q_{F_{\nu_1,\nu_2}}(x)$ and must be evaluated numerically [Abramowitz and Stegun 1970]. The mean and variance are

$$\begin{aligned} E(x) &= \frac{\nu_2}{\nu_2 - 2} & \nu_2 > 2 \\ \mathrm{var}(x) &= \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)} & \nu_2 > 4. \end{aligned} \tag{2.18}$$

Note that as $\nu_2 \to \infty$, $x \to x_1/\nu_1 \sim \chi^2_{\nu_1}/\nu_1$ since $x_2/\nu_2 \to 1$ (see Problem 2.3).

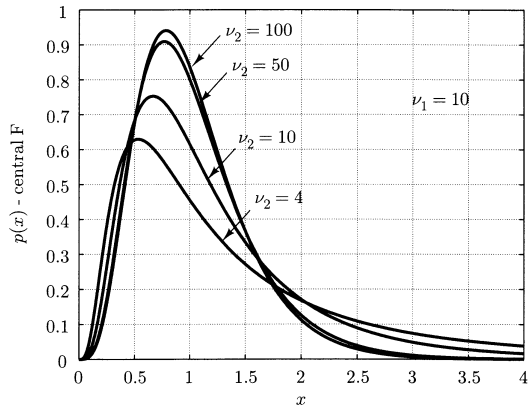


Figure 2.7. PDF for F random variable.

2.2.5 F (Noncentral)

The *noncentral* F PDF results from the ratio of a noncentral χ^2 random variable to a central χ^2 random variable. Specifically, if

$$x = \frac{x_1/\nu_1}{x_2/\nu_2}$$

where $x_1 \sim \chi'^2_{\nu_1}(\lambda)$ and $x_2 \sim \chi^2_{\nu_2}$ and x_1, x_2 are independent, then x has the noncentral F PDF. It is denoted by $F'_{\nu_1,\nu_2}(\lambda)$ as a noncentral F PDF with ν_1 numerator and ν_2 denominator degrees of freedom and noncentrality parameter λ. Its PDF in infinite series form is

$$\begin{aligned} p(x) &= \exp\left(-\frac{\lambda}{2}\right) \sum_{k=0}^\infty \frac{(\frac{\lambda}{2})^k}{k!} \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\frac{1}{2}\nu_1+k}}{B\left(\frac{\nu_1+2k}{2}, \frac{\nu_2}{2}\right)} \\ &\quad \cdot x^{\frac{\nu_1}{2}+k-1}\left(1 + \frac{\nu_1}{\nu_2}x\right)^{-\frac{1}{2}(\nu_1+\nu_2)-k}. \end{aligned} \tag{2.19}$$

Some examples of the PDF can be found in [Johnson and Kotz 1995]. For $\lambda = 0$ this reduces to the central F PDF (let $k = 0$ in (2.19)). Its mean and variance are

$$E(x) = \frac{\nu_2(\nu_1 + \lambda)}{\nu_1(\nu_2 - 2)} \qquad \nu_2 > 2$$

$$\mathrm{var}(x) = 2\left(\frac{\nu_2}{\nu_1}\right)^2 \frac{(\nu_1+\lambda)^2 + (\nu_1+2\lambda)(\nu_2-2)}{(\nu_2-2)^2(\nu_2-4)} \quad \nu_2 > 4. \tag{2.20}$$

The right-tail probability is denoted by $Q_{F'_{\nu_1,\nu_2}(\lambda)}(x)$ and requires numerical evaluation [Patnaik 1949]. Also, note that as $\nu_2 \to \infty$, $F'_{\nu_1,\nu_2}(\lambda) \to \chi'^2_{\nu_1}(\lambda)$ (see Problem 2.3).

2.2.6 Rayleigh

The *Rayleigh* PDF is obtained as the PDF of $x = \sqrt{x_1^2 + x_2^2}$, where $x_1 \sim \mathcal{N}(0, \sigma^2)$, $x_2 \sim \mathcal{N}(0,\sigma^2)$, and x_1, x_2 are independent. Its PDF is

$$p(x) = \begin{cases} \dfrac{x}{\sigma^2}\exp\left(-\dfrac{1}{2\sigma^2}x^2\right) & x > 0 \\ 0 & x < 0. \end{cases} \tag{2.21}$$

It is shown in Figure 2.8 for $\sigma^2 = 1$. The mean and variance are

$$\begin{aligned} E(x) &= \sqrt{\frac{\pi\sigma^2}{2}} \\ \mathrm{var}(x) &= \left(2 - \frac{\pi}{2}\right)\sigma^2. \end{aligned} \tag{2.22}$$

The right-tail probability is easily found as

$$\int_x^\infty p(t)dt = \exp\left(-\frac{x^2}{2\sigma^2}\right). \tag{2.23}$$

The Rayleigh PDF is related to the χ_2^2 PDF since if x is a Rayleigh random variable, then $x = \sqrt{\sigma^2 y}$, where $y \sim \chi_2^2$. As a result, the right-tail probabilities can be related as

$$\begin{aligned} \Pr\{x > \sqrt{\gamma'}\} &= \Pr\left\{x/\sqrt{\sigma^2} > \sqrt{\gamma'/\sigma^2}\right\} \\ &= \Pr\left\{\sqrt{y} > \sqrt{\gamma'/\sigma^2}\right\} \\ &= \Pr\{y > \gamma'/\sigma^2\} \\ &= Q_{\chi_2^2}\left(\frac{\gamma'}{\sigma^2}\right) \end{aligned}$$

or

$$\Pr\{x > \gamma\} = Q_{\chi_2^2}\left(\frac{\gamma^2}{\sigma^2}\right)$$

which produces (2.23) since $Q_{\chi_2^2}(x) = \exp(-x/2)$.

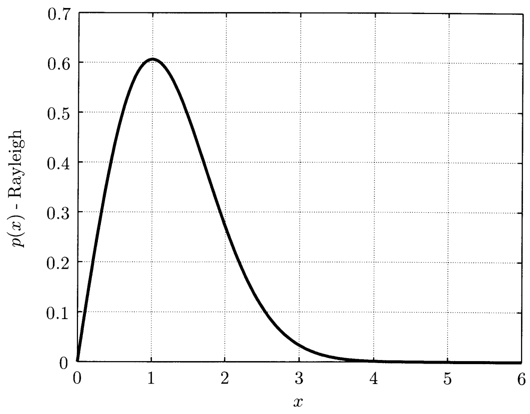


Figure 2.8. PDF for Rayleigh random variable ($\sigma^2 = 1$).

2.2.7 Rician

The *Rician* PDF is obtained as the PDF of $x = \sqrt{x_1^2 + x_2^2}$, where $x_1 \sim \mathcal{N}(\mu_1, \sigma^2)$, $x_2 \sim \mathcal{N}(\mu_2, \sigma^2)$, and x_1, x_2 are independent. Its PDF is (see Problem 7.19)

$$p(x) = \begin{cases} \dfrac{x}{\sigma^2}\exp\left[-\dfrac{1}{2\sigma^2}(x^2+\alpha^2)\right] I_0\left(\dfrac{\alpha x}{\sigma^2}\right) & x > 0 \\ 0 & x < 0 \end{cases} \tag{2.24}$$

where $\alpha^2 = \mu_1^2 + \mu_2^2$ and $I_0(u)$ is given by (2.13) with $r = 0$ or

$$\begin{aligned} I_0(u) &= \frac{1}{\pi}\int_0^\pi \exp(u\cos\theta)d\theta \\ &= \int_0^{2\pi} \exp(u\cos\theta)\frac{d\theta}{2\pi}. \end{aligned} \tag{2.25}$$

Some examples are given in Figure 2.9 for $\sigma^2 = 1$. For $\alpha^2 = 0$ it reduces to the Rayleigh PDF. Its moments are expressable in terms of confluent hypergeometric functions, which can be found in [Rice 1948, McDonough and Whalen 1995]. The right-tail probability can be shown to be related to that of the noncentral χ^2 random variable and must be evaluated numerically (see also Problem 7.20). To do so we proceed as follows:

$$\Pr\{x > \sqrt{\gamma'}\} = \Pr\left\{\sqrt{\frac{x_1^2 + x_2^2}{\sigma^2}} > \sqrt{\frac{\gamma'}{\sigma^2}}\right\}$$

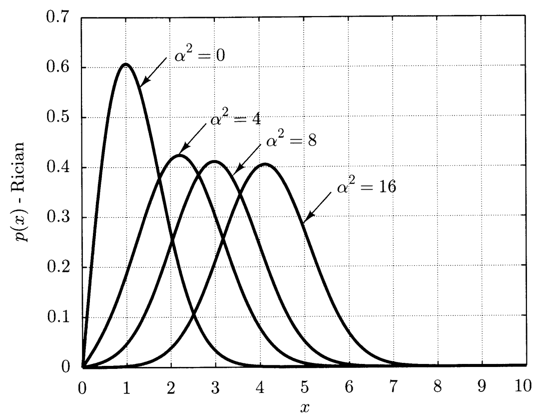


Figure 2.9. PDF for Rician random variable ($\sigma^2 = 1$).

$$= \Pr\left\{\frac{x_1^2 + x_2^2}{\sigma^2} > \frac{\gamma'}{\sigma^2}\right\}$$
$$= Q_{\chi_2'^2(\lambda)}\left(\frac{\gamma'}{\sigma^2}\right)$$

or

$$\Pr\{x > \gamma\} = Q_{\chi_2'^2(\lambda)}\left(\frac{\gamma^2}{\sigma^2}\right) \tag{2.26}$$

where $\lambda = (\mu_1^2 + \mu_2^2)/\sigma^2$. As a result, we can use the MATLAB program Qchipr2.m listed in Appendix 2D to determine the right-tail probability of a Rician random variable.

2.3 Quadratic Forms of Gaussian Random Variables

It frequently occurs that we require the PDF of $y = \mathbf{x}^T\mathbf{A}\mathbf{x}$, where $\mathbf{A}$ is a symmetric $n \times n$ matrix and $\mathbf{x}$ is an $n \times 1$ Gaussian vector with $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$. In general, this is a difficult problem. However, there are a number of special cases that will be useful later. They are:

1. If $\mathbf{A} = \mathbf{C}^{-1}$ and $\boldsymbol{\mu} = \mathbf{0}$, then

$$\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x} \sim \chi_n^2. \tag{2.27}$$

2. If $\mathbf{A} = \mathbf{C}^{-1}$ and $\boldsymbol{\mu} \neq \mathbf{0}$, then

$$\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x} \sim \chi_n'^2(\lambda) \tag{2.28}$$

where $\lambda = \boldsymbol{\mu}^T\mathbf{C}^{-1}\boldsymbol{\mu}$.

3. If $\mathbf{A}$ is idempotent (or $\mathbf{A}^2 = \mathbf{A}$) and of rank r, $\mathbf{C} = \mathbf{I}$, and $\boldsymbol{\mu} = \mathbf{0}$, then

$$\mathbf{x}^T\mathbf{A}\mathbf{x} \sim \chi_r^2. \tag{2.29}$$

The proofs of these results are outlined in Problems 2.4–2.6. Many other results of this kind may be found in [Graybill 1983].

2.4 Asymptotic Gaussian PDF

The multivariate Gaussian PDF was defined in (2.5). In general, it requires the evaluation of the determinant and inverse of the covariance matrix $\mathbf{C}$. In this section we describe an approximation to the determinant and inverse that is based on an asymptotic (for large data records) eigenanalysis of the covariance matrix. It is valid when $\mathbf{x}$ is a data vector from a *zero mean wide sense stationary (WSS) Gaussian random process*. In particular, assume that $\mathbf{x} = [x[0]\, x[1] \ldots x[N-1]]^T$, where $x[n]$ is a zero mean WSS Gaussian random process. Then, the covariance matrix becomes for $i = 0, 1, \ldots, N-1; j = 0, 1, \ldots, N-1$ is

$$\begin{aligned}[\mathbf{C}]_{ij} &= E(x[i]x[j]) \\ &= r_{xx}[i-j].\end{aligned}$$

As an example, if $N = 4$, we have

$$\begin{aligned}\mathbf{C} &= \begin{bmatrix} r_{xx}[0] & r_{xx}[-1] & r_{xx}[-2] & r_{xx}[-3] \\ r_{xx}[1] & r_{xx}[0] & r_{xx}[-1] & r_{xx}[-2] \\ r_{xx}[2] & r_{xx}[1] & r_{xx}[0] & r_{xx}[-1] \\ r_{xx}[3] & r_{xx}[2] & r_{xx}[1] & r_{xx}[0] \end{bmatrix} \\ &= \begin{bmatrix} r_{xx}[0] & r_{xx}[1] & r_{xx}[2] & r_{xx}[3] \\ r_{xx}[1] & r_{xx}[0] & r_{xx}[1] & r_{xx}[2] \\ r_{xx}[2] & r_{xx}[1] & r_{xx}[0] & r_{xx}[1] \\ r_{xx}[3] & r_{xx}[2] & r_{xx}[1] & r_{xx}[0] \end{bmatrix} = \mathbf{R}\end{aligned}$$

since $r_{xx}[-k] = r_{xx}[k]$. The covariance matrix reduces to an *autocorrelation matrix*, which we denote by $\mathbf{R}$, and is seen to be a symmetric Toeplitz matrix (see Appendix 1). As such, as $N \to \infty$, the eigenvalues λ_i and eigenvectors $\mathbf{v}_i$ are easily found.

Letting $P_{xx}(f)$ denote the power spectral density (PSD) of $x[n]$ we have that as $N \to \infty$

$$\begin{aligned} \lambda_i &= P_{xx}(f_i) \\ \mathbf{v}_i &= \frac{1}{\sqrt{N}} [\ 1 \ \exp(j2\pi f_i) \ \exp(j4\pi f_i) \ \ldots \ \exp(j2\pi(N-1)f_i)\ ]^T \end{aligned} \tag{2.30}$$

for $i = 0, 1, \ldots, N-1$ and $f_i = i/N$. The eigenvalues are equally spaced samples of the PSD over the frequency interval [0,1] and the eigenvectors are the discrete Fourier transform (DFT) vectors. (See Problem 2.8 for an example of the determination of the exact eigenvalues.) The approximation will be a good one if the data record length N is much larger than the correlation time of $x[n]$. The latter is defined to be the effective length of the autocorrelation function (ACF) or letting M be that length, we require $r_{xx}[k] \approx 0$ for $k > M$ (see also Problem 2.9). The derivation of these results can be found in [Gray 1972, Fuller 1976, Brockwell and Davis 1987]. We next give a heuristic justification. Consider a process whose ACF satisfies $r_{xx}[k] = 0$ for $|k| \geq 2$. This is termed a moving average process of order one (see Appendix 1). Then, an eigenvector $\mathbf{v} = [v_0\, v_1 \ldots v_{N-1}]^T$ of $\mathbf{R}$ must satisfy

$$\begin{bmatrix} r_{xx}[0] & r_{xx}[-1] & 0 & 0 & \ldots & 0 \\ r_{xx}[1] & r_{xx}[0] & r_{xx}[-1] & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & 0 & r_{xx}[1] & r_{xx}[0] & r_{xx}[-1] \\ 0 & 0 & \ldots & 0 & r_{xx}[1] & r_{xx}[0] \end{bmatrix} \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_{N-1} \end{bmatrix} = \lambda \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_{N-1} \end{bmatrix}$$

or

$$\begin{aligned} r_{xx}[0]v_0 + r_{xx}[-1]v_1 &= \lambda v_0 \\ r_{xx}[1]v_0 + r_{xx}[0]v_1 + r_{xx}[-1]v_2 &= \lambda v_1 \\ &\ \vdots \\ r_{xx}[1]v_{N-3} + r_{xx}[0]v_{N-2} + r_{xx}[-1]v_{N-1} &= \lambda v_{N-2} \\ r_{xx}[1]v_{N-2} + r_{xx}[0]v_{N-1} &= \lambda v_{N-1}. \end{aligned}$$

Ignoring the first and last equations, we have

$$r_{xx}[1]v_{n-1} + r_{xx}[0]v_n + r_{xx}[-1]v_{n+1} = \lambda v_n \qquad n = 1, 2, \ldots N-2. \tag{2.31}$$

This homogeneous difference equation is satisfied for $1 \leq n \leq N-2$ by $v_n = \exp(j2\pi f n)$ since then

$$\begin{aligned} & r_{xx}[1]\exp(-j2\pi f)\exp(j2\pi f n) + r_{xx}[0]\exp(j2\pi f n) \\ & \qquad + r_{xx}[-1]\exp(j2\pi f)\exp(j2\pi f n) = \lambda \exp(j2\pi f n) \end{aligned}$$

where

$$\begin{aligned} \lambda &= r_{xx}[1]\exp(-j2\pi f) + r_{xx}[0] + r_{xx}[-1]\exp(j2\pi f) \\ &= \sum_{k=-\infty}^{\infty} r_{xx}[k]\exp(-j2\pi f k) = P_{xx}(f). \end{aligned}$$

Apart from the $n = 0$ and $n = N-1$ equations, the eigenvector is seen to be $[1\ \exp(j2\pi f) \ldots \exp(j2\pi f(N-1))]^T$. If we choose $f_i = i/N$ for $i = 0, 1, \ldots, N-1$, then we can generate N eigenvectors

$$\mathbf{v}_i = \frac{1}{\sqrt{N}} [\ 1 \ \ \exp(j2\pi f_i) \ \ \exp(j4\pi f_i) \ \ \ldots \ \ \exp(j2\pi(N-1)f_i)\ ]^T$$

that are orthonormal as required (see Problem 2.10) and whose corresponding eigenvalues are $\lambda_i = P_{xx}(f_i)$. Clearly, the fact that the first and last equations are in error will not matter as $N \to \infty$. (See Problem 2.8 for the exact eigenvalues.) Also, note that the eigenvectors have been chosen to be complex, even though $\mathbf{R}$ is a real matrix. We could also have represented the eigenvectors in terms of *real* sines and cosines (see Problem 2.11), although the complex representation is much simpler. Since $\mathbf{v}_{N-i} = \mathbf{v}_i^*$ and $P_{xx}(f_{N-i}) = P_{xx}(f_i)$, the complex eigendecomposition

$$\mathbf{R} = \sum_{i=0}^{N-1} \lambda_i \mathbf{v}_i \mathbf{v}_i^H \tag{2.32}$$

where H denotes the complex conjugate transpose will actually produce a real matrix. See also Problem 2.12 for a simple heuristic derivation of this decomposition.

With the approximation of (2.30) we can now evaluate the determinant and inverse of $\mathbf{R}$ quite simply. It follows that (see Problem 2.13)

$$\det(\mathbf{R}) = \prod_{i=0}^{N-1} \lambda_i = \prod_{i=0}^{N-1} P_{xx}(f_i) \tag{2.33}$$

$$\mathbf{R}^{-1} = \sum_{i=0}^{N-1} \frac{1}{\lambda_i} \mathbf{v}_i \mathbf{v}_i^H = \sum_{i=0}^{N-1} \frac{1}{P_{xx}(f_i)} \mathbf{v}_i \mathbf{v}_i^H. \tag{2.34}$$

These expressions will be useful in approximating detectors. Additionally, they can be used to determine an asymptotic form for the Gaussian PDF under the assumption that $x[n]$ is zero mean and WSS. From (2.5) with $\mathbf{x} = [x[0]\, x[1] \ldots x[N-1]]^T$ and $\boldsymbol{\mu} = \mathbf{0}$ we have

$$\ln p(\mathbf{x}) = -\frac{N}{2}\ln 2\pi - \frac{1}{2}\ln\det(\mathbf{R}) - \frac{1}{2}\mathbf{x}^T\mathbf{R}^{-1}\mathbf{x}$$

and using (2.33) and (2.34)

$$\begin{aligned}\ln p(\mathbf{x}) &= -\frac{N}{2}\ln 2\pi - \frac{1}{2}\ln\prod_{i=0}^{N-1} P_{xx}(f_i) - \frac{1}{2}\mathbf{x}^T\sum_{i=0}^{N-1}\frac{1}{P_{xx}(f_i)}\mathbf{v}_i\mathbf{v}_i^H\mathbf{x} \\ &= -\frac{N}{2}\ln 2\pi - \frac{1}{2}\sum_{i=0}^{N-1}\ln P_{xx}(f_i) - \frac{1}{2}\sum_{i=0}^{N-1}\frac{1}{P_{xx}(f_i)}|\mathbf{v}_i^H\mathbf{x}|^2.\end{aligned}$$

But

$$|\mathbf{v}_i^H\mathbf{x}|^2 = \frac{1}{N}\left|\sum_{n=0}^{N-1} x[n]\exp(-j2\pi f_i n)\right|^2 = I(f_i)$$

where $I(f)$ is termed the *periodogram* (see [Kay 1988], [Kay-I 1993, pg. 80]). Hence

$$\ln p(\mathbf{x}) = -\frac{N}{2}\ln 2\pi - \frac{1}{2}\sum_{i=0}^{N-1}\left(\ln P_{xx}(f_i) + \frac{I(f_i)}{P_{xx}(f_i)}\right). \tag{2.35}$$

An equivalent form is

$$\ln p(\mathbf{x}) = -\frac{N}{2}\ln 2\pi - \frac{N}{2}\sum_{i=0}^{N-1}\left(\ln P_{xx}(f_i) + \frac{I(f_i)}{P_{xx}(f_i)}\right)\frac{1}{N}$$

and as $N \to \infty$, this becomes

$$\ln p(\mathbf{x}) = -\frac{N}{2}\ln 2\pi - \frac{N}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}}\left(\ln P_{xx}(f) + \frac{I(f)}{P_{xx}(f)}\right)df. \tag{2.36}$$

This was derived in [Kay-I 1993, Appendix 3D] using an alternative approach. The reader may also wish to refer to [Kay-I 1993, Section 15.9] for the approximate eigenanalysis in the complex data case.

2.5 Monte Carlo Performance Evaluation

When we are unable to determine the probability that a random variable exceeds a given value by analytical means or by numerical evaluation of a closed-form expression, we must resort to a Monte Carlo computer simulation. The analogous approach for the evaluation of the properties of an estimator was described in [Kay-I 1993, pp. 205–207]. In detection problems we wish to evaluate the probability that a random variable or statistic T exceeds a threshold γ or $\Pr\{T > \gamma\}$. As an example, if we observe the data set $\{x[0], x[1], \ldots, x[N-1]\}$ where $x[n] \sim \mathcal{N}(0, \sigma^2)$ and the $x[n]$'s are IID, we might wish to evaluate

$$\Pr\left\{\frac{1}{N}\sum_{n=0}^{N-1} x[n] > \gamma\right\}.$$

For this simple example, we can easily verify that

$$T = \frac{1}{N}\sum_{n=0}^{N-1} x[n] \sim \mathcal{N}(0, \sigma^2/N)$$

and therefore

$$\Pr\{T > \gamma\} = Q\left(\frac{\gamma}{\sqrt{\sigma^2/N}}\right). \tag{2.37}$$

Assuming, however, that we could not determine the probability either analytically or numerically, we could use a computer simulation to determine $\Pr\{T > \gamma\}$ as follows.

Data Generation:

1. Generate N independent $\mathcal{N}(0, \sigma^2)$ random variables. In MATLAB this is easily done by using the statement

```
x=sqrt(var)*randn(N,1)
```

   to generate an $N \times 1$ column vector composed of the realizations of the random variable $x[n]$, where var is the variance σ^2.
2. Compute $T = (1/N)\sum_{n=0}^{N-1} x[n]$ for the realization of random variables.
3. Repeat the procedure M times to yield M realizations of T or $\{T_1, T_2, \ldots, T_M\}$.

Probability Evaluation:

1. Count the number of T_i's that exceed γ and call this M_γ.
2. Estimate the probability $\Pr\{T > \gamma\}$ as $\hat{P} = M_\gamma/M$.

Note that this probability is actually an *estimated probability*, and hence the use of a hat. The choice of M, the number of realizations, will affect the results so that M should be gradually increased until the computed probability appears to converge. If the true probability is small, then M_γ may be quite small. For example, if $\Pr\{T > \gamma\} = 10^{-6}$, then only about 1 of $M = 10^6$ realizations will exceed γ. In such a case M will have to be much larger than 10^6 to ensure that the probability is accurately estimated. It is shown in Appendix 2A that if a relative absolute error of

$$\epsilon = \frac{|\hat{P} - P|}{P}$$

is desired for $100(1-\alpha)\%$ of the time, then we should choose M to satisfy

$$M \geq \frac{\left[Q^{-1}(\alpha/2)\right]^2 (1-P)}{\epsilon^2 P} \tag{2.38}$$

where P is the probability that is being estimated. The required number of realizations is valid for determining $\Pr\{T > \gamma\}$ using the Monte Carlo realizations $\{T_1, T_2, \ldots, T_M\}$, where the realizations are obtained from the generation of *independent* random variables. The random variables T_i need not be Gaussian in general, but only IID. As an example, if we wish to determine $\Pr\{T > 1\}$, which can be shown to be 0.16, with a relative absolute error of $\epsilon = 0.01$ (1%) for 95% of the time, then

$$M \geq \frac{\left[Q^{-1}(0.025)\right]^2 (1-0.16)}{(0.01)^2 0.16} \approx 2 \times 10^5.$$

When this approach is impractical, one can use *importance sampling* to reduce the computation [Mitchell 1981].

As an example of the Monte Carlo evaluation of (2.37) let $N = 10$ and $\sigma^2 = 10$. In Appendix 2E we list the MATLAB program montecarlo.m that determines $\Pr\{T > \gamma\}$. We plot the results versus γ as well as the true right-tail probability which is $Q(\gamma)$ from (2.37) in Figure 2.10. The number of realizations was chosen to be $M = 1000$ in Figure 2.10a and $M = 10,000$ in Figure 2.10b. The true right-tail probability as given by (2.37) is shown as a dashed curve while the Monte Carlo simulation result is shown as a solid curve. The slight discrepancy for $M = 1000$ is attributed to statistical error since for $M = 10,000$ the agreement is much better.

References

Abramowitz, M., I.A. Stegun, *Handbook of Mathematical Functions*, Dover Pub., New York, 1970.

Brockwell, P.J., R.A. Davis, *Time Series: Theory and Methods*, Springer-Verlag, New York, 1987.

Fuller, W.A., *Introduction to Statistical Time Series*, J. Wiley, New York, 1976.

Gray, R.M. "On the Asymptotic Eigenvalue Distribution of Toeplitz Matrices," *IEEE Trans. Inform. Theory*, Vol. IT–18, pp. 725–730, Nov. 1972.

Graybill, F.A., *Matrices with Applications in Statistics*, Wadsworth, California, 1983.

Johnson, N.L., S. Kotz, N. Balakrishnan, *Continuous Univariate Distributions*, Vols. 1,2, J. Wiley, New York, 1995.

Kay, S.M., *Modern Spectral Estimation: Theory and Application*, Prentice-Hall, Englewood Cliffs, N.J., 1988.

Kendall, Sir M., A. Stuart, *The Advanced Theory of Statistics*, Vols. 1–3, Macmillan, New York, 1976–1979.

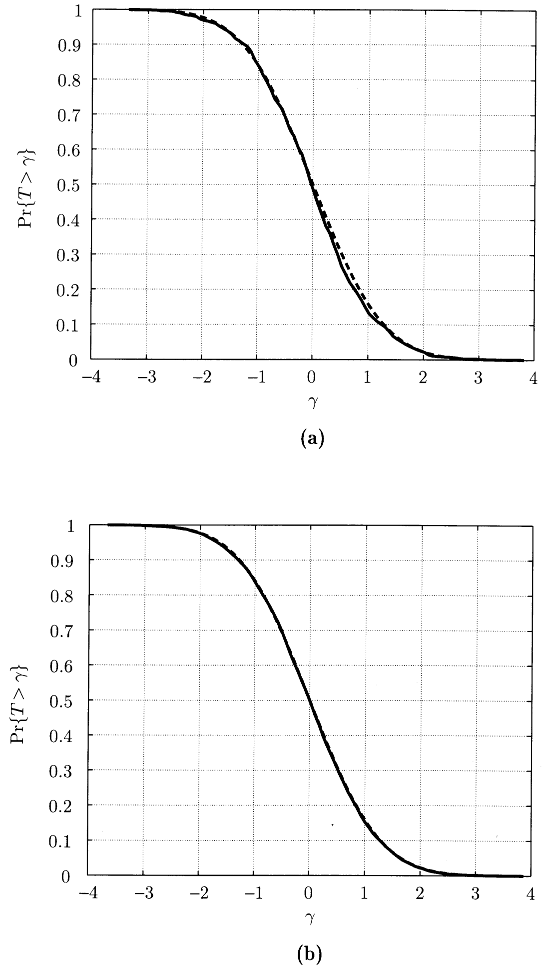


Figure 2.10. Monte Carlo computer simulation of $\Pr\{T > \gamma\}$ (a) $M = 1000$ (b) $M = 10,000$.

McDonough, R.N., A.T. Whalen, *Detection of Signals in Noise*, Academic Press, New York, 1995.

Mitchell, R.L., "Importance Sampling Applied to Simulation of False Alarm Statistics," *IEEE Trans. Aerosp. Elect. Syst.*, Vol. AES–17, pp. 15–24, Jan. 1981.

Mitchell, R.L., J.F. Walker, "Recursive Methods for Computing Detection Probabilities," *IEEE Trans. Aerosp. Elect. Syst.*, Vol. AES–7, pp. 671–676, July 1971.

Patnaik, P.B., "The Non-central χ^2- and F- distributions and their applications," *Biometrika*, Vol. 36, pp. 202–232, 1949.

Rice, S.O., "Statistical Properties of a Sinewave Plus Random Noise," *Bell System Tech. Jour.*, Jan. 1948.

Problems

2.1 Show that if $T \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\Pr\{T > \gamma\} = Q\left(\frac{\gamma - \mu}{\sigma}\right).$$

2.2 Derive (2.4) by noting that

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}t} t \exp\left(-\frac{1}{2}t^2\right) dt$$

and using integration by parts. Also, explain why the approximation improves as x increases.

2.3 Show that as $\nu_2 \to \infty$, an F_{ν_1,ν_2} random variable becomes a $\chi^2_{\nu_1}/\nu_1$ random variable.

2.4 Verify (2.27) by letting $\mathbf{C}^{-1} = \mathbf{D}^T\mathbf{D}$, where $\mathbf{D}$ is a whitening filter matrix (see [Kay-I 1993, Chapter 4]).

2.5 Verify (2.28) by the same approach as in Problem 2.4.

2.6 Verify (2.29) by noting that if $\mathbf{A}$ is an idempotent matrix of rank r, it is a symmetric matrix with orthonormal eigenvectors and r eigenvalues equal to one and the remaining ones equal to zero.

2.7 In this problem we show that if $\mathbf{R}$ is an $N \times N$ autocorrelation matrix corresponding to the nonwhite PSD $P_{xx}(f)$, then the eigenvalues of $\mathbf{R}$ satisfy

$$P_{xx}(f)_{\text{MIN}} < \lambda < P_{xx}(f)_{\text{MAX}}$$

where $P_{xx}(f)_{\text{MIN}}$ and $P_{xx}(f)_{\text{MAX}}$ are the minimum and maximum values of the PSD of $x[n]$, respectively. To do so note that

$$\lambda_{\text{MAX}} = \max_{\mathbf{u}} \frac{\mathbf{u}^T\mathbf{R}\mathbf{u}}{\mathbf{u}^T\mathbf{u}}$$

where $\mathbf{u} = [u[0]\, u[1] \ldots u[N-1]]^T$ and show that

$$\begin{aligned} \mathbf{u}^T\mathbf{R}\mathbf{u} &= \int_{-\frac{1}{2}}^{\frac{1}{2}} |U(f)|^2 P_{xx}(f) df \\ \mathbf{u}^T\mathbf{u} &= \int_{-\frac{1}{2}}^{\frac{1}{2}} |U(f)|^2 df \end{aligned}$$

where $U(f) = \sum_{n=0}^{N-1} u[n] \exp(-j2\pi f n)$. Then, conclude that

$$\lambda \le \lambda_{\text{MAX}} < P_{xx}(f)_{\text{MAX}}.$$

Similarly, show that $\lambda > P_{xx}(f)_{\text{MIN}}$ by utilizing

$$\lambda_{\text{MIN}} = \min_{\mathbf{u}} \frac{\mathbf{u}^T\mathbf{R}\mathbf{u}}{\mathbf{u}^T\mathbf{u}}.$$

2.8 In this problem we derive the exact eigenvalues for an autocorrelation matrix whose elements are

$$[\mathbf{R}]_{mn} = r_{xx}[m-n]$$

where

$$r_{xx}[k] = \begin{cases} \sigma^2(1 + b^2[1]) & k = 0 \\ \sigma^2 b[1] & k = 1 \\ 0 & k \ge 2. \end{cases}$$

This is a moving average process of order one (see [Kay 1988]). Its PSD is easily shown to be

$$P_{xx}(f) = r_{xx}[0] + 2r_{xx}[1]\cos 2\pi f.$$

The equation to be solved for the eigenvalues is $(\mathbf{R} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$, where $\mathbf{v}$ is an eigenvector. Equivalently, for $r_{xx}[1] \ne 0$ (since $P_{xx}(f)$ is assumed to be nonwhite), we need to solve

$$\left(\frac{\mathbf{R} - \lambda\mathbf{I}}{r_{xx}[1]}\right)\mathbf{v} = \mathbf{0}.$$

Letting $c = (r_{xx}[0]-\lambda)/r_{xx}[1]$, we have to find the solutions of $\mathbf{Av} = \mathbf{0}$, where

$$\mathbf{A} = \frac{\mathbf{R}-\lambda\mathbf{I}}{r_{xx}[1]} = \begin{bmatrix} c & 1 & 0 & 0 & \ldots & 0 \\ 1 & c & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & c & 1 \\ 0 & 0 & 0 & \ldots & 1 & c \end{bmatrix}.$$

Using the results from Problem 2.7 show that for $r_{xx}[1] > 0$

$$-2 < \frac{r_{xx}[0] - P_{xx}(f)_{\mathrm{MAX}}}{r_{xx}[1]} < c < \frac{r_{xx}[0] - P_{xx}(f)_{\mathrm{MIN}}}{r_{xx}[1]} < 2.$$

Similary, for $r_{xx}[1] < 0$ we have that $-2 < c < 2$. The equations to be solved $\mathbf{Av} = \mathbf{0}$ lead to the homogeneous equation set

$$\begin{aligned} v_1 + cv_0 &= 0 \\ v_n + cv_{n-1} + v_{n-2} &= 0 \qquad n = 2,3,\ldots,N-1 \\ v_{N-2} + cv_{N-1} &= 0 \end{aligned}$$

where $\mathbf{v} = [v_0\, v_1 \ldots v_{N-1}]^T$. Since $|c| < 2$, the solutions of the homogeneous difference equation for $n = 2,3,\ldots,N-1$ are given by z^n, where $z = \exp(\pm j\theta)$ and $c = -2\cos\theta$. Hence, the solution is $v_n = A\exp(j\theta n) + A^*\exp(-j\theta n)$ for any complex A and some θ. Next show that θ must satisy

$$\theta = \frac{2\pi m}{2(N+1)} \qquad m = 0,1,\ldots,2N+1$$

for the first and last equations to hold. (Note that $\cos\theta$ is periodic with period $2(N+1)$). For distinct eigenvalues and hence orthogonal eigenvectors, as well as to satisfy the constraint $|c| < 2$, we choose

$$m = \begin{cases} 1,3,5,\ldots,N,N+3,N+5,\ldots,2N & \text{for } N \text{ odd} \\ 1,3,5,\ldots,N,N+2,N+4,\ldots,2N & \text{for } N \text{ even.} \end{cases}$$

Finally, show that the eigenvalues are given by

$$r_{xx}[0] + 2r_{xx}[1]\cos\left[\frac{2\pi m}{2(N+1)}\right].$$

Note that $\lambda_i \approx P_{xx}(f_i)$ for large N.

2.9 The correlation time of a WSS random process is defined as the value of M for which $|r_{xx}[M]|/r_{xx}[0]$ is suitably small, say 0.001. For the WSS random process whose ACF is $r_{xx}[k] = 0.9^{|k|}$, determine the correlation time. Next, consider the process $x[n] = A$ where $A \sim \mathcal{N}(0,\sigma_A^2)$. Show that the process is WSS. What is the correlation time of the process? Does this process have the asymptotic eigendecomposition of (2.30)?

2.10 The normalized DFT exponential vectors are given by

$$\mathbf{v}_i = \frac{1}{\sqrt{N}}\left[\, 1 \;\; \exp(j2\pi f_i) \;\; \ldots \;\; \exp(j2\pi f_i(N-1)) \,\right]^T$$

for $i = 0,1,\ldots,N-1$ and where $f_i = i/N$. Prove that they are orthonormal or that

$$\mathbf{v}_m^H\mathbf{v}_n = \begin{cases} 1 & m = n \\ 0 & m \neq n. \end{cases}$$

2.11 In this problem we obtain a set of *real* eigenvectors for an autocorrelation matrix $\mathbf{R}$ as $N \to \infty$. We assume that N is even to simplify the results. Using (2.30) we first express the asymptotic eigenvectors as $\mathbf{v}_i = (\mathbf{c}_i + j\mathbf{s}_i)/\sqrt{N}$, where

$$\begin{aligned} \mathbf{c}_i &= [\, 1 \;\; \cos(2\pi f_i) \;\; \ldots \;\; \cos[2\pi f_i(N-1)] \,]^T \\ \mathbf{s}_i &= [\, 0 \;\; \sin(2\pi f_i) \;\; \ldots \;\; \sin[2\pi f_i(N-1)] \,]^T \end{aligned}$$

and $f_i = i/N$. Noting that $\mathbf{s}_0 = \mathbf{s}_{N/2} = \mathbf{0}$ and $\mathbf{c}_{N-i} = \mathbf{c}_i$, $\mathbf{s}_{N-i} = -\mathbf{s}_i$, show that

$$\begin{aligned} \mathbf{R} &= \sum_{i=0}^{N-1} P_{xx}(f_i)\mathbf{v}_i\mathbf{v}_i^H \\ &= P_{xx}(f_0)\frac{\mathbf{c}_0\mathbf{c}_0^T}{N} + \sum_{i=1}^{\frac{N}{2}-1} P_{xx}(f_i)\frac{\mathbf{c}_i\mathbf{c}_i^T + \mathbf{s}_i\mathbf{s}_i^T}{N/2} + P_{xx}(f_{N/2})\frac{\mathbf{c}_{N/2}\mathbf{c}_{N/2}^T}{N}. \end{aligned}$$

Next show that the set of vectors

$$\left\{ \frac{\mathbf{c}_0}{\sqrt{N}}, \frac{\mathbf{c}_1}{\sqrt{N/2}}, \frac{\mathbf{s}_1}{\sqrt{N/2}}, \ldots, \frac{\mathbf{c}_{N/2-1}}{\sqrt{N/2}}, \frac{\mathbf{s}_{N/2-1}}{\sqrt{N/2}}, \frac{\mathbf{c}_{N/2}}{\sqrt{N}} \right\}$$

is orthonormal using the results of Problem 2.10 and hence are the real eigenvectors. Finally, determine the corresponding eigenvalues.

2.12 A simple heuristic derivation of the asymptotic eigendecomposition of an autocorrelation matrix is examined in this problem. First argue that

$$[\mathbf{R}]_{mn} = r_{xx}[m-n] \approx \frac{1}{N}\sum_{i=0}^{N-1} P_{xx}(f_i)\exp[j2\pi f_i(m-n)]$$

where $f_i = i/N$ and the approximation holds for large N. Then show that this is equivalent to

$$[\mathbf{R}]_{mn} = \sum_{i=0}^{N-1} P_{xx}(f_i)[\mathbf{v}_i]_m[\mathbf{v}_i]_n^*$$

where

$$\mathbf{v}_i = \frac{1}{\sqrt{N}}[\ 1\ \ \exp(j2\pi f_i)\ \ \ldots\ \ \exp(j2\pi f_i(N-1))\ ]^T$$

and note that the $\mathbf{v}_i$'s are orthonormal (see Problem 2.10). Finally, derive (2.32) and identify the eigenvalues and eigenvectors.

2.13 Show that if $\mathbf{R}$ has the eigendecomposition given by (2.32), then

$$\begin{aligned}\det(\mathbf{R}) &= \prod_{i=0}^{N-1} \lambda_i \\ \mathbf{R}^{-1} &= \sum_{i=0}^{N-1} \frac{1}{\lambda_i}\mathbf{v}_i\mathbf{v}_i^H.\end{aligned}$$

2.14 Modify the MATLAB program montecarlo.m in Appendix 2E to determine $\Pr\{(1/N)\sum_{n=0}^{N-1} x^2[n] > 5\}$ if $x[n] \sim \mathcal{N}(0,5)$, $N = 50$, and the samples are IID. Compare your results against the theoretical probability. Hint: Use the central limit theorem.

2.15 A detection probability P_D is to be determined via a Monte Carlo computer simulation. It is known that $P_D \geq 0.8$. If the relative absolute error is to be no more than 0.01 for 95% of the time, how many Monte Carlo trials are required?

Appendix 2A

Number of Required Monte Carlo Trials

We estimate $P = \Pr\{T > \gamma\}$ as $\hat{P} = M_\gamma/M$, where M is the total number of trials (or realizations) and M_γ is the number of trials for which $T > \gamma$. We first determine the PDF of $\hat{P}$. Define a random variable ξ_i as follows

$$\xi_i = \begin{cases} 1 & \text{if } T_i > \gamma \\ 0 & \text{if } T_i < \gamma \end{cases}$$

where T_i is the outcome of the ith trial. The random variable ξ_i is a Bernoulli random variable with the probability of success ($\xi_i = 1$) as P. Thus, $E(\xi_i) = P$ and a natural estimator of P is the sample mean or

$$\hat{P} = \frac{1}{M}\sum_{i=1}^{M} \xi_i.$$

Since the ξ_i's are IID, we can invoke the central limit theorem to assert that $\hat{P}$ is approximately Gaussian for large M. Its mean is

$$\begin{aligned}E(\hat{P}) &= \frac{1}{M}\sum_{i=1}^{M} E(\xi_i) \\ &= \frac{1}{M}\sum_{i=1}^{M} 1 \cdot \Pr\{T_i > \gamma\} = P\end{aligned}$$

and its variance is

$$\mathrm{var}(\hat{P}) = \mathrm{var}\left(\frac{1}{M}\sum_{i=1}^{M} \xi_i\right)$$

$$= \frac{\mathrm{var}(\xi_i)}{M}$$

since the ξ_i's are IID and hence uncorrelated and all have the same variance. But

$$\begin{aligned}\mathrm{var}(\xi_i) &= E(\xi_i^2) - E^2(\xi_i) \\ &= 1^2 \cdot \Pr\{T_i > \gamma\} - (1 \cdot \Pr\{T_i > \gamma\})^2 \\ &= P - P^2 = P(1-P)\end{aligned}$$

and thus

$$\mathrm{var}(\hat{P}) = \frac{P(1-P)}{M}$$

and finally we have

$$\hat{P} \overset{a}{\sim} \mathcal{N}\left(P, \frac{P(1-P)}{M}\right).$$

The relative error $e = (\hat{P} - P)/P$ then has the PDF

$$e = \frac{\hat{P} - P}{P} \overset{a}{\sim} \mathcal{N}\left(0, \frac{1-P}{MP}\right).$$

To guarantee that this is not more than ϵ in absolute value for $100(1-\alpha)\%$ of the time we require

$$\Pr\{|e| > \epsilon\} \leq \alpha$$

or

$$2\Pr\{e > \epsilon\} \leq \alpha$$

which results in

$$2Q\left(\frac{\epsilon}{\sqrt{\frac{1-P}{MP}}}\right) \leq \alpha.$$

Solving for M produces the requirement

$$M \geq \frac{[Q^{-1}(\alpha/2)]^2(1-P)}{\epsilon^2 P}.$$

Appendix 2B

Normal Probability Paper

To construct normal probability paper for which $Q(x)$ will appear as a straight line, we first choose the vertical axis range. Typically, we choose the interval [0.0001, 0.9999]. The vertical axis is labeled with the values $\{y_1, y_2, \ldots, y_L\}$, where $y_1 = 0.0001$ and $y_L = 0.9999$. Also, L is chosen to be odd and $y_i = 1 - y_{L+1-i}$ so that $y_{(L+1)/2} = 0.5$. For the MATLAB program that follows $L = 11$ and

$$\mathbf{y} = [0.0001\ 0.001\ 0.01\ 0.1\ 0.3\ 0.5\ 0.7\ 0.9\ 0.99\ 0.999\ 0.9999]^T.$$

The horizontal axis will be the interval $[-Q^{-1}(y_1), Q^{-1}(y_1)]$ as shown in Figure 2.11. For $y = Q(x)$ to appear as a straight line we must warp the y values so that they become $Y_i = mQ^{-1}(y_i) + 1/2$. The slope m is found from Figure 2.11 as

$$m = \frac{2y_1 - 1}{2Q^{-1}(y_1)}.$$

To plot the point (x_i, y_i) we convert it to $(x_i, Y_i) = (x_i, mQ^{-1}(y_i) + 1/2)$. To obtain the true y values (before warping) we put tick marks at $Y_i = mQ^{-1}(y_i) + 1/2$ but label them as y_i. The MATLAB program plotprob.m listed below plots the right-tail probability on normal probability paper. If desired, the Gaussian right-tail probability $Q(x)$ can also be plotted for comparison.

plotprob.m

```
  function plotprob(x,Qx,xlab,ylab,nor)
% This program plots a set of right-tail probabilities
% (complementary cumulative distribution function) on normal
% probability paper.
%
% Input Parameters:
%
```

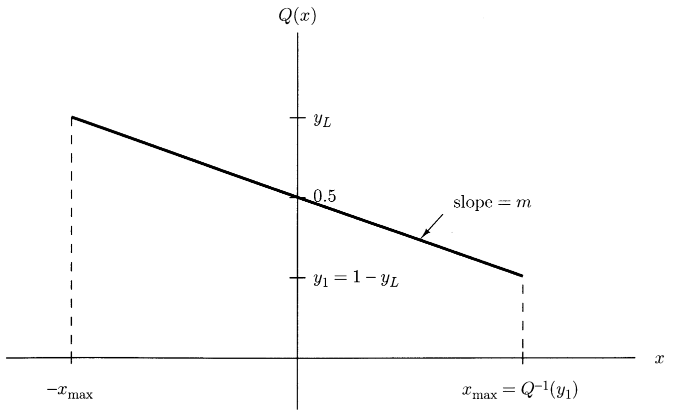


Figure 2.11. Construction of a normal probability paper.

```
%    x     - Real column vector of x values for desired
%            right-tail probabilities
%    Qx    - Real column vector of corresponding
%            right-tail probabilities
%    xlab - Label for x axis (enclose in single quotes)
%    ylab - Label for y axis (enclose in single quotes)
%    nor -  Set equal to 1 for Gaussian right-tail probability
%            to be plotted for comparison, and 0 otherwise
%
% Verification Test Case:
%
% The inputs x=[-5:0.01:5]';Qx=0.5*Q(x)+0.5*Q(x/sqrt(2));
% xlab='x'; ylab='Right-tail Probability';nor=1;
% should produce Figure 2.4.
%
% Set up y-axis values.
  y=[0.0001 0.001 0.01 0.1 0.3 0.5 0.7 0.9 0.99 0.999 0.9999]';
% Set up x-axis limits.
  xmax=Qinv(min(y));
% Warp y values for plotting.
  m=(2*min(y)-1)/(2*xmax);
  Y=m*Qinv(Qx)+0.5;
% Check to see if Q(x) (Gaussian right-tail probability)
% is to be plotted.
```



```
  if nor==1
    xnor=[-xmax:0.01:xmax]';
    ynor=m*xnor+0.5;
    plot(x,Y,'-',xnor,ynor,'--')
  else
    plot(x,Y)
  end
  xlabel(xlab)
  ylabel(ylab)
  axis([-xmax xmax min(y) max(y)]);
% Determine y tick mark locations by warping.
  Ytick=m*Qinv(y)+0.5;
% Set up y axis labels.
  t=['0.0001';' 0.001';'  0.01';'   0.1';'   0.3';'   0.5';...
     '   0.7';'   0.9';'  0.99';' 0.999';'0.9999'];
  set(gca,'Ytick',Ytick)
  set(gca,'Yticklabels',t)
  grid
```

Appendix 2C

MATLAB Program to Compute Gaussian Right-Tail Probability and its Inverse

Q.m

```
  function y=Q(x)
% This program computes the right-tail probability
% (complementary cumulative distribution function) for
% a N(0,1) random variable.
%
% Input Parameters:
%
%   x - Real column vector of x values
%
% Output Parameters:
%
%   y - Real column vector of right-tail probabilities
%
% Verification Test Case:
%
% The input x=[0 1 2]'; should produce y=[0.5 0.1587 0.0228]'.
%
 y=0.5*erfc(x/sqrt(2));
```


Qinv.m

```
  function y=Qinv(x)
% This program computes the inverse Q function or the value
% which is exceeded by a N(0,1) random variable with a
% probability of x.
%
% Input Parameters:
%
%   x - Real column vector of right-tail probabilities
%       (in interval [0,1])
%
% Output Parameters:
%
%   y - Real column vector of values of random variable
%
% Verification Test Case:
%
% The input x=[0.5 0.1587 0.0228]'; should produce
% y=[0 0.9998 1.9991]'.
%
  y=sqrt(2)*erfinv(1-2*x);
```

Appendix 2D

MATLAB Program to Compute Central and Noncentral χ^2 Right-Tail Probability

The MATLAB program listed in this appendix computes the right-tail probability of a $\chi_\nu'^2(\lambda)$ random variable. To obtain the result for a central χ^2 random variable or for χ_ν^2, let $\lambda = 0$. The algorithm is based on a similar one described in [Mitchell and Walker 1971].

From (2.15) we have

$$p(x) = \sum_{k=0}^{\infty} \frac{\exp(-\lambda/2)(\lambda/2)^k}{k!} \frac{x^{\nu/2+k-1}\exp(-\frac{1}{2}x)}{2^{\nu/2+k}\Gamma(\nu/2+k)}.$$

The right-tail probability is

$$\begin{aligned} Q_{\chi_\nu'^2(\lambda)}(x) &= \int_x^\infty p(t)dt \\ &= \sum_{k=0}^{\infty} \frac{\exp(-\lambda/2)(\lambda/2)^k}{k!} \int_x^\infty \frac{t^{\nu/2+k-1}\exp(-\frac{1}{2}t)}{2^{\nu/2+k}\Gamma(\nu/2+k)}dt. \end{aligned}$$

But the integral is just $Q_{\chi^2_{\nu+2k}}(x)$, the right-tail probability for a central $\chi^2_{\nu+2k}$ random variable and is given by (2.10) and (2.11). Hence

$$Q_{\chi_\nu'^2(\lambda)}(x) = \sum_{k=0}^{\infty} \frac{\exp(-\lambda/2)(\lambda/2)^k}{k!} Q_{\chi^2_{\nu+2k}}(x). \tag{2D.1}$$

We consider two cases, depending upon whether ν is even or odd. First assume ν is even. Then, $\nu + 2k$ is also even and (2.10) applies. Letting $m = \nu + 2k$ we have

$$Q_{\chi^2_m}(x) = \exp(-x/2) \sum_{l=0}^{\frac{m}{2}-1} \frac{(x/2)^l}{l!} \qquad m \geq 2.$$

To develop a recursion for this term we observe that

$$\begin{aligned} Q_{\chi^2_m}(x) &= \exp(-x/2) \sum_{l=0}^{\frac{m-2}{2}-1} \frac{(x/2)^l}{l!} + \exp(-x/2)\frac{(x/2)^{\frac{m}{2}-1}}{(\frac{m}{2}-1)!} \\ &= Q_{\chi^2_{m-2}}(x) + g(x,m). \end{aligned}$$

But

$$g(x,m) = \frac{x/2}{\frac{m}{2}-1} \frac{\exp(-x/2)(x/2)^{\frac{m-2}{2}-1}}{\left(\frac{m-2}{2}-1\right)!}$$

so that

$$g(x,m) = g(x,m-2)\frac{x}{m-2}$$

and the recursion is initialized by

$$Q_{\chi^2_2}(x) = g(x,2) = \exp(-x/2).$$

Now for ν odd, $m = \nu + 2k$ will also be odd, and (2.11) applies. Thus,

$$Q_{\chi^2_m}(x) = \begin{cases} 2Q(\sqrt{x}) & m = 1 \\ 2Q(\sqrt{x}) + \underbrace{\frac{\exp(-\frac{1}{2}x)}{\sqrt{\pi}} \sum_{l=1}^{\frac{m-1}{2}} \frac{(l-1)!(2x)^{l-\frac{1}{2}}}{(2l-1)!}}_{Q'_{\chi^2_m}(x)} & m \geq 3. \end{cases}$$

and assuming that $\nu \geq 3$ so that $m = \nu + 2k \geq 3$, we have from (2D.1)

$$\begin{aligned} Q_{\chi_\nu'^2(\lambda)}(x) &= \sum_{k=0}^{\infty} \frac{\exp(-\lambda/2)(\lambda/2)^k}{k!} \left[2Q(\sqrt{x}) + Q'_{\chi^2_{\nu+2k}}(x)\right] \\ &= 2Q(\sqrt{x}) + \sum_{k=0}^{\infty} \frac{\exp(-\lambda/2)(\lambda/2)^k}{k!} Q'_{\chi^2_{\nu+2k}}(x) \end{aligned}$$

since

$$\sum_{k=0}^{\infty} \frac{(\lambda/2)^k}{k!} = \exp(\lambda/2).$$

To develop a recursion for $Q'_{\chi^2_m}(x)$, assuming m odd

$$\begin{aligned}
Q'_{\chi^2_m}(x) &= \frac{\exp(-x/2)}{\sqrt{\pi}} \sum_{l=1}^{\frac{m-1}{2}} \frac{(l-1)!}{(2l-1)!}(2x)^{l-\frac{1}{2}} \\
&= \frac{\exp(-x/2)}{\sqrt{\pi}} \sum_{l=1}^{\frac{m-2-1}{2}} \frac{(l-1)!}{(2l-1)!}(2x)^{l-\frac{1}{2}} \\
&\quad + \frac{\exp(-x/2)}{\sqrt{\pi}} \frac{\left(\frac{m-1}{2}-1\right)!}{\left[2\left(\frac{m-1}{2}\right)-1\right]!}(2x)^{\frac{m-1}{2}-\frac{1}{2}} \\
&= Q'_{\chi^2_{m-2}}(x) + g(x,m).
\end{aligned}$$

But

$$\begin{aligned}
g(x,m) &= \frac{\exp(-x/2)}{\sqrt{\pi}} \frac{\left(\frac{m-3}{2}\right)!}{(m-2)!}(2x)^{\frac{m-2}{2}} \\
&= 2x\frac{\left(\frac{m-3}{2}\right)}{(m-2)(m-3)} \cdot \frac{\exp(-x/2)}{\sqrt{\pi}} \frac{\left(\frac{m-5}{2}\right)!}{(m-4)!}(2x)^{\frac{m-4}{2}} \\
&= g(x,m-2)\frac{x}{m-2}
\end{aligned}$$

where the recursion is initialized with

$$Q'_{\chi^2_3}(x) = g(x,3) = \sqrt{\frac{2x}{\pi}}\exp(-x/2).$$

If $\nu = 1$, then for $k = 0$, we have $m = \nu + 2k = 1$. In this case

$$Q_{\chi^2_m}(x) = 2Q(\sqrt{x})$$

so that

$$Q'_{\chi^2_1}(x) = 0.$$

To determine how many terms of the sum are required, we choose the first $M+1$ terms so that the truncation error is less than ϵ. To find M we have from (2D.1)

$$\epsilon = \sum_{k=M+1}^{\infty} \frac{\exp(-\lambda/2)(\lambda/2)^k}{k!} Q_{\chi^2_{\nu+2k}}(x).$$

But $Q_{\chi^2_{\nu+2k}}(x) < 1$ so that

$$\begin{aligned}
\epsilon &< \sum_{k=M+1}^{\infty} \frac{\exp(-\lambda/2)(\lambda/2)^k}{k!} \\
&= \exp(-\lambda/2)\sum_{k=0}^{\infty} \frac{(\lambda/2)^k}{k!} - \sum_{k=1}^{M} \frac{\exp(-\lambda/2)(\lambda/2)^k}{k!} \\
&= 1 - \exp(-\lambda/2)\sum_{k=0}^{M} \frac{(\lambda/2)^k}{k!}.
\end{aligned}$$

To maintain an error of less than ϵ we require

$$\sum_{k=0}^{M} \frac{(\lambda/2)^k}{k!} > \exp(-\lambda/2)(1-\epsilon). \tag{2D.2}$$

For a given ϵ and λ, the value of M can be found from (2D.2).

Qchipr2.m

```
  function P=Qchipr2(nu,lambda,x,epsilon)
%
% This program computes the right-tail probability
% of a central or noncentral chi-squared PDF.
%
% Input Parameters:
%
%   nu      = Degrees of freedom (1,2,3,etc.)
%   lambda  = Noncentrality parameter (must be positive),
%             set = 0 for central chi-squared PDF
%   x       = Real scalar value of random variable
%   epsilon = maximum allowable error (should be a small
%             number such as 1e-5) due to truncation of the
%             infinite sum
%
% Output Parameters:
%
%   P       = right-tail probability or the probability that
%             the random variable exceeds the given value
%             (1 - CDF)
%
% Verification Test Case:
%
```

```
%    The inputs nu=1, lambda=2, x=0.5, epsilon=0.0001
%    should produce P=0.7772.
%    The inputs nu=5, lambda=6, x=10, epsilon=0.0001
%    should produce P=0.5063.
%    The inputs nu=8, lambda=10, x=15, epsilon=0.0001
%    should produce P=0.6161.
%
%  Determine how many terms in sum to be used (find M).
   t=exp(lambda/2)*(1-epsilon);
   sum=1;
   M=0;
   while sum < t
     M=M+1;
     sum=sum+((lambda/2)^M)/prod(1:M);
   end
%  Use different algorithms for nu even or odd.
   if (nu/2-floor(nu/2)) == 0  % nu is even.
%  Compute k=0 term of sum.
%  Compute Qchi2_nu(x).
%  Start recursion with Qchi2_2(x).
     Q2=exp(-x/2); g=Q2;
     for m=4:2:nu  %  If nu=2, loop will be omitted.
       g=g*x/(m-2);
       Q2=Q2+g;
     end
%  Finish computation of k=0 term.
     P=exp(-lambda/2)*Q2;
%  Compute remaining terms of sum.
     for k=1:M
       m=nu+2*k;
       g=g*x/(m-2); Q2=Q2+g;
       arg=(exp(-lambda/2)*(lambda/2)^k)/prod(1:k);
       P=P+arg*Q2;
     end
   else  % nu is odd.
%  Compute k=0 term of sum.
     P=2*Q(sqrt(x));
%  Start recursion with Qchi2p_3(x).
     Q2p=sqrt(2*x/pi)*exp(-x/2); g=Q2p;
     if nu >1
       for m=5:2:nu  %  If nu=3, loop will be omitted.
         g=g*x/(m-2);
         Q2p=Q2p+g;
       end
       P=P+exp(-lambda/2)*Q2p;
```



```
%  Compute remaining terms of sum.
       for k=1:M
         m=nu+2*k;
         g=g*x/(m-2); Q2p=Q2p+g;
         arg=(exp(-lambda/2)*(lambda/2)^k)/prod(1:k);
         P=P+arg*Q2p;
       end
     else
%  If nu=1, the k=0 term is just Qchi2_1(x)=2Q(sqrt(x)).
%  Add the k=0 and k=1 terms.
       P=P+exp(-lambda/2)*(lambda/2)*Q2p;
%  Compute remaining terms.
       for k=2:M
         m=nu+2*k;
         g=g*x/(m-2); Q2p=Q2p+g;
         arg=(exp(-lambda/2)*(lambda/2)^k)/prod(1:k);
         P=P+arg*Q2p;
       end
     end
   end
```

Appendix 2E

MATLAB Program for Monte Carlo Computer Simulation

montecarlo.m

```
% This program is a Monte Carlo computer simulation that was
% used to generate Figure 2.10a.
%
% Set seed of random number generator to initial value.
  randn('seed',0);
% Set up values of variance, data record length, and number
% of realizations.
  var=10;
  N=10;
  M=1000;
% Dimension array of realizations.
  T=zeros(M,1);
% Compute realizations of the sample mean.
  for i=1:M
    x=sqrt(var)*randn(N,1);
    T(i)=mean(x);
  end
% Set number of values of gamma.
  ngam=100;
% Set up gamma array.
  gammamin=min(T);
  gammamax=max(T);
  gamdel=(gammamax-gammamin)/ngam;
  gamma=[gammamin:gamdel:gammamax]';
% Dimension P (the Monte Carlo estimate) and Ptrue
```



```
% (the theoretical or true probability).
  P=zeros(length(gamma),1);Ptrue=P;
% Determine for each gamma how many realizations exceeded
% gamma (Mgam) and use this to estimate the probability.
  for i=1:length(gamma)
    clear Mgam;
    Mgam=find(T>gamma(i));
    P(i)=length(Mgam)/M;
  end
% Compute the true probability.
  Ptrue=Q(gamma/(sqrt(var/N)));
  plot(gamma,P,'-',gamma,Ptrue,'--')
  xlabel('gamma')
  ylabel('P(T>gamma)')
  grid
```

Chapter 3

Statistical Decision Theory I

3.1 Introduction

In this chapter we lay the basic statistical groundwork for the design of detectors of signals in noise. The approaches follow directly from the theory of hypothesis testing. In particular, we address the *simple hypothesis* testing problem in which the PDF for each assumed hypothesis is completely known. A much more complicated problem arises when the PDF has unknown parameters. We defer that discussion until Chapter 6. The primary approaches to simple hypothesis testing are the *classical* approach based on the Neyman-Pearson theorem and the *Bayesian* approach based on minimization of the Bayes risk. In many ways these approaches are analogous to the classical and Bayesian methods of statistical estimation theory. The particular method employed depends upon our willingness to incorporate prior knowledge about the probabilities of occurrence of the various hypotheses. The choice of an appropriate approach is therefore dictated by the problem at hand. Sonar and radar systems typically use the Neyman-Pearson criterion while communications and pattern recognition systems employ the Bayes risk.

3.2 Summary

The detector that maximizes the probability of detection for a given probability of false alarm is the likelihood ratio test of (3.3) as specified by the Neyman-Pearson theorem. The threshold is found from the false alarm constraint. For the mean-shifted Gauss-Gauss hypothesis testing problem, the detection performance is summarized by (3.10) and is monotonic with the deflection coefficient of (3.9). The performance of a detector can also be displayed using the receiver operating characteristics as discussed in Section 3.4. In some detection problems a subset of the data may be discarded as being irrelevant to a decision. This is discussed in Section 3.5, with the condition for irrelevant data being (3.11). To minimize the probability of decision error as given by (3.12), we should employ the detector of (3.13), where the

threshold is now determined by the prior probabilities of the hypotheses. For equal prior probabilities of the hypotheses, the detector becomes the maximum likelihood detector of (3.14). More generally, the optimal decision rule that minimizes the probability of error is given by the maximum a posteriori probability detector of (3.16). A generalization of the minimum probability of error criterion is the Bayes risk as discussed in Section 3.7 with the detector given by (3.18). For multiple hypothesis testing the decision rule for minimizing the Bayes risk is given by (3.21). Specializing the result to the minimum probability of error criterion leads to the maximum a posteriori probability detector of (3.22) and the maximum likelihood detector of (3.24) for multiple hypothesis testing.

3.3 Neyman-Pearson Theorem

In discussing the Neyman-Pearson (NP) approach to signal detection we will center our discussion around a simple example of hypothesis testing. Assume that we observe a realization of a random variable whose PDF is either $\mathcal{N}(0,1)$ or $\mathcal{N}(1,1)$. The notation $\mathcal{N}(\mu,\sigma^2)$ denotes a Gaussian PDF with mean μ and variance σ^2. We must therefore determine if $\mu = 0$ or $\mu = 1$ based on a single observation $x[0]$. Each possible value of μ can be thought of as a hypothesis so that our problem is to choose among two competing hypotheses. These are summarized as follows:

$$\begin{aligned}\mathcal{H}_0 &: \mu = 0 \\ \mathcal{H}_1 &: \mu = 1\end{aligned} \tag{3.1}$$

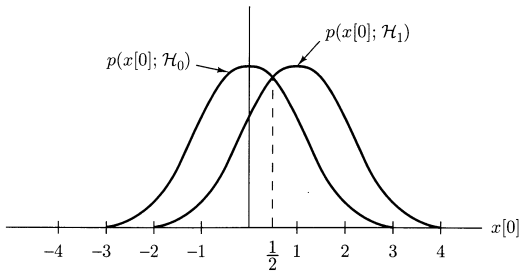


Figure 3.1. PDFs for hypothesis testing problem.

where $\mathcal{H}_0$ is referred to as the *null hypothesis* and $\mathcal{H}_1$ as the *alternative hypothesis*. This problem is known as a *binary* hypothesis test since we must choose between *two* hypotheses. The PDFs under each hypothesis are shown in Figure 3.1, with the difference in means causing the PDF under $\mathcal{H}_1$ to be shifted to the right. On the basis of a single sample it is difficult to determine which PDF generated it.

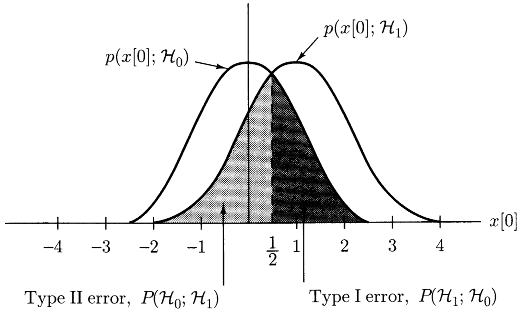


Figure 3.2. Possible hypothesis testing errors and their probabilities.

However, a reasonable approach might be to decide $\mathcal{H}_1$ if $x[0] > 1/2$. This is because if $x[0] > 1/2$, the observed sample is more *likely* if $\mathcal{H}_1$ is true. Or if $x[0] > 1/2$, we have from Figure 3.1 that $p(x[0];\mathcal{H}_1) > p(x[0];\mathcal{H}_0)$. Our detector then compares the observed datum value with $1/2$, the latter being called the *threshold*. Note that with this scheme we can make two types of errors. If we decide $\mathcal{H}_1$ but $\mathcal{H}_0$ is true, we make a *Type I error*. On the other hand, if we decide $\mathcal{H}_0$ but $\mathcal{H}_1$ is true, we make a *Type II error*. These errors are illustrated in Figure 3.2. The notation $P(\mathcal{H}_i;\mathcal{H}_j)$ indicates the probability of deciding $\mathcal{H}_i$ when $\mathcal{H}_j$ is true. For example, $P(\mathcal{H}_1;\mathcal{H}_0) = \Pr\{x[0] > 1/2; \mathcal{H}_0\}$ and is shown as the darker area. These two errors are unavoidable to some extent but may be traded off against each other. To do so we need only change the threshold as shown in Figure 3.3. Clearly, the Type I error probability ($P(\mathcal{H}_1;\mathcal{H}_0)$) is decreased at the expense of increasing the Type II error probability ($P(\mathcal{H}_0;\mathcal{H}_1)$). *It is not possible to reduce both error probabilities simultaneously.* A typical approach then in designing an optimal detector is to hold one error probability fixed while minimizing the other. We choose to constrain $P(\mathcal{H}_1;\mathcal{H}_0)$ to a fixed value, say α. If we view the problem of (3.1) as an attempt to distinguish between the hypotheses

$$\begin{array}{l}\mathcal{H}_0 : x[0] = w[0] \\ \mathcal{H}_1 : x[0] = s[0] + w[0]\end{array}$$

where $s[0] = 1$ and $w[0] \sim \mathcal{N}(0,1)$, then we have the signal detection problem. Deciding $\mathcal{H}_1$ when $\mathcal{H}_0$ is true can be thought of as a false alarm. As a result, $P(\mathcal{H}_1;\mathcal{H}_0)$ is referred to as the *probability of false alarm* and is denoted by P_{FA}. Usually this is a small value, say 10^{-8}, in keeping with the disastrous effects that may ensue. For example, if we falsely say an enemy aircraft is present, we may initiate an attack. To design the optimal detector we then seek to minimize the other error $P(\mathcal{H}_0;\mathcal{H}_1)$ or equivalently to maximize $1 - P(\mathcal{H}_0;\mathcal{H}_1)$. The latter is just

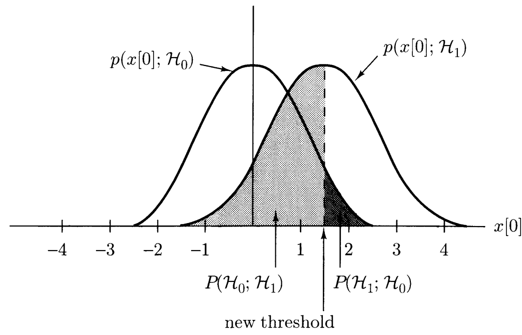


Figure 3.3. Trading off errors by adjusting threshold.

$P(\mathcal{H}_1;\mathcal{H}_1)$ and in keeping with the signal detection problem is called the *probability of detection*. It is denoted by P_D. This setup is termed the *Neyman-Pearson* (NP) approach to hypothesis testing or to signal detection. In summary, we wish to *maximize* $P_D = P(\mathcal{H}_1;\mathcal{H}_1)$ subject to the constraint $P_{FA} = P(\mathcal{H}_1;\mathcal{H}_0) = \alpha$.

Returning to the previous example we can constrain P_{FA} by choosing the threshold γ since

$$\begin{aligned} P_{FA} &= P(\mathcal{H}_1;\mathcal{H}_0) \\ &= \Pr\{x[0] > \gamma; \mathcal{H}_0\} \\ &= \int_\gamma^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt \\ &= Q(\gamma). \end{aligned}$$

As an example, if $P_{FA} = 10^{-3}$, we have $\gamma = 3$. We therefore decide $\mathcal{H}_1$ if $x[0] > 3$. Furthermore, with this choice we have

$$\begin{aligned} P_D &= P(\mathcal{H}_1;\mathcal{H}_1) \\ &= \Pr\{x[0] > \gamma; \mathcal{H}_1\} \\ &= \int_\gamma^\infty \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(t-1)^2\right] dt \\ &= Q(\gamma - 1) = Q(2) = 0.023. \end{aligned}$$

The question arises as to whether $P_D = 0.023$ is the maximum P_D for this problem. Our choice of the detector that decides $\mathcal{H}_1$ if $x[0] > \gamma$ was just a guess. Might there be a better approach?

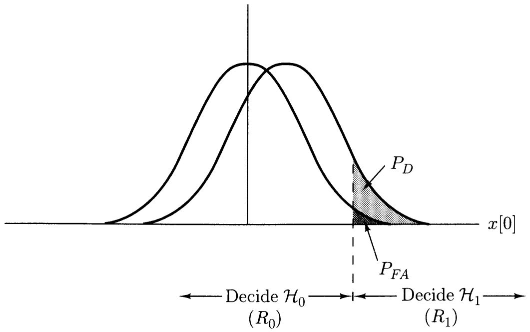


Figure 3.4. Decision regions and probabilities.

Before answering this question we first describe the operation of a detector in more general terms. The goal of a detector is to decide either $\mathcal{H}_0$ or $\mathcal{H}_1$ based on an observed set of data $\{x[0], x[1], \ldots, x[N-1]\}$. This is a mapping from each possible data set value into a decision. For the previous example the *decision regions* are shown in Figure 3.4. A detector then may be thought of as a mapping from the data values into a decision. In particular, let R_1 be the set of values in R^N that map into the decision $\mathcal{H}_1$ or

$$R_1 = \{\mathbf{x} : \text{decide } \mathcal{H}_1 \text{ or reject } \mathcal{H}_0\}.$$

This region is termed the *critical region* in statistics. The set of points in R^N that map into the decision $\mathcal{H}_0$ is the complement set of R_1 or $R_0 = \{\mathbf{x} : \text{decide } \mathcal{H}_0$ or reject $\mathcal{H}_1\}$. Clearly, $R_0 \cup R_1 = R^N$ since R_0 and R_1 *partition* the data space. For the previous example the critical region was $x[0] > 3$. The P_{FA} constraint then becomes

$$P_{FA} = \int_{R_1} p(\mathbf{x}; \mathcal{H}_0) d\mathbf{x} = \alpha. \tag{3.2}$$

In statistics, α is termed the *significance level* or *size* of the test. Now there are many sets R_1 that satisfy (3.2) (see Problem 3.2). Our goal is to choose the one that maximizes

$$P_D = \int_{R_1} p(\mathbf{x}; \mathcal{H}_1) d\mathbf{x}.$$

In statistics, P_D is called the *power* of the test and the critical region that attains the maximum power is the *best critical region.* See Table 3.1 for a summary of the statistical terminology and the engineering equivalents.

The NP theorem tells us how to choose R_1 if we are given $p(\mathbf{x}; \mathcal{H}_0)$, $p(\mathbf{x}; \mathcal{H}_1)$, and α.

Statisticians	Engineers
Test statistic ($T(\mathbf{x})$) and threshold (γ)	Detector
Null hypothesis ($\mathcal{H}_0$)	Noise only hypothesis
Alternative hypothesis ($\mathcal{H}_1$)	Signal + noise hypothesis
Critical region	Signal present decision region
Type I error (decide $\mathcal{H}_1$ when $\mathcal{H}_0$ true)	False alarm (FA)
Type II error (decide $\mathcal{H}_0$ when $\mathcal{H}_1$ true)	Miss (M)
Level of significance or size of test (α)	Probability of false alarm (P_{FA})
Probability of Type II error (β)	Probability of miss (P_M)
Power of test ($1-\beta$)	Probability of detection (P_D)

Table 3.1. Cross-Reference of Statistical Terms for Binary Hypothesis Testing

Theorem 3.1 (Neyman-Pearson) *To maximize P_D for a given $P_{FA} = \alpha$ decide $\mathcal{H}_1$ if*

$$L(\mathbf{x}) = \frac{p(\mathbf{x}; \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > \gamma \tag{3.3}$$

where the threshold γ is found from

$$P_{FA} = \int_{\{\mathbf{x}: L(\mathbf{x}) > \gamma\}} p(\mathbf{x}; \mathcal{H}_0) d\mathbf{x} = \alpha.$$

The proof is given in Appendix 3A. The function $L(\mathbf{x})$ is termed the *likelihood ratio* since it indicates for each value of $\mathbf{x}$ the likelihood of $\mathcal{H}_1$ versus the likelihood of $\mathcal{H}_0$. The entire test of (3.3) is called the *likelihood ratio test* (LRT). We next illustrate the NP test with some examples.

Example 3.1 - Introductory Example (continued)

For the hypothesis test of (3.1) we can easily find the NP test. Assume that we require $P_{FA} = 10^{-3}$. Then, from (3.3) we decide $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x}; \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} = \frac{\frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x[0]-1)^2\right]}{\frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}x^2[0]\right]} > \gamma$$

or

$$\exp\left[-\frac{1}{2}(x^2[0] - 2x[0] + 1 - x^2[0])\right] > \gamma$$

or finally

$$\exp\left(x[0] - \frac{1}{2}\right) > \gamma. \tag{3.4}$$

At this point we could determine γ from the false alarm constraint

$$P_{FA} = \Pr\left\{\exp\left(x[0] - \frac{1}{2}\right) > \gamma; \mathcal{H}_0\right\} = 10^{-3}.$$

This would require us to find the PDF of $\exp(x[0]-1/2)$. A much simpler approach is to note that the inequality of (3.4) is not changed if we take logarithms of both sides. This is because the logarithm is a monotonically increasing function (see Problem 3.3). Alternatively, since $\gamma > 0$, we can let $\gamma = \exp(\beta)$ so that we decide $\mathcal{H}_1$ if

$$\exp\left(x[0] - \frac{1}{2}\right) > \exp(\beta)$$

or

$$x[0] > \beta + \frac{1}{2} = \ln\gamma + \frac{1}{2}.$$

Letting $\gamma' = \ln\gamma + 1/2$ we decide $\mathcal{H}_1$ if $x[0] > \gamma'$. To explicitly find γ' (or equivalently γ) we use the P_{FA} constraint

$$\begin{aligned} P_{FA} = \Pr\{x[0] > \gamma'; \mathcal{H}_0\} &= 10^{-3} \\ \int_{\gamma'}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt &= 10^{-3} \end{aligned}$$

so that $\gamma' = 3$. The NP test is to decide $\mathcal{H}_1$ if $x[0] > 3$. Thus, the detector of the previous example is indeed optimum in the NP sense in that it maximizes P_D. As before we find P_D as follows

$$\begin{aligned} P_D &= \Pr\{x[0] > 3; \mathcal{H}_1\} \\ &= \int_3^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(t-1)^2\right] dt = 0.023. \end{aligned}$$

Note that the detection performance is poor. Although we have satisfied our false alarm constraint, we will only detect the signal a small fraction of the time. To improve the detection performance we can increase P_{FA}, employing the usual tradeoff. For example, if $P_{FA} = 0.5$, then the threshold is found from

$$0.5 = \int_{\gamma'}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt$$

as $\gamma' = 0$. Then

$$\begin{aligned} P_D &= \int_{\gamma'}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(t-1)^2\right] dt \\ &= \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(t-1)^2\right] dt \end{aligned}$$

$$\begin{aligned} &= Q\left(\frac{0-1}{1}\right) = Q(-1) \\ &= 1 - Q(1) = 0.84. \end{aligned}$$

(Recall that if $x \sim \mathcal{N}(\mu, \sigma^2)$, the right-tail probability for a threshold γ' is $Q((\gamma' - \mu)/\sigma)$. See Chapter 2.) By changing the threshold we can trade off P_{FA} and P_D. This point is discussed further in the next section.

◇

Example 3.2 - DC Level in WGN

Now consider the more general signal detection problem

$$\begin{aligned} \mathcal{H}_0 &: x[n] = w[n] \qquad n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 &: x[n] = A + w[n] \quad n = 0, 1, \ldots, N-1 \end{aligned}$$

where the signal is $s[n] = A$ for $A > 0$ and $w[n]$ is WGN with variance σ^2. The previous example is just a special case where $A = 1$, $N = 1$, and $\sigma^2 = 1$. Also, note that the current problem is actually a test of the mean of a multivariate Gaussian PDF. This is because under $\mathcal{H}_0$, $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ while under $\mathcal{H}_1$, $\mathbf{x} \sim \mathcal{N}(A\mathbf{1}, \sigma^2\mathbf{I})$, where $\mathbf{1}$ is the vector of all ones. Hence, we have equivalently

$$\begin{aligned} \mathcal{H}_0 &: \boldsymbol{\mu} = \mathbf{0} \\ \mathcal{H}_1 &: \boldsymbol{\mu} = A\mathbf{1}. \end{aligned}$$

We will often use this *parameter test of the PDF* interpretation in describing a signal detection problem. Now the NP detector decides $\mathcal{H}_1$ if

$$\frac{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right]}{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right]} > \gamma.$$

Taking the logarithm of both sides results in

$$-\frac{1}{2\sigma^2}\left(-2A\sum_{n=0}^{N-1}x[n] + NA^2\right) > \ln\gamma$$

which simplifies to

$$\frac{A}{\sigma^2}\sum_{n=0}^{N-1}x[n] > \ln\gamma + \frac{NA^2}{2\sigma^2}.$$

Since $A > 0$, we have finally

$$\frac{1}{N}\sum_{n=0}^{N-1} x[n] > \frac{\sigma^2}{NA}\ln\gamma + \frac{A}{2} = \gamma'. \tag{3.5}$$

The NP detector compares the *sample mean* $\bar{x} = (1/N)\sum_{n=0}^{N-1} x[n]$ to a threshold γ'. This is intuitively reasonable since $\bar{x}$ may be thought of as an estimate of A. If the estimate is large and positive, then the signal is probably present. How large the estimate must be before we are willing to declare that a signal is present depends upon our concern that noise only may cause a large estimate. To avoid this possibility we adjust γ' to control P_{FA}, with larger threshold values reducing P_{FA} (as well as P_D).

To determine the detection performance we first note that the test statistic $T(\mathbf{x}) = (1/N)\sum_{n=0}^{N-1} x[n]$ is Gaussian under each hypothesis. The means and variances are

$$\begin{aligned} E(T(\mathbf{x});\mathcal{H}_0) &= E\left(\frac{1}{N}\sum_{n=0}^{N-1} w[n]\right) \\ &= \frac{1}{N}\sum_{n=0}^{N-1} E(w[n]) \\ &= 0. \end{aligned}$$

Similarly, $E(T(\mathbf{x});\mathcal{H}_1) = A$ and

$$\begin{aligned} \mathrm{var}(T(\mathbf{x});\mathcal{H}_0) &= \mathrm{var}\left(\frac{1}{N}\sum_{n=0}^{N-1} w[n]\right) \\ &= \frac{1}{N^2}\sum_{n=0}^{N-1} \mathrm{var}(w[n]) \\ &= \frac{\sigma^2}{N}. \end{aligned}$$

Similarly, $\mathrm{var}(T(\mathbf{x});\mathcal{H}_1) = \sigma^2/N$ where we have noted that the noise samples are uncorrelated. Thus,

$$T(\mathbf{x}) \sim \begin{cases} \mathcal{N}(0, \frac{\sigma^2}{N}) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(A, \frac{\sigma^2}{N}) & \text{under } \mathcal{H}_1. \end{cases}$$

We have then

$$\begin{aligned} P_{FA} &= \Pr\{T(\mathbf{x}) > \gamma'; \mathcal{H}_0\} \\ &= Q\left(\frac{\gamma'}{\sqrt{\sigma^2/N}}\right) \end{aligned} \tag{3.6}$$

and

$$\begin{aligned} P_D &= \Pr\{T(\mathbf{x}) > \gamma'; \mathcal{H}_1\} \\ &= Q\left(\frac{\gamma' - A}{\sqrt{\sigma^2/N}}\right). \end{aligned} \tag{3.7}$$

We can relate P_D to P_{FA} more directly by noting that the Q function is monotonically decreasing since $1 - Q$ is a CDF, which is monotonically increasing. Thus, Q has an inverse that we denote as Q^{-1}. As a result, the threshold is found from (3.6) as

$$\gamma' = \sqrt{\frac{\sigma^2}{N}}Q^{-1}(P_{FA})$$

and

$$\begin{aligned} P_D &= Q\left(\frac{\sqrt{\sigma^2/N}Q^{-1}(P_{FA}) - A}{\sqrt{\sigma^2/N}}\right) \\ &= Q\left(Q^{-1}(P_{FA}) - \sqrt{\frac{NA^2}{\sigma^2}}\right). \end{aligned} \tag{3.8}$$

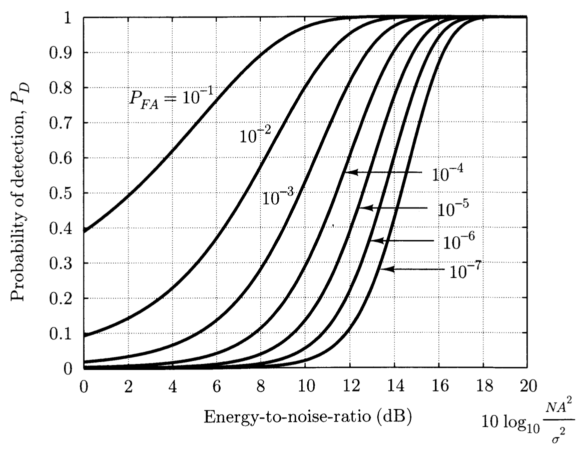


Figure 3.5. Detection performance for DC level in WGN.

It is seen that for a given P_{FA} the detection performance increases monotonically with NA^2/σ^2, which is the *signal energy-to-noise ratio* (ENR). An alternative interpretation is explored in Problem 3.5. The detection performance is shown in Figure 3.5 for various values of P_{FA}. It is sometimes convenient to display the detection curves on normal probability paper (see Chapter 2). This has the effect of straightening the curves when plotted versus $\sqrt{\text{ENR}}$ as shown in Figure 3.6. The advantage is an easier reading of the required ENR for a given P_D, especially for P_D's near one. The disadvantage is that the abscissa values are not in decibels (dB), which is customary in engineering. We will usually employ the former approach. $\diamond$

The previous example illustrates a particularly useful hypothesis testing problem called the *mean-shifted Gauss-Gauss* problem. We observe the value of a test statistic T and decide $\mathcal{H}_1$ if $T > \gamma'$ and $\mathcal{H}_0$ otherwise. The PDF of T is assumed to be

$$T \sim \begin{cases} \mathcal{N}(\mu_0, \sigma^2) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(\mu_1, \sigma^2) & \text{under } \mathcal{H}_1 \end{cases}$$

where $\mu_1 > \mu_0$. Hence, we wish to decide between the two hypotheses that differ by a shift in the mean of T. In the previous example $T = \bar{x}$. For this type of detector

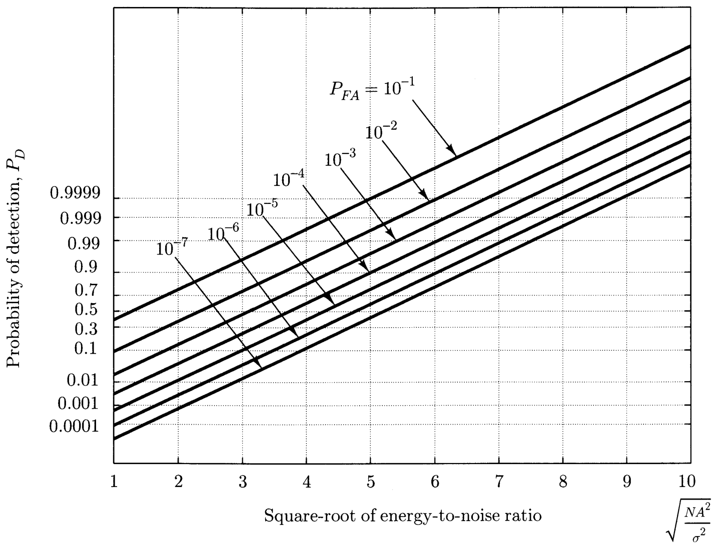


Figure 3.6. Detection performance for DC level in WGN-normal probability paper.

the detection performance is totally characterized by the *deflection coefficient* d^2. This is defined as

$$\begin{aligned} d^2 &= \frac{(E(T;\mathcal{H}_1) - E(T;\mathcal{H}_0))^2}{\text{var}(T;\mathcal{H}_0)} \qquad (3.9) \\ &= \frac{(\mu_1 - \mu_0)^2}{\sigma^2}. \end{aligned}$$

In the case when $\mu_0 = 0$, $d^2 = \mu_1^2/\sigma^2$ may be interpreted as a signal-to-noise ratio (SNR). To verify the dependence of detection performance on d^2 we have that

$$\begin{aligned} P_{FA} &= \Pr\{T > \gamma'; \mathcal{H}_0\} \\ &= Q\left(\frac{\gamma' - \mu_0}{\sigma}\right) \\ P_D &= \Pr\{T > \gamma'; \mathcal{H}_1\} \\ &= Q\left(\frac{\gamma' - \mu_1}{\sigma}\right) \\ &= Q\left(\frac{\mu_0 + \sigma Q^{-1}(P_{FA}) - \mu_1}{\sigma}\right) \\ &= Q\left(Q^{-1}(P_{FA}) - \left(\frac{\mu_1 - \mu_0}{\sigma}\right)\right) \end{aligned}$$

and using (3.9) we have

$$P_D = Q\left(Q^{-1}(P_{FA}) - \sqrt{d^2}\right) \qquad (3.10)$$

since $\mu_1 > \mu_0$. *The detection performance is therefore monotonic with the deflection coefficient.* We end this section with another example.

Example 3.3 - Change in Variance

This hypothesis testing example illustrates that a change in the variance of a Gaussian statistic can be used to distinguish between two hypotheses. We observe $x[n]$ for $n = 0, 1, \ldots, N-1$, where the $x[n]$'s are independent and identically distributed (IID). The latter qualification means that the first-order PDF for each $x[n]$ is the same. Assume that $x[n] \sim \mathcal{N}(0, \sigma_0^2)$ under $\mathcal{H}_0$ and $x[n] \sim \mathcal{N}(0, \sigma_1^2)$ under $\mathcal{H}_1$, where $\sigma_1^2 > \sigma_0^2$. Then the NP test is to decide $\mathcal{H}_1$ if

$$\frac{\frac{1}{(2\pi\sigma_1^2)^{\frac{N}{2}}} \exp\left(-\frac{1}{2\sigma_1^2}\sum_{n=0}^{N-1} x^2[n]\right)}{\frac{1}{(2\pi\sigma_0^2)^{\frac{N}{2}}} \exp\left(-\frac{1}{2\sigma_0^2}\sum_{n=0}^{N-1} x^2[n]\right)} > \gamma.$$

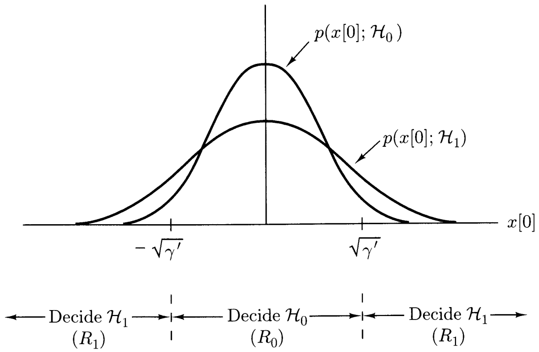


Figure 3.7. Decision regions for change in variance hypothesis test.

Taking logarithms of both sides we have

$$-\frac{1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)\sum_{n=0}^{N-1} x^2[n] > \ln\gamma + \frac{N}{2}\ln\frac{\sigma_1^2}{\sigma_0^2}.$$

Since $\sigma_1^2 > \sigma_0^2$, we have

$$\frac{1}{N}\sum_{n=0}^{N-1} x^2[n] > \gamma'$$

where

$$\gamma' = \frac{\frac{2}{N}\ln\gamma + \ln\frac{\sigma_1^2}{\sigma_0^2}}{\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}}.$$

The test statistic is just an estimate of the variance. We decide $\mathcal{H}_1$ if the power in the observed samples is large enough. In particular, if $N = 1$ we have a detector that decides $\mathcal{H}_1$ if $x^2[0] > \gamma'$ or equivalently if $|x[0]| > \sqrt{\gamma'}$. The decision regions are shown in Figure 3.7 and are seen to be plausible. The performance of this detector is examined in Problem 3.9 for $N = 2$ and in more generality in Chapter 5, where we discuss the energy detector.

◇

Note that for the DC level in WGN and the change in variance examples we distinguish between two hypotheses whose PDFs have different parameter values. We do so by estimating the parameter and comparing the estimated value to a threshold. This is not merely a coincidence but is due to the presence of a sufficient statistic [Kay-I 1993, Chapter 5]. In particular, assume that we observe

$\mathbf{x} = [x[0]\, x[1] \ldots x[N-1]]^T$ from a PDF that is parameterized by θ. The PDF is denoted by $p(\mathbf{x}; \theta)$. (In the DC level in WGN example $\theta = A$.) We wish to test for the value of θ as

$$\begin{aligned}\mathcal{H}_0 &: \theta = \theta_0\\ \mathcal{H}_1 &: \theta = \theta_1.\end{aligned}$$

If a sufficient statistic exists for θ, then by the Neyman-Fisher factorization theorem [Kay-I 1993, Chapter 5] we can express the PDF as

$$p(\mathbf{x}; \theta) = g(T(\mathbf{x}), \theta)h(\mathbf{x})$$

where $T(\mathbf{x})$ is a sufficient statistic for θ. The NP test, which is

$$\frac{p(\mathbf{x}; \theta_1)}{p(\mathbf{x}; \theta_0)} > \gamma$$

then becomes

$$\frac{g(T(\mathbf{x}), \theta_1)}{g(T(\mathbf{x}), \theta_0)} > \gamma.$$

Clearly, the test will depend on the data only through $T(\mathbf{x})$. In the DC level in WGN example it can be shown (see Problem 3.10) that the sufficient statistic is $T(\mathbf{x}) = (1/N)\sum_{n=0}^{N-1} x[n]$ while in the change in variance example $T(\mathbf{x}) = (1/N)\sum_{n=0}^{N-1} x^2[n]$ is a sufficient statistic. In essence the sufficient statistic summarizes all the relevant information in the data about θ that is needed to make a decision. (See also Problem 3.11.) Furthermore, if $T(\mathbf{x})$ is an unbiased estimator of θ, then the detector will be based on an *estimate* of the unknown parameter. Unfortunately, sufficient statistics do not always exist, as our final example illustrates.

Example 3.4 - DC Level in NonGaussian Noise

Assume that under $\mathcal{H}_0$ we observe N IID samples $x[n] = w[n]$ for $n = 0, 1, \ldots, N-1$ from the noise PDF $p(w[n])$ while under $\mathcal{H}_1$ we observe $x[n] = A + w[n]$ for $n = 0, 1, \ldots, N-1$. Thus, under $\mathcal{H}_0$ we have

$$p(\mathbf{x}; \mathcal{H}_0) = \prod_{n=0}^{N-1} p(x[n])$$

and under $\mathcal{H}_1$ we have

$$p(\mathbf{x}; \mathcal{H}_1) = \prod_{n=0}^{N-1} p(x[n] - A).$$

The NP detector decides $\mathcal{H}_1$ if

$$\frac{\prod_{n=0}^{N-1} p(x[n] - A)}{\prod_{n=0}^{N-1} p(x[n])} > \gamma.$$

If the PDF of the noise is a *Gaussian mixture*

$$p(w[n]) = \frac{1}{2}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}w^2[n]\right) + \frac{1}{2}\frac{1}{\sqrt{4\pi}}\exp\left(-\frac{1}{4}w^2[n]\right)$$

then the detector becomes

$$\frac{\prod_{n=0}^{N-1}\frac{1}{2}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(x[n]-A)^2\right) + \frac{1}{2}\frac{1}{\sqrt{4\pi}}\exp\left(-\frac{1}{4}(x[n]-A)^2\right)}{\prod_{n=0}^{N-1}\frac{1}{2}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x^2[n]\right) + \frac{1}{2}\frac{1}{\sqrt{4\pi}}\exp\left(-\frac{1}{4}x^2[n]\right)} > \gamma.$$

No further simplification is possible due to the lack of a sufficient statistic for A. We will explore the nonGaussian detection problem further in Chapter 10. ◇

3.4 Receiver Operating Characteristics

An alternative way of summarizing the detection performance of a NP detector is to plot P_D versus P_{FA}. As an example, for the DC level in WGN we have from (3.6), (3.7), and (3.8)

$$\begin{aligned} P_{FA} &= Q\left(\frac{\gamma'}{\sqrt{\sigma^2/N}}\right) \\ P_D &= Q\left(\frac{\gamma'-A}{\sqrt{\sigma^2/N}}\right) \end{aligned}$$

and

$$P_D = Q\left(Q^{-1}(P_{FA}) - \sqrt{d^2}\right)$$

where $d^2 = NA^2/\sigma^2$. The latter is shown in Figure 3.8 for $d^2 = 1$. Each point on the curve corresponds to a value of (P_{FA}, P_D) for a *given* threshold γ'. By adjusting γ' any point on the curve may be obtained. As expected as γ' increases, P_{FA} decreases but so does P_D and vice-versa. This type of performance summary is called the *receiver operating characteristic* (ROC). The ROC should always be above the 45° line (shown dashed in Figure 3.8). This is because the 45° ROC can be attained by a detector that bases its decision on flipping a coin, ignoring all the data. Consider the detector that decides $\mathcal{H}_1$ if a flipped coin comes up a head, where $\Pr\{\text{head}\} = p$. For a tail outcome we decide $\mathcal{H}_0$. Then,

$$\begin{aligned} P_{FA} &= \Pr\{\text{head}; \mathcal{H}_0\} \\ P_D &= \Pr\{\text{head}; \mathcal{H}_1\}. \end{aligned}$$

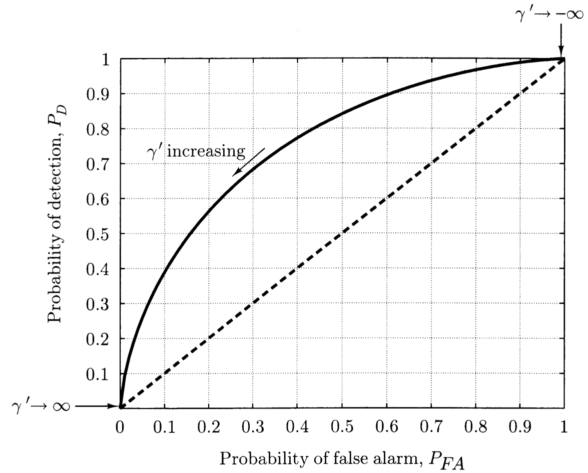


Figure 3.8. Receiver operating characteristics for DC level in WGN ($d^2 = 1$).

But the probability of obtaining a head does not depend upon which hypothesis is true and so $P_{FA} = P_D = p$. This detector then generates the point (p,p) on the ROC. To generate the other points on the 45° line we need only use coins with different p.

As the deflection coefficient increases, a family of ROCs is generated as shown in Figure 3.9. For $d^2 \to \infty$ the ideal ROC is attained or $P_D = 1$ for any P_{FA} (see also Problem 3.12). For $d^2 \to 0$, the 45° lower bound is attained. Other properties of the ROC are discussed in Problem 3.13.

3.5 Irrelevant Data

In many signal detection problems one must decide which data are relevant to the detection problem and which may be discarded. As an example, for a DC level in WGN assume that we observe some extra or reference noise samples $w_R[n]$ for $n = 0, 1, \ldots, N-1$. This could be the output of a second sensor, which is incapable of passing the DC signal. Hence, the observed data set is $\{x[0], x[1], \ldots, x[N-1], w_R[0], w_R[1], \ldots, w_R[N-1]\}$ or in vector form $[\mathbf{x}^T\ \mathbf{w}_R^T]^T$. It might at first appear that $\mathbf{w}_R$ is irrelevant to the detection problem, but that could be a hasty conclusion. If, for example, $x[n] = w[n]$ under $\mathcal{H}_0$, $x[n] = A + w[n]$ for $A > 0$ under $\mathcal{H}_1$, and $w_R[n] = w[n]$ under either hypothesis, then the reference noise samples $w_R[n]$ could be used to *cancel* the corrupting noise $w[n]$. In particular, a detector that decides

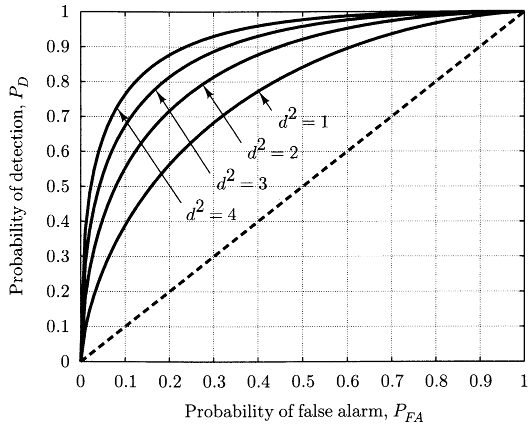


Figure 3.9. Family of receiver operating characteristics for DC level in WGN.

$\mathcal{H}_1$ if

$$T = x[0] - w_R[0] > \frac{A}{2}$$

would yield perfect detection. Under $\mathcal{H}_0$, $T = 0$, while under $\mathcal{H}_1$, $T = A$. Of course, this is an extreme case of perfect statistical dependence. At the other extreme, if $\mathbf{w}_R$ is independent of $\mathbf{x}$ under either hypothesis, then $\mathbf{w}_R$ is irrelevant to the problem. An example of this condition is encountered in the following problem. We observe $\{x[0], x[1], \ldots, x[N-1], x[N], \ldots, x[2N-1]\}$ or $\mathbf{x} = [\mathbf{x}_1^T\,\mathbf{x}_2^T]^T$ where $\mathbf{x}_1$ denotes the first N samples and $\mathbf{x}_2$ the remaining ones. Then, consider the problem

$$\begin{aligned}
\mathcal{H}_0 : x[n] &= w[n] \qquad n = 0, 1, \ldots, 2N-1 \\
\mathcal{H}_1 : x[n] &= \begin{cases} A + w[n] & n = 0, 1, \ldots, N-1 \\ w[n] & n = N, N+1, \ldots, 2N-1 \end{cases}
\end{aligned}$$

where $w[n]$ is WGN with variance σ^2. The noise samples outside the signal interval $[0, N-1]$ are irrelevant and can be discarded since they are independent of the data samples within the interval. This may also be verified by examining the NP test that decides $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x}_1, \mathbf{x}_2; \mathcal{H}_1)}{p(\mathbf{x}_1, \mathbf{x}_2; \mathcal{H}_0)} > \gamma$$

which becomes

$$\frac{\displaystyle\prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x[n] - A)^2\right] \prod_{n=N}^{2N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}x^2[n]\right]}{\displaystyle\prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}x[n]^2\right] \prod_{n=N}^{2N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}x^2[n]\right]} > \gamma$$

or finally

$$\frac{p(\mathbf{x}_1; \mathcal{H}_1)}{p(\mathbf{x}_1; \mathcal{H}_0)} > \gamma$$

so that $\mathbf{x}_2$ is irrelevant to the detection problem. Thus, in practice, *for detection of signals in WGN we can limit the observation interval to the signal interval.* If, however, the noise is correlated, then for best performance we should also include noise samples from outside the signal interval in our detector.

The preceding discussion can be generalized using the NP theorem. The likelihood ratio is

$$\begin{aligned}
L(\mathbf{x}_1, \mathbf{x}_2) &= \frac{p(\mathbf{x}_1, \mathbf{x}_2; \mathcal{H}_1)}{p(\mathbf{x}_1, \mathbf{x}_2; \mathcal{H}_0)} \\
&= \frac{p(\mathbf{x}_2 | \mathbf{x}_1; \mathcal{H}_1) p(\mathbf{x}_1; \mathcal{H}_1)}{p(\mathbf{x}_2 | \mathbf{x}_1; \mathcal{H}_0) p(\mathbf{x}_1; \mathcal{H}_0)}.
\end{aligned}$$

It follows that if

$$p(\mathbf{x}_2 | \mathbf{x}_1; \mathcal{H}_1) = p(\mathbf{x}_2 | \mathbf{x}_1; \mathcal{H}_0) \tag{3.11}$$

then $L(\mathbf{x}_1, \mathbf{x}_2) = L(\mathbf{x}_1)$ and $\mathbf{x}_2$ is irrelevant to the detection problem. A special case occurs when $\mathbf{x}_1$ and $\mathbf{x}_2$ are independent under either hypothesis and the PDF of $\mathbf{x}_2$ does not depend on the hypothesis. Then, (3.11) holds since $p(\mathbf{x}_2; \mathcal{H}_1) = p(\mathbf{x}_2; \mathcal{H}_0)$. The DC level in WGN with extra noise samples is an example. See also Problems 3.14 and 3.15.

3.6 Minimum Probability of Error

In some detection problems one can reasonably assign probabilities to the various hypotheses. In doing so, we express a prior belief in the likelihood of the hypotheses. An example is in digital communications in which the transmission of a "0" or "1" is equally likely. Then, it is reasonable to assign equal probabilities to $\mathcal{H}_0$ ("0" sent) and $\mathcal{H}_1$ ("1" sent). We say that $P(\mathcal{H}_0) = P(\mathcal{H}_1) = 1/2$, where $P(\mathcal{H}_0)$, $P(\mathcal{H}_1)$ are the *prior probabilities* of the respective hypotheses. In other applications, such as sonar or radar, this is not possible. If one is attempting to detect an enemy submarine, then the likelihood of its appearance can usually not be determined. This type of approach, where we assign prior probabilities, is the Bayesian approach to hypothesis testing. It is completely analogous to the Bayesian philosophy of estimation theory in which a prior PDF is assigned to an unknown parameter.

With the Bayesian paradigm we can define a *probability of error* P_e as

$$\begin{aligned} P_e &= \Pr\{\text{decide } \mathcal{H}_0, \mathcal{H}_1 \text{ true}\} + \Pr\{\text{decide } \mathcal{H}_1, \mathcal{H}_0 \text{ true}\} \\ &= P(\mathcal{H}_0|\mathcal{H}_1)P(\mathcal{H}_1) + P(\mathcal{H}_1|\mathcal{H}_0)P(\mathcal{H}_0) \end{aligned} \tag{3.12}$$

where $P(\mathcal{H}_i|\mathcal{H}_j)$ is the *conditional* probability that indicates the probability of deciding $\mathcal{H}_i$ when $\mathcal{H}_j$ is true. Note the slight distinction between $P(\mathcal{H}_i;\mathcal{H}_j)$ of the NP approach and $P(\mathcal{H}_i|\mathcal{H}_j)$ of the Bayesian approach. The former is the probability of deciding $\mathcal{H}_i$ if $\mathcal{H}_j$ *is true* with no probabilistic meaning assigned to the likelihood that $\mathcal{H}_j$ is true. The latter assumes that the outcome of a probabilistic experiment is observed to be $\mathcal{H}_j$ and that the probability of deciding $\mathcal{H}_i$ is *conditioned* on that outcome. Using the P_e criterion, the two errors are weighted appropriately to yield an overall error measure. Our goal will be to design a detector that minimizes P_e.

The derivation for the minimum P_e detector as a special case of the more general Bayesian detector is given in Appendix 3B. It is shown there that we should decide $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x}|\mathcal{H}_1)}{p(\mathbf{x}|\mathcal{H}_0)} > \frac{P(\mathcal{H}_0)}{P(\mathcal{H}_1)} = \gamma. \tag{3.13}$$

Similar to the NP test we compare the *conditional* likelihood ratio to a threshold. Here, however, the threshold is determined by the prior probabilities. If, as is commonly the case, the prior probabilities are equal, we decide $\mathcal{H}_1$ if

$$p(\mathbf{x}|\mathcal{H}_1) > p(\mathbf{x}|\mathcal{H}_0). \tag{3.14}$$

Equivalently, we choose the hypothesis with the larger conditional likelihood or the one that maximizes $p(\mathbf{x}|\mathcal{H}_i)$ for $i = 0, 1$. This is called the *maximum likelihood* (ML) detector. (Actually, we should term this the maximum *conditional* likelihood. We defer to common usage in not doing so.) An example follows.

Example 3.5 - DC Level in WGN - Minimum P_e Criterion

We have the detection problem

$$\begin{aligned} \mathcal{H}_0 &: x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 &: x[n] = A + w[n] & n = 0, 1, \ldots, N-1 \end{aligned}$$

where $A > 0$ and $w[n]$ is WGN with variance σ^2. If this is a digital communication problem where we transmit either $s_0[n] = 0$ or $s_1[n] = A$ (called an on-off keyed (OOK) communication system), it is reasonable to assume $P(\mathcal{H}_0) = P(\mathcal{H}_1) = 1/2$. The receiver that minimizes P_e is given by (3.13) with $\gamma = 1$. Hence, we decide $\mathcal{H}_1$

if

$$\frac{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right]}{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right]} > 1.$$

Taking logarithms yields

$$-\frac{1}{2\sigma^2}\left(-2A\sum_{n=0}^{N-1}x[n] + NA^2\right) > 0$$

or we decide $\mathcal{H}_1$ if $\bar{x} > A/2$. This is the same form of the detector as for the NP criterion except for the threshold (and, of course, the performance). To determine P_e we use (3.12) and note that

$$\bar{x} \sim \begin{cases} \mathcal{N}(0, \frac{\sigma^2}{N}) & \text{conditioned on } \mathcal{H}_0 \\ \mathcal{N}(A, \frac{\sigma^2}{N}) & \text{conditioned on } \mathcal{H}_1. \end{cases}$$

Thus

$$\begin{aligned} P_e &= \frac{1}{2}\left[P(\mathcal{H}_0|\mathcal{H}_1) + P(\mathcal{H}_1|\mathcal{H}_0)\right] \\ &= \frac{1}{2}\left[\Pr\{\bar{x} < A/2|\mathcal{H}_1\} + \Pr\{\bar{x} > A/2|\mathcal{H}_0\}\right] \\ &= \frac{1}{2}\left[\left(1 - Q\left(\frac{A/2 - A}{\sqrt{\sigma^2/N}}\right)\right) + Q\left(\frac{A/2}{\sqrt{\sigma^2/N}}\right)\right] \end{aligned}$$

and since $Q(-x) = 1 - Q(x)$, we have finally

$$P_e = Q\left(\sqrt{\frac{NA^2}{4\sigma^2}}\right). \tag{3.15}$$

The probability of error decreases monotonically with NA^2/σ^2, which is, of course, the deflection coefficient.

◇

Another form of the minimum P_e detector follows directly from (3.13). We decide $\mathcal{H}_1$ if

$$p(\mathbf{x}|\mathcal{H}_1)P(\mathcal{H}_1) > p(\mathbf{x}|\mathcal{H}_0)P(\mathcal{H}_0).$$

But from Bayes rule we have that

$$P(\mathcal{H}_i|\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{H}_i)P(\mathcal{H}_i)}{p(\mathbf{x})}$$

where the denominator $p(\mathbf{x})$ does not depend on the true hypothesis. In fact, $p(\mathbf{x})$ is just a normalizing factor that can be written as

$$p(\mathbf{x}) = p(\mathbf{x}|\mathcal{H}_0)P(\mathcal{H}_0) + p(\mathbf{x}|\mathcal{H}_1)P(\mathcal{H}_1).$$

As a result we decide $\mathcal{H}_1$ if

$$P(\mathcal{H}_1|\mathbf{x}) > P(\mathcal{H}_0|\mathbf{x}) \tag{3.16}$$

or we choose the hypothesis whose a posteriori (after the data are observed) probability is maximum. This detector, which minimizes P_e for any prior probability, is termed the *maximum a posteriori probability* (MAP) detector. Of course, for equal prior probabilities, the MAP detector reduces to the ML detector. The decision regions for the DC level in WGN with $N = 1$, $A = 1$, $\sigma^2 = 1$ are shown in Figure 3.10 for various prior probabilities (see Problem 3.16).

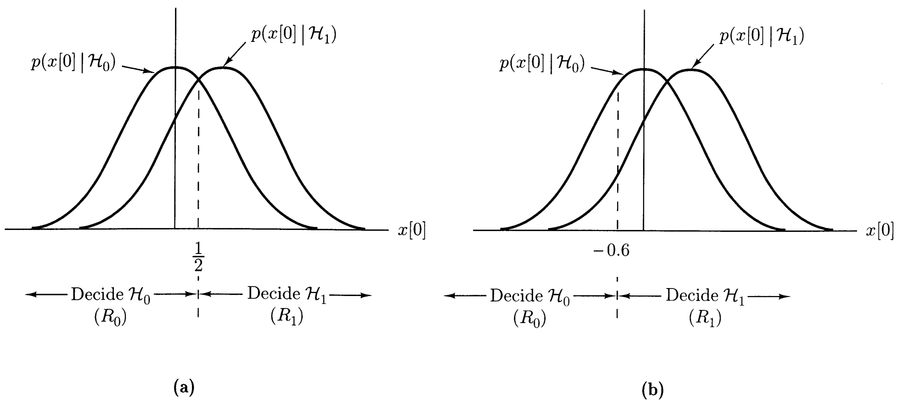


Figure 3.10. Effect of prior probability on decision regions (a) MAP detector with $P(\mathcal{H}_0) = P(\mathcal{H}_1) = 1/2$ (b) MAP detector with $P(\mathcal{H}_0) = 1/4$, $P(\mathcal{H}_1) = 3/4$.

3.7 Bayes Risk

A generalization of the minimum P_e criterion assigns costs to each type of error. Suppose that we wish to design a system to automatically inspect a machine part. The result of the inspection is either to use the part in a product if it is deemed satisfactory or else to discard it. We could set up the hypothesis test

$$\begin{aligned}\mathcal{H}_0 &: \text{part is defective}\\ \mathcal{H}_1 &: \text{part is satisfactory}\end{aligned}$$

and assign costs to the errors. Let C_{ij} be the cost if we decide $\mathcal{H}_i$ but $\mathcal{H}_j$ is true. For example, we would probably want $C_{10} > C_{01}$. If we decide the part is satisfactory but it proves to be defective, the entire product may be defective and we incur a large cost (C_{10}). If, however, we decide that the part is defective when it is not, we incur the smaller cost of the part only (C_{01}). Once costs have been assigned, the decision rule is based on minimizing the expected cost or *Bayes risk* $\mathcal{R}$ defined as

$$\mathcal{R} = E(C) = \sum_{i=0}^{1}\sum_{j=0}^{1} C_{ij}P(\mathcal{H}_i|\mathcal{H}_j)P(\mathcal{H}_j). \tag{3.17}$$

Usually, if no error is made, we do not assign a cost so that $C_{00} = C_{11} = 0$. However, for convenience we will retain the more general form. Also, note that if $C_{00} = C_{11} = 0$, $C_{10} = C_{01} = 1$, then $\mathcal{R} = P_e$.

Under the reasonable assumption that $C_{10} > C_{00}$, $C_{01} > C_{11}$, the detector that minimizes the Bayes risk is to decide $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x}|\mathcal{H}_1)}{p(\mathbf{x}|\mathcal{H}_0)} > \frac{(C_{10} - C_{00})P(\mathcal{H}_0)}{(C_{01} - C_{11})P(\mathcal{H}_1)} = \gamma. \tag{3.18}$$

See Appendix 3B for the proof. Once again, the conditional likelihood ratio is compared to a threshold.

3.8 Multiple Hypothesis Testing

We now consider the case where we wish to distinguish between M hypotheses, where $M > 2$. Such a problem arises quite frequently in communications, in which one of M signals must be detected. Also, pattern recognition systems make extensive use of the results in distinguishing between different patterns. In addition to signal detection, this problem also goes by the name of *classification* or *discrimination*. Although an NP criterion can be formulated for the M-ary hypothesis test, it seems to seldom be used in practice. The interested reader should consult [Lehmann 1959] for further details. More commonly the minimum P_e criterion or its generalization, the Bayes risk, is employed. We now consider the latter.

Assume that we now wish to decide among the M possible hypotheses $\{\mathcal{H}_0, \mathcal{H}_1, \ldots, \mathcal{H}_{M-1}\}$. The cost assigned to the decision to choose $\mathcal{H}_i$ when $\mathcal{H}_j$ is true is denoted by C_{ij}. The expected cost or Bayes risk becomes

$$\mathcal{R} = \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} C_{ij}P(\mathcal{H}_i|\mathcal{H}_j)P(\mathcal{H}_j). \tag{3.19}$$

For the particular assignment

$$C_{ij} = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases} \tag{3.20}$$

we have that $\mathcal{R} = P_e$. The decision rule that minimizes $\mathcal{R}$ is derived in Appendix 3C. There it is shown that we should choose the hypothesis that *minimizes*

$$C_i(\mathbf{x}) = \sum_{j=0}^{M-1} C_{ij} P(\mathcal{H}_j|\mathbf{x}) \tag{3.21}$$

over $i = 0, 1, \ldots, M-1$. To determine the decision rule that minimizes P_e we use (3.20). Then

$$\begin{aligned} C_i(\mathbf{x}) &= \sum_{\substack{j=0 \\ j \neq i}}^{M-1} P(\mathcal{H}_j|\mathbf{x}) \\ &= \sum_{j=0}^{M-1} P(\mathcal{H}_j|\mathbf{x}) - P(\mathcal{H}_i|\mathbf{x}). \end{aligned}$$

Since the first term is independent of i, $C_i(\mathbf{x})$ is minimized by *maximizing* $P(\mathcal{H}_i|\mathbf{x})$. Thus, the minimum P_e decision rule is to decide $\mathcal{H}_k$ if

$$P(\mathcal{H}_k|\mathbf{x}) > P(\mathcal{H}_i|\mathbf{x}) \qquad i \neq k. \tag{3.22}$$

As in the binary case we seek to maximize the a posteriori probability. This is the M-ary maximum a posteriori probability (MAP) decision rule. If, however, the prior probabilities are equal, then

$$\begin{aligned} P(\mathcal{H}_i|\mathbf{x}) &= \frac{p(\mathbf{x}|\mathcal{H}_i)P(\mathcal{H}_i)}{p(\mathbf{x})} \\ &= \frac{p(\mathbf{x}|\mathcal{H}_i)\frac{1}{M}}{p(\mathbf{x})} \end{aligned} \tag{3.23}$$

and to maximize $P(\mathcal{H}_i|\mathbf{x})$ we need only maximize $p(\mathbf{x}|\mathcal{H}_i)$. Hence, for equal prior probabilities we decide $\mathcal{H}_k$ if

$$p(\mathbf{x}|\mathcal{H}_k) > p(\mathbf{x}|\mathcal{H}_i) \qquad i \neq k. \tag{3.24}$$

This is the M-ary maximum likelihood (ML) decision rule.

Finally, observe from (3.23) that to maximize $P(\mathcal{H}_i|\mathbf{x})$ we can equivalently maximize $p(\mathbf{x}|\mathcal{H}_i)P(\mathcal{H}_i)$ since $p(\mathbf{x})$ does not depend on i. Equivalently then, the MAP rule maximizes

$$\ln p(\mathbf{x}|\mathcal{H}_i) + \ln P(\mathcal{H}_i).$$

An example follows.

Example 3.6 - Multiple DC Levels in WGN

Assume that we have the three hypotheses

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = -A + w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_2 : x[n] = A + w[n] & n = 0, 1, \ldots, N-1 \end{array}$$

where $A > 0$ and $w[n]$ is WGN with variance σ^2. Furthermore, if the prior probabilities are equal or $P(\mathcal{H}_0) = P(\mathcal{H}_1) = P(\mathcal{H}_2) = 1/3$, then the ML decision rule applies. Consider first the simple case of $N = 1$. We then have the PDFs shown in Figure 3.11. By symmetry it is clear from (3.24) that to minimize P_e we should decide $\mathcal{H}_0$ if $x[0] < -A/2$, $\mathcal{H}_1$ if $-A/2 < x[0] < A/2$, and $\mathcal{H}_2$ if $x[0] > A/2$. For multiple samples ($N > 1$) we cannot just plot the multivariate PDFs and observe the regions over which each one yields the maximum. Instead we need to derive a test statistic. The conditional PDF is

$$p(\mathbf{x}|\mathcal{H}_i) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n] - A_i)^2\right]$$

where $A_0 = -A$, $A_1 = 0$, $A_2 = A$. To maximize $p(\mathbf{x}|\mathcal{H}_i)$ we can equivalently minimize

$$D_i^2 = \sum_{n=0}^{N-1}(x[n] - A_i)^2.$$

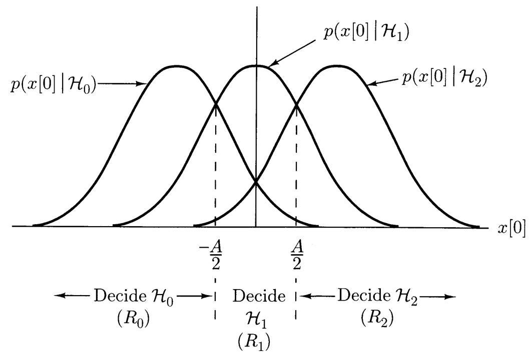


Figure 3.11. Decision regions for multiple DC levels in WGN ($N = 1$).

Using a slight manipulation we express D_i^2 as

$$\begin{aligned} D_i^2 &= \sum_{n=0}^{N-1}(x[n]-\bar{x}+\bar{x}-A_i)^2 \\ &= \sum_{n=0}^{N-1}(x[n]-\bar{x})^2+2(\bar{x}-A_i)\sum_{n=0}^{N-1}(x[n]-\bar{x})+N(\bar{x}-A_i)^2 \\ &= \sum_{n=0}^{N-1}(x[n]-\bar{x})^2+N(\bar{x}-A_i)^2. \end{aligned}$$

It is now apparent that to minimize D_i^2 we choose $\mathcal{H}_i$ for which A_i is closest to $\bar{x}$. Hence, we decide

$$\begin{array}{ll} \mathcal{H}_0 & \text{if } \bar{x} < -A/2 \\ \mathcal{H}_1 & \text{if } -A/2 < \bar{x} < A/2 \\ \mathcal{H}_2 & \text{if } \bar{x} > A/2. \end{array}$$

This detector is a special case of the minimum distance receiver discussed in Chapter 4. Also, it is an example of an M-ary pulse amplitude modulation (PAM) communication signaling scheme for $M = 3$. The reader may also wish to refer to Problem 3.19 in which the use of sufficient statistics is shown to greatly simplify this problem.

Next we determine the minimum P_e. Note that in the binary case there were only two types of errors. Here, we have six types and in general there are $M^2 - M = M(M-1)$ error types. It is therefore simpler to determine $1 - P_e = P_c$, where P_c is the probability of a *correct* decision. Thus

$$\begin{aligned} P_c &= \sum_{i=0}^{2} P(\mathcal{H}_i|\mathcal{H}_i)P(\mathcal{H}_i) \\ &= \frac{1}{3}\sum_{i=0}^{2} P(\mathcal{H}_i|\mathcal{H}_i) \\ &= \frac{1}{3}\left[\Pr\{\bar{x} < -A/2|\mathcal{H}_0\} + \Pr\{-A/2 < \bar{x} < A/2|\mathcal{H}_1\} + \Pr\{\bar{x} > A/2|\mathcal{H}_2\}\right]. \end{aligned}$$

Since $\bar{x} \sim \mathcal{N}(A_i, \sigma^2/N)$ (conditioned on $\mathcal{H}_i$), we have

$$\begin{aligned} P_c &= \frac{1}{3}\left[1 - Q\left(\frac{-\frac{A}{2}+A}{\sqrt{\sigma^2/N}}\right) + Q\left(\frac{-\frac{A}{2}}{\sqrt{\sigma^2/N}}\right) - Q\left(\frac{\frac{A}{2}}{\sqrt{\sigma^2/N}}\right) + Q\left(\frac{\frac{A}{2}-A}{\sqrt{\sigma^2/N}}\right)\right] \\ &= 1 - \frac{4}{3}Q\left(\sqrt{\frac{NA^2}{4\sigma^2}}\right) \end{aligned}$$

so that

$$P_e = \frac{4}{3}Q\left(\sqrt{\frac{NA^2}{4\sigma^2}}\right). \tag{3.25}$$

Note that P_e has increased over the binary case (see (3.15)) since the detector must decide among more hypotheses. The general M-ary case is discussed in Problem 3.20.

References

Kendall, Sir M., A. Stuart, *The Advanced Theory of Statistics*, Vol. 2, Macmillan, New York, 1979.

Lehmann, E.L., *Testing Statistical Hypotheses*, J. Wiley, New York, 1959.

Problems

3.1 Determine the NP test for distinguishing between the hypotheses $\mathcal{H}_0 : \mu = 0$ versus $\mathcal{H}_1 : \mu = 1$ based on the observed sample $x[0] \sim \mathcal{N}(\mu, 1)$. Then, find the Type I error (P_{FA}) and the Type II error ($P_M = 1 - P_D$, where P_M is the probability of a *miss*). Finally, plot P_M versus P_{FA}.

3.2 For the introductory example of Section 3.3 find two other critical regions to satisfy $P_{FA} = 10^{-3}$.

3.3 Show that if $g(x)$ is a monotonically increasing function of x, then $L(\mathbf{x}) > \gamma$ if and only if $g(L(\mathbf{x})) > g(\gamma)$. By monotonically increasing, we mean that the function $g(x)$ satisfies $g(x_2) > g(x_1)$ if and only if $x_2 > x_1$.

3.4 For the DC level in WGN detection problem assume that we wish to have $P_{FA} = 10^{-4}$ and $P_D = 0.99$. If the SNR is $10\log_{10} A^2/\sigma^2 = -30$ dB, determine the necessary number of samples N.

3.5 For the DC level in WGN detection problem consider the detector that decides $\mathcal{H}_1$ if $\bar{x} > \gamma'$. Since $\bar{x}$ is an estimator of A or $\hat{A} = \bar{x}$, a measure of the estimation accuracy is $E^2(\hat{A})/\text{var}(\hat{A})$. Relate this quantity to the signal ENR.

3.6 Modify Example 3.2 so that now $A < 0$. Determine the NP detector and its detection performance to show that it is the same as for $A > 0$. Hint: Use $Q^{-1}(x) = -Q^{-1}(1-x)$.

3.7 We observe the IID samples $x[n]$ for $n = 0, 1, \ldots, N-1$ from the Rayleigh PDF

$$p(x[n]) = \frac{x[n]}{\sigma^2}\exp\left(-\frac{1}{2}\frac{x^2[n]}{\sigma^2}\right).$$

Derive the NP test for the hypothesis testing problem

$$\begin{array}{l}\mathcal{H}_0 : \sigma^2 = \sigma_0^2 \\ \mathcal{H}_1 : \sigma^2 = \sigma_1^2 > \sigma_0^2.\end{array}$$

3.8 Find the NP test to distinguish between the hypotheses that a single sample $x[0]$ is observed from the possible PDFs

$$\begin{array}{l}\mathcal{H}_0 : p(x[0]) = \frac{1}{2}\exp\left(-|x[0]|\right) \\ \mathcal{H}_1 : p(x[0]) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x^2[0]\right).\end{array}$$

Show the decision regions. Hint: You will need to solve a quadratic inequality.

3.9 For Example 3.3 let $N = 2$ and show that the detection performance is summarized by

$$P_D = (P_{FA})^{\frac{\sigma_0^2}{\sigma_1^2}}.$$

Note that P_D is monotonically increasing with σ_1^2/σ_0^2.

3.10 For the DC level in WGN detection problem discussed in Example 3.2 show that $\bar{x}$ is a sufficient statistic for A. Use the identity

$$\sum_{n=0}^{N-1}(x[n] - A)^2 = \sum_{n=0}^{N-1} x^2[n] - 2AN\bar{x} + NA^2$$

to effect the factorization.

3.11 The exponential family of PDFs is defined by

$$p(x;\theta) = \exp\left[A(\theta)B(x) + C(x) + D(\theta)\right]$$

where θ is a parameter. Show that the Gaussian PDF with parameter μ or

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

is a special case of this family. Next for the exponential family show that a sufficient statistic for θ is $T(\mathbf{x}) = \sum_{n=0}^{N-1} B(x[n])$ based on the observed data $x[n]$ for $n = 0, 1, \ldots, N-1$. Using this result, find the sufficient statistic for μ in the Gaussian PDF. The reader may also wish to refer to [Kay-I 1993, Problems 5.14, 5.15].

3.12 Design a *perfect detector* for the problem

$$\begin{array}{l}\mathcal{H}_0 : x[0] \sim \mathcal{U}[-c, c] \\ \mathcal{H}_1 : x[0] \sim \mathcal{U}[1-c, 1+c]\end{array}$$

where $c > 0$ and $\mathcal{U}[a,b]$ denotes a uniform PDF on the interval $[a,b]$, by choosing c. A perfect detector has $P_{FA} = 0$ and $P_D = 1$.

3.13 Prove that the ROC is a concave function over the interval [0,1]. A concave function is one for which

$$\alpha g(x_1) + (1-\alpha)g(x_2) \le g(\alpha x_1 + (1-\alpha)x_2)$$

for $0 \le \alpha \le 1$ and any two points x_1 and x_2. To do so consider two points on the ROC $(p_1, P_D(p_1))$ and $(p_2, P_D(p_2))$ and find P_D for a *randomized test*. A randomized test first flips a coin with $\Pr\{\text{head}\} = \alpha$. If the outcome is a head, we employ the detector whose performance is $(p_1, P_D(p_1))$. Otherwise, we employ the detector whose performance is $(p_2, P_D(p_2))$. We decide $\mathcal{H}_1$ if the chosen detector decides $\mathcal{H}_1$. Hint: For a given P_{FA} the detection performance of the randomized detector must be less than or equal to that of the NP detector.

3.14 Consider the hypothesis testing problem

$$\mathbf{x} \sim \begin{cases} \mathcal{N}\left(\begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\right) & \text{under } \mathcal{H}_0 \\ \mathcal{N}\left(\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\right) & \text{under } \mathcal{H}_1\end{cases}$$

where $\mathbf{x} = [x[0]\, x[1]]^T$ is observed. Find the NP test statistic (do not evaluate the threshold) and explain what happens if $\rho = 0$.

3.15 Consider the detection of a signal $s[n]$ embedded in WGN with variance σ^2 based on the observed samples $x[n]$ for $n = 0, 1, \ldots, 2N-1$. The signal is given by

$$s[n] = \begin{cases} A & n = 0, 1, \ldots, N-1 \\ 0 & n = N, N+1, \ldots, 2N-1\end{cases}$$

under $\mathcal{H}_0$ and by

$$s[n] = \begin{cases} A & n = 0, 1, \ldots, N-1 \\ 2A & n = N, N+1, \ldots, 2N-1\end{cases}$$

under $\mathcal{H}_1$. Assume that $A > 0$ and find the NP detector as well as its detection performance. Explain the operation of the detector.

3.16 In Example 3.5 find the optimal detector if the prior probability $P(\mathcal{H}_1)$ is arbitrary. For $N = 1$, $A = 1$, and $\sigma^2 = 1$, find the detector, including the threshold, for the prior probabilities $P(\mathcal{H}_0) = P(\mathcal{H}_1) = 1/2$ and also for $P(\mathcal{H}_0) = 1/4$, $P(\mathcal{H}_1) = 3/4$. Explain your results.

3.17 Assume that we wish to distinguish between the hypotheses $\mathcal{H}_0 : \mathbf{x} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ and $\mathcal{H}_1 : \mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2\mathbf{I})$ based on $\mathbf{x} = [x[0]\, x[1]]^T$. If $P(\mathcal{H}_0) = P(\mathcal{H}_1)$, find the decision regions that minimize P_e. Hint: Show that the decision region boundary is a line that is the perpendicular bisector of the line segment from $\mathbf{0}$ to $\boldsymbol{\mu}$.

3.18 Find the MAP decision rule for

$$\begin{array}{l}\mathcal{H}_0 : x[0] \sim \mathcal{N}(0,1) \\ \mathcal{H}_1 : x[0] \sim \mathcal{N}(0,2)\end{array}$$

if $P(\mathcal{H}_0) = 1/2$ and also if $P(\mathcal{H}_0) = 3/4$. Display the decision regions in each case and explain.

3.19 For the simple hypothesis testing problem of this chapter one can show that an NP test based on the *sufficient statistic* is equivalent to an NP test based on the *original data* [Kendall and Stuart 1979]. In this problem we show how this result simplifies the derivation of a detector. Consider Example 3.6 and recall the result of Problem 3.10 that the sufficient statistic for the DC level is just the sample mean $\bar{x}$. Find the ML detector based on the observed value of the sample mean.

3.20 A general M-ary PAM communication system transmits one of M DC levels. Let the levels be $\{0, \pm A, \pm 2A, \ldots, \pm(M-1)A/2\}$ for M odd. The received data $x[n]$ for $n = 0, 1, \ldots, N-1$ will be one of the DC levels embedded in WGN with variance σ^2. Using the concept of sufficient statistics from Problem 3.19 find the ML detector. Then, show that the minimum P_e is

$$P_e = \frac{2M-2}{M} Q\left(\sqrt{\frac{NA^2}{4\sigma^2}}\right)$$

for equal prior probabilities.

3.21 Design a minimum P_e detector to decide among the hypotheses whose PDFs are

$$\begin{aligned} p(x[0]|\mathcal{H}_0) &= \frac{1}{2}\exp\left(-|x[0]+1|\right) \\ p(x[0]|\mathcal{H}_1) &= \frac{1}{2}\exp\left(-|x[0]|\right) \\ p(x[0]|\mathcal{H}_2) &= \frac{1}{2}\exp\left(-|x[0]-1|\right) \end{aligned}$$

assuming equal prior probabilities. Also, find the minimum P_e.

Appendix 3A

Neyman-Pearson Theorem

We use Lagrangian multipliers to maximize P_D for a given P_{FA}. Forming the Lagrangian

$$\begin{aligned} F &= P_D + \lambda(P_{FA} - \alpha) \\ &= \int_{R_1} p(\mathbf{x}; \mathcal{H}_1) d\mathbf{x} + \lambda\left(\int_{R_1} p(\mathbf{x}; \mathcal{H}_0) d\mathbf{x} - \alpha\right) \\ &= \int_{R_1} \left(p(\mathbf{x}; \mathcal{H}_1) + \lambda p(\mathbf{x}; \mathcal{H}_0)\right) d\mathbf{x} - \lambda\alpha. \end{aligned}$$

To maximize F we should include $\mathbf{x}$ in R_1 if the integrand is positive for that value of $\mathbf{x}$ or if

$$p(\mathbf{x}; \mathcal{H}_1) + \lambda p(\mathbf{x}; \mathcal{H}_0) > 0. \tag{3A.1}$$

When $p(\mathbf{x}; \mathcal{H}_1) + \lambda p(\mathbf{x}; \mathcal{H}_0) = 0$, $\mathbf{x}$ may be included in either R_0 or R_1. Since the probability of this occurrence is zero (assuming that the PDFs are continuous), we need not concern ourselves with this case. Hence, the $>$ sign in (3A.1) and the subsequent results can be replaced with $\geq$ if desired. We choose to retain the $>$ sign in our development. We thus decide $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x}; \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > -\lambda.$$

The Lagrangian multiplier is found from the constraint and must satisfy $\lambda < 0$. Otherwise, we decide $\mathcal{H}_1$ if the likelihood ratio $p(\mathbf{x}; \mathcal{H}_1)/p(\mathbf{x}; \mathcal{H}_0)$ exceeds a negative number. Since the likelihood ratio is always nonnegative, we would always decide $\mathcal{H}_1$, irrespective of the hypothesis, resulting in $P_{FA} = 1$. We let $\gamma = -\lambda$ so that finally we decide $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x}; \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > \gamma$$

where the threshold $\gamma > 0$ is found from $P_{FA} = \alpha$.

Appendix 3B

Minimum Bayes Risk Detector - Binary Hypothesis

We minimize $\mathcal{R}$ as given by (3.17). Note that if $C_{00} = C_{11} = 0$, $C_{01} = C_{10} = 1$, then $\mathcal{R} = P_e$ and so our derivation also applies to the minimum P_e problem. Now

$$\mathcal{R} = \sum_{i=0}^{1}\sum_{j=0}^{1} C_{ij} P(\mathcal{H}_i|\mathcal{H}_j)P(\mathcal{H}_j).$$

Let $R_1 = \{\mathbf{x} : \text{decide } \mathcal{H}_1\}$ be the critical region and R_0 denote its complement (decide $\mathcal{H}_0$). Then,

$$\begin{aligned}\mathcal{R} &= C_{00}P(\mathcal{H}_0)\int_{R_0} p(\mathbf{x}|\mathcal{H}_0)d\mathbf{x} + C_{01}P(\mathcal{H}_1)\int_{R_0} p(\mathbf{x}|\mathcal{H}_1)d\mathbf{x} \\ &\quad + C_{10}P(\mathcal{H}_0)\int_{R_1} p(\mathbf{x}|\mathcal{H}_0)d\mathbf{x} + C_{11}P(\mathcal{H}_1)\int_{R_1} p(\mathbf{x}|\mathcal{H}_1)d\mathbf{x}.\end{aligned}$$

But

$$\int_{R_0} p(\mathbf{x}|\mathcal{H}_i)d\mathbf{x} = 1 - \int_{R_1} p(\mathbf{x}|\mathcal{H}_i)d\mathbf{x}$$

since R_1 and R_0 partition the entire space. Using this we have

$$\begin{aligned}\mathcal{R} &= C_{01}P(\mathcal{H}_1) + C_{00}P(\mathcal{H}_0) \\ &\quad + \int_{R_1} \left[(C_{10}P(\mathcal{H}_0) - C_{00}P(\mathcal{H}_0))p(\mathbf{x}|\mathcal{H}_0)\right. \\ &\qquad \left. + (C_{11}P(\mathcal{H}_1) - C_{01}P(\mathcal{H}_1))p(\mathbf{x}|\mathcal{H}_1)\right] d\mathbf{x}.\end{aligned}$$

We include $\mathbf{x}$ in R_1 only if the integrand is negative or we decide $\mathcal{H}_1$ if

$$(C_{10} - C_{00})P(\mathcal{H}_0)p(\mathbf{x}|\mathcal{H}_0) < (C_{01} - C_{11})P(\mathcal{H}_1)p(\mathbf{x}|\mathcal{H}_1).$$

Assuming $C_{10} > C_{00}$, $C_{01} > C_{11}$, we have finally

$$\frac{p(\mathbf{x}|\mathcal{H}_1)}{p(\mathbf{x}|\mathcal{H}_0)} > \frac{(C_{10} - C_{00})P(\mathcal{H}_0)}{(C_{01} - C_{11})P(\mathcal{H}_1)} = \gamma.$$

Appendix 3C

Minimum Bayes Risk Detector - Multiple Hypotheses

We use a slightly different approach than in Appendix 3B. From (3.19)

$$\mathcal{R} = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} C_{ij} P(\mathcal{H}_i|\mathcal{H}_j) P(\mathcal{H}_j).$$

Let $R_i = \{\mathbf{x} : \text{decide } \mathcal{H}_i\}$, where the R_i's for $i = 0, 1, \ldots, M-1$ partition the space, so that

$$\begin{aligned}
\mathcal{R} &= \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} C_{ij} \int_{R_i} p(\mathbf{x}|\mathcal{H}_j) P(\mathcal{H}_j) d\mathbf{x} \\
&= \sum_{i=0}^{M-1} \int_{R_i} \sum_{j=0}^{M-1} C_{ij} p(\mathbf{x}|\mathcal{H}_j) P(\mathcal{H}_j) d\mathbf{x} \\
&= \sum_{i=0}^{M-1} \int_{R_i} \sum_{j=0}^{M-1} C_{ij} P(\mathcal{H}_j|\mathbf{x}) p(\mathbf{x}) d\mathbf{x}.
\end{aligned}$$

Let $C_i(\mathbf{x}) = \sum_{j=0}^{M-1} C_{ij} P(\mathcal{H}_j|\mathbf{x})$ be the average cost of deciding $\mathcal{H}_i$ if $\mathbf{x}$ is observed. Then

$$\mathcal{R} = \sum_{i=0}^{M-1} \int_{R_i} C_i(\mathbf{x}) p(\mathbf{x}) d\mathbf{x}.$$

Now, each $\mathbf{x}$ must be assigned to *one and only one* of the R_i partitions. The contribution to $\mathcal{R}$ if we assign $\mathbf{x}$ to R_1, for example, is $C_1(\mathbf{x})p(\mathbf{x})d\mathbf{x}$. If, however,

we assign $\mathbf{x}$ to R_2, the cost contribution is $C_2(\mathbf{x})p(\mathbf{x})d\mathbf{x}$. Generalizing, we should assign $\mathbf{x}$ to R_k if

$$C_i(\mathbf{x}) = \sum_{j=0}^{M-1} C_{ij} P(\mathcal{H}_j|\mathbf{x})$$

is minimum for $i = k$. Hence, we decide $\mathcal{H}_i$ for which

$$\sum_{j=0}^{M-1} C_{ij} P(\mathcal{H}_j|\mathbf{x})$$

is minimum.

Chapter 4

Deterministic Signals

4.1 Introduction

The problem described in this chapter is the detection of a known signal in Gaussian noise. This is perhaps the simplest detection problem encountered because the resultant hypothesis test is a simple versus simple hypothesis. As discussed in Chapter 3, the optimal test is well known. If the probability of detection is to be maximized subject to a constant probability of false alarm, then the Neyman-Pearson criterion is employed, while to minimize the average cost, the Bayesian risk criterion is used. Furthermore, the resulting test statistic is a linear function of the data due to the Gaussian noise assumption, and therefore the performance of the detector is easily determined. The detector that evolves from these assumptions is termed the *matched filter*. It has found extensive use in applications in which the signal is under the designer's control so that the known signal assumption is valid. The salient example is in coherent communication systems.

4.2 Summary

The Neyman-Pearson detector for a known signal in WGN is the replica-correlator as given by (4.3). Alternatively, the matched filter implementation is given by (4.5) for use in (4.4). The detection performance is summarized in (4.14). When the Gaussian noise is not white, the Neyman-Pearson detector is given by (4.16) and its performance by (4.18). For the binary communication problem the optimal receiver is the minimum distance receiver given by (4.20). Its probability of error is determined by (4.25). In the M-ary communication case the optimal receiver is again a minimum distance receiver as summarized by (4.26). Its probability of error is found for orthogonal signals as (4.28). The general classical linear model is described in Section 4.6. For the case of known model parameters the Neyman-Pearson detector is given by (4.29) and is shown to be a special case of the detector for a known signal in colored noise. In Section 4.7 we apply the theory to multiple

orthogonal signaling for communications, which illustrates the concept of channel capacity, and to pattern recognition for images.

4.3 Matched Filters

4.3.1 Development of Detector

We begin our discussion of optimal detection approaches by considering the problem of detecting a *known deterministic* signal in white Gaussian noise (WGN). The Neyman-Pearson (NP) criterion will be used, but as discussed in Chapter 3, the resulting test statistic will be identical to that obtained using the Bayesian risk criterion. Only the threshold and detection performance will differ. The detection problem is to distinguish between the hypotheses

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = s[n] + w[n] & n = 0, 1, \ldots, N-1 \end{array} \tag{4.1}$$

where the signal $s[n]$ is assumed known and $w[n]$ is WGN with variance σ^2. WGN is defined as a zero mean Gaussian noise process with autocorrelation function (ACF)

$$\begin{aligned} r_{ww}[k] &= E(w[n]w[n+k]) \\ &= \sigma^2\delta[k] \end{aligned}$$

where $\delta[k]$ is the discrete-time delta function ($\delta[k] = 1$ if $k = 0$ and $\delta[k] = 0$ for $k \neq 0$). Such a model can be derived from a continuous-time setup as described in [Kay-I 1993, pg. 54].

The NP detector decides $\mathcal{H}_1$ if the likelihood ratio exceeds a threshold or

$$L(\mathbf{x}) = \frac{p(\mathbf{x}; \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > \gamma \tag{4.2}$$

where $\mathbf{x} = [x[0]\, x[1] \ldots x[N-1]]^T$. Since

$$\begin{aligned} p(\mathbf{x}; \mathcal{H}_1) &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-s[n])^2\right] \\ p(\mathbf{x}; \mathcal{H}_0) &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right] \end{aligned}$$

we have

$$L(\mathbf{x}) = \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{n=0}^{N-1}(x[n]-s[n])^2 - \sum_{n=0}^{N-1}x^2[n]\right)\right] > \gamma.$$

Taking the logarithm (a monotonically increasing transformation) of both sides does not change the inequality (see Section 3.3) so that

$$l(\mathbf{x}) = \ln L(\mathbf{x}) = -\frac{1}{2\sigma^2}\left(\sum_{n=0}^{N-1}(x[n]-s[n])^2 - \sum_{n=0}^{N-1}x^2[n]\right) > \ln\gamma.$$

We decide $\mathcal{H}_1$ if

$$\frac{1}{\sigma^2}\sum_{n=0}^{N-1}x[n]s[n] - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}s^2[n] > \ln\gamma.$$

Since $s[n]$ is known (and thus not a function of the data), we may incorporate the energy term into the threshold to yield

$$T(\mathbf{x}) = \sum_{n=0}^{N-1}x[n]s[n] > \sigma^2\ln\gamma + \frac{1}{2}\sum_{n=0}^{N-1}s^2[n].$$

Calling the right-hand-side of the inequality a new threshold γ', we decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \sum_{n=0}^{N-1}x[n]s[n] > \gamma'. \tag{4.3}$$

This is our NP detector and as expected consists of a test statistic $T(\mathbf{x})$ (a function of the data) and a threshold γ', which is chosen to satisfy $P_{FA} = \alpha$ for a given α. We now determine the test statistic for some simple examples.

Example 4.1 - DC Level in WGN

Assume that $s[n] = A$ for some known level A, where $A > 0$. Then from (4.3), $T(\mathbf{x}) = A\sum_{n=0}^{N-1}x[n]$. An equivalent detector divides $T(\mathbf{x})$ by NA to decide $\mathcal{H}_1$ if

$$T'(\mathbf{x}) = \frac{1}{NA}T(\mathbf{x}) = \frac{1}{N}\sum_{n=0}^{N-1}x[n] = \bar{x} > \gamma''$$

where $\gamma'' = \gamma'/NA$. But this is just the sample mean detector discussed in Chapter 3. Its performance has also been described there. Note that if $A < 0$, the inequality reverses and we decide $\mathcal{H}_1$ if $\bar{x} < \gamma''$.

◇

Example 4.2 - Damped Exponential in WGN

Now let $s[n] = r^n$ for $0 < r < 1$. From (4.3) the test statistic is

$$T(\mathbf{x}) = \sum_{n=0}^{N-1}x[n]r^n$$

and is seen to weight the earlier samples more heavily than the later ones. This is due to the fact that the signal is decaying as n increases while the noise power remains constant. The signal-to-noise ratio (SNR) for the nth sample is $s^2[n]/\sigma^2 = r^{2n}/\sigma^2$, which decreases with n. The detection performance can easily be determined as described in Section 4.3.2.

◇

In general, the test statistic of (4.3) weights the data samples according to the value of the signal. More weight is reserved for the larger signal samples. Even negative signal samples are weighted in the same manner because by multiplying $x[n]$ by $s[n]$, negative samples yield a positive contribution to the sum. The detector of (4.3) is referred to as a *correlator* or *replica-correlator* since we correlate the received data with a replica of the signal. The detector is shown in Figure 4.1a. An alternative interpretation of the test statistic is based on relating the correlation process to the effect of a finite impulse response (FIR) filter on the data. Specifically, if $x[n]$ is the input to an FIR filter with impulse response $h[n]$, where $h[n]$ is nonzero

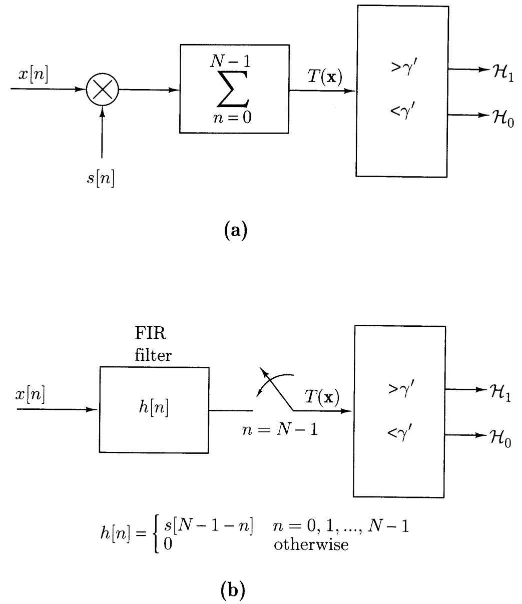


Figure 4.1. Neyman-Pearson detector for deterministic signal in white Gaussian noise (a) Replica-correlator (b) Matched filter.

for $n = 0, 1, \ldots, N-1$, then the output at time n is

$$y[n] = \sum_{k=0}^{n} h[n-k]x[k] \tag{4.4}$$

for $n \geq 0$. (For $n < 0$ the output is zero since we assume $x[n]$ is also nonzero only over the interval $[0, N-1]$.) If we let the impulse response be a "flipped around" version of the signal or

$$h[n] = s[N-1-n] \qquad n = 0, 1, \ldots, N-1 \tag{4.5}$$

then

$$y[n] = \sum_{k=0}^{n} s[N-1-(n-k)]x[k].$$

Now the output of the filter at time $n = N-1$ is

$$y[N-1] = \sum_{k=0}^{N-1} s[k]x[k]$$

which with a change of summation variable is identical to the replica-correlator. This implementation of the NP detector is shown in Figure 4.1b and is known as a *matched filter*. The filter impulse response is *matched* to the signal. Figure 4.2 shows some examples. The matched filter impulse response is obtained by flipping $s[n]$ about $n = 0$ and shifting to the right by $N-1$ samples.

Although the test statistic is obtained by sampling the matched filter output at time $n = N-1$, it is instructive to view the entire output of the matched filter. For the DC level signal shown in Figure 4.2b, the signal output is given via the convolution sum in Figure 4.3. Note that the signal output attains a maximum at the sampling time. This is true in general (see Problem 4.2). When noise is present, the maximum may be perturbed but it should be intuitively clear that the best detection performance will be obtained by sampling at $n = N-1$. If, however, the signal does not begin at $n = 0$, but we assume that it does and use the corresponding matched filter, poor detection performance may be obtained. An example is given in Problem 4.3. Thus, for signals with *unknown arrival times*, we cannot use the matched filter in its present form. In Chapter 7 we will see how to modify it to circumvent this problem.

The matched filter may also be viewed in the frequency domain. From (4.4) we have

$$y[n] = \int_{-\frac{1}{2}}^{\frac{1}{2}} H(f)X(f)\exp(j2\pi fn)df \tag{4.6}$$

where $H(f), X(f)$ are the discrete-time Fourier transforms of $h[n]$ and $x[n]$, respectively. But from (4.5), $H(f) = \mathcal{F}\{s[N-1-n]\}$, where $\mathcal{F}$ denotes the discrete-time Fourier transform. This may be shown to be

$$H(f) = S^*(f)\exp[-j2\pi f(N-1)] \tag{4.7}$$

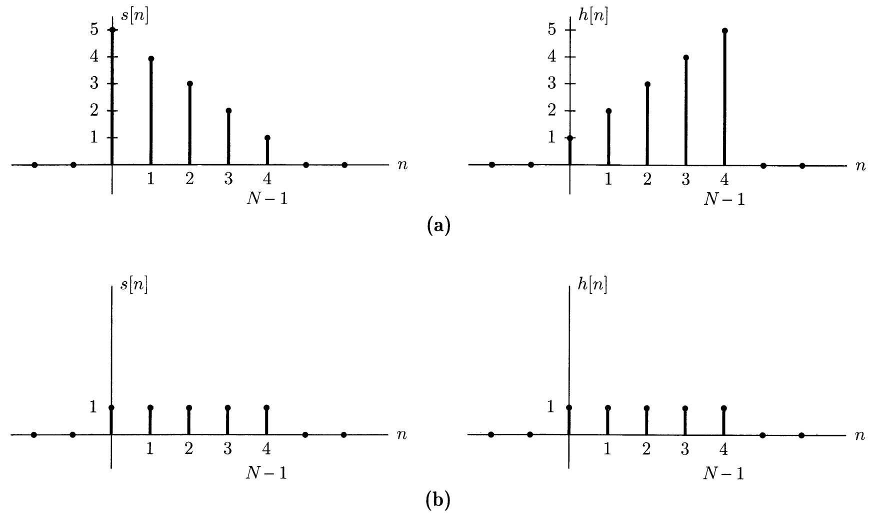


Figure 4.2. Examples of matched filter impulse response ($N = 5$).

so that (4.6) becomes

$$y[n] = \int_{-\frac{1}{2}}^{\frac{1}{2}} S^*(f)X(f)\exp[j2\pi f(n-(N-1))]df. \tag{4.8}$$

Sampling the output at $n = N-1$ produces

$$y[N-1] = \int_{-\frac{1}{2}}^{\frac{1}{2}} S^*(f)X(f)df \tag{4.9}$$

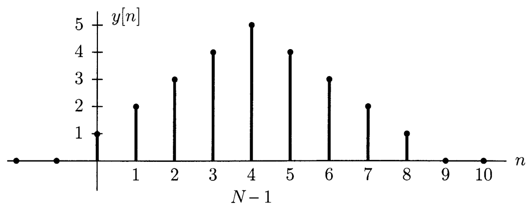


Figure 4.3. Matched filter signal output for DC level signal input ($N = 5$).

which could also have been obtained from the correlator implementation by using Parseval's theorem. Note from (4.7) that the matched filter emphasizes the bands where there is more signal power. Also, from (4.9) (or equivalently from (4.3)), when the noise is absent or $X(f) = S(f)$, the matched filter output is just the signal *energy*.

Another property of the matched filter is that *it maximizes the SNR at the output of an FIR filter*. In other words, we consider all detectors of the form of Figure 4.1b but with an arbitrary $h[n]$ over $[0, N-1]$ and zero otherwise. If we define the output SNR as

$$\begin{aligned}\eta &= \frac{E^2(y[N-1];\mathcal{H}_1)}{\operatorname{var}(y[N-1];\mathcal{H}_1)} \\ &= \frac{\left(\sum_{k=0}^{N-1} h[N-1-k]s[k]\right)^2}{E\left[\left(\sum_{k=0}^{N-1} h[N-1-k]w[k]\right)^2\right]}\end{aligned} \tag{4.10}$$

then the matched filter of (4.5) maximizes (4.10). To show this let $\mathbf{s} = [s[0]\, s[1] \dots s[N-1]]^T$, $\mathbf{h} = [h[N-1]\, h[N-2] \dots h[0]]^T$ and $\mathbf{w} = [w[0]\, w[1] \dots w[N-1]]^T$. Then

$$\begin{aligned}\eta &= \frac{\left(\mathbf{h}^T\mathbf{s}\right)^2}{E\left[\left(\mathbf{h}^T\mathbf{w}\right)^2\right]} = \frac{\left(\mathbf{h}^T\mathbf{s}\right)^2}{\mathbf{h}^T E(\mathbf{w}\mathbf{w}^T)\mathbf{h}} \\ &= \frac{\left(\mathbf{h}^T\mathbf{s}\right)^2}{\mathbf{h}^T\sigma^2\mathbf{I}\mathbf{h}} = \frac{1}{\sigma^2}\frac{\left(\mathbf{h}^T\mathbf{s}\right)^2}{\mathbf{h}^T\mathbf{h}}.\end{aligned}$$

By the Cauchy-Schwarz inequality (see Appendix 1)

$$(\mathbf{h}^T\mathbf{s})^2 \le (\mathbf{h}^T\mathbf{h})(\mathbf{s}^T\mathbf{s})$$

with equality if and only if $\mathbf{h} = c\mathbf{s}$, where c is any constant. Hence

$$\eta \le \frac{1}{\sigma^2}\mathbf{s}^T\mathbf{s}$$

with equality if and only if $\mathbf{h} = c\mathbf{s}$. The maximum output SNR is attained for (letting $c = 1$)

$$h[N-1-n] = s[n] \qquad n = 0, 1, \dots, N-1$$

or equivalently

$$h[n] = s[N-1-n] \qquad n = 0, 1, \dots, N-1$$

which is the matched filter. Note that the maximum SNR is $\eta_{\max} = \mathbf{s}^T\mathbf{s}/\sigma^2 = \mathcal{E}/\sigma^2$, where $\mathcal{E}$ is the signal energy. One would expect that the performance of the matched filter detector would increase monotonically with $\eta_{\max}$. This will be seen to be the case in the next section. For the detection problem of a known deterministic signal in WGN, the NP criterion and the maximum SNR criterion both lead to the matched filter detector. Since we know that the NP criterion produces an optimal detector, the maximum SNR criterion also does so for these model assumptions. However, for nonGaussian noise, as an example, the matched filter is not optimal in the NP sense but may still be said to maximize the SNR at the output of a *linear* FIR filter. (Actually, the matched filter more generally maximizes the SNR at the output of *any* linear filter, i.e., even ones with an infinite impulse response (see Problem 4.4).) This is because in the nonGaussian noise case the NP detector is not linear. For moderate deviations of the noise PDF from Gaussian, however, the matched filter may still be a good approximation. See Chapter 10 for a further discussion.

4.3.2 Performance of Matched Filter

We now determine the detection performance. Specifically, we will derive P_D for a given P_{FA}. Using the replica-correlator form we decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]s[n] > \gamma'.$$

Under either hypothesis $x[n]$ is Gaussian and since $T(\mathbf{x})$ is a linear combination of Gaussian random variables, $T(\mathbf{x})$ is also Gaussian. Let $E(T;\mathcal{H}_i)$ and $\operatorname{var}(T;\mathcal{H}_i)$ denote the expected value and variance of $T(\mathbf{x})$ under $\mathcal{H}_i$. Then

$$\begin{aligned}E(T;\mathcal{H}_0) &= E\left(\sum_{n=0}^{N-1} w[n]s[n]\right) = 0 \\ E(T;\mathcal{H}_1) &= E\left(\sum_{n=0}^{N-1} (s[n]+w[n])s[n]\right) = \mathcal{E} \\ \operatorname{var}(T;\mathcal{H}_0) &= \operatorname{var}\left(\sum_{n=0}^{N-1} w[n]s[n]\right) \\ &= \sum_{n=0}^{N-1} \operatorname{var}(w[n])s^2[n] \\ &= \sigma^2\sum_{n=0}^{N-1} s^2[n] = \sigma^2\mathcal{E}\end{aligned}$$

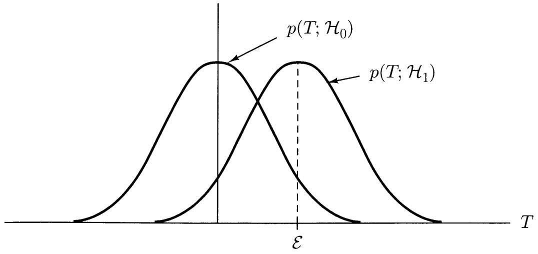


Figure 4.4. PDFs of matched filter test statistic.

where we have used the fact that the $w[n]$'s are uncorrelated. Similarly, $\mathrm{var}(T;\mathcal{H}_1) = \sigma^2\mathcal{E}$. Thus,

$$T \sim \begin{cases} \mathcal{N}(0,\sigma^2\mathcal{E}) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(\mathcal{E},\sigma^2\mathcal{E}) & \text{under } \mathcal{H}_1 \end{cases} \tag{4.11}$$

which is shown in Figure 4.4. Note that the scaled test statistic $T' = T/\sqrt{\sigma^2\mathcal{E}}$ has the PDF

$$T' \sim \begin{cases} \mathcal{N}(0,1) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(\sqrt{\mathcal{E}/\sigma^2},1) & \text{under } \mathcal{H}_1. \end{cases}$$

It follows then that the detection performance must increase with $\sqrt{\mathcal{E}/\sigma^2}$ or with $\mathcal{E}/\sigma^2$ since as $\mathcal{E}/\sigma^2$ increases the PDFs retain the same shape but move further apart. To verify this we have from (4.11)

$$\begin{aligned} P_{FA} &= \Pr\{T > \gamma';\mathcal{H}_0\} \\ &= Q\left(\frac{\gamma'}{\sqrt{\sigma^2\mathcal{E}}}\right) \end{aligned} \tag{4.12}$$

$$\begin{aligned} P_D &= \Pr\{T > \gamma';\mathcal{H}_1\} \\ &= Q\left(\frac{\gamma' - \mathcal{E}}{\sqrt{\sigma^2\mathcal{E}}}\right) \end{aligned} \tag{4.13}$$

where

$$\begin{aligned} Q(x) &= \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt \\ &= 1 - \Phi(x) \end{aligned}$$

and $\Phi(x)$ is the CDF for a $\mathcal{N}(0,1)$ random variable. Now, since a CDF is a monotonically increasing function, $Q(x)$ must be monotonically decreasing and so has an

inverse function $Q^{-1}(\cdot)$. (See Chapter 2 for an evaluation of $Q(\cdot)$ and $Q^{-1}(\cdot)$.) This allows us to write (4.12) as

$$\gamma' = \sqrt{\sigma^2\mathcal{E}}\,Q^{-1}(P_{FA}).$$

Substituting in (4.13) produces

$$\begin{aligned} P_D &= Q\left(\frac{\sqrt{\sigma^2\mathcal{E}}\,Q^{-1}(P_{FA})}{\sqrt{\sigma^2\mathcal{E}}} - \sqrt{\frac{\mathcal{E}}{\sigma^2}}\right) \\ &= Q\left(Q^{-1}(P_{FA}) - \sqrt{\frac{\mathcal{E}}{\sigma^2}}\right). \end{aligned} \tag{4.14}$$

As asserted, as $\eta = \mathcal{E}/\sigma^2$ increases, the argument of $Q(\cdot)$ decreases, and P_D increases. The detection performance is summarized in Figure 4.5. The key parameter is the SNR at the matched filter output, which is the energy-to-noise ratio (ENR) or $\mathcal{E}/\sigma^2$. These detection curves are identical to those in Figure 3.5. The only difference is that for the DC level in WGN, the energy is NA^2 so that $\eta = NA^2/\sigma^2$. It is apparent from the curves that to improve the detection performance we can always increase P_{FA} and/or increase the ENR. The ENR is increased by increasing the signal energy, either by increasing the signal level (A) or signal duration (N) in the case of a DC level signal. *The shape of the signal does not affect the detection performance.* For example, the two signals shown in Figure 4.6 both yield the

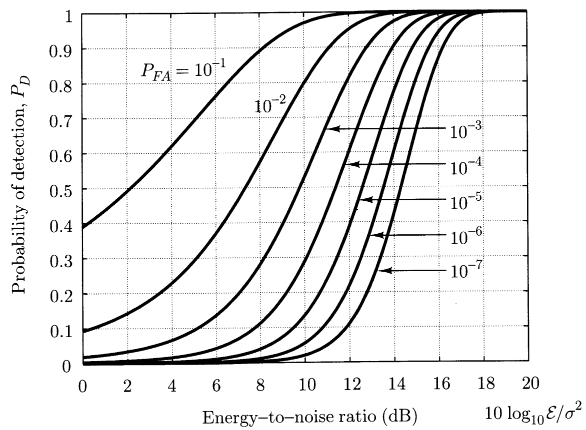


Figure 4.5. Detection performance of matched filter.

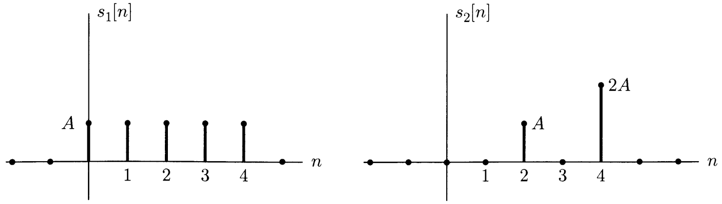


Figure 4.6. Signals yielding same detection performance.

same detection performance since their energies are equal. As we will see shortly, the signal shape does become an important design consideration when the noise is *colored.*

The use of a matched filter for signal detection leads to the concept of *processing gain.* Processing gain can be viewed as the advantage of making a decision based on a test statistic, which is an optimal combination of the data, as opposed to a decision based on a direct examination of the data. It is defined as the SNR of the optimal test statistic, i.e, the *processed data,* divided by the SNR of a single data sample, i.e., the unprocessed data. As an example, consider a DC level in WGN. If one were to attempt to detect the signal based on only a single sample, then the performance would be given by (4.14) with

$$\eta_{\text{in}} = \frac{\mathcal{E}}{\sigma^2} = \frac{A^2}{\sigma^2}.$$

By using a matched filter to process N samples the performance improves to

$$\eta_{\text{out}} = \frac{\mathcal{E}}{\sigma^2} = \frac{NA^2}{\sigma^2}.$$

The improvement in SNR or $\eta_{\text{out}}/\eta_{\text{in}}$ is termed the processing gain (PG) or

$$\begin{aligned} PG &= 10\log_{10}\frac{\eta_{\text{out}}}{\eta_{\text{in}}} \\ &= 10\log_{10} N \qquad \text{dB.} \end{aligned}$$

For example, from Figure 4.5 the ENR required to attain $P_D = 0.5$ for a $P_{FA} = 10^{-3}$ is about 10 dB. If we required our detector to have $P_D = 0.95$, the ENR must be increased by 4 dB. This is attainable if we increase N by a factor of 2.5, since then the PG will increase by 4 dB. Processing gain considerations are important in designing sonar/radar systems. See also Problem 4.9 for a further discussion.

Lastly, the matched filter test statistic has the PDF of the mean-shifted Gauss-Gauss problem ($\mathcal{N}(\mu_1, \sigma^2)$ versus $\mathcal{N}(\mu_2, \sigma^2)$). Thus, the deflection coefficient characterizes the detection performance as described in Chapter 3. Recall that the

definition of this coefficient is

$$d^2 = \frac{(E(T;\mathcal{H}_1) - E(T;\mathcal{H}_0))^2}{\text{var}(T;\mathcal{H}_0)}. \tag{4.15}$$

From (4.11) this is $d^2 = \mathcal{E}/\sigma^2$ or just the SNR at the matched filter output. Not surprising then, the performance as given by (4.14) increases monotonically with d^2. It should be emphasized that only for the mean-shifted Gauss-Gauss problem does the deflection coefficient completely characterize the detection performance. However, it is often used for *approximate* detection performance evaluation for other detection problems.

4.4 Generalized Matched Filters

The matched filter is an optimal detector for a known signal in WGN. In many situations, however, the noise is more accurately modeled as *correlated* noise. Thus, we now assume that $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$, where $\mathbf{C}$ is the covariance matrix. If the noise is modeled as wide sense stationary (WSS), then $\mathbf{C}$ has the special form of a symmetric Toeplitz matrix (see Appendix 1). This is because for a zero mean WSS random process

$$[\mathbf{C}]_{mn} = \text{cov}(w[m], w[n]) = E(w[m]w[n]) = r_{ww}[m-n].$$

Hence, the elements along any NW to SE diagonal of $\mathbf{C}$ are the same. For nonstationary noise $\mathbf{C}$ will be an arbitrary covariance matrix.

For WGN we assumed that the received data samples were observed over the signal interval $[0, N-1]$. (It is assumed that the signal is zero outside the interval $[0, N-1]$.) The data samples outside this interval are irrelevant since the noise outside the interval is independent of the noise inside the interval and so can be discarded. Hence, in the presence of WGN there is no loss in detection performance by assuming the observation interval to be $[0, N-1]$ (see also Problem 4.4). However, for a signal embedded in correlated noise this is not the case. Improved performance may be obtained by incorporating data samples outside the signal interval into the detector. Problem 4.5 discusses a somewhat contrived example where perfect detection is possible by doing so. The reader may also wish to refer to Problem 4.14 in which the optimal detector is described for an infinite observation interval. In the discussion to follow we assume that the observed data samples are $\mathbf{x} = [x[0]\, x[1] \ldots x[N-1]]^T$. In practice, larger data intervals appear to be seldom used due to the nonstationarity of the noise samples. However, the reader should bear in mind that improved detection performance may be possible in some situations.

To determine the NP detector we again determine the likelihood ratio test (LRT) with

$$p(\mathbf{x};\mathcal{H}_1) = \frac{1}{(2\pi)^{\frac{N}{2}}\det^{\frac{1}{2}}(\mathbf{C})}\exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{s})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{s})\right]$$

$$p(\mathbf{x};\mathcal{H}_0) = \frac{1}{(2\pi)^{\frac{N}{2}}\det^{\frac{1}{2}}(\mathbf{C})}\exp\left[-\frac{1}{2}\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x}\right]$$

where we have noted that under $\mathcal{H}_0$, $\mathbf{x}\sim\mathcal{N}(\mathbf{0},\mathbf{C})$ and under $\mathcal{H}_1$, $\mathbf{x}\sim\mathcal{N}(\mathbf{s},\mathbf{C})$. In the WGN case $\mathbf{C}=\sigma^2\mathbf{I}$ and the PDFs reduce to the ones in Section 4.3.1. Now, we decide $\mathcal{H}_1$ if

$$l(\mathbf{x}) = \ln\frac{p(\mathbf{x};\mathcal{H}_1)}{p(\mathbf{x};\mathcal{H}_0)} > \ln\gamma.$$

But

$$\begin{aligned} l(\mathbf{x}) &= -\frac{1}{2}\left[(\mathbf{x}-\mathbf{s})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{s}) - \mathbf{x}^T\mathbf{C}^{-1}\mathbf{x}\right] \\ &= -\frac{1}{2}\left[\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x} - 2\mathbf{x}^T\mathbf{C}^{-1}\mathbf{s} + \mathbf{s}^T\mathbf{C}^{-1}\mathbf{s} - \mathbf{x}^T\mathbf{C}^{-1}\mathbf{x}\right] \\ &= \mathbf{x}^T\mathbf{C}^{-1}\mathbf{s} - \frac{1}{2}\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s} \end{aligned}$$

or by incorporating the non-data-dependent term into the threshold we decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \mathbf{x}^T\mathbf{C}^{-1}\mathbf{s} > \gamma'. \tag{4.16}$$

Note that for WGN $\mathbf{C}=\sigma^2\mathbf{I}$ the detector reduces to

$$\frac{\mathbf{x}^T\mathbf{s}}{\sigma^2} > \gamma'$$

or

$$\mathbf{x}^T\mathbf{s} = \sum_{n=0}^{N-1} x[n]s[n] > \gamma''$$

as before. The detector of (4.16) is referred to as a *generalized replica-correlator* or *matched filter*. It may be viewed as a replica-correlator where the replica is the modified signal $\mathbf{s}'=\mathbf{C}^{-1}\mathbf{s}$. Then

$$T(\mathbf{x}) = \mathbf{x}^T\mathbf{C}^{-1}\mathbf{s} = \mathbf{x}^T\mathbf{s}'$$

so that we correlate with the modified signal. An example follows.

Example 4.3 - Uncorrelated Noise with Unequal Variances

If $w[n]\sim\mathcal{N}(0,\sigma_n^2)$ and the $w[n]$'s are uncorrelated, then $\mathbf{C}=\mathrm{diag}(\sigma_0^2,\sigma_1^2,\ldots,\sigma_{N-1}^2)$ and $\mathbf{C}^{-1}=\mathrm{diag}(1/\sigma_0^2,1/\sigma_1^2,\ldots,1/\sigma_{N-1}^2)$. Hence, from (4.16) we decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \sum_{n=0}^{N-1}\frac{x[n]s[n]}{\sigma_n^2} > \gamma'.$$

It is seen that if a data sample has a small variance, then we weight its contribution to the sum more heavily. Alternatively, under $\mathcal{H}_1$ we have

$$\begin{aligned} T(\mathbf{x}) &= \sum_{n=0}^{N-1}\frac{w[n]+s[n]}{\sigma_n}\frac{s[n]}{\sigma_n} \\ &= \sum_{n=0}^{N-1}\left(w'[n]+\frac{s[n]}{\sigma_n}\right)\frac{s[n]}{\sigma_n} \end{aligned}$$

where the noise samples have been equalized or *prewhitened* since $\mathbf{C}_{w'}=\mathbf{I}$. Thus, the generalized matched filter first prewhitens the noise samples. In doing so, it also distorts the signal, if present, to be $s'[n]=s[n]/\sigma_n$. As a result, after prewhitening the detector correlates against the distorted signal. The generalized matched filter can be expressed as

$$T(\mathbf{x}') = \sum_{n=0}^{N-1} x'[n]s'[n]$$

where $x'[n]=x[n]/\sigma_n$. It can be thought of as a prewhitener, which is then followed by a correlator or matched filter to the distorted signal $s'[n]$.

◇

More generally, for any $\mathbf{C}$ that is positive definite, it can be shown that $\mathbf{C}^{-1}$ exists and is also positive definite. Consequently, it may be factored as $\mathbf{C}^{-1}=\mathbf{D}^T\mathbf{D}$, where $\mathbf{D}$ is a nonsingular $N\times N$ matrix. For the previous example, $\mathbf{D}=\mathrm{diag}(1/\sigma_0,1/\sigma_1,\ldots,1/\sigma_{N-1})$. Thus, the test statistic becomes

$$\begin{aligned} T(\mathbf{x}) &= \mathbf{x}^T\mathbf{C}^{-1}\mathbf{s} \\ &= \mathbf{x}^T\mathbf{D}^T\mathbf{D}\mathbf{s} \\ &= \mathbf{x}'^T\mathbf{s}' \end{aligned}$$

where $\mathbf{x}'=\mathbf{D}\mathbf{x}$ and $\mathbf{s}'=\mathbf{D}\mathbf{s}$. The prewhitening form of the generalized matched filter is shown in Figure 4.7. To show that the linear transformation $\mathbf{D}$, called a *prewhitening matrix*, does indeed produce WGN, let $\mathbf{w}'=\mathbf{D}\mathbf{w}$. Then

$$\begin{aligned} \mathbf{C}_{w'} &= E(\mathbf{w}'\mathbf{w}'^T) = E(\mathbf{D}\mathbf{w}\mathbf{w}^T\mathbf{D}^T) \\ &= \mathbf{D}E(\mathbf{w}\mathbf{w}^T)\mathbf{D}^T = \mathbf{D}\mathbf{C}\mathbf{D}^T \\ &= \mathbf{D}(\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T = \mathbf{I}. \end{aligned}$$

Alternatively, the generalized matched filter can be derived using a prewhitening approach as described in Problem 4.11.

If the data record length N is large and the noise is WSS, then the generalized matched filter may be approximated as shown in Problem 4.12. The test statistic

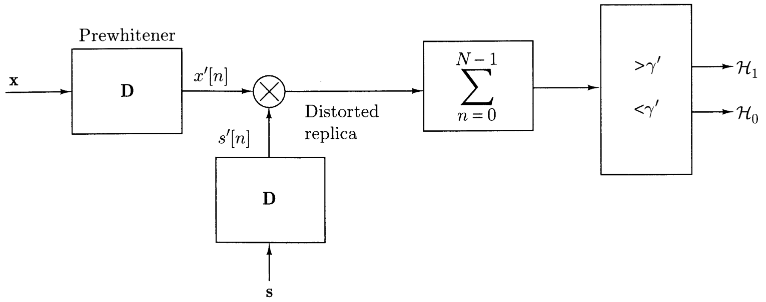


Figure 4.7. Generalized matched filter as prewhitener plus replica-correlator (matched filter).

then becomes

$$T(\mathbf{x}) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{X(f)S^*(f)}{P_{ww}(f)} df \tag{4.17}$$

where $P_{ww}(f)$ is the PSD of the noise. The whitening effect of $\mathbf{C}^{-1}$ is replaced by the frequency weighting $1/P_{ww}(f)$ in the test statistic. Again it is evident that the frequency bands of importance are those for which the noise PSD is small or the SNR is high (see also (4.19)).

4.4.1 Performance of Generalized Matched Filter

The generalized matched filter decides $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \mathbf{x}^T\mathbf{C}^{-1}\mathbf{s} > \gamma'.$$

Under either hypothesis the test statistic is Gaussian, being a linear transformation of $\mathbf{x}$. Determining the moments we have

$$\begin{aligned}
E(T;\mathcal{H}_0) &= E(\mathbf{w}^T\mathbf{C}^{-1}\mathbf{s}) = 0 \\
E(T;\mathcal{H}_1) &= E\left[(\mathbf{s}+\mathbf{w})^T\mathbf{C}^{-1}\mathbf{s}\right] = \mathbf{s}^T\mathbf{C}^{-1}\mathbf{s} \\
\operatorname{var}(T;\mathcal{H}_0) &= E\left[(\mathbf{w}^T\mathbf{C}^{-1}\mathbf{s})^2\right] = E(\mathbf{s}^T\mathbf{C}^{-1}\mathbf{w}\mathbf{w}^T\mathbf{C}^{-1}\mathbf{s}) \\
&= \mathbf{s}^T\mathbf{C}^{-1}E(\mathbf{w}\mathbf{w}^T)\mathbf{C}^{-1}\mathbf{s} = \mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}
\end{aligned}$$

where we have used $\mathbf{C}^{-1^T} = \mathbf{C}^{T^{-1}} = \mathbf{C}^{-1}$. Also

$$\begin{aligned}
\operatorname{var}(T;\mathcal{H}_1) &= E\left[\left(\mathbf{x}^T\mathbf{C}^{-1}\mathbf{s} - E(\mathbf{x}^T\mathbf{C}^{-1}\mathbf{s})\right)^2\right]
\end{aligned}$$

$$\begin{aligned}
&= E\left[\left((\mathbf{x}-E(\mathbf{x}))^T\mathbf{C}^{-1}\mathbf{s}\right)^2\right] \\
&= E\left[(\mathbf{w}^T\mathbf{C}^{-1}\mathbf{s})^2\right] = \operatorname{var}(T;\mathcal{H}_0).
\end{aligned}$$

As in the white noise case we can easily find P_{FA} and P_D. From Chapter 3

$$P_D = Q\left(Q^{-1}(P_{FA}) - \sqrt{d^2}\right)$$

where d^2 is the deflection coefficient of (4.15). The latter is easily shown to be $d^2 = \mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}$ so that finally

$$P_D = Q\left(Q^{-1}(P_{FA}) - \sqrt{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}\right). \tag{4.18}$$

For $\mathbf{C} = \sigma^2\mathbf{I}$ we have (4.14). Note that the probability of detection increases monotonically with $\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}$, not with $\mathcal{E}/\sigma^2$. In the WGN case the signal shape was unimportant, only the signal energy mattered. Now, however, the signal can be designed to maximize $\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}$ and hence P_D. An example follows.

Example 4.4 - Signal Design for Uncorrelated Noise with Unequal Variances

We continue Example 4.3. Since $\mathbf{C}^{-1} = \operatorname{diag}(1/\sigma_0^2, 1/\sigma_1^2, \ldots, 1/\sigma_{N-1}^2)$ we have

$$\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s} = \sum_{n=0}^{N-1} \frac{s^2[n]}{\sigma_n^2}.$$

If we attempt to maximize this by choice of $s[n]$, we arrive at the conclusion that we should let $s[n] \to \infty$. Clearly, a practical constraint is necessary. We choose to constrain the signal energy or we let $\sum_{n=0}^{N-1} s^2[n] = \mathcal{E}$. Then using Lagrangian multipliers to effect the maximization, we maximize

$$F = \sum_{n=0}^{N-1} \frac{s^2[n]}{\sigma_n^2} + \lambda\left(\mathcal{E} - \sum_{n=0}^{N-1} s^2[n]\right).$$

Taking partial derivatives yields

$$\begin{aligned}
\frac{\partial F}{\partial s[k]} &= \frac{2s[k]}{\sigma_k^2} - 2\lambda s[k] \\
&= 2s[k]\left(\frac{1}{\sigma_k^2} - \lambda\right) \\
&= 0 \qquad\qquad k = 0, 1, \ldots, N-1.
\end{aligned}$$

Since λ is a constant, we can have $1/\sigma_k^2 - \lambda = 0$ for at most a single k (assuming all the σ_n^2's are distinct). Assume this condition is met for $k = j$. Then, $s[k] = 0$ for the remaining samples or for $k \neq j$. To determine j we note that under these conditions

$$\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s} = \frac{s^2[j]}{\sigma_j^2} = \frac{\mathcal{E}}{\sigma_j^2}.$$

We choose j for which σ_j^2 is *minimum*. In other words, we choose to concentrate the signal energy in the sample that has the minimum noise variance. In this way the deflection coefficient is maximized.

◇

We now generalize this result by considering an arbitrary noise covariance matrix. The optimal signal is chosen by maximizing $\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}$ subject to the fixed energy constraint $\mathbf{s}^T\mathbf{s} = \mathcal{E}$. Again, making use of Lagrangian multipliers we seek to maximize

$$F = \mathbf{s}^T\mathbf{C}^{-1}\mathbf{s} + \lambda(\mathcal{E} - \mathbf{s}^T\mathbf{s}).$$

Using the identities (see Appendix 1)

$$\begin{aligned}\frac{\partial \mathbf{b}^T\mathbf{x}}{\partial \mathbf{x}} &= \mathbf{b}\\ \frac{\partial \mathbf{x}^T\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} &= 2\mathbf{A}\mathbf{x}\end{aligned}$$

for $\mathbf{A}$ a symmetric matrix, we have

$$\frac{\partial F}{\partial \mathbf{s}} = 2\mathbf{C}^{-1}\mathbf{s} - 2\lambda\mathbf{s} = \mathbf{0}$$

or

$$\mathbf{C}^{-1}\mathbf{s} = \lambda\mathbf{s}.$$

Hence, $\mathbf{s}$ is an eigenvector of $\mathbf{C}^{-1}$. The eigenvector that should be chosen is the one that maximizes $\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}$. But

$$\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s} = \mathbf{s}^T\lambda\mathbf{s} = \lambda\mathcal{E}.$$

Thus, we should choose $\mathbf{s}$ as the eigenvector of $\mathbf{C}^{-1}$ whose corresponding eigenvalue, i.e. λ, is maximum. (Recall that the eigenvalues of a positive definite covariance matrix are all real and positive.) Alternatively, since $\mathbf{C}^{-1}\mathbf{s} = \lambda\mathbf{s}$ implies that $\mathbf{C}\mathbf{s} = (1/\lambda)\mathbf{s}$, *we should choose the signal as the eigenvector of* $\mathbf{C}$ *that has the minimum eigenvalue.* If the eigenvalues of $\mathbf{C}$ are distinct, then this choice yields the unique optimal signal. Thus, if $\mathbf{v}_{\min}$ is the eigenvector associated with the minimum eigenvalue, then we choose $\mathbf{s} = \sqrt{\mathcal{E}}\mathbf{v}_{\min}$. The scaling enforces the energy constraint since the eigenvector is assumed to be scaled to have unity length. An example follows.

Example 4.5 - Signal Design for Correlated Noise

Assume that

$$\mathbf{C} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

where ρ is the correlation coefficient, which satisfies $|\rho| \leq 1$. Then, by solving the characteristic equation $(1-\lambda)^2 - \rho^2 = 0$ we find that $\lambda_1 = 1+\rho$, $\lambda_2 = 1-\rho$. The eigenvectors $\mathbf{v}_i$ are easily found from $(\mathbf{C} - \lambda_i\mathbf{I})\mathbf{v} = \mathbf{0}$ or

$$\begin{bmatrix} 1-\lambda_i & \rho \\ \rho & 1-\lambda_i \end{bmatrix}\begin{bmatrix} [\mathbf{v}]_1 \\ [\mathbf{v}]_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

This can be shown to produce

$$\begin{aligned}\mathbf{v}_1 &= \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}\\ \mathbf{v}_2 &= \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}.\end{aligned}$$

Assuming $\rho > 0$, so that the two noise samples are positively correlated, λ_2 is the minimum eigenvalue. Hence, the optimal signal is

$$\mathbf{s} = \sqrt{\mathcal{E}}\mathbf{v}_2 = \sqrt{\frac{\mathcal{E}}{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

The test statistic then becomes

$$\begin{aligned}T(\mathbf{x}) &= \mathbf{x}^T\mathbf{C}^{-1}\mathbf{s}\\ &= \mathbf{x}^T\mathbf{C}^{-1}\sqrt{\mathcal{E}}\mathbf{v}_2\\ &= \sqrt{\mathcal{E}}\mathbf{x}^T\frac{1}{\lambda_2}\mathbf{v}_2\\ &= \frac{\sqrt{\frac{\mathcal{E}}{2}}}{1-\rho}(x[0]-x[1]).\end{aligned}$$

By subtracting the two data samples we are effectively canceling the noise (due to the positive correlation). However, the signal contribution will not be canceled since $s[1] = -s[0]$. Also, the deflection coefficient for this example becomes

$$d^2 = \mathbf{s}^T\mathbf{C}^{-1}\mathbf{s} = \mathcal{E}\mathbf{v}_2^T\mathbf{C}^{-1}\mathbf{v}_2$$

$$
\begin{aligned}
&= \mathcal{E}\mathbf{v}_2^T \frac{1}{\lambda_2}\mathbf{v}_2 = \frac{\mathcal{E}}{\lambda_2} \\
&= \frac{\mathcal{E}}{1-\rho}
\end{aligned}
$$

and as $\rho \to 1$, $d^2 \to \infty$ since the noise is perfectly canceled.

◇

Finally, consider the case of a large data record and WSS noise. The detection performance of the test statistic of (4.17) can be shown to be given by (4.18) but with $d^2 = \mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}$ replaced by

$$d^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{|S(f)|^2}{P_{ww}(f)}df. \tag{4.19}$$

From the deflection coefficient it follows that to maximize detection performance one should concentrate the signal energy in the frequency band where the noise PSD is minimum. See also Problem 4.13.

4.5 Multiple Signals

4.5.1 Binary Case

The problem we have been addressing is the detection of a known signal in noise. Sonar/radar systems make extensive use of the matched filter detector. In a communication system, however, the problem is slightly different. We typically transmit one of M signals. At the receiver there is no question as to whether a signal is present but only which one. The problem then is one of classification (although we still generally refer to it as detection), which we now discuss.

We begin with the binary detector and later generalize the results. Mathematically, we have the following hypothesis testing problem

$$
\begin{aligned}
\mathcal{H}_0 &: x[n] = s_0[n] + w[n] \quad n = 0, 1, \ldots, N-1 \\
\mathcal{H}_1 &: x[n] = s_1[n] + w[n] \quad n = 0, 1, \ldots, N-1
\end{aligned}
$$

where $s_0[n], s_1[n]$ are known deterministic signals and $w[n]$ is WGN with variance σ^2. Because we are considering the communication problem, in which each type of error is equally undesirable, it makes sense to choose the minimum probability of error criterion. We then decide $\mathcal{H}_1$ if (see Chapter 3)

$$\frac{p(\mathbf{x}|\mathcal{H}_1)}{p(\mathbf{x}|\mathcal{H}_0)} > \gamma = \frac{P(\mathcal{H}_0)}{P(\mathcal{H}_1)} = 1$$

where we have assumed equal prior probabilities of transmitting $s_0[n]$ and $s_1[n]$. We choose the hypothesis having the larger conditional likelihood, i.e., the larger

$p(\mathbf{x}|\mathcal{H}_i)$. As discussed previously, this is the ML rule. Since

$$p(\mathbf{x}|\mathcal{H}_i) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-s_i[n])^2\right]$$

we decide $\mathcal{H}_i$ for which

$$D_i^2 = \sum_{n=0}^{N-1}(x[n]-s_i[n])^2 \tag{4.20}$$

is minimum. This is referred to as a *minimum distance receiver*. If we consider the data and signal samples as vectors in R^N, then

$$
\begin{aligned}
D_i^2 &= (\mathbf{x}-\mathbf{s}_i)^T(\mathbf{x}-\mathbf{s}_i) \\
&= ||\mathbf{x}-\mathbf{s}_i||^2
\end{aligned}
$$

where $||\boldsymbol{\xi}||^2 = \sum_{i=0}^{N-1}\xi_i^2$ is the squared Euclidean norm in R^N. Hence, *we choose the hypothesis whose signal vector is closest to* $\mathbf{x}$.

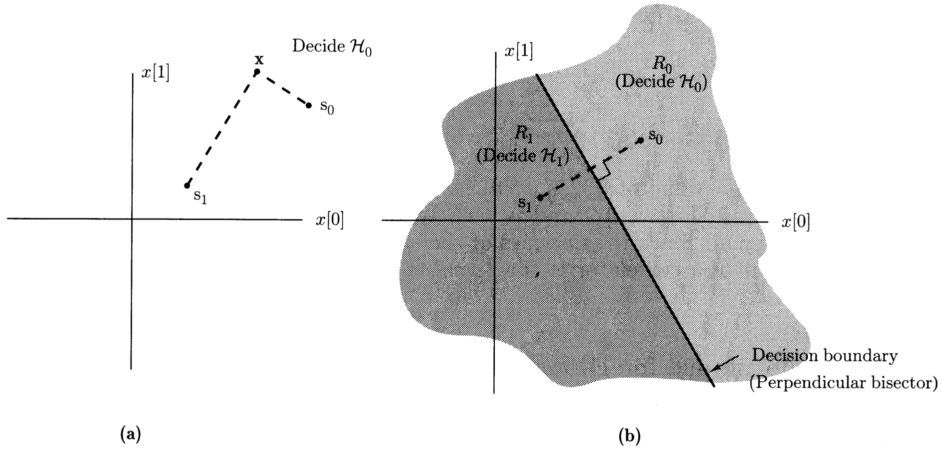


Figure 4.8. Example of decision regions (a) Distance comparisons (b) Decision regions.

Example 4.6 - Minimum Distance Receiver

For sake of illustration consider the case of $N = 2$. For all values of $\mathbf{x} = [x[0]\, x[1]]^T$ we wish to determine if $\mathbf{s}_0$ or $\mathbf{s}_1$ was sent. A typical situation is shown in Figure 4.8a, in which $\mathbf{s}_0$ and $\mathbf{s}_1$ and a given value of $\mathbf{x}$ are depicted. We assign $\mathbf{x}$ to $\mathcal{H}_0$ if $||\mathbf{x}-\mathbf{s}_0||$ is smaller than $||\mathbf{x}-\mathbf{s}_1||$ and if not, to $\mathcal{H}_1$. As shown in Figure 4.8a, it is clear that we should assign $\mathbf{x}$ to $\mathcal{H}_0$. More generally, the minimum distance receiver divides up the plane into two regions that are separated by the perpendicular bisector of the line segment joining the two signal vectors as shown in Figure 4.8b. To the right of this line, called the *decision boundary*, we decide $\mathcal{H}_0$ and to the left we decide $\mathcal{H}_1$. Hence, we have a simple mapping from the received data samples into a decision. In higher dimensions (for $N > 2$) the decision boundary is still the perpendicular bisector of the line segment joining $\mathbf{s}_0$ and $\mathbf{s}_1$. In Problem 4.18 we outline the proof for the more general case.

◇

The minimum distance receiver can also be expressed in a more familiar form. Since

$$D_i^2 = \sum_{n=0}^{N-1} x^2[n] - 2\sum_{n=0}^{N-1} x[n]s_i[n] + \sum_{n=0}^{N-1} s_i^2[n]$$

we decide $\mathcal{H}_i$ for which

$$\begin{aligned} T_i(\mathbf{x}) &= \sum_{n=0}^{N-1} x[n]s_i[n] - \frac{1}{2}\sum_{n=0}^{N-1} s_i^2[n] \\ &= \sum_{n=0}^{N-1} x[n]s_i[n] - \frac{1}{2}\mathcal{E}_i \end{aligned} \tag{4.21}$$

is *maximum*. (The value of $\sum_{n=0}^{N-1} x^2[n]$ is the same for $i = 0$ or $i = 1$.) This receiver is shown in Figure 4.9 and is seen to incorporate replica-correlators for each assumed signal. Since the energies of the signals may be different (recall that the signal output of a correlator is the signal energy), a bias term is necessary as compensation.

4.5.2 Performance for Binary Case

We now determine P_e for the ML receiver of (4.21). In doing so we will be particularly interested in the question of signal design. As might be expected from Figure 4.8, the best performance occurs when $\mathbf{s}_0$ and $\mathbf{s}_1$ are farthest apart. Of course, in practice we must impose a finite energy constraint on the chosen signals. Now the probability of error P_e is

$$P_e = P(\mathcal{H}_1|\mathcal{H}_0)P(\mathcal{H}_0) + P(\mathcal{H}_0|\mathcal{H}_1)P(\mathcal{H}_1)$$

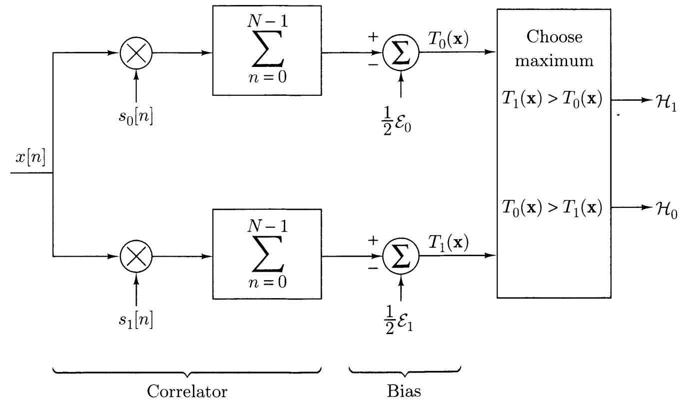


Figure 4.9. Minimum probability of error receiver for binary signal detection.

where $P(\mathcal{H}_i|\mathcal{H}_j)$ is the conditional probability of deciding $\mathcal{H}_i$ given that $\mathcal{H}_j$ is true. Assuming that the prior probabilities are equal, we have

$$\begin{aligned} P_e &= \frac{1}{2}\left[P(\mathcal{H}_1|\mathcal{H}_0) + P(\mathcal{H}_0|\mathcal{H}_1)\right] \\ &= \frac{1}{2}\left[\Pr\{T_1(\mathbf{x}) > T_0(\mathbf{x})|\mathcal{H}_0\} + \Pr\{T_0(\mathbf{x}) > T_1(\mathbf{x})|\mathcal{H}_1\}\right] \\ &= \frac{1}{2}\left[\Pr\{T_1(\mathbf{x}) - T_0(\mathbf{x}) > 0|\mathcal{H}_0\} + \Pr\{T_0(\mathbf{x}) - T_1(\mathbf{x}) > 0|\mathcal{H}_1\}\right]. \end{aligned} \tag{4.22}$$

Let $T(\mathbf{x}) = T_1(\mathbf{x}) - T_0(\mathbf{x})$. Then, from (4.21)

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n](s_1[n] - s_0[n]) - \frac{1}{2}(\mathcal{E}_1 - \mathcal{E}_0)$$

which is a Gaussian random variable conditioned on either hypothesis. The moments are

$$\begin{aligned} E(T|\mathcal{H}_0) &= \sum_{n=0}^{N-1} s_0[n](s_1[n] - s_0[n]) - \frac{1}{2}(\mathcal{E}_1 - \mathcal{E}_0) \\ &= \sum_{n=0}^{N-1} s_0[n]s_1[n] - \frac{1}{2}\sum_{n=0}^{N-1} s_0^2[n] - \frac{1}{2}\sum_{n=0}^{N-1} s_1^2[n] \end{aligned}$$

$$
\begin{aligned}
&= -\frac{1}{2}\sum_{n=0}^{N-1}(s_1[n]-s_0[n])^2 \\
&= -\frac{1}{2}||\mathbf{s}_1-\mathbf{s}_0||^2.
\end{aligned}
$$

Similarly

$$E(T|\mathcal{H}_1) = \frac{1}{2}||\mathbf{s}_1-\mathbf{s}_0||^2 = -E(T|\mathcal{H}_0).$$

Also

$$
\begin{aligned}
\operatorname{var}(T|\mathcal{H}_0) &= \operatorname{var}\left(\sum_{n=0}^{N-1} x[n](s_1[n]-s_0[n])\middle|\mathcal{H}_0\right) \\
&= \sum_{n=0}^{N-1}\operatorname{var}(x[n])(s_1[n]-s_0[n])^2 \\
&= \sigma^2||\mathbf{s}_1-\mathbf{s}_0||^2.
\end{aligned}
$$

Similarly

$$\operatorname{var}(T|\mathcal{H}_1) = \sigma^2||\mathbf{s}_1-\mathbf{s}_0||^2 = \operatorname{var}(T|\mathcal{H}_0).$$

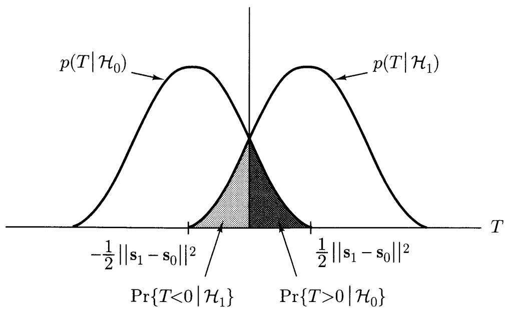


Figure 4.10. Errors for binary signal detection.

The PDF of T conditioned on either hypothesis is shown in Figure 4.10. It is clear that the errors are the same because of the inherent receiver symmetry. Thus, from (4.22)

$$P_e = \Pr\{T(\mathbf{x}) > 0|\mathcal{H}_0\}$$

where conditioned on $\mathcal{H}_0$, $T \sim \mathcal{N}(-\frac{1}{2}||\mathbf{s}_1-\mathbf{s}_0||^2, \sigma^2||\mathbf{s}_1-\mathbf{s}_0||^2)$ so that

$$P_e = Q\left(\frac{\frac{1}{2}||\mathbf{s}_1-\mathbf{s}_0||^2}{\sqrt{\sigma^2||\mathbf{s}_1-\mathbf{s}_0||^2}}\right)$$

or finally

$$P_e = Q\left(\frac{1}{2}\sqrt{\frac{||\mathbf{s}_1-\mathbf{s}_0||^2}{\sigma^2}}\right). \tag{4.23}$$

As asserted earlier, as $||\mathbf{s}_1-\mathbf{s}_0||^2$ increases, P_e decreases, as expected. However, in choosing the signals in an attempt to minimize P_e, we must impose an energy constraint. This is because the average power is usually limited (due to FCC regulations or the physics of the transmitting device) and the time duration is limited (due to the need to transmit symbols at a given rate). We therefore constrain the *average* signal energy or $\bar{\mathcal{E}} = \frac{1}{2}(\mathcal{E}_0+\mathcal{E}_1)$, which assumes equal prior probabilities. Then

$$
\begin{aligned}
||\mathbf{s}_1-\mathbf{s}_0||^2 &= \mathbf{s}_1^T\mathbf{s}_1 - 2\mathbf{s}_1^T\mathbf{s}_0 + \mathbf{s}_0^T\mathbf{s}_0 \\
&= 2\bar{\mathcal{E}} - 2\mathbf{s}_1^T\mathbf{s}_0 \\
&= 2\bar{\mathcal{E}}(1-\rho_s)
\end{aligned}
$$

where we define ρ_s as

$$\rho_s = \frac{\mathbf{s}_1^T\mathbf{s}_0}{\frac{1}{2}(\mathbf{s}_1^T\mathbf{s}_1+\mathbf{s}_0^T\mathbf{s}_0)}. \tag{4.24}$$

Note that the interpretation of ρ_s is as a *signal correlation coefficient* since $|\rho_s| \le 1$ (see Problem 4.20). For instance, if $\mathbf{s}_1^T\mathbf{s}_0 = 0$ or the signal vectors are orthogonal, then $\rho_s = 0$. Thus, the performance of a binary communication system is given by

$$P_e = Q\left(\sqrt{\frac{\bar{\mathcal{E}}(1-\rho_s)}{2\sigma^2}}\right). \tag{4.25}$$

To minimize P_e subject to an average energy constraint, we should choose the signals to ensure $\rho_s = -1$. Some examples follow.

Example 4.7 - Phase Shift Keying

In phase shift keying (PSK) (also called *coherent* PSK since we assume perfect knowledge of the signal) we transmit a sinusoid with one of two phases or

$$
\begin{aligned}
s_0[n] &= A\cos 2\pi f_0 n \\
s_1[n] &= A\cos(2\pi f_0 n + \pi) = -A\cos 2\pi f_0 n
\end{aligned}
$$

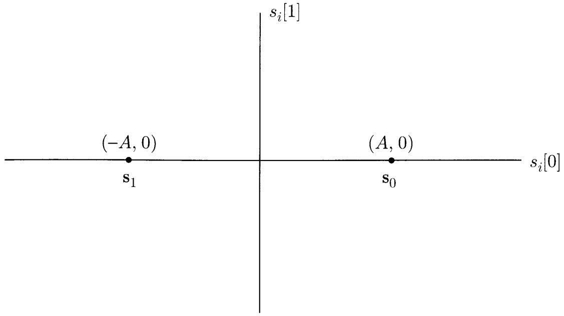


Figure 4.11. Signal representation for coherent PSK ($f_0 = 0.25$).

for $n = 0, 1, \ldots, N-1$ so that $\mathbf{s}_1 = -\mathbf{s}_0$. For $N = 2$ the signals appear as shown in Figure 4.11 for $f_0 = 0.25$. It is easily shown that $\rho_s = -1$ so that P_e is minimized. These signals are said to be *antipodal.* Note that each signal has the same energy $\mathcal{E} \approx NA^2/2$ so that $\bar{\mathcal{E}} = \mathcal{E}$. From (4.25) we have $P_e = Q(\sqrt{\mathcal{E}/\sigma^2})$, which is plotted in Figure 4.12 versus the ENR, $\mathcal{E}/\sigma^2$. For a typical error rate of 10^{-8} we require an average ENR of 15 dB.

◇

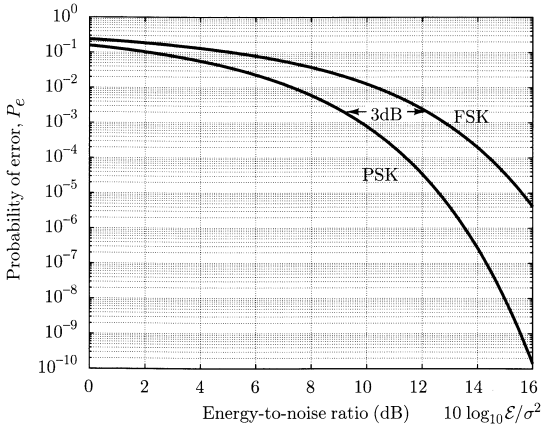


Figure 4.12. Performance of typical binary signaling schemes.

Example 4.8 - Frequency Shift Keying

In frequency shift keying (FSK) (or coherent FSK) we transmit a sinusoid with one of two frequencies or

$$\begin{aligned} s_0[n] &= A\cos 2\pi f_0 n \\ s_1[n] &= A\cos 2\pi f_1 n \end{aligned}$$

for $n = 0, 1, \ldots, N-1$. The signal correlation depends on the frequency spacing. For $|f_1 - f_0| \gg 1/(2N)$, the signals are approximately orthogonal (see Problem 4.21) and further can be shown to have approximately the same energy $\mathcal{E} \approx NA^2/2$. Thus, from (4.25) we have $P_e = Q\left(\sqrt{\mathcal{E}/(2\sigma^2)}\right)$. The performance is shown in Figure 4.12. Comparing PSK and FSK we see that the average energy of the FSK system must be twice that of the PSK system to have the same error probability. Others choices of the frequencies will yield slightly better performance (see Problem 4.22).

◇

4.5.3 M-ary Case

If we now choose to transmit one of M signals $\{s_0[n], s_1[n], \ldots, s_{M-1}[n]\}$ with equal prior probabilities, then as before we should choose $\mathcal{H}_i$ for which $p(\mathbf{x}|\mathcal{H}_i)$ is maximum. The optimal receiver is again a minimum distance receiver and thus we choose $\mathcal{H}_k$ if

$$T_k(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]s_k[n] - \frac{1}{2}\mathcal{E}_k \tag{4.26}$$

is the maximum statistic of $\{T_0(\mathbf{x}), T_1(\mathbf{x}), \ldots, T_{M-1}(\mathbf{x})\}$. The optimal receiver is shown in Figure 4.13.

To determine the error probability in general is difficult. This is because an error occurs if any of the $M-1$ statistics exceeds the one associated with the true hypothesis. Alternatively, an error occurs if the maximum of the other $M-1$ statistics exceeds the true one. Finding the PDF of the maximum of a number of statistics is a problem in *order statistics* [Kendall and Stuart 1979]. For nonindependent random variables this is a mathematically intractable problem. For independent random variables, on the other hand, one can find the PDF quite readily. In the current problem, if the signals are *orthogonal,* then the statistics will be independent. This is because, conditioned on any hypothesis, the statistics are jointly Gaussian random variables, and with the orthogonality assumption, they are uncorrelated and hence independent. To verify this we have from (4.26) that conditioned on any hypothesis say $\mathcal{H}_l$

$$\operatorname{cov}(T_i, T_j|\mathcal{H}_l) = E\left(\sum_{m=0}^{N-1} w[m]s_i[m] \sum_{n=0}^{N-1} w[n]s_j[n]\right)$$

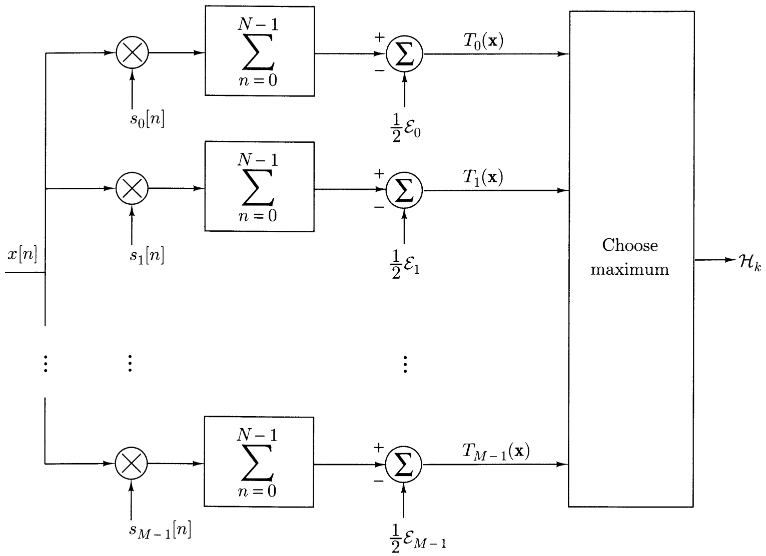


Figure 4.13. Minimum probability of error receiver for M-ary detection.

$$\begin{aligned} &= \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} E(w[m]w[n])s_i[m]s_j[n] \\ &= \sigma^2 \sum_{n=0}^{N-1} s_i[n]s_j[n] = 0 \qquad \text{for } i \neq j \end{aligned}$$

where the last step follows from the signal orthogonality assumption. Now, to simplify matters assume *equal signal energies* so that $\mathcal{E}_i = \mathcal{E}$. We note that an error is committed if $\mathcal{H}_i$ is the true hypothesis, but T_i is not the maximum. Therefore

$$P_e = \sum_{i=0}^{M-1} \Pr\{T_i < \max(T_0, \dots, T_{i-1}, T_{i+1}, \dots, T_{M-1}) | \mathcal{H}_i\} \, P(\mathcal{H}_i).$$

By symmetry all of the conditional error probabilities in the above sum are the same (see also Figure 4.10 for the $M = 2$ case) and hence

$$P_e = \Pr\{T_0 < \max(T_1, T_2, \dots, T_{M-1}) | \mathcal{H}_0\}.$$

But conditioned on $\mathcal{H}_0$

$$\begin{aligned} T_i(\mathbf{x}) &= \sum_{n=0}^{N-1} x[n]s_i[n] - \frac{1}{2}\mathcal{E} \\ &\sim \begin{cases} \mathcal{N}(\frac{1}{2}\mathcal{E}, \sigma^2\mathcal{E}) & \text{for } i = 0 \\ \mathcal{N}(-\frac{1}{2}\mathcal{E}, \sigma^2\mathcal{E}) & \text{for } i \neq 0 \end{cases} \end{aligned} \tag{4.27}$$

as may readily be verified. Thus

$$\begin{aligned} P_e &= 1 - \Pr\{T_0 > \max(T_1, T_2, \dots, T_{M-1}) | \mathcal{H}_0\} \\ &= 1 - \Pr\{T_1 < T_0, T_2 < T_0, \dots, T_{M-1} < T_0 | \mathcal{H}_0\} \\ &= 1 - \int_{-\infty}^{\infty} \Pr\{T_1 < t, T_2 < t, \dots, T_{M-1} < t | T_0 = t, \mathcal{H}_0\} p_{T_0}(t) dt \\ &= 1 - \int_{-\infty}^{\infty} \prod_{i=1}^{M-1} \Pr\{T_i < t | \mathcal{H}_0\} p_{T_0}(t) dt \end{aligned}$$

where the last step follows from the independence of the T_i's. Now from (4.27) we have

$$P_e = 1 - \int_{-\infty}^{\infty} \Phi^{M-1}\left(\frac{t + \frac{1}{2}\mathcal{E}}{\sqrt{\sigma^2\mathcal{E}}}\right) \frac{1}{\sqrt{2\pi\sigma^2\mathcal{E}}} \exp\left[-\frac{1}{2\sigma^2\mathcal{E}}\left(t - \frac{1}{2}\mathcal{E}\right)^2\right] dt$$

where $\Phi(\cdot)$ is the CDF for a $\mathcal{N}(0,1)$ random variable. Letting $u = (t + \frac{1}{2}\mathcal{E})/\sqrt{\sigma^2\mathcal{E}}$ the probability of error can finally be expressed as

$$P_e = 1 - \int_{-\infty}^{\infty} \Phi^{M-1}(u) \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(u - \sqrt{\frac{\mathcal{E}}{\sigma^2}}\right)^2\right] du \tag{4.28}$$

and is seen to depend on the ENR, $\mathcal{E}/\sigma^2$. This is plotted in Figure 4.14 for various values of M. Note that as M increases so does the error probability. This occurs because the receiver must distinguish between more signals whose intersignal distances do not increase. To see this we must first realize that for M orthogonal signals we require $N \geq M$. As an example, for $N = M = 2$ and $N = M = 3$, we have the signal space shown in Figure 4.15. The distances between the signals are the same for $M = 2$ and $M = 3$, since the signal energy does not increase with increasing N. In Figure 4.15 each signal has energy $\mathcal{E} = 1$. As M increases, we must choose from among a larger set of signals and therefore, P_e must increase with M. We will return to this example in Section 4.7 when we consider the communication problem in more detail.

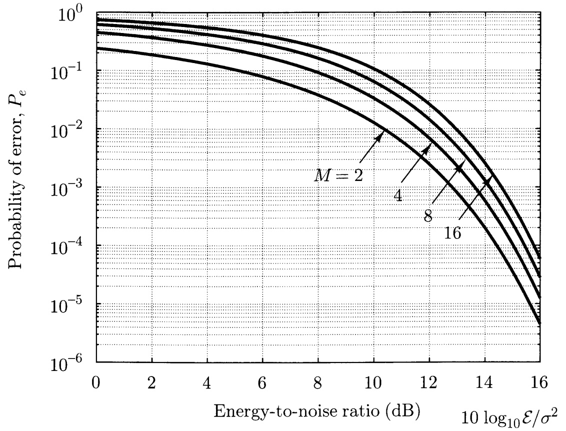
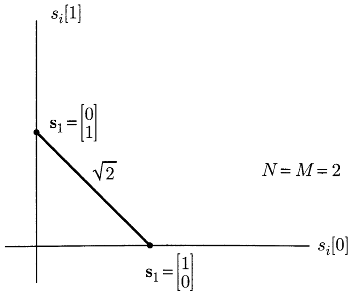
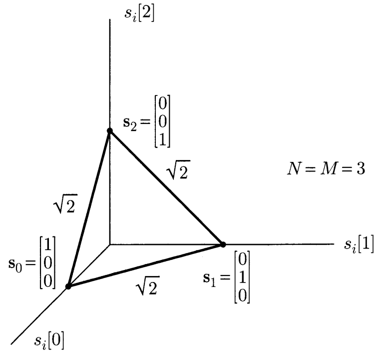


Figure 4.14. Performance of M-ary orthogonal signal communication system.



(a)



(b)

Figure 4.15. Illustration of increasing probability of error with increasing number of orthogonal signals.

4.6 Linear Model

In [Kay-I 1993, Chapter 4] the linear model was introduced. Owing to its mathematical simplicity and general applicability to real-world problems, it is a frequently

used model. To briefly review, the classical general linear model assumes that the data vector $\mathbf{x}$ can be written as

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

where $\mathbf{x}$ is an $N \times 1$ vector of received data samples or $\mathbf{x} = [x[0]\,x[1]\dots x[N-1]]^T$, $\mathbf{H}$ is a known $N \times p$ full rank matrix with $N > p$ and is termed the observation matrix, $\boldsymbol{\theta}$ is a $p \times 1$ vector of parameters, which may or may not be known, and $\mathbf{w}$ is an $N \times 1$ random vector or $\mathbf{w} = [w[0]\,w[1]\dots w[N-1]]^T$ with PDF $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$. For the present discussion $\boldsymbol{\theta}$ is assumed known under $\mathcal{H}_1$ with its value denoted by $\boldsymbol{\theta}_1$. We may view $\mathbf{s} = \mathbf{H}\boldsymbol{\theta}_1$ as a known deterministic signal. Under $\mathcal{H}_0$ we have that $\boldsymbol{\theta} = \mathbf{0}$ so that there is no signal present. Later we will address the detection problem under the assumption that $\boldsymbol{\theta}$ is unknown (see Chapter 7). Many examples of this type of modeling have been given in [Kay-I 1993]. A simple example is the DC level in WGN, which is a special case of the linear model with

$$\begin{aligned}
\mathbf{H} &= [1\,1\dots 1]^T \\
\boldsymbol{\theta}_1 &= A \\
\mathbf{C} &= \sigma^2\mathbf{I}.
\end{aligned}$$

In applying the linear model to the detection problem we must decide if $\mathbf{s} = \mathbf{H}\boldsymbol{\theta}_1$ is present or not. Hence, we have

$$\begin{aligned}
&\mathcal{H}_0 : \mathbf{x} = \mathbf{w} \\
&\mathcal{H}_1 : \mathbf{x} = \mathbf{H}\boldsymbol{\theta}_1 + \mathbf{w}.
\end{aligned}$$

The NP detector immediately follows by letting $\mathbf{s} = \mathbf{H}\boldsymbol{\theta}_1$ in the detector of (4.16). We then decide $\mathcal{H}_1$ if

$$\begin{aligned}
T(\mathbf{x}) &= \mathbf{x}^T\mathbf{C}^{-1}\mathbf{s} \\
&= \mathbf{x}^T\mathbf{C}^{-1}\mathbf{H}\boldsymbol{\theta}_1 > \gamma'.
\end{aligned} \tag{4.29}$$

The performance is given by (4.18) with the appropriate substitution for $\mathbf{s}$. An example illustrates the detector.

Example 4.9 - Sinusoidal Detection

Assume that we wish to detect a sinusoid $s[n] = A\cos(2\pi f_0 n + \phi)$ in WGN. Rewriting $s[n]$ as $s[n] = A\cos\phi\cos 2\pi f_0 n - A\sin\phi\sin 2\pi f_0 n$ and letting $a = A\cos\phi$ and $b = -A\sin\phi$, we then have the detection problem

$$\begin{aligned}
&\mathcal{H}_0 : x[n] = w[n] && n = 0, 1, \dots, N-1 \\
&\mathcal{H}_1 : x[n] = a\cos 2\pi f_0 n + b\sin 2\pi f_0 n + w[n] && n = 0, 1, \dots, N-1
\end{aligned}$$

where $w[n]$ is WGN with variance σ^2. Under $\mathcal{H}_1$ we have the linear model with

$$\begin{aligned}
\mathbf{H} &= \begin{bmatrix} 1 & 0 \\ \cos 2\pi f_0 & \sin 2\pi f_0 \\ \vdots & \vdots \\ \cos[2\pi f_0(N-1)] & \sin[2\pi f_0(N-1)] \end{bmatrix} \\
\boldsymbol{\theta}_1 &= \begin{bmatrix} a \\ b \end{bmatrix} \\
\mathbf{C} &= \sigma^2\mathbf{I}.
\end{aligned}$$

Note that $\mathbf{H}$ is full rank for $0 < f_0 < 1/2$ since the columns are linearly independent (see Problem 4.26). The NP test statistic from (4.29) is

$$T(\mathbf{x}) = \frac{1}{\sigma^2}\mathbf{x}^T\mathbf{H}\boldsymbol{\theta}_1$$

or scaling the statistic we have

$$T'(\mathbf{x}) = \frac{2}{N}\mathbf{x}^T\mathbf{H}\boldsymbol{\theta}_1 = \left(\frac{2}{N}\mathbf{H}^T\mathbf{x}\right)^T\boldsymbol{\theta}_1.$$

But

$$\frac{2}{N}\mathbf{H}^T\mathbf{x} = \begin{bmatrix} \frac{2}{N}\sum_{n=0}^{N-1} x[n]\cos 2\pi f_0 n \\ \frac{2}{N}\sum_{n=0}^{N-1} x[n]\sin 2\pi f_0 n \end{bmatrix}$$

which is recognized as an estimator of $[a\, b]^T$ (see Problem 4.27 and [Kay-I 1993, pp. 88–90]). Letting the estimator be $\hat{\boldsymbol{\theta}} = [\hat{a}\,\hat{b}]^T$, we have

$$T'(\mathbf{x}) = \hat{a}a + \hat{b}b.$$

This is just a correlation of the true $\boldsymbol{\theta}$ under $\mathcal{H}_1$, which is $\boldsymbol{\theta}_1 = [a\, b]^T$, with the estimate $\hat{\boldsymbol{\theta}}$. Clearly, if $\mathcal{H}_0$ is true, $\hat{a} \approx E(\hat{a}) \approx 0$ and $\hat{b} \approx E(\hat{b}) \approx 0$ so that $T'(\mathbf{x}) \approx 0$. If, however, $\mathcal{H}_1$ is true, then $\hat{a} \approx E(\hat{a}) \approx a$ and $\hat{b} \approx E(\hat{b}) \approx b$ so that $T'(\mathbf{x}) \approx a^2 + b^2$. (If $f_0 = k/N$ for $k = 1, 2, \ldots, N/2 - 1$, then $E(\hat{a}) = a$ and $E(\hat{b}) = b$). The latter is proportional to signal power. In summary, for the linear model we correlate the $\hat{\boldsymbol{\theta}}$ parameter estimate (the "data") with the true value of $\boldsymbol{\theta}$ under $\mathcal{H}_1$ (the "signal").

◇

To generalize these results we recall that the minimum variance unbiased (MVU) estimator of $\boldsymbol{\theta}$ for the general linear model is [Kay-I 1993, pg. 97]

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x}.$$

Then, using (4.29) we have

$$\begin{aligned}
T(\mathbf{x}) &= \mathbf{x}^T\mathbf{C}^{-1}\mathbf{H}\boldsymbol{\theta}_1 \\
&= \left[(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x}\right]^T\boldsymbol{\theta}_1 \\
&= \left[(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})\hat{\boldsymbol{\theta}}\right]^T\boldsymbol{\theta}_1 \\
&= \hat{\boldsymbol{\theta}}^T(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})\boldsymbol{\theta}_1.
\end{aligned}$$

Also, recall that the covariance matrix of the MVU estimator $\hat{\boldsymbol{\theta}}$ is $\mathbf{C}_{\hat{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}$. Hence, we decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \hat{\boldsymbol{\theta}}^T\mathbf{C}_{\hat{\theta}}^{-1}\boldsymbol{\theta}_1 > \gamma'. \tag{4.30}$$

Referring to (4.16) we now observe that the test statistic is identical to that obtained for the detection of a known signal $\mathbf{s}$ in correlated noise with covariance $\mathbf{C}$ if we make the correspondences

$$\begin{aligned}
\mathbf{x} &\to \hat{\boldsymbol{\theta}} \\
\mathbf{C} &\to \mathbf{C}_{\hat{\theta}} \\
\mathbf{s} &\to \boldsymbol{\theta}_1.
\end{aligned}$$

This is not merely a coincidence but stems from the properties of the linear model. In Appendix 4A we explore this further. Note also that the signal detection problem is equivalent to

$$\begin{aligned}
\mathcal{H}_0 &: \boldsymbol{\theta} = \boldsymbol{\theta}_0 = \mathbf{0} \\
\mathcal{H}_1 &: \boldsymbol{\theta} = \boldsymbol{\theta}_1
\end{aligned}$$

where $\mathbf{s} = \mathbf{H}\boldsymbol{\theta}_1$. This is just a parameter test of the PDF.

Finally, the detection performance is easily found from (4.18) as

$$P_D = Q\left(Q^{-1}(P_{FA}) - \sqrt{\boldsymbol{\theta}_1^T\mathbf{C}_{\hat{\theta}}^{-1}\boldsymbol{\theta}_1}\right). \tag{4.31}$$

4.7 Signal Processing Examples

We now apply some of the previous results to the communication problem and the general problem of pattern recognition.

Example 4.10 - Channel Capacity in Communications

In this example we explore the use of block coding to show how nearly errorless digital communication is possible. A fundamental result is the concept of *channel capacity*, which is the maximum rate in bit/sec for errorless communication. Consider a stream of binary digits that are to be transmitted, with one bit occuring every T seconds. A binary signaling scheme would transmit one of two possible signals every T seconds. For example, the coherent PSK format would transmit $s_0(t) = A\cos 2\pi F_0 t$ for a "0" and $s_1(t) = -A\cos 2\pi F_0 t$ for a "1". (Here we use the capital letter F to denote continuous-time frequency or frequency in Hz.) Instead, let us consider what happens if we store L bits, which represent one of 2^L different messages, and then transmit one of 2^L signals. For example, if $L = 2$ we store 2 bits and then transmit one of the four signals $s_i(t) = A\cos 2\pi F_i t$ for $i = 0, 1, 2, 3$, as in an FSK scheme for example. The receiver must then decide which one of the $M = 2^L$ signals was sent. Since we must wait for the L bits to accumulate before a transmission can be made, the signal that is transmitted can now be LT seconds long. We will assume that the signals are orthogonal, as for example, in an M-ary FSK communication system. In such a system the signals will need to be spaced at least $1/LT$ Hz apart in frequency to satisfy the orthogonality assumption. This requires approximately $2^L/LT$ Hz of bandwidth.

We have already derived the error probability for M equal energy orthogonal signals as given by (4.28). It should be emphasized that P_e refers to the error in detecting which one of the M signals was sent or equivalently which one of the 2^L bit sequences was encoded. Thus, P_e is the *message error* probability, where the message consists of L bits, *not* the *bit error* probability. Now, the energy of the transmitted continuous-time signal will be L times that of the binary scheme. Thus, if the available transmitter power is P watts, the energy of the transmitted continuous-time signal is LPT. In the discrete-time domain this yields approximately

$$\mathcal{E} = \sum_{n=0}^{N-1} s^2(n\Delta) \approx \frac{LPT}{\Delta}$$

where $1/\Delta$ is the sampling rate. The probability of message error is then from (4.28)

$$P_e = 1 - \int_{-\infty}^{\infty} \Phi^{M-1}(u) \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(u - \sqrt{\frac{LPT}{\Delta\sigma^2}}\right)^2\right] du. \tag{4.32}$$

We plot this versus the ENR per bit or $\mathcal{E}/(\sigma^2 L) = PT/(\Delta\sigma^2)$. In Figure 4.16 we see that as L increases, P_e decreases, as long as the ENR per bit exceeds some threshold. In fact as $L \to \infty$, $P_e \to 0$ if we are above the threshold and $P_e \to 1$ if we are below the threshold. The threshold can be shown to be (see Problem 4.29)

$$\frac{PT}{\Delta\sigma^2} = 2\ln 2 \tag{4.33}$$

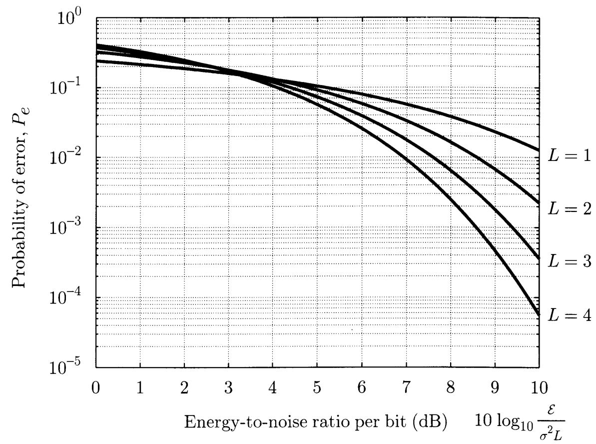


Figure 4.16. Probability of error for block encoding.

and hence we require

$$\frac{PT}{\Delta\sigma^2} > 2\ln 2$$

in order to drive the error probability to zero. Letting $R = 1/T$ be the transmission rate in bits/sec, we must ensure that the transmission rate satisfies

$$R < \frac{P}{2\sigma^2\Delta\ln 2} \quad \text{bits/sec.}$$

The upper limit on transmission rate is called the *channel capacity*, denoted by C_∞, and is seen to depend on SNR. The SNR is $P/(\sigma^2\Delta)$, where we note that the frequency band is assumed to be $[-1/(2\Delta), 1/(2\Delta)]$ Hz. In fact,

$$C_\infty = \frac{P}{2\sigma^2\Delta\ln 2}$$

is a special case of the famous Shannon channel capacity formula for a bandlimited channel [Cover and Thomas 1991]. If B denotes the low-pass bandwidth in Hz and the continuous-time noise PSD is

$$P_{ww}(F) = \begin{cases} \frac{N_0}{2} & |F| < B \\ 0 & |F| > B \end{cases},$$

then the channel capacity is

$$\begin{aligned} C &= B\log_2(1+\mathrm{SNR}) \\ &= B\log_2\left(1+\frac{P}{N_0 B}\right). \end{aligned}$$

Now as $L \to \infty$, we have that $B \to \infty$ since an orthogonal signaling scheme such as FSK requires a bandwidth of $B = 2^L/LT$ Hz. As $B \to \infty$, by L'Hospital's rule we have

$$\begin{aligned} C_\infty &= \lim_{B\to\infty} B\log_2\left(1+\frac{P}{N_0 B}\right) \\ &= \lim_{B\to\infty} \frac{\log_2\left(1+\frac{P}{N_0 B}\right)}{1/B} \\ &= \frac{P}{N_0 \ln 2}. \end{aligned}$$

Since the noise PSD is $N_0/2$ and the bandwidth is $B = 1/(2\Delta)$, the total noise power is $\sigma^2 = N_0 B = N_0/(2\Delta)$ and hence

$$C_\infty = \frac{P}{2\sigma^2 \Delta \ln 2}$$

as asserted. The subscript ∞ in C_∞ is now seen to indicate the channel capacity of an *infinite bandwidth* channel.

Intuitively, as L increases we can increase the energy of each signal since for L time slots the energy is LPT. However, as L increases the receiver must also now distinguish between more signals. For high enough SNR the improvement in ENR due to increasing L more than outweighs the detection degradation due to a larger number of signals.

◇

Example 4.11 - Pattern Recognition (Classification)

In pattern recognition one attempts to determine which pattern from a class of patterns is present [Fukunaga 1990]. For example, in computer vision applications it is important to determine the location of an object in a recorded image. If the object has a different gray level than the background, then recognition of the object can be accomplished by distinguishing between the two levels. More generally, we might wish to distinguish between M different gray levels. In a typical image the gray levels will not fall into M distinct classes but will only "on the average" differ. Thus, the measured pixel values could be modeled as a random vector because they will depend on lighting, orientation of the recording device, and many other uncontrollable

factors, as well as the innate variability of the image itself. Hence, we choose to model the pixel values as the random vector $\mathbf{x}$, where $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_i, \sigma^2\mathbf{I})$. This model will be valid for a region in the image for which the gray level is essentially constant. In practice, this assumption of *local stationarity* will be approximately satisfied if the region is small enough. The different classes are distinguished by their mean value $\boldsymbol{\mu}_i$. We term the measured pixel values the *feature vector* since it is this information that allows us to discriminate between classes. We also assume that the variability σ^2 of the measured pixel values is the same for all of the classes. Now, using a Bayesian model for the decision process and assigning prior probabilities $P(\mathcal{H}_i)$ to each of the M classes, we can design an optimal classifier. We need only employ the MAP decision rule (see Chapter 3) and assign $\mathbf{x}$ to class k if

$$P(\mathcal{H}_k|\mathbf{x}) > P(\mathcal{H}_i|\mathbf{x}) \qquad i = 0, 1, \ldots, k-1, k+1, \ldots, M-1.$$

This approach will minimize the probability of error. If the prior probabilities are equal (each gray level is equally likely), then the MAP rule becomes the ML rule (see Chapter 3), and we assign $\mathbf{x}$ to class k if $p(\mathbf{x}|\mathcal{H}_k)$ is maximum. But as was already shown in Section 4.5.1 for the $M = 2$ case, this produces a minimum distance receiver. More generally, for M classes we obtain a similar result and assign $\mathbf{x}$ to $\mathcal{H}_k$ if

$$D_k^2 = ||\mathbf{x} - \boldsymbol{\mu}_k||^2$$

is minimum over all D_i^2 for $i = 0, 1, \ldots, M-1$.

As an example, consider the synthetic image of Figure 4.17, which consists of a 50×50 pixel array of four gray levels. The levels are 1 (black), 2 (dark gray), 3 (light gray), and 4 (white). A noisy image is shown in Figure 4.18, which was obtained by adding WGN with $\sigma^2 = 0.5$ to each pixel. It is desired to classify each pixel in the image as one of the four gray levels. To do so a reasonable approach is to classify each pixel based on the observed level of that pixel and its *neighbors*. This assumes that the level does not change over a small window (local stationarity). Let $x[m,n]$ denote the observed pixel value at location $[m,n]$ (of Figure 4.18). Then, for a 3×3 window, for example, we base our decision on the data samples

$$\mathbf{x}[m,n] = \begin{bmatrix} x[m-1,n+1] & x[m,n+1] & x[m+1,n+1] \\ x[m-1,n] & x[m,n] & x[m+1,n] \\ x[m-1,n-1] & x[m,n-1] & x[m+1,n-1] \end{bmatrix}.$$

To classify pixel $[m,n]$ we must compute

$$D_i^2[m,n] = ||\mathbf{x}[m,n] - \mu_i \mathbf{1}\mathbf{1}^T||_F^2 \qquad i = 0, 1, 2, 3$$

where $\mu_i = i + 1$, $\mathbf{1} = [111]^T$, and $||\cdot||_F$ denotes the Frobenius norm of a matrix. The latter is defined as

$$||\mathbf{A}||_F^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} a^2[i,j]$$

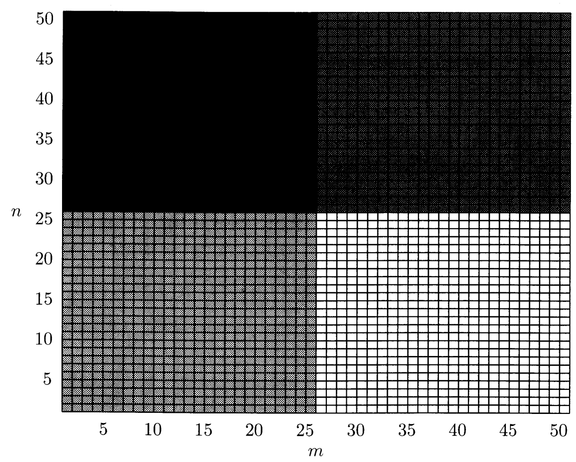


Figure 4.17. Synthetic image.

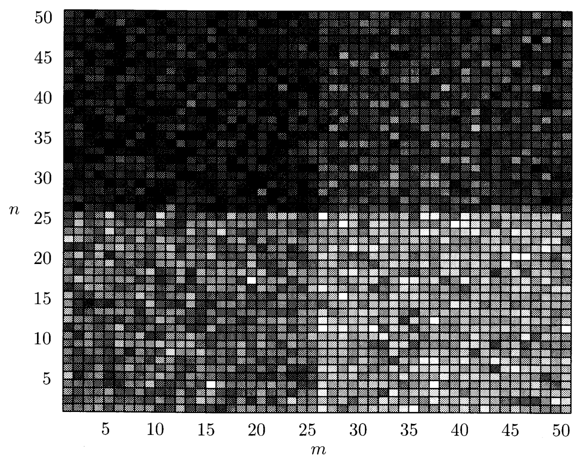


Figure 4.18. Noise corrupted image.

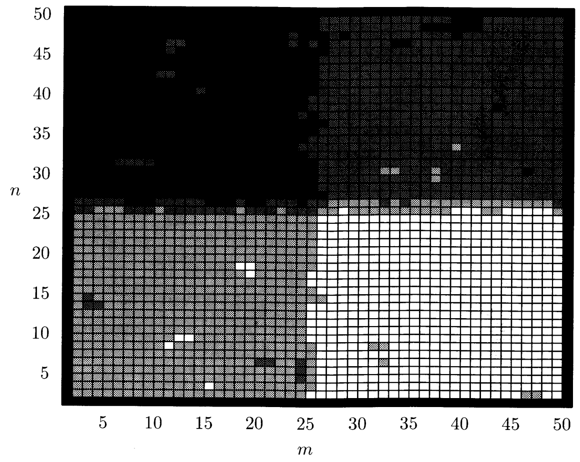


Figure 4.19. Classification using 3×3 window.

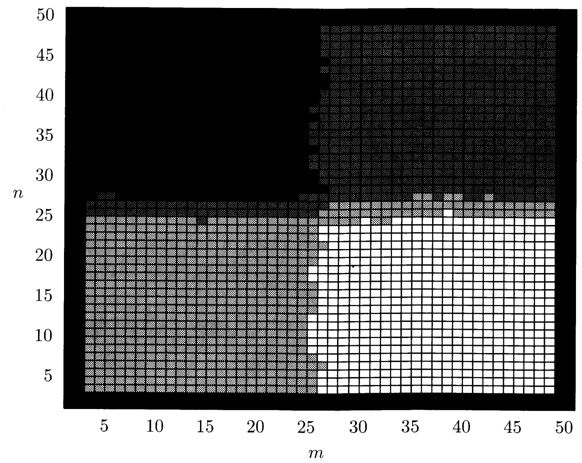


Figure 4.20. Classification using 5×5 window.

for an $N \times N$ matrix $\mathbf{A}$, where the $[i,j]$ element is $a[i,j]$. Next we assign $\mathbf{x}[m,n]$ to class k if $D_k^2[m,n]$ is minimum. The results of applying a 3×3 window are shown in Figure 4.19. The errors contained within the blocks are due to noise. The errors along the edges (boundaries between the blocks) are due to the change in mean gray level within the window. Recall that we have assumed all pixels within the window have the same mean gray level μ_i. This is not true at the boundaries of the blocks. Also, the black border, which is one pixel wide, is a result of our decision to arbitrarily classify the pixels at the edges of the image as black. This is because all of the neighbors are not available for these pixels. If we increase the window size to 5×5, then the noise errors are reduced as seen in Figure 4.20. However, the edge errors increase since the boundary between the blocks is now 2 or 3 pixels wide. Clearly, for the best "noise smoothing" the window should be large, but for good edge extraction, the window should be small. In practice, a tradeoff must be made between these two conflicting requirements.

$\diamond$

References

Cover, T.A., J.A. Thomas, *Elements of Information Theory*, J. Wiley, New York, 1991.

Fukunaga, K., *Introduction to Statistical Pattern Recognition*, 2nd ed., Academic Press, New York, 1990.

Kendall, Sir M., A. Stuart, *The Advanced Theory of Statistics*, Vol. 2, Macmillan, New York, 1979.

Problems

4.1 If $s[n] = (-1)^n$ for $n = 0,1,2,3$ and is zero otherwise, find the impulse response of the matched filter and the matched filter signal output for all time.

4.2 Using (4.8) show that the signal output of the matched filter is maximized by sampling the output at $n = N-1$.

4.3 Consider the detection of the signal $s[n] = A\cos 2\pi f_0 n$ $(0 < f_0 < 1/2)$ for $n = 0,1,\ldots,N-1$ in the presence of WGN. Find the signal output of a matched filter at time $n = N-1$. Next assume that the signal is delayed by $n_0 > 0$ samples so that we receive $s[n-n_0]$. Applying the same matched filter as before, find the signal output at $n = N-1$ as a function of n_0. You may assume that N is large enough so that the average value of a sinusoid is zero when averaged over a few periods.

4.4 Assume that we wish to detect a known signal $s[n]$ which is nonzero only for $n = 0,1,\ldots,N-1$ in WGN with variance σ^2. We now allow our observation interval to be infinite or we observe $x[n]$ for $-\infty < n < \infty$. Let $h[n]$ be the impulse response of a linear shift invariant (LSI) filter. Then, the output $y[n]$ to an input $x[n]$ is given by

$$y[n] = \sum_{k=-\infty}^{\infty} h[n-k]x[k].$$

If the filter output is sampled at $n = N-1$, the output SNR can be defined as

$$\eta = \frac{\left(\displaystyle\sum_{k=-\infty}^{\infty} h[N-1-k]s[k]\right)^2}{E\left[\left(\displaystyle\sum_{k=-\infty}^{\infty} h[N-1-k]w[k]\right)^2\right]}.$$

Show that η is maximized by choosing $h[n]$ as the matched filter or $h[n] = s[N-1-n]$ for $n = 0,1,\ldots,N-1$ and $h[n] = 0$ otherwise. Thus, the use of noise samples outside the signal interval does not improve the detection performance if the noise samples are uncorrelated. Hint: Use the assumption that $s[k] = 0$ for k not in the interval $[0, N-1]$.

4.5 In Problem 4.4 it was shown that the matched filter was optimum even for an infinite observation interval. In this problem we show that this is not generally the case when the noise is correlated. Repeat Problem 4.4 but now assume that the noise samples $\{w[0], w[1], \ldots, w[N-1]\}$ are WGN, and that outside the $[0, N-1]$ interval the noise is periodic. Thus, $w[n] = w[n+N]$. Find the output SNR η (use the expression for Problem 4.4) if the LSI filter is

$$h[n] = \begin{cases} 1 & \text{for } n = 0,1,\ldots,N-1 \\ -1 & n = N, N+1, \ldots, 2N-1 \\ 0 & \text{otherwise} \end{cases}$$

and the output is sampled at $n = N-1$. Explain your results.

4.6 It is desired to detect the known signal $s[n] = Ar^n$ for $n = 0,1,\ldots,N-1$ in WGN with variance σ^2. Find the NP detector and its detection performance. Explain what happens as $N \to \infty$ for $0 < r < 1$, $r = 1$, and $r > 1$.

4.7 A radar signal $s[n] = A\cos 2\pi f_0 n$ for $n = 0,1,\ldots,N-1$ is received embedded in WGN with variance $\sigma^2 = 1$. A detector is to be designed that maintains $P_{FA} = 10^{-8}$. If $f_0 = 0.25$ and $N = 25$, find the probability of detection versus A.

4.8 We wish to design a signal for the best detection performance in WGN. Two competing signals are proposed. They are

$$\begin{array}{ll} s_1[n] = A & n = 0, 1, \ldots, N-1 \\ s_2[n] = A(-1)^n & n = 0, 1, \ldots, N-1. \end{array}$$

Which one will yield the better detection performance?

4.9 Consider the detection of $s[n] = A\cos 2\pi f_0 n$ for $n = 0, 1, \ldots, N-1$ in the presence of WGN with variance σ^2. Define the input SNR as the average power of a signal sample to the noise power. This is approximately $\eta_{\text{in}} = (A^2/2)/\sigma^2$. Find the output SNR of a matched filter and hence the PG. Next determine the frequency response of the matched filter and plot its magnitude as N increases. Explain why the matched filter improves the detectability of a sinusoid. Assume that $0 < f_0 < 1/2$ and N is large. Hint: You will need the result that

$$\sum_{n=0}^{N-1} \exp(j\alpha n) = \exp\left[j(N-1)\alpha/2\right] \frac{\sin\frac{N\alpha}{2}}{\sin\frac{\alpha}{2}}.$$

4.10 Find the matrix prewhitener $\mathbf{D}$ for the covariance matrix

$$\mathbf{C} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

Hint: See Example 4.5 and use the eigenanalysis decomposition $\mathbf{V}^T\mathbf{C}\mathbf{V} = \mathbf{\Lambda}$, where $\mathbf{V} = [\mathbf{v}_1\, \mathbf{v}_2]$, $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2)$. Also, recall that $\mathbf{V}^T = \mathbf{V}^{-1}$.

4.11 Consider the detection of a deterministic signal $\mathbf{s}$ in noise $\mathbf{w}$ with $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$. If $\mathbf{C}^{-1} = \mathbf{D}^T\mathbf{D}$, where $\mathbf{D}$ is invertible, we can equivalently base our detector on the transformed data vector $\mathbf{y} = \mathbf{D}\mathbf{x}$. Find the NP detector by evaluating the LRT based on $\mathbf{y}$.

4.12 From Chapter 2 we know that the eigenvalues of an $N \times N$ symmetric Toeplitz covariance matrix $\mathbf{C}$ for a WSS random process $w[n]$ with PSD $P_{ww}(f)$ approach, as $N \to \infty$,

$$\lambda_i = P_{ww}\left(\frac{i}{N}\right) \qquad i = 0, 1, \ldots, N-1$$

and the corresponding eigenvectors approach

$$\mathbf{v}_i = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & \exp(j2\pi\frac{i}{N}) & \exp(j2\pi\frac{2i}{N}) & \ldots & \exp(j2\pi\frac{(N-1)i}{N}) \end{bmatrix}^T$$

for $i = 0, 1, \ldots, N-1$. Using standard properties of eigenvalues and eigenvectors show that

$$\mathbf{x}^T\mathbf{C}^{-1}\mathbf{s} \approx \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{X(f)S^*(f)}{P_{ww}(f)} df$$

for large N, where $S(f) = \sum_{n=0}^{N-1} s[n]\exp(-j2\pi fn)$ and $X(f) = \sum_{n=0}^{N-1} x[n]\exp(-j2\pi fn)$.

4.13 Using the same asymptotic properties of the eigenvalues and eigenvectors of $\mathbf{C}$ as in Problem 4.12 show that

$$d^2 = \mathbf{s}^T\mathbf{C}^{-1}\mathbf{s} \approx \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{|S(f)|^2}{P_{ww}(f)} df$$

for large N.

4.14 In this problem we determine the optimal linear filter for detecting a known signal in colored WSS Gaussian noise based on *an infinite observation interval.* Thus, we use all the noise samples from outside the signal interval. The detection problem then becomes

$$\begin{array}{rcll} \mathcal{H}_0 : x[n] & = & w[n] & -\infty < n < \infty \\ \mathcal{H}_1 : x[n] & = & \begin{cases} s[n] + w[n] & 0 \le n \le N-1 \\ w[n] & \text{otherwise} \end{cases} \end{array}$$

where $w[n]$ is Gaussian noise with a PSD $P_{ww}(f)$. The output SNR of an LSI filter at time $n = N-1$ is

$$\eta = \frac{\left(\displaystyle\sum_{k=-\infty}^{\infty} h[k]s[N-1-k]\right)^2}{E\left[\left(\displaystyle\sum_{k=-\infty}^{\infty} h[k]w[N-1-k]\right)^2\right]}.$$

Note that since the filter processes data over the infinite observation interval $-\infty < n < \infty$, it will be noncausal in general. Show that η may be written in the frequency domain as

$$\eta = \frac{\left(\displaystyle\int_{-\frac{1}{2}}^{\frac{1}{2}} H(f)S(f)\exp[j2\pi f(N-1)]df\right)^2}{\displaystyle\int_{-\frac{1}{2}}^{\frac{1}{2}} |H(f)|^2 P_{ww}(f) df}.$$

Next, use the Cauchy-Schwarz inequality (see Appendix 1) or

$$\left|\int_{-\frac{1}{2}}^{\frac{1}{2}} g(f)h(f)df\right|^2 \le \int_{-\frac{1}{2}}^{\frac{1}{2}} |g(f)|^2 df \int_{-\frac{1}{2}}^{\frac{1}{2}} |h(f)|^2 df$$

with equality if and only if $g(f) = ch^*(f)$ for c a complex constant to prove that η is maximized for

$$H(f) = \frac{S^*(f)\exp[-j2\pi f(N-1)]}{P_{ww}(f)}.$$

4.15 Find the NP detector and its performance for the detection of a known signal $s[n] = A$ for $n = 0, 1, \ldots, N-1$, where $A > 0$, in correlated Gaussian noise. The $N \times 1$ noise vector is characterized by $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$, where $\mathbf{C} = \sigma^2 \mathrm{diag}(1, r, \ldots, r^{N-1})$ and $r > 0$. What happens as $N \to \infty$?

4.16 If in Problem 4.8 we have instead colored WSS Gaussian noise with ACF $r_{ww}[k] = P + \sigma^2\delta[k]$ and $P > 0$ so that the covariance matrix is

$$\mathbf{C} = \sigma^2\mathbf{I} + P\mathbf{1}\mathbf{1}^T$$

find the better signal. Do so by determining the deflection coefficient $d^2 = \mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}$. Explain your results in terms of the PSD of the noise.

4.17 A known signal is to be detected in WSS Gaussian autoregressive noise with PSD

$$P_{ww}(f) = \frac{\sigma^2}{|1 + a\exp(-j2\pi f)|^2}.$$

Find the asymptotic NP detector (4.17) and show that it can be written approximately as

$$T(\mathbf{x}) = \frac{1}{\sigma^2}\sum_{n=1}^{N-1}(x[n] + ax[n-1])(s[n] + as[n-1])$$

where it is assumed that $x[n]$ for $n = 0, 1, \ldots, N-1$ are observed. Explain intuitively the operation of the statistic in terms of prewhitening and matched filtering (correlating). Hint: Note that the noise is whitened by the LSI filter with the system function $\mathcal{A}(z) = 1 + az^{-1}$.

4.18 For a minimum distance receiver for binary communications prove that the decision boundary in R^N is the perpendicular bisector of the line joining $\mathbf{s}_0$ and $\mathbf{s}_1$. Do this by finding the set of points for which $||\mathbf{x} - \mathbf{s}_0|| = ||\mathbf{x} - \mathbf{s}_1||$.

4.19 A binary communication system with $N = 2$ employs $\mathbf{s}_0 = [1 \ -1]^T$ and $\mathbf{s}_1 = [1 \ 1]^T$. The received signal is embedded in WGN with variance $\sigma^2 = 1$. Draw the decision regions in R^2 to minimize the probability of error. Do *not* assume that $P(\mathcal{H}_0) = P(\mathcal{H}_1)$. Explain your results.

4.20 Prove that the signal correlation coefficient defined in (4.24) satisfies $|\rho_s| \leq 1$.

4.21 Show that the signals $s_0[n] = A\cos 2\pi f_0 n$ and $s_1[n] = A\cos 2\pi f_1 n$ for $n = 0, 1, \ldots, N-1$ are approximately orthogonal if $|f_1 - f_0| \gg 1/(2N)$ and $0 < f_0 < 1/2$, $0 < f_1 < 1/2$. Hint: You can assume that N is large and you will need the result

$$\begin{aligned}
\frac{1}{N}\sum_{n=0}^{N-1}\cos\alpha n &= \frac{1}{N}\mathrm{Re}\left(\sum_{n=0}^{N-1}\exp(j\alpha n)\right) \\
&= \frac{1}{N}\mathrm{Re}\left(\exp\left[j(N-1)\alpha/2\right]\frac{\sin\frac{N\alpha}{2}}{\sin\frac{\alpha}{2}}\right) \\
&\approx \frac{\sin N\alpha}{2N\sin\frac{\alpha}{2}}.
\end{aligned}$$

4.22 For a coherent FSK system find the difference in frequencies that minimizes the probability of error. Assume that N is large. Hint: First show that

$$\rho_s \approx \frac{\sin 2\pi(f_1 - f_0)N}{2N\sin\pi(f_1 - f_0)}$$

using the hint in Problem 4.21.

4.23 An on-off keyed (OOK) communication system uses $s_0[n] = 0$ and $s_1[n] = A\cos 2\pi f_1 n$ for $n = 0, 1, \ldots, N-1$. Find the probability of error performance if the signal is embedded in WGN with variance σ^2. Compare the OOK system to coherent PSK and coherent FSK on the basis of equal peak power or the same A. Assume that N is large.

4.24 In a pulse amplitude modulation (PAM) communication system we transmit one of M levels so that

$$s_i[n] = A_i \qquad n = 0, 1, \ldots, N-1$$

for $i = 0, 1, \ldots, M-1$. If P_e is to be minimized and each signal is equally likely to be transmitted, find the optimal receiver for WGN of variance σ^2. Also, find the minimum P_e if $M = 2$. If the average signal energy is to be constrained, how should A_0 and A_1 be chosen?

4.25 Show that (4.28) reduces to

$$P_e = Q\left(\sqrt{\frac{\mathcal{E}}{2\sigma^2}}\right)$$

(which is (4.25) with $\rho_s = 0$ and $\bar{\mathcal{E}} = \mathcal{E}$) for $M = 2$. Hint: You will need to transform the variables for the double integral by letting $v = u - \sqrt{\mathcal{E}/\sigma^2}$ and also

$$\begin{bmatrix} t' \\ v' \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} t \\ v \end{bmatrix}.$$

The latter transformation has the effect of rotating the axes by 45°.

4.26 Prove that the matrix $\mathbf{H}$ of Example 4.9 is full rank.

4.27 Show that in Example 4.9 $[\hat{a}\,\hat{b}]^T$ is an estimator of $[a\,b]^T$ by determining the mean of the estimator when $\mathcal{H}_1$ is true. Assume N is large.

4.28 Assume that we wish to detect a line

$$s[n] = A + Bn \qquad n = 0, 1, \ldots, N-1$$

in WGN of variance σ^2, where A, B are known. Show that the data may be written in the form of the linear model. Next, determine the NP detector and its performance.

4.29 Verify that the threshold ENR per bit for errorless communication as $L \to \infty$ is given by (4.33). To do so first rewrite (4.32) as

$$P_e = 1 - \int_{-\infty}^{\infty} \Phi^{M-1}\left(v + \sqrt{\frac{PT\ln M}{\Delta\sigma^2 \ln 2}}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}v^2\right) dv.$$

Then, show that

$$\lim_{M\to\infty} \ln \Phi^{M-1}\left(v + \sqrt{\frac{PT\ln M}{\Delta\sigma^2 \ln 2}}\right) = \begin{cases} -\infty & \frac{PT}{\Delta\sigma^2} < 2\ln 2 \\ 0 & \frac{PT}{\Delta\sigma^2} > 2\ln 2 \end{cases}$$

and hence $P_e \to 1$ if $PT/(\Delta\sigma^2) < 2\ln 2$ and $P_e \to 0$ if $PT/(\Delta\sigma^2) > 2\ln 2$. Hint: Treat M as a continuous variable and apply L'Hospitals rule.

4.30 Prove that the $\mathbf{B}$ matrix given by (4A.1) has rank $N-1$. To do so, omit the last column to obtain the $(N-1)\times(N-1)$ matrix $\mathbf{B}' = \mathbf{I}_{N-1} - (1/N)\mathbf{1}_{N-1}\mathbf{1}_{N-1}^T$. Then show that $\det(\mathbf{B}') \neq 0$. Hint: Use the identity

$$\det(\mathbf{I}_K + \mathbf{A}_{KL}\mathbf{B}_{LK}) = \det(\mathbf{I}_L + \mathbf{B}_{LK}\mathbf{A}_{KL})$$

where the dimensions of the matrices are explicitly shown.

4.31 For the classical general linear model prove that $\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x}$ is a sufficient statistic for $\boldsymbol{\theta}$. You can apply the Neyman-Fisher factorization theorem [Kay-I 1993, pg. 117] once you have verified the following identity

$$(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{H}\boldsymbol{\theta}) = (\mathbf{x}-\mathbf{H}\hat{\boldsymbol{\theta}})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{H}\hat{\boldsymbol{\theta}}) + (\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^T\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}).$$

Appendix 4A

Reduced Form of the Linear Model

Consider the general linear model $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$. The performance of any detector is unchanged if we base it on the transformed data $\mathbf{y} = \mathbf{A}\mathbf{x}$, where $\mathbf{A}$ is an invertible $N \times N$ matrix. We let $\mathbf{A}$ be the partitioned matrix

$$\mathbf{A} = \begin{bmatrix} (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1} \\ \mathbf{B} \end{bmatrix} = \begin{bmatrix} p \times N \\ (N-p) \times N \end{bmatrix}$$

where $\mathbf{B}\mathbf{H} = \mathbf{0}$. The rows of $\mathbf{B}$ are orthogonal to the columns of $\mathbf{H}$. Since there are p linearly independent columns of $\mathbf{H}$, the rows of $\mathbf{B}$ span the $(N-p)$ dimensional complement subspace. Clearly, $\mathbf{B}$ is not unique. As an example, for the DC level in WGN we have that $\mathbf{H} = \mathbf{1}, \mathbf{C} = \sigma^2\mathbf{I}$, and $\mathbf{B}\mathbf{H} = \mathbf{0}$ implies that the sum of the elements in each row of $\mathbf{B}$ is zero. One possibility is the $(N-1)\times N$ matrix

$$\mathbf{B} = \begin{bmatrix} \frac{N-1}{N} & -\frac{1}{N} & -\frac{1}{N} & \cdots & -\frac{1}{N} \\ -\frac{1}{N} & \frac{N-1}{N} & -\frac{1}{N} & \cdots & -\frac{1}{N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -\frac{1}{N} & \cdots & -\frac{1}{N} & \frac{N-1}{N} & -\frac{1}{N} \end{bmatrix} \tag{4A.1}$$

which is the submatrix formed from the first $(N-1)$ rows of $\mathbf{I}_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^T$. (The subscripts on $\mathbf{I}$ and $\mathbf{1}$ indicate their dimension.) One can show that the rows of $\mathbf{B}$ are linearly independent (see Problem 4.30). Now we have that

$$\begin{aligned} \mathbf{y} = \mathbf{A}\mathbf{x} &= \begin{bmatrix} (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1} \\ \mathbf{B} \end{bmatrix} (\mathbf{H}\boldsymbol{\theta} + \mathbf{w}) \\ &= \begin{bmatrix} (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}(\mathbf{H}\boldsymbol{\theta} + \mathbf{w}) \\ \mathbf{B}(\mathbf{H}\boldsymbol{\theta} + \mathbf{w}) \end{bmatrix} \end{aligned}$$

$$= \begin{bmatrix} \boldsymbol{\theta} + (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{w} \\ \mathbf{B}\mathbf{w} \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}.$$

We see that the signal information (whether $\boldsymbol{\theta} = \mathbf{0}$ or $\boldsymbol{\theta} = \boldsymbol{\theta}_1$) is contained in the first p elements of $\mathbf{y}$. The remaining elements do not depend on the signal. Furthermore, the two random vectors $\mathbf{y}_1$ and $\mathbf{y}_2$ are independent since they are jointly Gaussian and

$$\begin{aligned} E\left[(\mathbf{y}_1 - \boldsymbol{\theta})\mathbf{y}_2^T\right] &= (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}E(\mathbf{w}\mathbf{w}^T)\mathbf{B}^T \\ &= (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{B}^T \\ &= (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}(\mathbf{B}\mathbf{H})^T = \mathbf{0} \end{aligned}$$

since $\mathbf{BH} = \mathbf{0}$. We may therefore discard $\mathbf{y}_2$ in any hypothesis testing problem concerning $\boldsymbol{\theta}$. Since the covariance matrix for $\mathbf{y}_1 = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x}$ is $(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}$, we have

$$\mathbf{y}_1 \sim \mathcal{N}(\boldsymbol{\theta}, (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1})$$

and we note that $\mathbf{y}_1$ is the MVU estimator of $\boldsymbol{\theta}$ and $\mathbf{C}_{\hat{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}$. If a signal is present $\boldsymbol{\theta} = \boldsymbol{\theta}_1$ and if not $\boldsymbol{\theta} = \mathbf{0}$. Our hypothesis testing problem then reduces to testing whether the mean of a Gaussian random vector is zero or not. Specifically, we have

$$\begin{aligned} \mathcal{H}_0 &: \hat{\boldsymbol{\theta}} \sim \mathcal{N}(\mathbf{0}, (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}) \\ \mathcal{H}_1 &: \hat{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}_1, (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}). \end{aligned}$$

This is the detection problem for a known signal in correlated Gaussian noise with covariance matrix $(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}$. The NP detector is given by (4.16). With the appropriate substitutions we decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \hat{\boldsymbol{\theta}}^T\mathbf{C}_{\hat{\theta}}^{-1}\boldsymbol{\theta}_1 > \gamma'$$

which agrees with (4.30). In summary, for testing hypotheses about the $\boldsymbol{\theta}$ parameter of a linear model we may replace the N data samples $\mathbf{x}$ by the p data samples $\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x}$ and use the PDF for $\hat{\boldsymbol{\theta}}$ or $\hat{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}, (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1})$. Essentially, $\hat{\boldsymbol{\theta}}$ is a sufficient statistic for $\boldsymbol{\theta}$ (see Problem 4.31) so that there is no loss of information. Any decision problem can be based on $\hat{\boldsymbol{\theta}}$ with no degradation in performance [Kendall and Stuart 1979].

Chapter 5

Random Signals

5.1 Introduction

In the previous chapter we were able to detect signals in the presence of noise by detecting the change in the *mean* of a test statistic. This was because the signal was assumed deterministic, and hence its presence altered the mean of the received data. In some cases a signal is more appropriately modeled as a random process. An example is speech, for which the waveform of a given sound depends on the identity of the speaker, the context in which the sound is spoken, the health of the speaker, etc. It is therefore, unrealistic to assume that the signal is known. A better approach is to assume that the signal is a random process with a known covariance structure. In this chapter we examine the optimal detectors that result from the random signal model.

5.2 Summary

The NP detector for a zero mean, white Gaussian signal in WGN is the energy detector which is given by (5.1). Its performance is summarized by (5.2) and (5.3). Generalizing the signal to allow an arbitrary covariance matrix leads to the estimator-correlator of (5.5) and (5.6). An eigendecomposition of the signal covariance matrix reduces the detector to the canonical form of (5.9). The detection performance may be determined from (5.10) and (5.11). When the signal can be modeled using the Bayesian linear model, the estimator-correlator becomes (5.18). An important special case is the Rayleigh fading sinusoid. The NP detector is the periodogram statistic, which is given by (5.20). Its detection performance is succinctly given by (5.23). For FSK communication through a Rayleigh fading channel the optimal receiver is given by (5.24). The probability of error of (5.25) indicates the large degradation due to fading. If the signal to be detected can be modeled as a zero mean, WSS Gaussian random process, then the estimator-correlator can be approximated by (5.27) for large data records. The most general detector for a

Gaussian signal allows a nonzero signal mean. This assumption results in the NP detector for the general Gaussian problem as given by (5.30). Finally, an example of the modeling of a fading channel for wideband signals is described in Section 5.7. The model is a tapped delay line with random weights. For a pseudorandom noise input signal the NP detector becomes the incoherent multipath combiner of (5.35).

5.3 Estimator-Correlator

We will assume the signal to be a zero mean Gaussian random process with a known covariance. The noise is assumed to be WGN with known variance σ^2 and to be independent of the signal. Later, we will generalize our results to the case of a signal with a nonzero mean and noise with an arbitrary covariance matrix. An example of a simple detector for this type of signal modeling follows.

Example 5.1 - Energy Detector

We model the signal as a zero mean, white, WSS Gaussian random process with variance σ_s^2. Or we say $s[n]$ is WGN (although the term "noise" is actually a misnomer). We assume the noise $w[n]$ is WGN with variance σ^2 and is independent of the signal. The detection problem is to distinguish between the hypotheses

$$\begin{array}{lll}\mathcal{H}_0: x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1: x[n] = s[n] + w[n] & n = 0, 1, \ldots, N-1.\end{array}$$

A NP detector decides $\mathcal{H}_1$ if the likelihood ratio exceeds a threshold or if

$$L(\mathbf{x}) = \frac{p(\mathbf{x}; \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > \gamma.$$

But from our modeling assumptions $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ under $\mathcal{H}_0$ and thus, $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, (\sigma_s^2 + \sigma^2)\mathbf{I})$ under $\mathcal{H}_1$. As a result, we have

$$L(\mathbf{x}) = \frac{\frac{1}{[2\pi(\sigma_s^2 + \sigma^2)]^{\frac{N}{2}}} \exp\left[-\frac{1}{2(\sigma_s^2 + \sigma^2)} \sum_{n=0}^{N-1} x^2[n]\right]}{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n]\right]}$$

so that the log-likelihood ratio (LLR) becomes

$$\begin{aligned} l(\mathbf{x}) &= \frac{N}{2} \ln\left(\frac{\sigma^2}{\sigma_s^2 + \sigma^2}\right) - \frac{1}{2}\left(\frac{1}{\sigma_s^2 + \sigma^2} - \frac{1}{\sigma^2}\right) \sum_{n=0}^{N-1} x^2[n] \\ &= \frac{N}{2} \ln\left(\frac{\sigma^2}{\sigma_s^2 + \sigma^2}\right) + \frac{1}{2} \frac{\sigma_s^2}{\sigma^2(\sigma_s^2 + \sigma^2)} \sum_{n=0}^{N-1} x^2[n]. \end{aligned}$$

Hence, we decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} x^2[n] > \gamma'. \tag{5.1}$$

The NP detector computes the *energy* in the received data and compares it to a threshold. Hence, it is known as an *energy detector*. Intuitively, if the signal is present, the energy of the received data increases. In fact, the equivalent test statistic $T'(\mathbf{x}) = (1/N)\sum_{n=0}^{N-1} x^2[n]$ can be thought of as an estimator of the variance. Comparing this to a threshold recognizes that the variance under $\mathcal{H}_0$ is σ^2 but under $\mathcal{H}_1$ it increases to $\sigma_s^2 + \sigma^2$.

The detection performance can be found by noting that (see Chapter 2)

$$\begin{array}{ll} \dfrac{T(\mathbf{x})}{\sigma^2} \sim \chi_N^2 & \text{under } \mathcal{H}_0 \\ \dfrac{T(\mathbf{x})}{\sigma_s^2 + \sigma^2} \sim \chi_N^2 & \text{under } \mathcal{H}_1 \end{array}$$

the statistic being the sum of the squares of N IID Gaussian random variables. To find P_{FA} and P_D we recall that the right-tail probability for a χ_ν^2 random variable or (see Chapter 2)

$$Q_{\chi_\nu^2}(x) = \int_x^\infty p(t)dt$$

is

$$Q_{\chi_\nu^2}(x) = \begin{cases} 2Q(\sqrt{x}) & \nu = 1 \\ 2Q(\sqrt{x}) + \dfrac{\exp(-\frac{1}{2}x)}{\sqrt{\pi}} \displaystyle\sum_{k=1}^{\frac{\nu-1}{2}} \frac{(k-1)!(2x)^{k-\frac{1}{2}}}{(2k-1)!} & \nu > 1 \text{ and } \nu \text{ odd} \\ \exp(-\frac{1}{2}x) \displaystyle\sum_{k=0}^{\frac{\nu}{2}-1} \frac{(\frac{x}{2})^k}{k!} & \nu \text{ even.} \end{cases}$$

Hence from (5.1)

$$\begin{aligned} P_{FA} &= \Pr\{T(\mathbf{x}) > \gamma'; \mathcal{H}_0\} \\ &= \Pr\left\{\frac{T(\mathbf{x})}{\sigma^2} > \frac{\gamma'}{\sigma^2}; \mathcal{H}_0\right\} \\ &= Q_{\chi_N^2}\left(\frac{\gamma'}{\sigma^2}\right) \end{aligned} \tag{5.2}$$

and

$$\begin{aligned} P_D &= \Pr\{T(\mathbf{x}) > \gamma'; \mathcal{H}_1\} \\ &= Q_{\chi_N^2}\left(\frac{\gamma'}{\sigma_s^2 + \sigma^2}\right). \end{aligned} \tag{5.3}$$

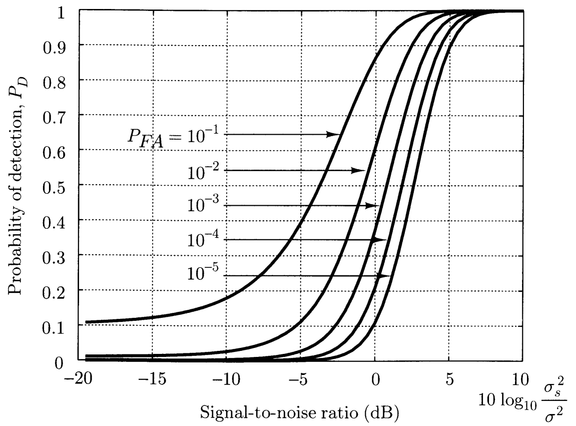


Figure 5.1. Energy detector performance ($N = 25$).

It should be noted that the threshold can be found from (5.2) as explained in Problem 5.1. Also, the detection performance increases monotonically with SNR, defined as σ_s^2/σ^2. To see this assume that for a given P_{FA} the argument in (5.2) is $\gamma'/\sigma^2 = \gamma''$. Then, from (5.3)

$$\begin{aligned} P_D &= Q_{\chi^2_N}\left(\frac{\gamma'/\sigma^2}{\sigma_s^2/\sigma^2+1}\right) \\ &= Q_{\chi^2_N}\left(\frac{\gamma''}{\sigma_s^2/\sigma^2+1}\right) \end{aligned}$$

and as σ_s^2/σ^2 increases, the argument of the $Q_{\chi^2_N}$ function decreases and thus P_D increases. The detection performance of the energy detector is given in Figure 5.1 for $N = 25$ (see also Problem 5.2 for an approximation).

◇

We now generalize the energy detector to signals with arbitrary covariance matrices. To this end we assume $s[n]$ is a zero mean, Gaussian random process with covariance matrix $\mathbf{C}_s$. As before, $w[n]$ is WGN with variance σ^2. Hence

$$\mathbf{x} \sim \begin{cases} \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(\mathbf{0}, \mathbf{C}_s + \sigma^2\mathbf{I}) & \text{under } \mathcal{H}_1. \end{cases}$$

The NP detector decides $\mathcal{H}_1$ if

$$L(\mathbf{x}) = \frac{\dfrac{1}{(2\pi)^{\frac{N}{2}}\det^{\frac{1}{2}}(\mathbf{C}_s+\sigma^2\mathbf{I})}\exp\left[-\dfrac{1}{2}\mathbf{x}^T(\mathbf{C}_s+\sigma^2\mathbf{I})^{-1}\mathbf{x}\right]}{\dfrac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left(-\dfrac{1}{2\sigma^2}\mathbf{x}^T\mathbf{x}\right)} > \gamma.$$

Taking logarithms and retaining only the data-dependent terms yields

$$-\frac{1}{2}\mathbf{x}^T\left[(\mathbf{C}_s+\sigma^2\mathbf{I})^{-1} - \frac{1}{\sigma^2}\mathbf{I}\right]\mathbf{x} > \gamma'$$

or

$$T(\mathbf{x}) = \sigma^2\mathbf{x}^T\left[\frac{1}{\sigma^2}\mathbf{I} - (\mathbf{C}_s+\sigma^2\mathbf{I})^{-1}\right]\mathbf{x} > 2\gamma'\sigma^2.$$

Using the matrix inversion lemma (see Appendix 1)

$$(\mathbf{A}+\mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{DA}^{-1}\mathbf{B}+\mathbf{C}^{-1})^{-1}\mathbf{DA}^{-1}$$

we have upon letting $\mathbf{A} = \sigma^2\mathbf{I}$, $\mathbf{B} = \mathbf{D} = \mathbf{I}$, $\mathbf{C} = \mathbf{C}_s$

$$(\sigma^2\mathbf{I}+\mathbf{C}_s)^{-1} = \frac{1}{\sigma^2}\mathbf{I} - \frac{1}{\sigma^4}\left(\frac{1}{\sigma^2}\mathbf{I}+\mathbf{C}_s^{-1}\right)^{-1} \tag{5.4}$$

so that

$$T(\mathbf{x}) = \mathbf{x}^T\left[\frac{1}{\sigma^2}\left(\frac{1}{\sigma^2}\mathbf{I}+\mathbf{C}_s^{-1}\right)^{-1}\right]\mathbf{x}.$$

Now let $\hat{\mathbf{s}} = (1/\sigma^2)((1/\sigma^2)\mathbf{I}+\mathbf{C}_s^{-1})^{-1}\mathbf{x}$. This may be also written as

$$\begin{aligned} \hat{\mathbf{s}} &= \frac{1}{\sigma^2}\left(\frac{1}{\sigma^2}\mathbf{I}+\mathbf{C}_s^{-1}\right)^{-1}\mathbf{x} \\ &= \frac{1}{\sigma^2}\left[\frac{1}{\sigma^2}(\mathbf{C}_s+\sigma^2\mathbf{I})\mathbf{C}_s^{-1}\right]^{-1}\mathbf{x} \\ &= \mathbf{C}_s(\mathbf{C}_s+\sigma^2\mathbf{I})^{-1}\mathbf{x}. \end{aligned}$$

Hence, we decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \mathbf{x}^T\hat{\mathbf{s}} > \gamma'' \tag{5.5}$$

or

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]\hat{s}[n]$$

where

$$\hat{\mathbf{s}} = \mathbf{C}_s(\mathbf{C}_s + \sigma^2\mathbf{I})^{-1}\mathbf{x}. \tag{5.6}$$

The NP detector correlates the received data with an *estimate* of the signal, i.e., $\hat{s}[n]$. It is therefore termed an *estimator-correlator*. Note that the test statistic is a quadratic form in the data and thus will not be a Gaussian random variable. (Recall that the energy detector was a scaled χ^2_N random variable.) We claim that $\hat{\mathbf{s}}$ is actually a Wiener filter estimator of the signal. It should be emphasized that although $s[n]$ is a random process, the interpretation of $\hat{\mathbf{s}}$ is as an estimate of the *given realization* of the signal. To see that $\hat{\mathbf{s}}$ is the Wiener filter estimator recall that if (see [Kay-I 1993, pg. 389]) $\boldsymbol{\theta}$ is an unknown random vector whose realization is to be estimated based on the data vector $\mathbf{x}$, and $\boldsymbol{\theta}$ and $\mathbf{x}$ are jointly Gaussian with zero mean, the minimum mean square error (MMSE) estimator is

$$\hat{\boldsymbol{\theta}} = \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{x} \tag{5.7}$$

where $\mathbf{C}_{\theta x} = E(\boldsymbol{\theta}\mathbf{x}^T)$ and $\mathbf{C}_{xx} = E(\mathbf{x}\mathbf{x}^T)$. Note that the MMSE estimator is *linear* due to the jointly Gaussian assumption. Here we have $\boldsymbol{\theta} = \mathbf{s}$ and $\mathbf{x} = \mathbf{s} + \mathbf{w}$, with $\mathbf{s}$ and $\mathbf{w}$ uncorrelated. The MMSE estimate of the signal realization is from (5.7)

$$\begin{aligned} \hat{\mathbf{s}} &= E\left[\mathbf{s}(\mathbf{s}+\mathbf{w})^T\right]\left(E\left[(\mathbf{s}+\mathbf{w})(\mathbf{s}+\mathbf{w})^T\right]\right)^{-1}\mathbf{x} \\ &= \mathbf{C}_s(\mathbf{C}_s + \sigma^2\mathbf{I})^{-1}\mathbf{x} \end{aligned}$$

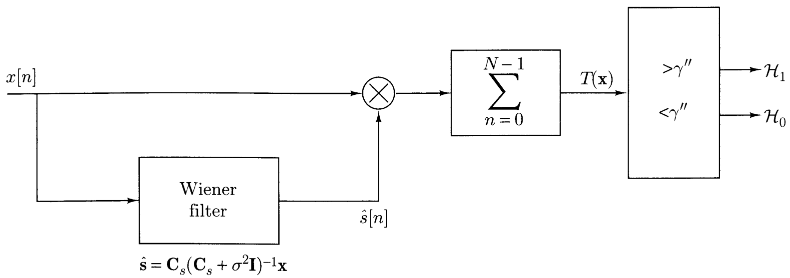


Figure 5.2. Estimator-correlator for detection of Gaussian random signal in white Gaussian noise.

which is identical to (5.6). (See also Problem 5.5 for an alternative derivation of the Wiener filter estimator.) The estimator-correlator is shown in Figure 5.2 and should be compared with the replica-correlator of Figure 4.1a. Some examples follow.

Example 5.2 - Energy Detector (continued)

If the signal is white, then $\mathbf{C}_s = \sigma_s^2\mathbf{I}$, and the signal estimator is

$$\begin{aligned} \hat{\mathbf{s}} &= \sigma_s^2\mathbf{I}(\sigma_s^2\mathbf{I} + \sigma^2\mathbf{I})^{-1}\mathbf{x} \\ &= \frac{\sigma_s^2}{\sigma_s^2 + \sigma^2}\mathbf{x} \end{aligned}$$

or

$$\hat{s}[n] = \frac{\sigma_s^2}{\sigma_s^2 + \sigma^2}x[n].$$

This is a zero memory filter that weights the received data by the fixed scale factor $\sigma_s^2/(\sigma_s^2 + \sigma^2)$. If $\sigma_s^2 \gg \sigma^2$, the weights are approximately one and for $\sigma_s^2 \ll \sigma^2$, they are near zero. See also [Kay-I 1993, pg. 402]. In this case, the detector simplifies by incorporating the known scale factor into the threshold. Thus, we decide $\mathcal{H}_1$ if

$$\sum_{n=0}^{N-1} x[n]\hat{s}[n] = \frac{\sigma_s^2}{\sigma_s^2 + \sigma^2}\sum_{n=0}^{N-1} x^2[n] > \gamma''$$

or

$$\sum_{n=0}^{N-1} x^2[n] > \frac{\gamma''(\sigma_s^2 + \sigma^2)}{\sigma_s^2}$$

as before.

◇

Example 5.3 - Correlated Signal

Now assume that $N = 2$ and

$$\mathbf{C}_s = \sigma_s^2\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

where ρ is the correlation coefficient between $s[0]$ and $s[1]$. From (5.5) and (5.6) the test statistic is

$$T(\mathbf{x}) = \mathbf{x}^T\mathbf{C}_s(\mathbf{C}_s + \sigma^2\mathbf{I})^{-1}\mathbf{x}.$$

But instead of expressing the test statistic in terms of $\mathbf{x}$, it is advantageous to let $\mathbf{y} = \mathbf{V}^T\mathbf{x}$, where

$$\mathbf{V} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}.$$

Then, since $\mathbf{V}^T = \mathbf{V}^{-1}$ ($\mathbf{V}$ is an orthogonal matrix) we have

$$\begin{aligned} T(\mathbf{x}) &= \mathbf{x}^T\mathbf{V}\mathbf{V}^T\mathbf{C}_s\mathbf{V}\mathbf{V}^{-1}(\mathbf{C}_s+\sigma^2\mathbf{I})^{-1}\mathbf{V}\mathbf{V}^T\mathbf{x} \\ &= (\mathbf{V}^T\mathbf{x})^T(\mathbf{V}^T\mathbf{C}_s\mathbf{V})\left[\mathbf{V}^{-1}(\mathbf{C}_s+\sigma^2\mathbf{I})\mathbf{V}\right]^{-1}\mathbf{V}^T\mathbf{x} \\ &= (\mathbf{V}^T\mathbf{x})^T(\mathbf{V}^T\mathbf{C}_s\mathbf{V})\left(\mathbf{V}^T\mathbf{C}_s\mathbf{V}+\sigma^2\mathbf{I}\right)^{-1}\mathbf{V}^T\mathbf{x}. \end{aligned}$$

Now $\mathbf{V}^T\mathbf{C}_s\mathbf{V} = \mathbf{\Lambda}_s$, where $\mathbf{\Lambda}_s$ is the *diagonal* matrix

$$\mathbf{\Lambda}_s = \sigma_s^2\begin{bmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{bmatrix}$$

as can be easily verified. The test statistic becomes

$$\begin{aligned} T(\mathbf{x}) &= \mathbf{y}^T\mathbf{\Lambda}_s(\mathbf{\Lambda}_s+\sigma^2\mathbf{I})^{-1}\mathbf{y} \\ &= \mathbf{y}^T\mathbf{A}\mathbf{y} \end{aligned} \tag{5.8}$$

where $\mathbf{A}$ is the diagonal matrix

$$\mathbf{A} = \begin{bmatrix} \dfrac{\sigma_s^2(1+\rho)}{\sigma_s^2(1+\rho)+\sigma^2} & 0 \\ 0 & \dfrac{\sigma_s^2(1-\rho)}{\sigma_s^2(1-\rho)+\sigma^2} \end{bmatrix}.$$

Thus, we have

$$T(\mathbf{x}) = \frac{\sigma_s^2(1+\rho)}{\sigma_s^2(1+\rho)+\sigma^2}y^2[0] + \frac{\sigma_s^2(1-\rho)}{\sigma_s^2(1-\rho)+\sigma^2}y^2[1].$$

The data are first linearly transformed from $\mathbf{x}$ to $\mathbf{y}$, after which a *weighted* energy detector is applied. Note that if $\rho = 0$, i.e., the signal is white, we have just

$$\begin{aligned} T(\mathbf{x}) &= \frac{\sigma_s^2}{\sigma_s^2+\sigma^2}(y^2[0]+y^2[1]) \\ &= \frac{\sigma_s^2}{\sigma_s^2+\sigma^2}(x^2[0]+x^2[1]) \end{aligned}$$

as in Example 5.2, since $\mathbf{y}^T\mathbf{y} = \mathbf{y}^T\mathbf{V}\mathbf{V}^T\mathbf{y} = \mathbf{x}^T\mathbf{x}$. It is interesting to note that the effect of the linear transformation is to decorrelate $\mathbf{x}$. To see this we have under $\mathcal{H}_1$

$$\begin{aligned} \mathbf{C}_y = E(\mathbf{y}\mathbf{y}^T) &= E(\mathbf{V}^T\mathbf{x}\mathbf{x}^T\mathbf{V}) \\ &= \mathbf{V}^T\mathbf{C}_x\mathbf{V} \\ &= \mathbf{V}^T(\mathbf{C}_s+\sigma^2\mathbf{I})\mathbf{V} \\ &= \mathbf{V}^T\mathbf{C}_s\mathbf{V}+\sigma^2\mathbf{I} \\ &= \mathbf{\Lambda}_s+\sigma^2\mathbf{I} \end{aligned}$$

which is a diagonal matrix. Similarly, under $\mathcal{H}_0$ we have $\mathbf{C}_y = \sigma^2\mathbf{I}$. Hence, $\mathbf{y}$ is composed of uncorrelated random variables, although of different variances. Because of the unequal variances, the energy detector weights the squares of $y[n]$ differently. ◇

The previous example has introduced the *canonical form* of the estimator-correlator. The astute reader will recognize that the decorrelation matrix $\mathbf{V}$ is just the modal matrix for $\mathbf{C}_s$ (see Example 4.5 and Appendix 1), which is the matrix whose columns are the eigenvectors of $\mathbf{C}_s$ (hence $\mathbf{V}^T = \mathbf{V}^{-1}$ since $\mathbf{C}_s$ is symmetric). Also, the diagonal elements of $\mathbf{\Lambda}_s$ are the corresponding eigenvalues of $\mathbf{C}_s$. Since the addition of a scaled identity matrix to $\mathbf{C}_s$ does not change the eigenvectors but only adds σ^2 to each eigenvalue, the modal matrix also decorrelates $\mathbf{C}_x = \mathbf{C}_s + \sigma^2\mathbf{I}$. More generally, let the eigendecomposition of the $N \times N$ covariance matrix $\mathbf{C}_s$ be

$$\mathbf{V}^T\mathbf{C}_s\mathbf{V} = \mathbf{\Lambda}_s$$

where $\mathbf{V} = [\mathbf{v}_0\, \mathbf{v}_1 \ldots \mathbf{v}_{N-1}]$ for $\mathbf{v}_i$ the ith eigenvector of $\mathbf{C}_s$ and $\mathbf{\Lambda}_s = \operatorname{diag}(\lambda_{s_0}, \lambda_{s_1}, \ldots, \lambda_{s_{N-1}})$ for λ_{s_i} the corresponding eigenvalue. (Since $\mathbf{C}_s$ is symmetric, λ_{s_n} is real, and because it is positive semidefinite, $\lambda_{s_n} \geq 0$.) Then, from (5.5), (5.6), and (5.8) the NP test statistic becomes

$$\begin{aligned} T(\mathbf{x}) &= \mathbf{x}^T\mathbf{C}_s(\mathbf{C}_s+\sigma^2\mathbf{I})^{-1}\mathbf{x} \\ &= \mathbf{y}^T\mathbf{\Lambda}_s(\mathbf{\Lambda}_s+\sigma^2\mathbf{I})^{-1}\mathbf{y} \\ &= \sum_{n=0}^{N-1}\frac{\lambda_{s_n}}{\lambda_{s_n}+\sigma^2}y^2[n]. \end{aligned} \tag{5.9}$$

This is the *canonical form* of the detector which is shown in Figure 5.3. The weighting coefficients $\lambda_{s_n}/(\lambda_{s_n}+\sigma^2)$ are actually Wiener filter weights in a transformed space (see Problem 5.8). For example, if $\lambda_{s_0} \gg \sigma^2$, the signal component of $\mathbf{x}$

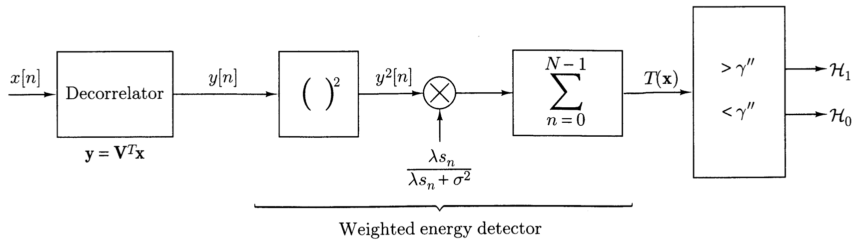


Figure 5.3. Canonical form for detection of Gaussian random signal in white Gaussian noise.

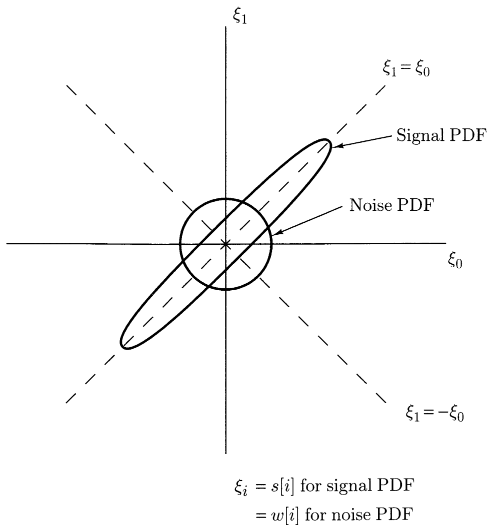


Figure 5.4. Contours of constant PDF for signal only and noise only.

along the $\mathbf{v}_0$ direction is much larger than that of the noise component. Hence, the contribution of $y[0]$ to $T(\mathbf{x})$ is weighted more heavily. Considering the previous example, if $\rho \approx 1$ and $\sigma_s^2 \gg \sigma^2$, then

$$\begin{aligned}\frac{\lambda_{s_0}}{\lambda_{s_0}+\sigma^2} &= \frac{\sigma_s^2(1+\rho)}{\sigma_s^2(1+\rho)+\sigma^2} \approx 1 \\ \frac{\lambda_{s_1}}{\lambda_{s_1}+\sigma^2} &= \frac{\sigma_s^2(1-\rho)}{\sigma_s^2(1-\rho)+\sigma^2} \approx 0.\end{aligned}$$

Thus, $y[0]$ is retained and $y[1]$ is essentially discarded. But then the component of $\mathbf{x}$ along $\mathbf{v}_0 = [1/\sqrt{2}\ 1/\sqrt{2}]^T$ is kept and that along $\mathbf{v}_1 = [1/\sqrt{2}\ \ -1/\sqrt{2}]^T$ is discarded. The reason for this behavior is illustrated in Figure 5.4, which for $\rho \approx 1$ shows the PDF of the signal to be concentrated along the line $\xi_1 = \xi_0$. The component of $\mathbf{x}$ along this line is likely to be much larger when the signal is present than the component along the orthogonal line $\xi_1 = -\xi_0$. For no signal there is no preferred direction. Alternatively, the SNR is larger for the $y[0]$ component than for the $y[1]$ component. Since $\mathbf{C}_y = \mathbf{\Lambda}_s + \sigma^2\mathbf{I}$ under $\mathcal{H}_1$, the SNRs of the $y[0]$ and

$y[1]$ components are, respectively

$$\begin{aligned}\eta_0^2 &= \frac{E(y_s^2[0])}{E(y_w^2[0])} = \frac{\lambda_{s_0}}{\sigma^2} = \frac{\sigma_s^2(1+\rho)}{\sigma^2} \approx \frac{2\sigma_s^2}{\sigma^2} \gg 1 \\ \eta_1^2 &= \frac{E(y_s^2[1])}{E(y_w^2[1])} = \frac{\lambda_{s_1}}{\sigma^2} = \frac{\sigma_s^2(1-\rho)}{\sigma^2} \approx 0.\end{aligned}$$

This explains the weighting of the energy detector samples. For the case of nonwhite noise the canonical detector is discussed in Problem 5.11.

The detection performance of the estimator-correlator is difficult to determine analytically. This is because the test statistic $T(\mathbf{x})$ as given by (5.9) is a *weighted* sum of independent χ_1^2 random variables. As a result, we do not obtain the simple scaled χ_N^2 PDFs as we did for the energy detector. In the latter case, since $s[n]$ is a white Gaussian process, we have $\mathbf{C}_s = \sigma_s^2\mathbf{I}$ and $\lambda_{s_n} = \sigma_s^2$ for $n = 0, 1, \ldots, N-1$. Thus, $T(\mathbf{x}) = (\sigma_s^2/(\sigma_s^2+\sigma^2))\sum_{n=0}^{N-1} y^2[n]$. In Appendix 5A it is shown that for the detector of (5.9) or equivalently (5.5), (5.6)

$$P_{FA} = \int_{\gamma''}^{\infty}\int_{-\infty}^{\infty}\prod_{n=0}^{N-1}\frac{1}{\sqrt{1-2j\alpha_n\omega}}\exp(-j\omega t)\frac{d\omega}{2\pi}dt \tag{5.10}$$

$$P_D = \int_{\gamma''}^{\infty}\int_{-\infty}^{\infty}\prod_{n=0}^{N-1}\frac{1}{\sqrt{1-2j\lambda_{s_n}\omega}}\exp(-j\omega t)\frac{d\omega}{2\pi}dt \tag{5.11}$$

where

$$\alpha_n = \frac{\lambda_{s_n}\sigma^2}{\lambda_{s_n}+\sigma^2}.$$

The inner integrals are seen to be inverse Fourier transforms (although with a negative sign for j) and are the PDFs of $T(\mathbf{x})$. Numerical evaluation is necessary in general. We next present an example in which a closed-form expression can be found (see also Example 5.5).

Example 5.4 - Paired Eigenvalues of Signal Covariance Matrix

Assume that N is even, so that $N = 2M$ and that

$$\mathbf{C}_s = \begin{bmatrix}\mathbf{C}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{C}_1\end{bmatrix} = \begin{bmatrix} M\times M & M\times M \\ M\times M & M\times M\end{bmatrix}.$$

Because of the block-diagonal form of $\mathbf{C}_s$, the signal $\mathbf{s}$ can be split into two independent subvectors or $\mathbf{s} = [\mathbf{s}_1^T\ \mathbf{s}_2^T]^T$ with each $M \times 1$ subvector having covariance $\mathbf{C}_1$. We may view the signal vector as being composed of two independent realizations of the M-length signal. This may occur if the signal vectors are obtained successively in time or obtained as the outputs of two sensors, as in array processing (see

Chapter 13). The covariance may be diagonalized by the modal matrix

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_1 \end{bmatrix}$$

where $\mathbf{V}_1^T\mathbf{C}_1\mathbf{V}_1 = \mathbf{\Lambda}_1$. The matrix $\mathbf{V}_1$ is the modal matrix, and $\mathbf{\Lambda}_1$ is the corresponding diagonal eigenvalue matrix for $\mathbf{C}_1$. As a result,

$$\mathbf{V}^T\mathbf{C}_s\mathbf{V} = \begin{bmatrix} \mathbf{\Lambda}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_1 \end{bmatrix}$$

and the eigenvalues of $\mathbf{C}_s$ occur in pairs. Letting $\lambda_{s_0}, \lambda_{s_1}, \ldots, \lambda_{s_{M-1}}$ be the eigenvalues of $\mathbf{C}_1$, and assuming they are distinct, we have from (5.10)

$$\begin{aligned} & \int_{-\infty}^{\infty} \prod_{n=0}^{N-1} \frac{1}{\sqrt{1-2j\alpha_n\omega}} \exp(-j\omega t)\frac{d\omega}{2\pi} \\ = & \int_{-\infty}^{\infty} \prod_{n=0}^{M-1} \frac{1}{1-2j\alpha_n\omega} \exp(-j\omega t)\frac{d\omega}{2\pi} \\ = & \mathcal{F}_{-t}^{-1}\left\{\prod_{n=0}^{M-1} \frac{1}{1-2j\alpha_n\omega}\right\} \end{aligned} \tag{5.12}$$

where $\mathcal{F}_{-t}^{-1}$ denotes the inverse Fourier transform evaluated at $-t$. Using a partial fraction expansion (recall that it is assumed there are no repeated signal eigenvalues for $\mathbf{C}_1$ and hence no repeated α_n's)

$$\prod_{n=0}^{M-1} \frac{1}{1-2j\alpha_n\omega} = \sum_{n=0}^{M-1} \frac{A_n}{1-2j\alpha_n\omega}$$

where

$$A_n = \prod_{\substack{i=0\\i\neq n}}^{M-1} \frac{1}{1-\frac{\alpha_i}{\alpha_n}}.$$

Also, it is easy to show that

$$\mathcal{F}_{-t}^{-1}\left\{\frac{1}{1-2j\alpha\omega}\right\} = \begin{cases} \frac{1}{2\alpha}\exp(-\frac{t}{2\alpha}) & t>0 \\ 0 & t<0 \end{cases} \tag{5.13}$$

so that (5.12) becomes

$$\mathcal{F}_{-t}^{-1}\left\{\prod_{n=0}^{M-1} \frac{1}{1-2j\alpha_n\omega}\right\} = \begin{cases} \sum_{n=0}^{M-1} \frac{A_n}{2\alpha_n}\exp\left(-\frac{t}{2\alpha_n}\right) & t>0 \\ 0 & t<0 \end{cases}$$

and

$$P_{FA} = \int_{\gamma''}^{\infty} \sum_{n=0}^{M-1} \frac{A_n}{2\alpha_n}\exp\left(-\frac{t}{2\alpha_n}\right) dt.$$

Finally, we have

$$P_{FA} = \sum_{n=0}^{M-1} A_n \exp\left(-\frac{\gamma''}{2\alpha_n}\right) \tag{5.14}$$

where

$$\begin{aligned} A_n &= \prod_{\substack{i=0\\i\neq n}}^{M-1} \frac{1}{1-\frac{\alpha_i}{\alpha_n}} \\ \alpha_n &= \frac{\lambda_{s_n}\sigma^2}{\lambda_{s_n}+\sigma^2} \end{aligned}$$

and similarly

$$P_D = \sum_{n=0}^{M-1} B_n \exp\left(-\frac{\gamma''}{2\lambda_{s_n}}\right) \tag{5.15}$$

where

$$B_n = \prod_{\substack{i=0\\i\neq n}}^{M-1} \frac{1}{1-\frac{\lambda_{s_i}}{\lambda_{s_n}}}.$$

Note that the form of $\mathbf{C}_s$ chosen is exactly that assumed for the complex Gaussian PDF if the real and imaginary random vectors are assumed to be independent (see [Kay-I 1993, pg. 507]). Hence, in the complex data case the detection performance of the estimator-correlator is simply found from (5.14) and (5.15) for any covariance matrix (see Problem 13.8).

◇

The estimator-correlator can be generalized when the observation noise $w[n]$ is Gaussian but not white. If the covariance matrix for the noise is $\mathbf{C}_w$, then the NP detector can be shown to be (see Problem 5.10)

$$T(\mathbf{x}) = \mathbf{x}^T\mathbf{C}_w^{-1}\hat{\mathbf{s}} \tag{5.16}$$

where

$$\hat{\mathbf{s}} = \mathbf{C}_s(\mathbf{C}_s+\mathbf{C}_w)^{-1}\mathbf{x}. \tag{5.17}$$

Now there is a prewhitener for the noise (the $\mathbf{C}_w^{-1}$ term) and the Wiener filter, or equivalently the MMSE estimator for $\mathbf{s}$, is modified appropriately. A canonical form for the detector is described in Problem 5.11.

5.4 Linear Model

The linear model in the context of deterministic signals was discussed in Chapter 4. For example, the DC level in WGN has the form $x[n] = w[n]$ under $\mathcal{H}_0$ and $x[n] = A + w[n]$ under $\mathcal{H}_1$ or

$$\mathbf{x} = \begin{cases} \mathbf{w} & \text{under } \mathcal{H}_0 \\ A\mathbf{1} + \mathbf{w} & \text{under } \mathcal{H}_1 \end{cases}$$

where A is a deterministic constant. In some situations A may be more accurately modeled as the realization of a random variable. For example, in detecting an edge of an object in an image we may not know the exact change in pixel value that signifies an edge. However, we may have some prior knowledge about the possible values that A may take on. If we model A as a random variable with $A \sim \mathcal{N}(0, \sigma_A^2)$ and A is independent of $w[n]$, then we have the *Bayesian linear model* described in [Kay-I 1993, Chapter 11]. Summarizing the model, we assume that

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

where $\mathbf{x} = [x[0]\, x[1] \ldots x[N-1]]^T$, $\mathbf{H}$ is a known $N \times p$ observation matrix, $\boldsymbol{\theta}$ is a $p \times 1$ random vector with $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_\theta)$, and $\mathbf{w}$ is a $N \times 1$ noise vector with $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ and independent of $\boldsymbol{\theta}$. (Actually, we previously described the Bayesian linear model in more generality by letting $E(\boldsymbol{\theta}) \neq \mathbf{0}$ and $\mathbf{C}_w$ being arbitrary. The present definition will be adequate for the more common signal detection problems.)

The detection problem becomes

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{x} = \mathbf{w} \\ \mathcal{H}_1 &: \mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}. \end{aligned}$$

To find the NP detector we can use our previous results and note that $\mathbf{s} = \mathbf{H}\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \mathbf{H}\mathbf{C}_\theta\mathbf{H}^T)$. Thus, the estimator-correlator decides $\mathcal{H}_1$ if (see (5.5) and (5.6))

$$T(\mathbf{x}) = \mathbf{x}^T\mathbf{C}_s(\mathbf{C}_s + \sigma^2\mathbf{I})^{-1}\mathbf{x} > \gamma''$$

or substituting for $\mathbf{C}_s$ we decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \mathbf{x}^T\mathbf{H}\mathbf{C}_\theta\mathbf{H}^T(\mathbf{H}\mathbf{C}_\theta\mathbf{H}^T + \sigma^2\mathbf{I})^{-1}\mathbf{x} > \gamma''. \tag{5.18}$$

This can also be shown to reduce to $T(\mathbf{x}) = \mathbf{x}^T\hat{\mathbf{s}} = \mathbf{x}^T\mathbf{H}\hat{\boldsymbol{\theta}}$, where $\hat{\boldsymbol{\theta}}$ is the MMSE estimator of $\boldsymbol{\theta}$ (see Problem 5.15). An important application of this result is the Rayleigh fading sinusoid.

Example 5.5 - Rayleigh Fading Sinusoid

This example has been previously described from an estimation viewpoint in [Kay-I 1993, Example 11.1] so that the reader may wish to refer to it. In the Rayleigh fading model we assume that when the signal is present we observe

$$x[n] = A\cos(2\pi f_0 n + \phi) + w[n] \qquad n = 0, 1, \ldots, N-1$$

where $0 < f_0 < 1/2$ and $w[n]$ is WGN with variance σ^2. The critical assumption concerns the modeling of the amplitude and phase. In a time-varying multipath environment a transmitted sinusoid will appear at the receiver as a narrowband random process as shown in Figure 5.5. This is because the signal will travel along different paths in reaching the receiver. The net effect is to cause constructive or destructive interference, resulting in an unpredictable amplitude and phase. This is typical of an underwater channel as encountered in the sonar problem or a troposcatter channel over which a communication system might operate. Furthermore,

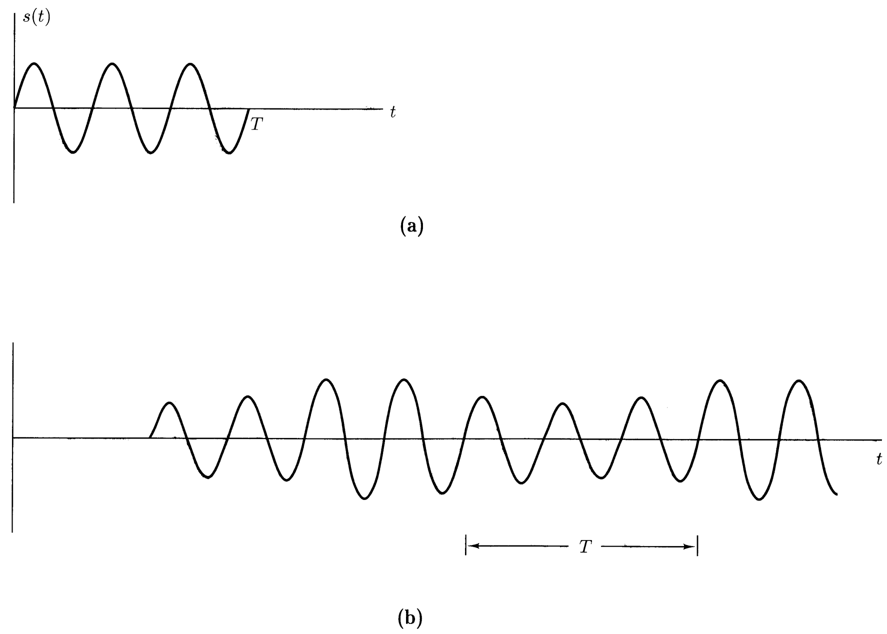


Figure 5.5. Typical input and output signals for multipath channel (a) Transmit signal (b) Received signal.

the propagation characteristics of the channel will change in time due to a moving transmitter and/or receiver or the nonstationary nature of the channel. This causes the amplitude and phase of the received sinusoid to change in time as depicted in Figure 5.5. If our observation interval T is short, then as shown in Figure 5.5, we can model the received signal as a pure sinusoid, albeit with unknown amplitude and phase. It is reasonable to assume that they are random variables, independent of each other. Instead of assigning a PDF to A and ϕ, however, it is more convenient to note that

$$s[n] = A\cos(2\pi f_0 n + \phi) = a\cos 2\pi f_0 n + b\sin 2\pi f_0 n$$

where $a = A\cos\phi, b = -A\sin\phi$ and assign a PDF to $[a\ b]^T$. Note that the signal model is now *linear* in the parameters a and b. For mathematical simplicity as well as an appeal to the central limit theorem (due to the superposition of the many multipath arrivals) we assume that

$$\boldsymbol{\theta} = \begin{bmatrix} a \\ b \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \sigma_s^2\mathbf{I})$$

and that $\boldsymbol{\theta}$ is independent of $w[n]$. It follows that $s[n]$ is a WSS Gaussian random process since

$$E(s[n]) = E(a)\cos 2\pi f_0 n + E(b)\sin 2\pi f_0 n = 0$$

and

$$\begin{aligned}
&E(s[n]s[n+k]) \\
&= E[(a\cos 2\pi f_0 n + b\sin 2\pi f_0 n)(a\cos 2\pi f_0(n+k) + b\sin 2\pi f_0(n+k))] \\
&= E(a^2\cos 2\pi f_0 n\cos 2\pi f_0(n+k)) + E(b^2\sin 2\pi f_0 n\sin 2\pi f_0(n+k)) \\
&= \sigma_s^2(\cos 2\pi f_0 n\cos 2\pi f_0(n+k) + \sin 2\pi f_0 n\sin 2\pi f_0(n+k)) \\
&= \sigma_s^2\cos 2\pi f_0 k
\end{aligned}$$

so that $s[n]$ has the ACF $r_{ss}[k] = \sigma_s^2\cos 2\pi f_0 k$. Furthermore, one can show that the PDF of $A = \sqrt{a^2+b^2}$ is Rayleigh or

$$p(A) = \begin{cases} \dfrac{A}{\sigma_s^2}\exp\left(-\dfrac{A^2}{2\sigma_s^2}\right) & A > 0 \\ 0 & A < 0 \end{cases}$$

and the PDF of $\phi = \arctan(-b/a)$ is $\mathcal{U}[0, 2\pi]$, and A and ϕ are independent of each other [Papoulis 1965]. The term "Rayleigh fading" refers to the amplitude PDF.

With these assumptions we now have the Bayesian linear model since $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$, where

$$\mathbf{H} = \begin{bmatrix} 1 & 0 \\ \cos 2\pi f_0 & \sin 2\pi f_0 \\ \vdots & \vdots \\ \cos 2\pi f_0(N-1) & \sin 2\pi f_0(N-1) \end{bmatrix}$$

$\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \sigma_s^2\mathbf{I})$ and $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$, with $\boldsymbol{\theta}$ independent of $\mathbf{w}$. The NP detector is easily found from (5.18) as

$$\begin{aligned}
T(\mathbf{x}) &= \mathbf{x}^T\mathbf{H}\mathbf{C}_\theta\mathbf{H}^T(\mathbf{H}\mathbf{C}_\theta\mathbf{H}^T + \sigma^2\mathbf{I})^{-1}\mathbf{x} \\
&= \sigma_s^2\mathbf{x}^T\mathbf{H}\mathbf{H}^T(\sigma_s^2\mathbf{H}\mathbf{H}^T + \sigma^2\mathbf{I})^{-1}\mathbf{x}.
\end{aligned}$$

Recalling the matrix inversion lemma (see Appendix 1)

$$(\mathbf{A} + \mathbf{B}\mathbf{C}\mathbf{D})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{D}\mathbf{A}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{D}\mathbf{A}^{-1}$$

and letting $\mathbf{A} = \sigma^2\mathbf{I}$, $\mathbf{B} = \sigma_s^2\mathbf{H}$, $\mathbf{C} = \mathbf{I}$, $\mathbf{D} = \mathbf{H}^T$, we have

$$T(\mathbf{x}) = \sigma_s^2\mathbf{x}^T\mathbf{H}\mathbf{H}^T\left[\frac{1}{\sigma^2}\mathbf{I} - \frac{1}{\sigma^4}\sigma_s^2\mathbf{H}\left(\frac{\sigma_s^2\mathbf{H}^T\mathbf{H}}{\sigma^2} + \mathbf{I}\right)^{-1}\mathbf{H}^T\right]\mathbf{x}.$$

But

$$\mathbf{H}^T\mathbf{H} = \begin{bmatrix} \sum_{n=0}^{N-1}\cos^2 2\pi f_0 n & \sum_{n=0}^{N-1}\cos 2\pi f_0 n\sin 2\pi f_0 n \\ \sum_{n=0}^{N-1}\cos 2\pi f_0 n\sin 2\pi f_0 n & \sum_{n=0}^{N-1}\sin^2 2\pi f_0 n \end{bmatrix}$$

and if we assume that N is large and $0 < f_0 < 1/2$ we have

$$\begin{aligned}
\sum_{n=0}^{N-1}\cos^2 2\pi f_0 n &= \sum_{n=0}^{N-1}\left(\frac{1}{2} + \frac{1}{2}\cos 4\pi f_0 n\right) \approx \frac{N}{2} \\
\sum_{n=0}^{N-1}\cos 2\pi f_0 n\sin 2\pi f_0 n &= \frac{1}{2}\sum_{n=0}^{N-1}\sin 4\pi f_0 n \approx 0 \\
\sum_{n=0}^{N-1}\sin^2 2\pi f_0 n &= \sum_{n=0}^{N-1}\left(\frac{1}{2} - \frac{1}{2}\cos 4\pi f_0 n\right) \approx \frac{N}{2}
\end{aligned}$$

so that $\mathbf{H}^T\mathbf{H} \approx (N/2)\mathbf{I}$. Thus

$$T(\mathbf{x}) = \sigma_s^2\mathbf{x}^T\mathbf{H}\mathbf{H}^T\left(\frac{1}{\sigma^2}\mathbf{I} - \frac{\sigma_s^2}{\sigma^4}\mathbf{H}\frac{1}{\frac{N\sigma_s^2}{2\sigma^2} + 1}\mathbf{I}\mathbf{H}^T\right)\mathbf{x}$$

$$
\begin{aligned}
&= \frac{\sigma_s^2}{\sigma^2}\mathbf{x}^T\mathbf{H}\mathbf{H}^T\mathbf{x} - \frac{\frac{N\sigma_s^4}{2\sigma^4}}{\frac{N\sigma_s^2}{2\sigma^2}+1}\mathbf{x}^T\mathbf{H}\mathbf{H}^T\mathbf{x} \\
&= \frac{c}{N}\mathbf{x}^T\mathbf{H}\mathbf{H}^T\mathbf{x}
\end{aligned}
$$

where $c = N\sigma_s^2/(N\sigma_s^2/2+\sigma^2) > 0$. Incorporating the positive constant c into the threshold results in

$$T'(\mathbf{x}) = \frac{1}{N}\mathbf{x}^T\mathbf{H}\mathbf{H}^T\mathbf{x} = \frac{1}{N}\left\|\mathbf{H}^T\mathbf{x}\right\|^2$$

where $\|\cdot\|$ denotes the Euclidean norm. Hence

$$T'(\mathbf{x}) = \frac{1}{N}\left\|\begin{bmatrix} \sum_{n=0}^{N-1} x[n]\cos 2\pi f_0 n \\ \sum_{n=0}^{N-1} x[n]\sin 2\pi f_0 n \end{bmatrix}\right\|^2$$

or

$$
\begin{aligned}
T'(\mathbf{x}) &= \frac{1}{N}\left[\left(\sum_{n=0}^{N-1} x[n]\cos 2\pi f_0 n\right)^2 + \left(\sum_{n=0}^{N-1} x[n]\sin 2\pi f_0 n\right)^2\right] && (5.19) \\
&= \frac{1}{N}\left|\sum_{n=0}^{N-1} x[n]\exp(-j2\pi f_0 n)\right|^2. && (5.20)
\end{aligned}
$$

We decide $\mathcal{H}_1$ if $T'(\mathbf{x}) > \gamma''/c = \gamma'''$. The detector is shown in Figure 5.6. In the first implementation as given by (5.19) we correlate the data with the cosine ("in phase") and sine ("in quadrature") replicas. Since the phase is random, one or both of these outputs (I or Q) will be large in magnitude if a signal is present. Since the sign of the correlator output can be positive or negative, we square the I and Q outputs and sum. (See also Problem 5.17 in which it is shown that incorrect knowledge of ϕ can lead to poor matched filter detection performance.) This type of detector is known as a *quadrature* matched filter or an *incoherent* matched filter. The second implementation is the periodogram detector. The Fourier transform of $x[n]$ is computed, which is followed by a magnitude-squared operation, a scaling by $1/N$, and a comparison to a threshold. Usually, the Fourier transform is implemented as a fast Fourier transform (FFT), which explains the widespread use of FFTs as narrowband detectors in radar and sonar systems. The periodogram detector can also be interpreted as a PSD estimator [Kay 1988]. As such, a large spike will be present at the known frequency f_0 when a signal is present. For an unknown signal frequency a similar detector is used (see Chapter 7).

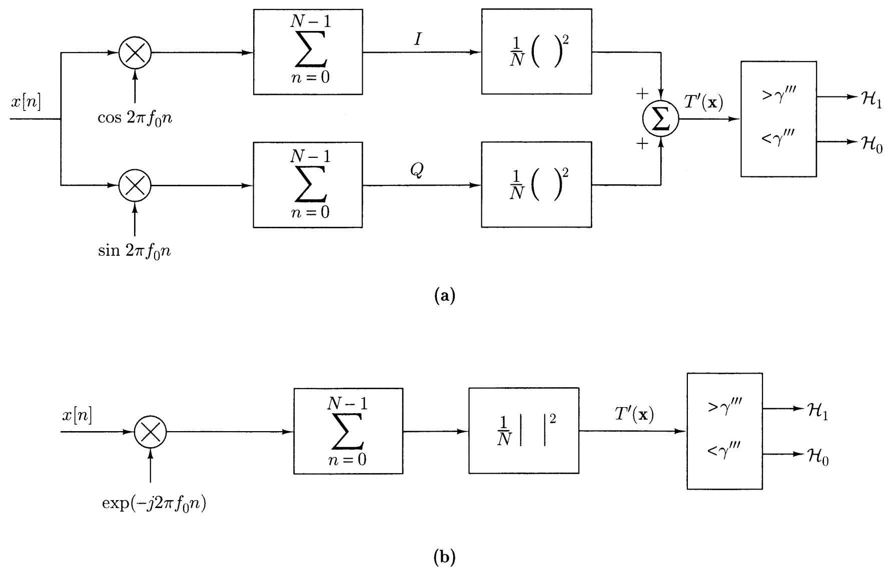


Figure 5.6. Equivalent detectors for Rayleigh fading sinusoid in white Gaussian noise (a) Quadrature or incoherent matched filter (b) Periodogram detector.

The performance of the quadrature matched filter is a straightforward application of our previous results. (In Problem 5.18 we show how to obtain the same results from first principles.) The key to using (5.10) and (5.11) is to be able to invert the Fourier transforms. We showed in Example 5.4 that for eigenvalues occurring in distinct pairs, this is particularly simple. Here, we have for the signal covariance matrix

$$
\begin{aligned}
\mathbf{C}_s = \mathbf{H}\mathbf{C}_\theta\mathbf{H}^T &= \mathbf{H}\sigma_s^2\mathbf{I}\mathbf{H}^T \\
&= \sigma_s^2\mathbf{H}\mathbf{H}^T \\
&= \sigma_s^2\begin{bmatrix}\mathbf{h}_0 & \mathbf{h}_1\end{bmatrix}\begin{bmatrix}\mathbf{h}_0^T \\ \mathbf{h}_1^T\end{bmatrix} \\
&= \frac{N\sigma_s^2}{2}\frac{\mathbf{h}_0}{\sqrt{N/2}}\frac{\mathbf{h}_0^T}{\sqrt{N/2}} + \frac{N\sigma_s^2}{2}\frac{\mathbf{h}_1}{\sqrt{N/2}}\frac{\mathbf{h}_1^T}{\sqrt{N/2}}.
\end{aligned}
$$

Letting $\lambda_{s_0} = \lambda_{s_1} = N\sigma_s^2/2$, $\mathbf{v}_0 = \mathbf{h}_0/\sqrt{N/2}$, and $\mathbf{v}_1 = \mathbf{h}_1/\sqrt{N/2}$ we see that

$$\mathbf{C}_s = \lambda_{s_0}\mathbf{v}_0\mathbf{v}_0^T + \lambda_{s_1}\mathbf{v}_1\mathbf{v}_1^T$$

where $\mathbf{v}_0^T\mathbf{v}_1 = (2/N)\sum_{n=0}^{N-1}\cos 2\pi f_0 n \sin 2\pi f_0 n \approx 0$ for large N. Thus, $\mathbf{v}_0, \mathbf{v}_1$ are approximately eigenvectors of $\mathbf{C}_s$ with corresponding eigenvalues $\lambda_{s_0}, \lambda_{s_1}$. The remaining $N-2$ eigenvalues are zero since $\mathbf{C}_s$ has rank two. It is easily verified that using $\lambda_{s_0} = \lambda_{s_1} = N\sigma_s^2/2, \lambda_{s_2} = \cdots \lambda_{s_{N-1}} = 0$ in (5.10) and noting (5.13) results in

$$\begin{aligned} P_{FA} &= \Pr\{T'(\mathbf{x}) > \gamma'''; \mathcal{H}_0\} \\ &= \Pr\{T(\mathbf{x}) > \gamma''; \mathcal{H}_0\} \\ &= \exp\left(-\frac{\gamma''}{2\alpha_0}\right) \end{aligned}$$

where

$$\alpha_0 = \frac{\lambda_{s_0}\sigma^2}{\lambda_{s_0} + \sigma^2} = \frac{N\sigma_s^2\sigma^2/2}{N\sigma_s^2/2 + \sigma^2} = c\sigma^2/2.$$

Since $\gamma'' = c\gamma'''$, this reduces to

$$P_{FA} = \exp\left(-\frac{\gamma'''}{\sigma^2}\right). \tag{5.21}$$

Also, from (5.11) and (5.13)

$$\begin{aligned} P_D &= \exp\left(-\frac{\gamma''}{2\lambda_{s_0}}\right) \\ &= \exp\left(-\frac{\gamma''}{N\sigma_s^2}\right) \end{aligned}$$

which results in

$$P_D = \exp\left(-\frac{\gamma'''}{N\sigma_s^2/2 + \sigma^2}\right). \tag{5.22}$$

To relate P_D to P_{FA} let $\bar{\eta} = NE(A^2/2)/\sigma^2 = N\sigma_s^2/\sigma^2$ be the average ENR. Note that the *expected* signal energy is $\bar{\mathcal{E}} = NE(A^2/2)$. The threshold is easily found as

$$\gamma''' = \sigma^2 \ln \frac{1}{P_{FA}}.$$

Using it in (5.22) produces

$$P_D = P_{FA}^{\frac{1}{1+\bar{\eta}/2}}. \tag{5.23}$$

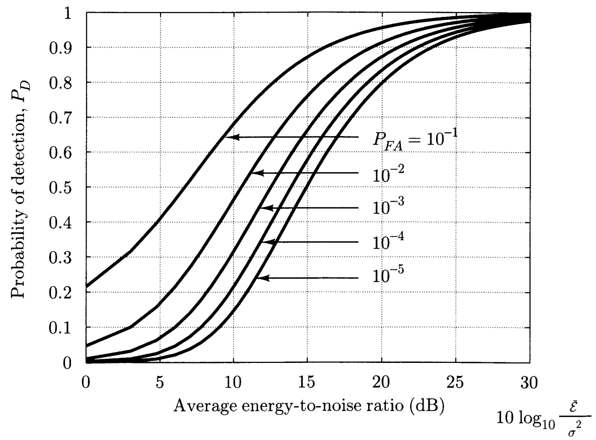


Figure 5.7. Detection performance for Rayleigh fading sinusoid.

The probability of detection increases very slowly with increasing $\bar{\eta} = \bar{\mathcal{E}}/\sigma^2$ as shown in Figure 5.7. This is because the sinusoidal amplitude has a Rayleigh PDF, which is shown in Figure 5.8 for various values of σ_s^2. Even as σ_s^2 increases, there is still a high probability that the sinusoidal amplitude will be small. This results in an *average* probability of detection that does not increase rapidly with σ_s^2.

◇

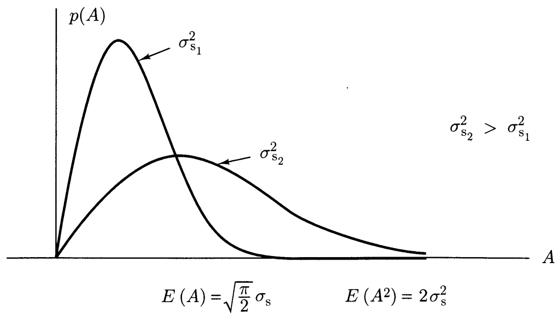


Figure 5.8. PDF for Rayleigh amplitude.

Example 5.6 - Incoherent FSK for a Multipath Channel

Another important example is that of *communication* through a multipath channel. In particular, we now consider the problem of detecting the signal $s_0[n] = \cos 2\pi f_0 n$ or $s_1[n] = \cos 2\pi f_1 n$ after propagation through a Rayleigh fading channel. It is assumed that $0 < f_0 < 1/2$, $0 < f_1 < 1/2$ and that N is large. Hence, at the channel output we have the detection problem

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = A_0 \cos(2\pi f_0 n + \phi_0) + w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = A_1 \cos(2\pi f_1 n + \phi_1) + w[n] & n = 0, 1, \ldots, N-1 \end{array}$$

where $w[n]$ is WGN with variance σ^2. We assume the Rayleigh fading model for the sinusoidal amplitudes and phases and furthermore that the PDF of $[A_0 \, \phi_0]^T$ is the same as that of $[A_1 \, \phi_1]^T$. We wish to design a receiver to minimize the probability of error P_e. For equal prior probabilities of transmission of each signal we should decide $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x}|\mathcal{H}_1)}{p(\mathbf{x}|\mathcal{H}_0)} > 1$$

or

$$\frac{\dfrac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}(\mathbf{C}_{s_1} + \sigma^2\mathbf{I})} \exp\left[-\dfrac{1}{2}\mathbf{x}^T(\mathbf{C}_{s_1} + \sigma^2\mathbf{I})^{-1}\mathbf{x}\right]}{\dfrac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}(\mathbf{C}_{s_0} + \sigma^2\mathbf{I})} \exp\left[-\dfrac{1}{2}\mathbf{x}^T(\mathbf{C}_{s_0} + \sigma^2\mathbf{I})^{-1}\mathbf{x}\right]} > 1.$$

Letting $\mathbf{H}_i$ be the $N \times 2$ observation matrix of the linear model based on the frequency f_i (see Example 5.5), we have $\mathbf{C}_{s_i} = \mathbf{H}_i \mathbf{C}_\theta \mathbf{H}_i^T = \sigma_s^2 \mathbf{H}_i \mathbf{H}_i^T$. Then, letting $\mathbf{I}_2$ be the 2×2 identity matrix, it follows that

$$\begin{aligned} \det(\mathbf{C}_{s_i} + \sigma^2\mathbf{I}) &= \det(\sigma_s^2 \mathbf{H}_i \mathbf{H}_i^T + \sigma^2 \mathbf{I}) \\ &= \det(\sigma_s^2 \mathbf{H}_i^T \mathbf{H}_i + \sigma^2 \mathbf{I}_2) \\ &\approx \det(\sigma_s^2 \frac{N}{2} \mathbf{I}_2 + \sigma^2 \mathbf{I}_2) \end{aligned}$$

for large N, and is independent of the signal frequency. We have used the identity of Problem 4.30, and the approximate orthogonality of the columns of $\mathbf{H}_i$. Hence, we have

$$\frac{\exp\left[-\frac{1}{2}\mathbf{x}^T(\mathbf{C}_{s_1} + \sigma^2\mathbf{I})^{-1}\mathbf{x}\right]}{\exp\left[-\frac{1}{2}\mathbf{x}^T(\mathbf{C}_{s_0} + \sigma^2\mathbf{I})^{-1}\mathbf{x}\right]} > 1$$

and using (5.4) produces

$$\frac{1}{\sigma^2}\mathbf{x}^T\mathbf{x} - \frac{1}{\sigma^4}\mathbf{x}^T\left(\frac{1}{\sigma^2}\mathbf{I} + \mathbf{C}_{s_1}^{-1}\right)^{-1}\mathbf{x} < \frac{1}{\sigma^2}\mathbf{x}^T\mathbf{x} - \frac{1}{\sigma^4}\mathbf{x}^T\left(\frac{1}{\sigma^2}\mathbf{I} + \mathbf{C}_{s_0}^{-1}\right)^{-1}\mathbf{x}$$

or

$$\mathbf{x}^T\hat{\mathbf{s}}_1 > \mathbf{x}^T\hat{\mathbf{s}}_0$$

where $\hat{\mathbf{s}}_0$, $\hat{\mathbf{s}}_1$ are given by (5.6) with $\mathbf{C}_s$ replaced by $\mathbf{C}_{s_0}$ and $\mathbf{C}_{s_1}$, respectively. But from Example 5.5, $T(\mathbf{x}) = \mathbf{x}^T\hat{\mathbf{s}}_0 = cI(f_0)$, where $I(f_0)$ is the periodogram as given by (5.20), and similarly for the other signal. Thus, we decide $\mathcal{H}_1$ if

$$I(f_1) > I(f_0). \tag{5.24}$$

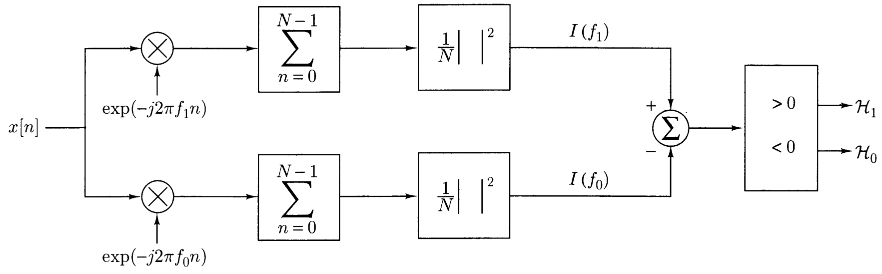


Figure 5.9. Incoherent FSK receiver for Rayleigh fading channel.

The receiver is shown in Figure 5.9. To determine P_e we assume that the frequencies are chosen so that $I(f_1)$ is unaffected by the signal at f_0 and vice-versa. This means that the Fourier transform of $s_0[n]$ has a negligible component at $f = f_1$. If the frequencies are widely separated or $|f_1 - f_0| \gg 1/N$, this assumption will be satisfied. Then, for equal prior probabilities of signal transmission

$$P_e = \frac{1}{2}(P(\mathcal{H}_1|\mathcal{H}_0) + P(\mathcal{H}_0|\mathcal{H}_1))$$

where $P(\mathcal{H}_i|\mathcal{H}_j)$ is the conditional probability of deciding $\mathcal{H}_i$ when $\mathcal{H}_j$ is true. Furthermore, by symmetry we have that $P(\mathcal{H}_0|\mathcal{H}_1) = P(\mathcal{H}_1|\mathcal{H}_0)$ and so

$$\begin{aligned} P_e &= P(\mathcal{H}_0|\mathcal{H}_1) \\ &= \Pr\{I(f_0) > I(f_1)|\mathcal{H}_1\} \\ &= \int_0^\infty \Pr\{I(f_1) < t | I(f_0) = t, \mathcal{H}_1\} p_{I(f_0)}(t|\mathcal{H}_1)\, dt \end{aligned}$$

where $p_{I(f_0)}(t|\mathcal{H}_1)$ is the conditional PDF of $I(f_0)$, conditioned on $\mathcal{H}_1$ having occured. But the statistics $I(f_0)$ and $I(f_1)$ are approximately independent for widely spaced frequencies. This is because under $\mathcal{H}_1$ there is no signal contribution to

$I(f_0)$, and also the noise outputs of the Fourier transform for a white noise input are independent for widely separated frequencies (see Problem 5.19). Then,

$$P_e = \int_0^\infty \Pr\{I(f_1) < t|\mathcal{H}_1\} p_{I(f_0)}(t|\mathcal{H}_1)dt.$$

But we can use the results of Example 5.5 to simplify matters. In particular, note that from (5.22)

$$\Pr\{I(f_1) < t|\mathcal{H}_1\} = 1 - P_D$$

where

$$P_D = \exp\left(-\frac{t}{N\sigma_s^2/2 + \sigma^2}\right).$$

Also, because $I(f_0)$ under $\mathcal{H}_1$ is due to noise only (the signal $A_1\cos(2\pi f_1 n + \phi_1)$ does not contribute to the periodogram at $f = f_0$), we have from (5.21)

$$\begin{aligned} p_{I(f_0)}(t|\mathcal{H}_1) &= \frac{d}{dt}\Pr\{I(f_0) < t|\text{noise only}\} \\ &= \frac{d}{dt}(1 - P_{FA}) \\ &= \frac{d}{dt}\left[1 - \exp(-t/\sigma^2)\right] \\ &= \frac{1}{\sigma^2}\exp(-t/\sigma^2) \end{aligned}$$

so that

$$\begin{aligned} P_e &= \int_0^\infty \left[1 - \exp\left(-\frac{t}{N\sigma_s^2/2 + \sigma^2}\right)\right]\frac{1}{\sigma^2}\exp(-t/\sigma^2)dt \\ &= \frac{1}{2 + \frac{\bar{\eta}}{2}} \end{aligned} \tag{5.25}$$

where $\bar{\eta} = \bar{\mathcal{E}}/\sigma^2 = NE(A_i^2/2)/\sigma^2$. In Figure 5.10 we compare the performance of the incoherent FSK receiver for a multipath channel with that of a coherent FSK receiver for a perfect channel. The latter is given by (see Example 4.8)

$$P_e = Q\left(\sqrt{\frac{\mathcal{E}}{2\sigma^2}}\right)$$

where $\mathcal{E}$ is the energy of the deterministic sinusoidal signal. Recall that the received signals are $A\cos 2\pi f_0 n$ and $A\cos 2\pi f_1 n$, whose energies are approximately equal and $\mathcal{E} \approx NA^2/2$. As indicated in Figure 5.10, the P_e reduces very slowly with increasing average ENR. As in the previous example, this is due to the distribution of Rayleigh amplitudes. In many communication systems this high error rate is unacceptable. To reduce it one must rely on *diversity* techniques as discussed in Section 5.7.

◇

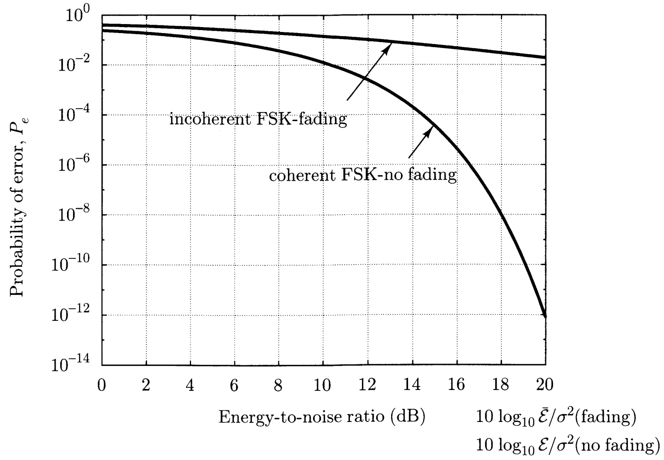


Figure 5.10. Comparison of performance for optimal receivers for perfect channel and Rayleigh fading channel.

5.5 Estimator-Correlator for Large Data Records

For the detection of signals that are WSS Gaussian random processes in WGN, the estimator-correlator can be approximated for large N by a detector that is based on the PSD of $s[n]$. The $N \times N$ matrix inversion required to estimate the signal as in (5.6) can be alleviated. To find the asymptotic form one can invoke the properties of $N \times N$ Toeplitz matrices as $N \to \infty$. This leads to an eigendecomposition in terms of the PSD values. See Problem 5.20 for a further discussion. An alternative approach that we now pursue is to start from first principles using the asymptotic form of the PDF. In [Kay-I 1993, Appendix 3D] it was shown that if $x[n]$ is a zero mean WSS Gaussian random process with PSD $P_{xx}(f)$, then the log-PDF can be approximated for large data records by

$$\ln p(\mathbf{x}) \approx -\frac{N}{2}\ln 2\pi - \frac{N}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \ln P_{xx}(f)df - \frac{N}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{I(f)}{P_{xx}(f)}df \tag{5.26}$$

where

$$I(f) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x[n]\exp(-j2\pi fn)\right|^2$$

is the periodogram. The NP detector decides $\mathcal{H}_1$ if

$$l(\mathbf{x}) = \ln p(\mathbf{x};\mathcal{H}_1) - \ln p(\mathbf{x};\mathcal{H}_0) > \gamma'.$$

But under $\mathcal{H}_0$, $x[n] = w[n]$ and $P_{xx}(f) = \sigma^2$, while under $\mathcal{H}_1$, $x[n] = s[n] + w[n]$ and $P_{xx}(f) = P_{ss}(f) + \sigma^2$ so that

$$\begin{aligned} l(\mathbf{x}) &= -\frac{N}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \ln(P_{ss}(f)+\sigma^2)df - \frac{N}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{I(f)}{P_{ss}(f)+\sigma^2}df \\ &\quad + \frac{N}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \ln\sigma^2 df + \frac{N}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{I(f)}{\sigma^2}df \\ &= -\frac{N}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \ln\left(\frac{P_{ss}(f)}{\sigma^2}+1\right)df + \frac{N}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} I(f)\left(\frac{1}{\sigma^2} - \frac{1}{P_{ss}(f)+\sigma^2}\right)df \\ &= -\frac{N}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \ln\left(\frac{P_{ss}(f)}{\sigma^2}+1\right)df + \frac{N}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} I(f)\frac{P_{ss}(f)}{(P_{ss}(f)+\sigma^2)\sigma^2}df. \end{aligned}$$

Combining the non-data-dependent term into the threshold and scaling, we decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = N\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{P_{ss}(f)}{P_{ss}(f)+\sigma^2} I(f)df > \gamma''. \tag{5.27}$$

To interpret this detector we observe that

$$H(f) = \frac{P_{ss}(f)}{P_{ss}(f)+\sigma^2} \tag{5.28}$$

is the Wiener filter frequency response $H(f)$ for an infinite length noncausal filter (see [Kay-I 1993, pg. 405]) and is a real function of frequency. Hence

$$\begin{aligned} T(\mathbf{x}) &= \int_{-\frac{1}{2}}^{\frac{1}{2}} H(f)X(f)X^*(f)df \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} X(f)(H(f)X(f))^* df \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} X(f)\hat{S}^*(f)df \end{aligned} \tag{5.29}$$

where $\hat{S}(f) = H(f)X(f)$ is the estimator of the Fourier transform of the signal. (See also (4.17) with $P_{ww}(f) = \sigma^2$ for a similar test statistic for a *known deterministic* signal.) Alternatively, using Parseval's theorem we have approximately

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]\hat{s}[n]$$

where $\hat{s}[n] = \mathcal{F}^{-1}\{H(f)X(f)\}$. For the large data record case we estimate the signal by passing the data through a Wiener filter. Then we *correlate* the signal estimate with the data in the frequency domain.

5.6 General Gaussian Detection

In Chapter 4 we considered the detection of a deterministic signal in WGN while in this chapter we have addressed the detection of a random signal in WGN. The most general signal assumption is to allow the signal to be composed of a deterministic component and a random component. The signal then can be modeled as a random process with the deterministic part corresponding to a nonzero mean and the random part corresponding to a zero mean random process with a given signal covariance matrix. Also, for generality the noise covariance can be assumed to be arbitrary. These assumptions lead to the *general Gaussian detection problem*, which mathematically is described as

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{x} = \mathbf{w} \\ \mathcal{H}_1 &: \mathbf{x} = \mathbf{s} + \mathbf{w} \end{aligned}$$

where $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_w)$, $\mathbf{s} \sim \mathcal{N}(\boldsymbol{\mu}_s, \mathbf{C}_s)$, and $\mathbf{s}$ and $\mathbf{w}$ are independent. Hence, the signal can be discriminated from the noise based on its mean and covariance differences. The NP detector decides $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x};\mathcal{H}_1)}{p(\mathbf{x};\mathcal{H}_0)} > \gamma$$

or

$$\frac{\dfrac{1}{(2\pi)^{\frac{N}{2}}\det^{\frac{1}{2}}(\mathbf{C}_s+\mathbf{C}_w)}\exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_s)^T(\mathbf{C}_s+\mathbf{C}_w)^{-1}(\mathbf{x}-\boldsymbol{\mu}_s)\right]}{\dfrac{1}{(2\pi)^{\frac{N}{2}}\det^{\frac{1}{2}}(\mathbf{C}_w)}\exp\left[-\frac{1}{2}\mathbf{x}^T\mathbf{C}_w^{-1}\mathbf{x}\right]} > \gamma.$$

Taking the logarithm, retaining only the data-dependent terms, and scaling produces the test statistic

$$\begin{aligned} T(\mathbf{x}) &= \mathbf{x}^T\mathbf{C}_w^{-1}\mathbf{x} - (\mathbf{x}-\boldsymbol{\mu}_s)^T(\mathbf{C}_s+\mathbf{C}_w)^{-1}(\mathbf{x}-\boldsymbol{\mu}_s) \\ &= \mathbf{x}^T\mathbf{C}_w^{-1}\mathbf{x} - \mathbf{x}^T(\mathbf{C}_s+\mathbf{C}_w)^{-1}\mathbf{x} \\ &\quad + 2\mathbf{x}^T(\mathbf{C}_s+\mathbf{C}_w)^{-1}\boldsymbol{\mu}_s - \boldsymbol{\mu}_s^T(\mathbf{C}_s+\mathbf{C}_w)^{-1}\boldsymbol{\mu}_s. \end{aligned}$$

But from the matrix inversion lemma

$$\mathbf{C}_w^{-1} - (\mathbf{C}_s+\mathbf{C}_w)^{-1} = \mathbf{C}_w^{-1}\mathbf{C}_s(\mathbf{C}_s+\mathbf{C}_w)^{-1}$$

so that by ignoring the nondata dependent term and scaling we have

$$T'(\mathbf{x}) = \mathbf{x}^T(\mathbf{C}_s+\mathbf{C}_w)^{-1}\boldsymbol{\mu}_s + \frac{1}{2}\mathbf{x}^T\mathbf{C}_w^{-1}\mathbf{C}_s(\mathbf{C}_s+\mathbf{C}_w)^{-1}\mathbf{x}. \tag{5.30}$$

The test statistic consists of a quadratic form as well as a linear form in $\mathbf{x}$. As special cases we have

1. $\mathbf{C}_s = \mathbf{0}$ or a deterministic signal with $\mathbf{s} = \boldsymbol{\mu}_s$. Then

$$T'(\mathbf{x}) = \mathbf{x}^T \mathbf{C}_w^{-1} \boldsymbol{\mu}_s$$

which is our prewhitener and matched filter.

2. $\boldsymbol{\mu}_s = \mathbf{0}$ or a random signal with $\mathbf{s} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_s)$. Then

$$\begin{aligned} T'(\mathbf{x}) &= \frac{1}{2}\mathbf{x}^T \mathbf{C}_w^{-1}\mathbf{C}_s(\mathbf{C}_s + \mathbf{C}_w)^{-1}\mathbf{x} \\ &= \frac{1}{2}\mathbf{x}^T \mathbf{C}_w^{-1}\hat{\mathbf{s}} \end{aligned}$$

where $\hat{\mathbf{s}} = \mathbf{C}_s(\mathbf{C}_s + \mathbf{C}_w)^{-1}\mathbf{x}$ is the MMSE estimator of $\mathbf{s}$. This is a prewhitener followed by an estimator-correlator.

An example follows.

Example 5.7 - Deterministic/Random Signal in WGN

Assume that we observe $x[n] = w[n]$ under $\mathcal{H}_0$, where $w[n]$ is WGN and $x[n] = s[n] + w[n]$ under $\mathcal{H}_1$, where $s[n] \sim \mathcal{N}(A, \sigma_s^2)$ and are IID and independent of $w[n]$. Thus, $s[n]$ can be thought of as the sum of a deterministic DC level (due to A) and a white Gaussian process with variance σ_s^2. The NP detector follows from (5.30) by letting $\mathbf{C}_w = \sigma^2\mathbf{I}$, $\boldsymbol{\mu}_s = A\mathbf{1}$, and $\mathbf{C}_s = \sigma_s^2\mathbf{I}$. Hence

$$\begin{aligned} T'(\mathbf{x}) &= \mathbf{x}^T(\sigma_s^2\mathbf{I} + \sigma^2\mathbf{I})^{-1}A\mathbf{1} + \frac{1}{2}\mathbf{x}^T\frac{1}{\sigma^2}\mathbf{I}\sigma_s^2\mathbf{I}(\sigma_s^2\mathbf{I} + \sigma^2\mathbf{I})^{-1}\mathbf{x} \\ &= \frac{NA}{\sigma_s^2 + \sigma^2}\bar{x} + \frac{1}{2}\frac{\sigma_s^2/\sigma^2}{\sigma_s^2 + \sigma^2}\sum_{n=0}^{N-1} x^2[n] \end{aligned}$$

which amounts to the sum of an averager, which attempts to discriminate on the basis of mean $\bar{x}$, and an energy detector, which attempts to discriminate on the basis of variance $(1/N)\sum_{n=0}^{N-1} x^2[n]$.

◇

A special case of the general Gaussian problem is the general Bayesian linear model in which $\mathbf{s} = \mathbf{H}\boldsymbol{\theta}$ with $\mathbf{H}$ an $N \times p$ matrix and $\boldsymbol{\theta}$ a $p \times 1$ random vector with $\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}_\theta, \mathbf{C}_\theta)$ and independent of $\mathbf{w}$. Then from (5.30) we have upon letting $\boldsymbol{\mu}_s = \mathbf{H}\boldsymbol{\mu}_\theta$ and $\mathbf{C}_s = \mathbf{H}\mathbf{C}_\theta\mathbf{H}^T$

$$T'(\mathbf{x}) = \mathbf{x}^T(\mathbf{H}\mathbf{C}_\theta\mathbf{H}^T + \mathbf{C}_w)^{-1}\mathbf{H}\boldsymbol{\mu}_\theta + \frac{1}{2}\mathbf{x}^T\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_\theta\mathbf{H}^T(\mathbf{H}\mathbf{C}_\theta\mathbf{H}^T + \mathbf{C}_w)^{-1}\mathbf{x}.$$

5.7 Signal Processing Example

5.7.1 Tapped Delay Line Channel Model

We now apply our previous results to the detection of a signal that has been transmitted through a multipath channel. This is a typical problem encountered in wireless communications [Pahlavan and Levesque 1994]. In contrast to the Rayleigh fading sinusoid example (see Example 5.5) we now generalize the problem to allow an arbitrary signal to be transmitted. Furthermore, we employ an explicit model for the channel effects. This model is the tapped delay line (TDL) or FIR filter as shown in Figure 5.11. To obtain an intuitive form of the detector we then assume the transmitted signal to be a pseudorandom noise (PRN) sequence. The reader should note that this type of signal is useful for communication in a multipath environment as the wide spectral content results in signal frequencies that fade independently. One may view this approach as that of *frequency diversity* [Proakis 1989]. Now the input-output description of the channel is

$$s[n] = \sum_{k=0}^{p-1} h[k]u[n-k] \tag{5.31}$$

as shown in Figure 5.11. The TDL weights $h[k]$ are usually unknown and may change with time as the channel characteristics change. If we assume that the weights change slowly in time, then over the signal interval we can assume that they are constant. In some situations such an assumption is not valid and so we must employ a linear *time varying* filter. The reader may wish to refer to [Kay-I 1993, Example 13.3] in which an estimator of the time varying weights is discussed and [Kay-I 1993, Example 4.3] in which the nonrandom TDL channel model is described. Although the tap weights are unknown, their average power can sometimes be determined on

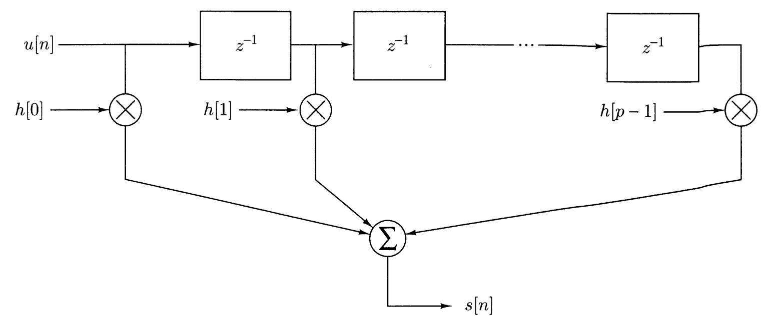


Figure 5.11. Tapped delay line channel model.

the basis of physical scattering considerations. With this in mind we assume the weights to be random variables with zero mean and $\text{var}(h[k]) = \sigma_k^2$. Furthermore, the scattering is assumed to be uncorrelated so that $\text{cov}(h[i], h[j]) = 0$ for $i \neq j$. Such a model is termed the *uncorrelated scattering* model. Also, the channel model is referred to as a *random linear time invariant channel model.* Finally, to complete the statistical description we assume that the tap weights are Gaussian random variables so that

$$\mathbf{h} = \begin{bmatrix} h[0] \\ h[1] \\ \vdots \\ h[p-1] \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_h) \tag{5.32}$$

where $\mathbf{C}_h = \text{diag}(\sigma_0^2, \sigma_1^2, \ldots, \sigma_{p-1}^2)$. In sonar the average power of the tap weights is referred to as the *range scattering function* [Van Trees 1971] and in communications as the *multipath delay profile* [Proakis 1989]. The transmitted signal $u[n]$ is assumed known. Note that the effect of the random channel is to cause the received signal $s[n]$ to also be random. The signal at the channel output is embedded in WGN of variance σ^2 so that

$$\begin{aligned} x[n] &= s[n] + w[n] \\ &= \sum_{k=0}^{p-1} h[k]u[n-k] + w[n]. \end{aligned}$$

At this point it is of interest to note that if the transmitted signal were a sinusoid, then the received signal would also be sinusoidal, although with a random amplitude and phase. The resulting model would be that discussed in Example 5.5 (see also Problem 5.24). If the input signal $u[n]$ is nonzero over the interval $[0, K-1]$, then the output signal is nonzero over the interval $[0, K+p-2]$. Choosing $N = K+p-1$ we have the usual detection problem

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = s[n] + w[n] & n = 0, 1, \ldots, N-1. \end{array}$$

Furthermore, because the channel model was chosen to be linear we have the Bayesian linear model or $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$. As an example, if $K = 2$ and $p = 4$ we have the $N \times p = (K+p-1) \times p = 5 \times 4$ matrix $\mathbf{H}$ from (5.31) for $n = 0, 1, \ldots, 4$ as

$$\mathbf{H} = \begin{bmatrix} u[0] & 0 & 0 & 0 \\ u[1] & u[0] & 0 & 0 \\ 0 & u[1] & u[0] & 0 \\ 0 & 0 & u[1] & u[0] \\ 0 & 0 & 0 & u[1] \end{bmatrix}$$

or $[\mathbf{H}]_{ij} = u[i-j]$ for $i = 0, 1, \ldots, N-1$; $j = 0, 1, \ldots, p-1$, and the $p \times 1 = 4 \times 1$ TDL weight vector

$$\boldsymbol{\theta} = [\ h[0] \quad h[1] \quad h[2] \quad h[3]\ ]^T.$$

Also, from (5.32) $\boldsymbol{\theta} = \mathbf{h} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_h)$ and we assume that $\boldsymbol{\theta}$ is independent of $\mathbf{w}$. In summary, the Bayesian linear model of Section 5.4 can be applied and hence the NP detector decides $\mathcal{H}_1$ if (see (5.18))

$$T(\mathbf{x}) = \mathbf{x}^T\mathbf{H}\mathbf{C}_h\mathbf{H}^T(\mathbf{H}\mathbf{C}_h\mathbf{H}^T + \sigma^2\mathbf{I})^{-1}\mathbf{x} > \gamma''. \tag{5.33}$$

As an illustration, consider the use of a PRN transmit signal. This sequence has the property that the temporal autocorrelation function is approximately (for large N) an impulse [MacWilliams and Sloane 1976] or

$$r_{uu}[k] = \frac{1}{K} \sum_{n=0}^{K-1-|k|} u[n]u[n+|k|] \approx \sigma_u^2\delta[k]$$

where $\sigma_u^2 = (1/K)\sum_{n=0}^{K-1} u^2[n] = \mathcal{E}/K$. As a result, since the columns of $\mathbf{H}$ are shifted versions of each other, they are approximately orthogonal. Hence,

$$\mathbf{H}^T\mathbf{H} \approx \mathcal{E}\mathbf{I}. \tag{5.34}$$

This allows us to simplifly (5.33) by using the matrix inversion lemma. Letting $\mathbf{A} = \sigma^2\mathbf{I}$, $\mathbf{B} = \mathbf{H}$, $\mathbf{C} = \mathbf{C}_h$, and $\mathbf{D} = \mathbf{H}^T$, we have

$$(\sigma^2\mathbf{I} + \mathbf{H}\mathbf{C}_h\mathbf{H}^T)^{-1} = \frac{1}{\sigma^2}\mathbf{I} - \frac{1}{\sigma^4}\mathbf{H}\left(\frac{\mathbf{H}^T\mathbf{H}}{\sigma^2} + \mathbf{C}_h^{-1}\right)^{-1}\mathbf{H}^T$$

so that

$$\begin{aligned} T(\mathbf{x}) &= \mathbf{x}^T\mathbf{H}\mathbf{C}_h\mathbf{H}^T\left[\frac{1}{\sigma^2}\mathbf{I} - \frac{1}{\sigma^4}\mathbf{H}\left(\frac{\mathbf{H}^T\mathbf{H}}{\sigma^2} + \mathbf{C}_h^{-1}\right)^{-1}\mathbf{H}^T\right]\mathbf{x} \\ &= \mathbf{x}^T\mathbf{H}\underbrace{\left[\frac{\mathbf{C}_h}{\sigma^2} - \frac{\mathcal{E}}{\sigma^4}\mathbf{C}_h\left(\frac{\mathcal{E}}{\sigma^2}\mathbf{I} + \mathbf{C}_h^{-1}\right)^{-1}\right]}_{\mathbf{E}}\mathbf{H}^T\mathbf{x} \end{aligned}$$

where we have used (5.34). But $\mathbf{E}$ is a diagonal matrix with $[i,i]$th element

$$[\mathbf{E}]_{ii} = \frac{1}{\mathcal{E} + \frac{\sigma^2}{\sigma_i^2}}$$

for $i = 0, 1, \ldots, p-1$. Also

$$\mathbf{H}^T\mathbf{x} = \begin{bmatrix} \sum_{n=0}^{K-1} x[n]u[n] \\ \sum_{n=1}^{K} x[n]u[n-1] \\ \vdots \\ \sum_{n=p-1}^{K+p-2} x[n]u[n-(p-1)] \end{bmatrix}$$

which is recognized as a correlation of the data $x[n]$ with each delayed replica of the input signal $u[n]$. Letting

$$z[k] = \sum_{n=k}^{K-1+k} x[n]u[n-k]$$

we finally have that

$$T(\mathbf{x}) = \sum_{k=0}^{p-1} \frac{1}{\mathcal{E} + \frac{\sigma^2}{\sigma_k^2}} z^2[k]$$

or we decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \sum_{k=0}^{p-1} \frac{\mathcal{E}\sigma_k^2}{\mathcal{E}\sigma_k^2 + \sigma^2} \left(\frac{z[k]}{\sqrt{\mathcal{E}}}\right)^2 > \gamma''. \tag{5.35}$$

The overall detector structure is shown in Figure 5.12. It may be viewed as an *optimal incoherent multipath combiner*. For strong paths or $\mathcal{E}\sigma_k^2 \gg \sigma^2$, we weight the squared-correlator outputs more heavily. The weighting coefficients are actually the Wiener filter weights. This follows from

$$\begin{aligned} z[k] &= \sum_{n=k}^{K-1+k} (s[n] + w[n])u[n-k] \\ &= \sum_{n=k}^{K-1+k} \left(\sum_{l=0}^{p-1} h[l]u[n-l] + w[n]\right) u[n-k] \\ &= \sum_{l=0}^{p-1} h[l] \sum_{n=k}^{K-1+k} u[n-l]u[n-k] + \sum_{n=k}^{K-1+k} w[n]u[n-k] \\ &\approx \mathcal{E}h[k] + \sum_{n=k}^{K-1+k} w[n]u[n-k] \end{aligned}$$

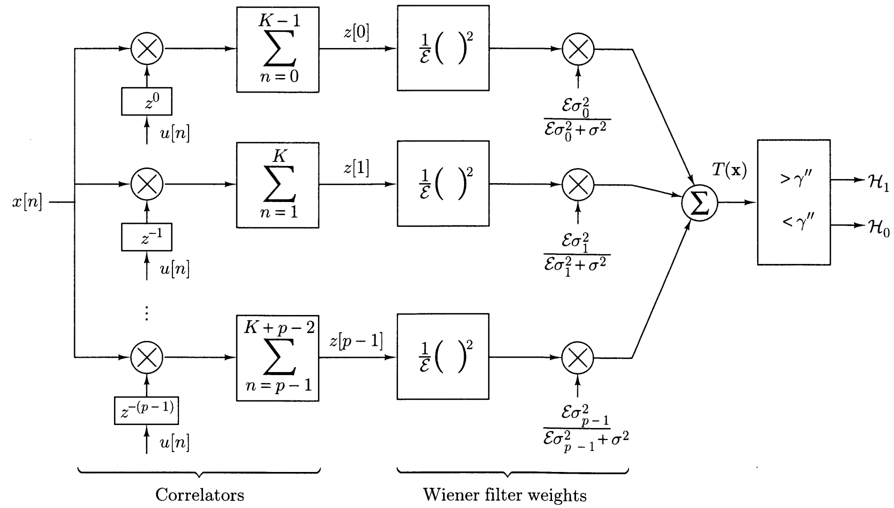


Figure 5.12. Optimal multipath combiner for random TDL channel with pseudorandom signal input.

due to the approximate orthogonality of the columns of $\mathbf{H}$. As a result,

$$\frac{z[k]}{\sqrt{\mathcal{E}}} = \sqrt{\mathcal{E}}h[k] + \frac{1}{\sqrt{\mathcal{E}}} \sum_{n=k}^{K-1+k} w[n]u[n-k] \tag{5.36}$$

and

$$\begin{aligned} \operatorname{var}(\sqrt{\mathcal{E}}h[k]) &= \mathcal{E}\sigma_k^2 \\ \operatorname{var}\left(\frac{1}{\sqrt{\mathcal{E}}} \sum_{n=k}^{K-1+k} w[n]u[n-k]\right) &= \frac{1}{\mathcal{E}}\sigma^2 \sum_{n=k}^{K-1+k} u^2[n-k] = \sigma^2. \end{aligned}$$

The interpretation of $\mathcal{E}\sigma_k^2/(\mathcal{E}\sigma_k^2 + \sigma^2)$ as the Wiener filter weight for the kth delay follows from these results.

The detection performance may be obtained by noting that $T(\mathbf{x})$ is in the canonical form of (5.9) and then using the results of Example 5.4. This assumes that the σ_k^2's occur in distinct pairs. As a simple example, let $p = 4$ with $\sigma_0^2 = \sigma_3^2 = 1/6$, $\sigma_1^2 = \sigma_2^2 = 1/3$. This delay scattering function is shown in Figure 5.13. Note that the *average* received signal energy is $\bar{\mathcal{E}} = E(\sum_{n=0}^{N-1} s^2[n])$. This can be evaluated as

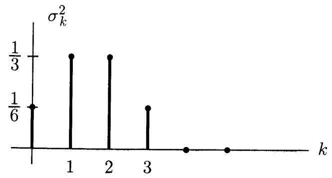


Figure 5.13. Delay scattering function or multipath profile.

follows:

$$\begin{aligned}\bar{\mathcal{E}} &= E(\mathbf{s}^T\mathbf{s}) = \mathrm{tr}[E(\mathbf{s}\mathbf{s}^T)] \\ &= \mathrm{tr}[E(\mathbf{H}\mathbf{h}\mathbf{h}^T\mathbf{H})] = \mathrm{tr}(\mathbf{H}\mathbf{C}_h\mathbf{H}^T) \\ &= \mathrm{tr}(\mathbf{C}_h\mathbf{H}^T\mathbf{H}) \approx \mathcal{E}\sum_{k=0}^{p-1}\sigma_k^2\end{aligned}$$

where we have used (5.31), (5.32), and (5.34). For the given tap weight variances of this example we have $\bar{\mathcal{E}} = \mathcal{E}$. It is easily shown from (5.36) that if we let $y[k] = z[k]/\sqrt{\mathcal{E}}$, then $\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ under $\mathcal{H}_0$ and $\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathcal{E}\mathbf{C}_h + \sigma^2\mathbf{I})$ under $\mathcal{H}_1$ so that the derivation in Appendix 5A applies with $\mathbf{C}_s = \mathcal{E}\mathbf{C}_h$. Thus, making use of (5.14) and (5.15) with $\lambda_{s_0} = \mathcal{E}\sigma_0^2$ and $\lambda_{s_1} = \mathcal{E}\sigma_1^2$ for $M = 2$, it follows that

$$P_{FA} = A_0 \exp\left(-\frac{\gamma''}{2\alpha_0}\right) + A_1 \exp\left(-\frac{\gamma''}{2\alpha_1}\right)$$

where

$$\begin{aligned}\alpha_0 &= \frac{\mathcal{E}\sigma_0^2\sigma^2}{\mathcal{E}\sigma_0^2 + \sigma^2} \\ \alpha_1 &= \frac{\mathcal{E}\sigma_1^2\sigma^2}{\mathcal{E}\sigma_1^2 + \sigma^2} \\ A_0 &= \frac{1}{1 - \frac{\alpha_1}{\alpha_0}} \\ A_1 &= \frac{1}{1 - \frac{\alpha_0}{\alpha_1}} = 1 - A_0.\end{aligned}$$

This simplifies to

$$P_{FA} = -\frac{\frac{\mathcal{E}}{\sigma^2} + 3}{3}\exp\left[-\frac{\gamma''\left(\frac{\mathcal{E}}{\sigma^2} + 6\right)}{2\mathcal{E}}\right] + \frac{\frac{\mathcal{E}}{\sigma^2} + 6}{3}\exp\left[-\frac{\gamma''\left(\frac{\mathcal{E}}{\sigma^2} + 3\right)}{2\mathcal{E}}\right]$$

or letting $\gamma''' = \gamma''/\mathcal{E}$ and noting that $\bar{\mathcal{E}} = \mathcal{E}$

$$P_{FA} = -\frac{\frac{\bar{\mathcal{E}}}{\sigma^2} + 3}{3}\exp\left[-\frac{\gamma'''\left(\frac{\bar{\mathcal{E}}}{\sigma^2} + 6\right)}{2}\right] + \frac{\frac{\bar{\mathcal{E}}}{\sigma^2} + 6}{3}\exp\left[-\frac{\gamma'''\left(\frac{\bar{\mathcal{E}}}{\sigma^2} + 3\right)}{2}\right]. \tag{5.37}$$

Also,

$$P_D = B_0 \exp\left(-\frac{\gamma''}{2\mathcal{E}\sigma_0^2}\right) + B_1 \exp\left(-\frac{\gamma''}{2\mathcal{E}\sigma_1^2}\right)$$

where

$$\begin{aligned}B_0 &= \frac{1}{1 - \frac{\sigma_1^2}{\sigma_0^2}} \\ B_1 &= \frac{1}{1 - \frac{\sigma_0^2}{\sigma_1^2}} = 1 - B_0.\end{aligned}$$

This simplifies to

$$P_D = -\exp(-3\gamma''') + 2\exp(-3\gamma'''/2). \tag{5.38}$$

The detection performance for the TDL channel is similar to that of the Rayleigh fading channel (see Figure 5.7). This is a slight degradation at lower average ENRs

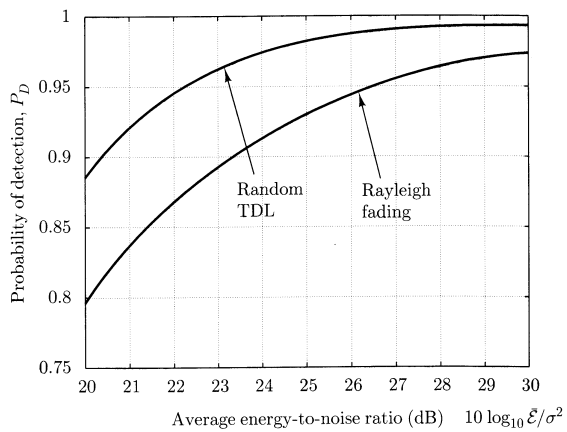


Figure 5.14. Detection performance for PRN signal transmitted via random TDL signal channel.

but a marked improvement at higher ones. As an example, in Figure 5.14, we plot P_D for both channels for $P_{FA} = 10^{-5}$. To find the threshold from (5.37) for each $\bar{\mathcal{E}}/\sigma^2$ we used a fixed point iteration (similar to Problem 5.1). We see that for the sample average ENR the detection performance is greatly improved. For example, for $P_D = 0.95$ we require 4.3 dB less average ENR for the PRN sequence. The difference can be attributed to the wideband signal that yields frequency diversity or equivalency the incoherent averaging of the resolvable multipaths.

References

Johnson, N.L., S. Kotz, N. Balakrishnan, *Continuous Univariate Distributions*, Vol. 1, J. Wiley, New York, 1995.

Kay, S.M., *Modern Spectral Estimation: Theory and Application*, Prentice-Hall, Englewood Cliffs, N.J., 1988.

MacWilliams, F.J., N.J. Sloane, "Pseudo-Random Sequences and Arrays," *Proc. IEEE*, Vol. 64, pp. 1715–1729, Dec. 1976.

Pahlavan, K, A.H. Levesque, "Wireless Data Communications," *Proc. IEEE*, Vol. 82, pp. 1398–1430, Sept. 1994.

Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1965.

Proakis, J.G., *Digital Communications, 2nd Ed.*, McGraw-Hill, New York, 1989.

Van Trees, H.L., *Detection, Estimation, and Modulation Theory*, Vol. III, J. Wiley, New York, 1971.

Problems

5.1 To find the threshold for an energy detector we use (5.2). For the case of N even (for N odd we can employ a similar approach)

$$P_{FA} = \exp\left(-\frac{\gamma'}{2\sigma^2}\right)\left[1 + \sum_{r=1}^{\frac{N}{2}-1} \frac{\left(\frac{\gamma'}{2\sigma^2}\right)^r}{r!}\right].$$

By letting $\gamma'' = \gamma'/2\sigma^2$ and rearranging terms we have

$$\gamma'' = -\ln P_{FA} + \ln\left[1 + \sum_{r=1}^{\frac{N}{2}-1} \frac{(\gamma'')^r}{r!}\right].$$

To solve for γ'' we can use the fixed point iteration

$$\gamma''_{k+1} = -\ln P_{FA} + \ln\left[1 + \sum_{r=1}^{\frac{N}{2}-1} \frac{(\gamma''_k)^r}{r!}\right].$$

For $P_{FA} = 10^{-3}$ and $N = 6$ find the threshold γ' by iterating with $\gamma''_0 = 1$.

5.2 A χ^2_N random variable can be thought of as the sum of the squares of N $\mathcal{N}(0,1)$ independent random variables. Therefore, for large N it can be approximated by a Gaussian random variable. This is due to the central limit theorem. First find an approximation for $Q_{\chi^2_N}(x)$ for N large. Then, based on this approximation verify that the performance of an energy detector is given by

$$P_D \approx Q\left(\frac{Q^{-1}(P_{FA}) - \sqrt{\frac{N}{2}}\frac{\sigma_s^2}{\sigma^2}}{\frac{\sigma_s^2}{\sigma^2} + 1}\right).$$

5.3 Find the NP detector for the problem of a random signal $s[n]$ with mean zero and covariance matrix $\mathbf{C}_s = \operatorname{diag}(\sigma^2_{s_0}, \sigma^2_{s_1}, \ldots, \sigma^2_{s_{N-1}})$ embedded in WGN with variance σ^2. Assume that the data samples observed are $x[n]$ for $n = 0, 1, \ldots, N-1$. Do not explicitly evaluate the threshold.

5.4 We wish to detect a random DC level A embedded in WGN with variance σ^2. Under $\mathcal{H}_0$, $x[n] = w[n]$ and under $\mathcal{H}_1$, $x[n] = A + w[n]$ for $n = 0, 1, \ldots, N-1$, where $A \sim \mathcal{N}(0, \sigma_A^2)$. Find the NP detector by first showing that the MMSE estimator of the signal is

$$\hat{\mathbf{s}} = \hat{A}\mathbf{1} = \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}}\bar{x}\mathbf{1}$$

where $\bar{x}$ is the sample mean and $\mathbf{1}$ is an $N \times 1$ vector of all ones. Do not explicitly evaluate the threshold. Hint: Use Woodbury's identity given in Appendix 1.

5.5 In this problem we prove that the *linear* MMSE estimator or Wiener filter is given by $\hat{\mathbf{s}} = \mathbf{C}_s(\mathbf{C}_s + \sigma^2\mathbf{I})^{-1}\mathbf{x}$. If $\mathbf{s}$ and $\mathbf{w}$ are jointly Gaussian, the MMSE estimator happens to be linear, and hence $\hat{\mathbf{s}}$ is also the MMSE estimator, as we have asserted. To do so let $\hat{\mathbf{s}} = \mathbf{W}\mathbf{x}$, where $\mathbf{W}$ is an $N \times N$ matrix. We choose $\mathbf{W}$ to minimize

$$J = E\left[(\boldsymbol{\alpha}^T(\mathbf{s} - \hat{\mathbf{s}}))^2\right]$$

for an arbitrary $N \times 1$ vector $\boldsymbol{\alpha}$. By minimizing this function we will minimize the MSE for each component of $\hat{\mathbf{s}}$ since, for example, if $\boldsymbol{\alpha} = [1\,0\,0\ldots0]^T$, then

$J = E[(s[0] - \hat{s}[0])^2]$. To effect the minimization we employ a calculus of variations approach. We let $\mathbf{W} = \mathbf{W}_{\text{opt}} + \epsilon\delta\mathbf{W}$, where $\mathbf{W}_{\text{opt}}$ is the optimal value of $\mathbf{W}$ and $\delta\mathbf{W}$ is an arbitrary $N \times N$ matrix variation. Then, since the optimal value of $\mathbf{W}$ minimizes J we must have

$$\left.\frac{\partial J}{\partial \epsilon}\right|_{\epsilon=0} = 0.$$

Show that this yields the result

$$\boldsymbol{\alpha}^T E[(\mathbf{s} - \mathbf{W}_{\text{opt}}\mathbf{x})\mathbf{x}^T]\delta\mathbf{W}^T\boldsymbol{\alpha} = 0.$$

Next note that $\boldsymbol{\alpha}$ is arbitrary as is $\delta\mathbf{W}$. Hence, it must be true that $E[(\mathbf{s} - \mathbf{W}_{\text{opt}}\mathbf{x})\mathbf{x}^T] = \mathbf{0}$, since for the proper choice of $\boldsymbol{\alpha}$ and $\delta\mathbf{W}$ we can generate any element of the matrix. For example, if $\boldsymbol{\alpha} = [1\,0\,0\ldots0]^T$ and $\delta\mathbf{W}^T\boldsymbol{\alpha} = [0\,1\,0\,0\ldots0]^T$, we have the [1, 2] element of $E[(\mathbf{s} - \mathbf{W}_{\text{opt}}\mathbf{x})\mathbf{x}^T]$. Finally, derive the Wiener filter.

5.6 We wish to detect the $N \times 1$ Gaussian random signal $\mathbf{s}$ with zero mean and covariance matrix $\mathbf{C}_s = (\sigma_s^2/2)(\mathbf{1}\mathbf{1}^T + \mathbf{1}^-\mathbf{1}^{-T})$ in WGN with variance σ^2. The vectors are defined as $\mathbf{1} = [1\,1\,1\ldots1]^T$ and $\mathbf{1}^- = [1\,(-1)\,1\ldots(-1)]^T$ (the latter assumes N is even). If $N = 10$, plot P_D versus σ_s^2/σ^2 for $P_{FA} = 10^{-2}$. Hint: The eigenvectors of $\mathbf{C}_s$ with nonzero eigenvalues are $\mathbf{1}/\sqrt{N}$ and $\mathbf{1}^-/\sqrt{N}$ and use (5.10) and (5.11).

5.7 Find the detection statistic in Example 5.3 if $\rho \to 1$. Explain why $y[1]$ is discarded. Hint: Show that $s[1] \to s[0]$ as $\rho \to 1$.

5.8 Let $\mathbf{W} = \mathbf{C}_s(\mathbf{C}_s + \sigma^2\mathbf{I})^{-1}$ be the Wiener filter matrix. Show that the effect of the filter on an input vector $\mathbf{x}$ is equivalent to first transforming $\mathbf{x}$ to $\mathbf{y} = \mathbf{V}^T\mathbf{x}$, filtering each component of $\mathbf{y}$ by $\lambda_{s_n}/(\lambda_{s_n} + \sigma^2)$ to form $\mathbf{y}'$, and then transforming back to the original space using $\mathbf{x}' = \mathbf{V}\mathbf{y}'$. The signal covariance matrix has the modal matrix $\mathbf{V}$ and eigenvalues λ_{s_n}.

5.9 For the canonical detector of (5.9) assume that $N = 4$ and that the signal covariance matrix has eigenvalues $\{2, 2, 1, 1\}$. Find P_{FA} and P_D if $\sigma^2 = 1$.

5.10 If under $\mathcal{H}_0$ we have $\mathbf{x} = \mathbf{w}$ with $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_w)$ and under $\mathcal{H}_1$ we have $\mathbf{x} = \mathbf{s} + \mathbf{w}$ with $\mathbf{s} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_s)$ and the same PDF for $\mathbf{w}$, show that the NP detector is given by (5.16) and (5.17). The signal and noise random vectors are assumed to be independent of each other.

5.11 A canonical form for the detector of (5.16), (5.17) or

$$T(\mathbf{x}) = \mathbf{x}^T\mathbf{C}_w^{-1}\mathbf{C}_s(\mathbf{C}_s + \mathbf{C}_w)^{-1}\mathbf{x}$$

can be derived by employing some results from linear algebra. To do so we need to diagonalize two symmetric matrices by a single linear transformation. In this problem we show how this is done. First define the square-root of the eigenvalue matrix of $\mathbf{C}_w$ as $\sqrt{\mathbf{\Lambda}_w} = \operatorname{diag}(\sqrt{\lambda_{w_0}}, \sqrt{\lambda_{w_1}}, \ldots, \sqrt{\lambda_{w_{N-1}}})$ and let $\mathbf{V}_w$ by the modal matrix. Show that $\mathbf{A}^T\mathbf{C}_w\mathbf{A} = \mathbf{I}$ if $\mathbf{A} = \mathbf{V}_w\sqrt{\mathbf{\Lambda}_w}^{-1}$. Next let the modal matrix of $\mathbf{B} = \mathbf{A}^T\mathbf{C}_s\mathbf{A}$ be $\mathbf{V}_B$ and the eigenvalue matrix be $\mathbf{\Lambda}_B$. Note that $\mathbf{B}$ is symmetric so that $\mathbf{V}_B$ is an orthogonal matrix. Then show that $T(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{B}(\mathbf{B} + \mathbf{I})^{-1}\mathbf{A}^T\mathbf{x}$ so that by letting $\mathbf{y} = \mathbf{V}_B^T\mathbf{A}^T\mathbf{x} = \mathbf{V}_B^T\sqrt{\mathbf{\Lambda}_w}^{-1}\mathbf{V}_w^T\mathbf{x}$ we have

$$T(\mathbf{y}) = \mathbf{y}^T\mathbf{\Lambda}_B(\mathbf{\Lambda}_B + \mathbf{I})^{-1}\mathbf{y}.$$

Consequently, we have the test statistic for the correlated noise case and

$$T(\mathbf{y}) = \sum_{n=0}^{N-1} \frac{\lambda_{B_n}}{\lambda_{B_n} + 1} y^2[n]$$

which is referred to as the canonical form. Note that the matrix $\mathbf{D} = \mathbf{A}\mathbf{V}_B = \mathbf{V}_w\sqrt{\mathbf{\Lambda}_w}^{-1}\mathbf{V}_B$ simultaneously diagonalizes $\mathbf{C}_s$ and $\mathbf{C}_w$ since $\mathbf{D}^T\mathbf{C}_s\mathbf{D} = \mathbf{\Lambda}_B$ and $\mathbf{D}^T\mathbf{C}_w\mathbf{D} = \mathbf{I}$.

5.12 We wish to detect the signal $s[n] = Ar^n$ for $n = 0, 1, \ldots, N-1$, where $0 < r < 1$ and $A \sim \mathcal{N}(0, \sigma_A^2)$ in WGN with variance σ^2. A and $w(n)$ are independent of each other. Find the NP test statistic. Hint: Use Woodbury's identity in Appendix 1.

5.13 A rank one signal is a random signal whose mean is zero and whose covariance matrix has rank one. As such the covariance matrix can be written as $\mathbf{C}_s = \mathbf{u}\mathbf{u}^T$, where $\mathbf{u}$ is an $N \times 1$ vector. Show that the signal $s[n] = Ah[n]$ for $n = 0, 1, \ldots, N-1$, where $h[n]$ is a deterministic sequence and A is a random variable with $E(A) = 0$ and $\operatorname{var}(A) = \sigma_A^2$ is a rank one signal.

5.14 Show the NP detector for a Gaussian rank one signal (defined in Problem 5.13 as a Gaussian random process whose covariance matrix is $\mathbf{C}_s = \sigma_A^2\mathbf{h}\mathbf{h}^T$) embedded in WGN with variance σ^2 can be written as

$$T'(\mathbf{x}) = \left(\sum_{n=0}^{N-1} x[n]h[n]\right)^2.$$

Also, determine P_{FA} and P_D. Hint: The test statistic is a scaled χ_1^2 random variable under $\mathcal{H}_0$ and $\mathcal{H}_1$.

5.15 Show that $\hat{\boldsymbol{\theta}} = \mathbf{C}_\theta\mathbf{H}^T(\mathbf{H}\mathbf{C}_\theta\mathbf{H}^T + \sigma^2\mathbf{I})^{-1}\mathbf{x}$ in (5.18) is the MMSE estimator of $\boldsymbol{\theta}$ by using (5.7).

5.16 Find $\hat{s}[n]$ for the Rayleigh fading sinusoid of Example 5.5. You may assume that N is large. Hint: Note that $T(\mathbf{x}) = \mathbf{x}^T\hat{\mathbf{s}}$, where $\hat{\mathbf{s}} = [\hat{s}[0]\ \hat{s}[1], \ldots, \hat{s}[N-1]]^T$.

5.17 Consider the problem of detecting the signal $s[n] = A\cos(2\pi f_0 n + \phi)$ in WGN with variance σ^2. Assume that A is known but that ϕ is not. The matched filter

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n] A \cos 2\pi f_0 n$$

is used, where it has been assumed, incorrectly, that $\phi = 0$. Find the deflection coefficient as a function of ϕ and discuss. You may assume that N is large.

5.18 Find P_{FA} and P_D for the Rayleigh fading sinusoid of Example 5.5 by deriving the PDF of (5.20). You should first show that for large N

$$\begin{aligned} I &= \sum_{n=0}^{N-1} x[n]\cos 2\pi f_0 n \sim \begin{cases} \mathcal{N}(0, \frac{N\sigma^2}{2}) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(0, \frac{N\sigma^2}{2} + \frac{N^2\sigma_s^2}{4}) & \text{under } \mathcal{H}_1 \end{cases} \\ Q &= \sum_{n=0}^{N-1} x[n]\sin 2\pi f_0 n \sim \begin{cases} \mathcal{N}(0, \frac{N\sigma^2}{2}) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(0, \frac{N\sigma^2}{2} + \frac{N^2\sigma_s^2}{4}) & \text{under } \mathcal{H}_1 \end{cases} \end{aligned}$$

and I and Q are uncorrelated under either hypothesis and hence independent (since they are jointly Gaussian).

5.19 In this problem we show that the Fourier transform of a segment of white noise is uncorrelated at widely separated frequencies. To do so let

$$\begin{aligned} W_c(f) &= \sum_{n=0}^{N-1} w[n]\cos 2\pi f n \\ W_s(f) &= -\sum_{n=0}^{N-1} w[n]\sin 2\pi f n \end{aligned}$$

and $\mathbf{W}(f) = [W_c(f)\ W_s(f)]^T$ and show that $E(\mathbf{W}(f_0)\mathbf{W}^T(f_1)) \approx \mathbf{0}$ if $|f_1 - f_0| \gg 1/N$. Then let $W(f) = \sum_{n=0}^{N-1} w[n]\exp(-j2\pi f n) = W_c(f) + jW_s(f)$, and show that $E(W^*(f_0)W(f_1)) \approx 0$. Note that if the noise is Gaussian, the Fourier transform outputs will be approximately independent. Hint: Use the hint in Problem 4.21.

5.20 Consider the estimator-correlator in its canonical form of (5.9). For N large, the eigendecomposition of $\mathbf{C}_s$ for $\mathbf{C}_s$ a symmetric Toeplitz matrix is (see Chapter 2)

$$\begin{aligned} \lambda_{s_i} &= P_{ss}(f_i) \\ \mathbf{v}_i &= \frac{1}{\sqrt{N}}[\ 1\quad \exp(j2\pi f_i) \ldots \exp(j2\pi f_i(N-1))\ ]^T \end{aligned}$$

for $i = 0, 1, \ldots, N-1$ and where $f_i = i/N$. Use this to show that (5.5) can be approximated by (5.27). Assume for simplicity that N is even, although the result holds for any N. Hint: Note that $P_{ss}(f_{N-i}) = P_{ss}(f_i)$. Also, $\mathbf{v}_{N-i} = \mathbf{v}_i^*$ so that $(\mathbf{v}_i^T\mathbf{x})^2 + (\mathbf{v}_{N-i}^T\mathbf{x})^2 = 2|\mathbf{v}_i^T\mathbf{x}|^2$.

5.21 Discuss the detector given by (5.27) if $P_{ss}(f) \ll \sigma^2$ for all f. Also, show that under this assumption

$$T(\mathbf{x}) = \frac{N}{\sigma^2}\sum_{k=-(N-1)}^{N-1} r_{ss}[k]\hat{r}_{xx}[k]$$

where $r_{ss}[k] = \mathcal{F}^{-1}\{P_{ss}(f)\}$ is the ACF of $s[n]$ and

$$\hat{r}_{xx}[k] = \frac{1}{N}\sum_{n=0}^{N-1-|k|} x[n]x[n+|k|].$$

Hint: First show that

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} I(f)\exp(j2\pi f k)df = \begin{cases} \hat{r}_{xx}[k] & |k| \le N-1 \\ 0 & |k| \ge N. \end{cases}$$

5.22 The asymptotic NP detector was defined in (5.27). If the signal has the PSD

$$P_{ss}(f) = \begin{cases} P_0 & 0 \le f \le \frac{1}{4} \\ 0 & \frac{1}{4} < f \le \frac{1}{2} \end{cases}$$

find the test statistic and explain the operation of the detector.

5.23 Find the NP detector for the general Gaussian problem where $\mathbf{C}_w$ is the covariance matrix of the noise and $\mathbf{C}_s = \eta\mathbf{C}_w$ for $\eta > 0$ is the covariance matrix of the signal. The means of the signal and noise are both zero. Next determine P_{FA} and P_D. Finally relate P_D to P_{FA} for $N = 2$. Hint: Note that if $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$, then $\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x} \sim \chi_N^2$, where $\mathbf{x}$ has dimension $N \times 1$ (see (2.27)).

5.24 In the signal processing example of Section 5.7, if $u[n] = \cos 2\pi f_0 n$ for $n \geq 0$, show that the signal at the channel output for $n \geq p-1$ is

$$A\cos(2\pi f_0 n + \phi) = a\cos 2\pi f_0 n + b \sin 2\pi f_0 n$$

where

$$\begin{aligned} a &= \sum_{k=0}^{p-1} h[k]\cos 2\pi f_0 k \\ b &= \sum_{k=0}^{p-1} h[k]\sin 2\pi f_0 k. \end{aligned}$$

Also, for large p show that

$$\begin{aligned} a &\sim \mathcal{N}\left(0, \frac{1}{2}\sum_{k=0}^{p-1}\sigma_k^2\right) \\ b &\sim \mathcal{N}\left(0, \frac{1}{2}\sum_{k=0}^{p-1}\sigma_k^2\right) \end{aligned}$$

and a and b are uncorrelated and hence independent (since they are jointly Gaussian). Hint: For the second part assume that the Fourier transform of the σ_k^2 sequence or of $\{\sigma_0^2, \sigma_1^2, \ldots, \sigma_{p-1}^2\}$ is small at $f = 2f_0$.

Appendix 5A

Detection Performance of the Estimator-Correlator

In this appendix we derive the detection performance for the estimator-correlator detector that decides $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \mathbf{x}^T\mathbf{C}_s(\mathbf{C}_s + \sigma^2\mathbf{I})^{-1}\mathbf{x} > \gamma''$$

where we assume $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ under $\mathcal{H}_0$ and $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_s + \sigma^2\mathbf{I})$ under $\mathcal{H}_1$. As shown previously (see (5.9))

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} \frac{\lambda_{s_n}}{\lambda_{s_n} + \sigma^2} y^2[n]$$

where $\mathbf{y} = \mathbf{V}^T\mathbf{x}$ for $\mathbf{V}$ the modal matrix of $\mathbf{C}_s$ and λ_{s_n} the nth eigenvalue of $\mathbf{C}_s$. Now under either hypothesis $\mathbf{y}$ is a Gaussian random vector, being a linear transformation of $\mathbf{x}$, and since $E(\mathbf{y}) = \mathbf{0}$ under either hypothesis

$$\mathbf{C}_y = E(\mathbf{y}\mathbf{y}^T) = E(\mathbf{V}^T\mathbf{x}\mathbf{x}^T\mathbf{V}) = \mathbf{V}^T\mathbf{C}_x\mathbf{V}$$

which becomes

$$\mathbf{C}_y = \begin{cases} \sigma^2\mathbf{I} & \text{under } \mathcal{H}_0 \\ \mathbf{\Lambda}_s + \sigma^2\mathbf{I} & \text{under } \mathcal{H}_1. \end{cases}$$

Thus

$$\mathbf{y} \sim \begin{cases} \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(\mathbf{0}, \mathbf{\Lambda}_s + \sigma^2\mathbf{I}) & \text{under } \mathcal{H}_1. \end{cases}$$

First consider the probability of false alarm.

$$P_{FA} = \Pr\{T(\mathbf{x}) > \gamma''; \mathcal{H}_0\}$$

$$= \Pr\left\{\sum_{n=0}^{N-1} \frac{\lambda_{s_n}}{\lambda_{s_n}+\sigma^2} y^2[n] > \gamma''; \mathcal{H}_0\right\}$$
$$= \Pr\left\{\sum_{n=0}^{N-1} \frac{\lambda_{s_n}\sigma^2}{\lambda_{s_n}+\sigma^2} z^2[n] > \gamma''; \mathcal{H}_0\right\}$$

where $z[n] = y[n]/\sigma$. We need to find the CDF of $T(\mathbf{x}) = \sum_{n=0}^{N-1} \alpha_n z^2[n]$, where the $z[n]$'s are IID $\mathcal{N}(0,1)$ random variables. We can use characteristic functions to accomplish this. If the characteristic function is defined as

$$\phi_x(\omega) = E[\exp(j\omega x)]$$

then by using the independence of the $z[n]$'s we have

$$\begin{aligned}
\phi_T(\omega) &= E[\exp(j\omega T)] \\
&= E\left[\exp\left(j\omega \sum_{n=0}^{N-1} \alpha_n z^2[n]\right)\right] \\
&= \prod_{n=0}^{N-1} E\left[\exp(j\omega\alpha_n z^2[n])\right] \\
&= \prod_{n=0}^{N-1} \phi_{z^2}(\alpha_n \omega).
\end{aligned}$$

The PDF of T is obtained as the inverse Fourier transform of the characteristic function or

$$p_T(t) = \begin{cases} \displaystyle\int_{-\infty}^{\infty} \phi_T(\omega)\exp(-j\omega t)\frac{d\omega}{2\pi} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

since $T \geq 0$ and hence $p_T(t) = 0$ for $t < 0$. Now $z^2[n] \sim \chi_1^2$ and the characteristic function can be shown to be [Johnson and Kotz 1995]

$$\phi_{\chi_1^2}(\omega) = \frac{1}{\sqrt{1-2j\omega}}.$$

Thus,

$$\phi_T(\omega) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{1-2j\alpha_n\omega}}$$

where

$$\alpha_n = \frac{\lambda_{s_n}\sigma^2}{\lambda_{s_n}+\sigma^2}$$

so that

$$P_{FA} = \int_{\gamma''}^{\infty}\int_{-\infty}^{\infty} \prod_{n=0}^{N-1} \frac{1}{\sqrt{1-2j\alpha_n\omega}} \exp(-j\omega t)\frac{d\omega}{2\pi} dt.$$

Similarly, we can show that

$$P_D = \int_{\gamma''}^{\infty}\int_{-\infty}^{\infty} \prod_{n=0}^{N-1} \frac{1}{\sqrt{1-2j\lambda_{s_n}\omega}} \exp(-j\omega t)\frac{d\omega}{2\pi} dt.$$

Chapter 6

Statistical Decision Theory II

6.1 Introduction

Previously, we had assumed complete knowledge of the PDFs under $\mathcal{H}_0$ and $\mathcal{H}_1$, allowing the design of optimal detectors. We now turn to the more realistic problem in which the PDF is not completely known. For example, the radar return from a target will be delayed by the propagation time of the signal through the medium. As a result, its arrival time is generally unknown. A communication receiver may not have perfect knowledge of the frequency of a transmitted signal. Thus, in each case the signal portion of the PDF will be unknown due to an unknown parameter. Similarly, the noise characteristics may not be known a priori. The noise might be reasonably modeled as white Gaussian noise but with an unknown variance. Such is the case in sonar in which the noise power depends upon environmental conditions, which are not known in advance. The design of good detectors when the PDFs have unknown parameters is therefore of great practical importance. In this chapter we review the common approaches to this hypothesis testing problem. In later chapters (Chapters 7–9) we will apply the theory to determine actual detectors for a wide range of practical problems. The material in this chapter and in Chapters 7–9 relies heavily on estimation theory. The reader may wish to review the referenced sections in [Kay-I 1993].

6.2 Summary

Example 6.1 illustrates the existence of a uniformly most powerful test, one that yields the maximum P_D for all values of the unknown parameter. When it exists, it is only for a one-sided parameter test. When such a test does not exist, we can use the Bayesian approach (6.10), which requires the specification of a prior PDF for the unknown parameters, or the generalized likelihood ratio test of (6.12). The Bayesian approach requires a multidimensional integration while the generalized likelihood ratio test requires the evaluation of the MLEs. The generalized likelihood ratio test

has the asymptotic PDF (as $N \to \infty$) given by (6.23) and (6.24). Other tests that are asymptotically equivalent to the generalized likelihood ratio test are the Wald test and the Rao test. They may be easier to compute in practice. For no nuisance parameters (parameters that are unknown but are of no interest) the Wald test is given by (6.30) and the Rao test by (6.31). The latter does not require an MLE evaluation. When nuisance parameters are present, the Wald and Rao tests are given by (6.34) and (6.35), respectively. The Rao test in this case only requires the MLE under $\mathcal{H}_0$ but not under $\mathcal{H}_1$. The asymptotic statistics of the Wald and Rao tests are identical to those of the generalized likelihood ratio test. For a one-sided scalar parameter test the locally most powerful test is given by (6.36) and its asymptotic performance by (6.37). This detector yields the maximum P_D for weak signals. When testing multiple hypotheses with unknown parameters, a test that generalizes the maximum likelihood rule is given by (6.40). Under certain conditions, it reduces to the minimum description length (6.41). These rules are especially useful for nested hypotheses, which occur in the fitting of models to data.

6.2.1 Summary of Composite Hypothesis Testing

The approaches to the *composite hypothesis testing* problem and their asymptotic performances are now summarized.

General Tests

The data $\mathbf{x}$ has the PDF $p(\mathbf{x};\boldsymbol{\theta}_0,\mathcal{H}_0)$ or $p(\mathbf{x}|\boldsymbol{\theta}_0;\mathcal{H}_0)$ under $\mathcal{H}_0$ and $p(\mathbf{x};\boldsymbol{\theta}_1,\mathcal{H}_1)$ or $p(\mathbf{x}|\boldsymbol{\theta}_1;\mathcal{H}_1)$ under $\mathcal{H}_1$. The forms of the PDFs as well as the dimensionalities of the unknown parameter vectors $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ may be different under each hypothesis.

1. Generalized likelihood ratio test (GLRT)

a. *Test*
Decide $\mathcal{H}_1$ if

$$L_G(\mathbf{x}) = \frac{p(\mathbf{x};\hat{\boldsymbol{\theta}}_1,\mathcal{H}_1)}{p(\mathbf{x};\hat{\boldsymbol{\theta}}_0,\mathcal{H}_0)} > \gamma$$

where $\hat{\boldsymbol{\theta}}_i$ is the MLE of $\boldsymbol{\theta}_i$ (maximizes $p(\mathbf{x};\boldsymbol{\theta}_i,\mathcal{H}_i)$).

b. *Asymptotic performance*
No general results.

2. Bayesian approach

a. *Test*
Decide $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x};\mathcal{H}_1)}{p(\mathbf{x};\mathcal{H}_0)} = \frac{\int p(\mathbf{x}|\boldsymbol{\theta}_1;\mathcal{H}_1)p(\boldsymbol{\theta}_1)d\boldsymbol{\theta}_1}{\int p(\mathbf{x}|\boldsymbol{\theta}_0;\mathcal{H}_0)p(\boldsymbol{\theta}_0)d\boldsymbol{\theta}_0} > \gamma$$

where $p(\mathbf{x};\mathcal{H}_i)$ is the unconditional data PDF, $p(\mathbf{x}|\boldsymbol{\theta}_i;\mathcal{H}_i)$ is the conditional data PDF and $p(\boldsymbol{\theta}_i)$ is the prior PDF.

b. *Asymptotic performance*
No general results.

Parameter Tests (Two-sided vector parameter)

No Nuisance Parameters

The PDF under $\mathcal{H}_0$ and $\mathcal{H}_1$ is the same except that the value of the unknown parameter vector is different. The PDF is denoted by $p(\mathbf{x};\boldsymbol{\theta})$. The hypothesis test is

$$\begin{aligned}\mathcal{H}_0 &: \boldsymbol{\theta} = \boldsymbol{\theta}_0\\ \mathcal{H}_1 &: \boldsymbol{\theta} \neq \boldsymbol{\theta}_0\end{aligned}$$

where $\boldsymbol{\theta}$ is $r \times 1$.

1. Generalized likelihood ratio test

a. *Test*
Decide $\mathcal{H}_1$ if

$$L_G(\mathbf{x}) = \frac{p(\mathbf{x};\hat{\boldsymbol{\theta}}_1)}{p(\mathbf{x};\boldsymbol{\theta}_0)} > \gamma$$

where $\hat{\boldsymbol{\theta}}_1$ is the MLE under $\mathcal{H}_1$ (maximizes $p(\mathbf{x};\boldsymbol{\theta})$).

b. *Asymptotic performance*

$$2\ln L_G(\mathbf{x}) \overset{a}{\sim} \begin{cases} \chi_r^2 & \text{under } \mathcal{H}_0 \\ \chi_r'^2(\lambda) & \text{under } \mathcal{H}_1 \end{cases}$$

where

$$\lambda = (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^T \mathbf{I}(\boldsymbol{\theta}_0)(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0).$$

$\mathbf{I}(\boldsymbol{\theta})$ denotes the Fisher information matrix and $\boldsymbol{\theta}_1$ is the true value under $\mathcal{H}_1$.

2. Wald test

a. *Test*
Decide $\mathcal{H}_1$ if

$$T_{\mathrm{W}}(\mathbf{x}) = (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0)^T \mathbf{I}(\hat{\boldsymbol{\theta}}_1)(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0) > \gamma.$$

b. *Asymptotic performance*
Same as GLRT.

3. Rao test

a. *Test*
Decide $\mathcal{H}_1$ if

$$T_{\mathrm{R}}(\mathbf{x}) = \left.\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|^T_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \mathbf{I}^{-1}(\boldsymbol{\theta}_0) \left.\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} > \gamma.$$

b. *Asymptotic performance*
Same as GLRT.

Nuisance Parameters

The PDF under $\mathcal{H}_0$ and $\mathcal{H}_1$ is the same except that the value of the unknown parameter vector is different. The parameter vector is $\boldsymbol{\theta} = [\boldsymbol{\theta}_r^T \ \boldsymbol{\theta}_s^T]^T$, where $\boldsymbol{\theta}_r$ is $r \times 1$ and $\boldsymbol{\theta}_s$ (the nuisance parameter vector) is $s \times 1$. The PDF is denoted by $p(\mathbf{x};\boldsymbol{\theta}_r,\boldsymbol{\theta}_s)$. The hypothesis test is

$$\begin{aligned}\mathcal{H}_0 &: \boldsymbol{\theta} = \boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s\\ \mathcal{H}_1 &: \boldsymbol{\theta} \neq \boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s\,.\end{aligned}$$

1. Generalized likelihood ratio test

a. *Test*
Decide $\mathcal{H}_1$ if

$$L_G(\mathbf{x}) = \frac{p(\mathbf{x};\hat{\boldsymbol{\theta}}_{r_1},\hat{\boldsymbol{\theta}}_{s_1})}{p(\mathbf{x};\boldsymbol{\theta}_{r_0},\hat{\boldsymbol{\theta}}_{s_0})} > \gamma$$

where $\hat{\boldsymbol{\theta}}_{r_1}, \hat{\boldsymbol{\theta}}_{s_1}$ is the MLE under $\mathcal{H}_1$ (unrestricted MLE, which is found by maximizing $p(\mathbf{x};\boldsymbol{\theta}_r,\boldsymbol{\theta}_s)$ over $\boldsymbol{\theta}_r,\boldsymbol{\theta}_s$), and $\boldsymbol{\theta}_{s_0}$ is the MLE under $\mathcal{H}_0$ (restricted MLE, which is found by maximizing $p(\mathbf{x};\boldsymbol{\theta}_{r_0},\boldsymbol{\theta}_s)$ over $\boldsymbol{\theta}_s$).

b. *Asymptotic performance*

$$2\ln L_G(\mathbf{x}) \overset{a}{\sim} \begin{cases} \chi_r^2 & \text{under } \mathcal{H}_0 \\ \chi_r'^2(\lambda) & \text{under } \mathcal{H}_1 \end{cases}$$

where

$$\begin{aligned}\lambda = {} & (\boldsymbol{\theta}_{r_1} - \boldsymbol{\theta}_{r_0})^T \Big[\mathbf{I}_{\theta_r\theta_r}(\boldsymbol{\theta}_{r_0},\boldsymbol{\theta}_s)\\ & - \mathbf{I}_{\theta_r\theta_s}(\boldsymbol{\theta}_{r_0},\boldsymbol{\theta}_s)\mathbf{I}^{-1}_{\theta_s\theta_s}(\boldsymbol{\theta}_{r_0},\boldsymbol{\theta}_s)\mathbf{I}_{\theta_s\theta_r}(\boldsymbol{\theta}_{r_0},\boldsymbol{\theta}_s)\Big](\boldsymbol{\theta}_{r_1} - \boldsymbol{\theta}_{r_0}).\end{aligned}$$

$\boldsymbol{\theta}_{r_1}$ is the true value under $\mathcal{H}_1$ and $\boldsymbol{\theta}_s$ is the true value, which is the same under either hypothesis. The Fisher information matrix is partitioned as

$$\begin{aligned}\mathbf{I}(\boldsymbol{\theta}) = \mathbf{I}(\boldsymbol{\theta}_r,\boldsymbol{\theta}_s) &= \begin{bmatrix} \mathbf{I}_{\theta_r\theta_r}(\boldsymbol{\theta}_r,\boldsymbol{\theta}_s) & \mathbf{I}_{\theta_r\theta_s}(\boldsymbol{\theta}_r,\boldsymbol{\theta}_s) \\ \mathbf{I}_{\theta_s\theta_r}(\boldsymbol{\theta}_r,\boldsymbol{\theta}_s) & \mathbf{I}_{\theta_s\theta_s}(\boldsymbol{\theta}_r,\boldsymbol{\theta}_s) \end{bmatrix}\\ &= \begin{bmatrix} r\times r & r\times s \\ s\times r & s\times s \end{bmatrix}.\end{aligned}$$

2. Wald test

a. *Test*
Decide $\mathcal{H}_1$ if

$$T_W(\mathbf{x}) = (\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0})^T \left(\left[\mathbf{I}^{-1}(\hat{\boldsymbol{\theta}}_1)\right]_{\theta_r\theta_r}\right)^{-1} (\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0}) > \gamma$$

where $\hat{\boldsymbol{\theta}}_1 = [\hat{\boldsymbol{\theta}}_{r_1}^T \, \hat{\boldsymbol{\theta}}_{s_1}^T]^T$ is the MLE under $\mathcal{H}_1$ and

$$\left[\mathbf{I}^{-1}(\boldsymbol{\theta})\right]_{\theta_r\theta_r} = \left(\mathbf{I}_{\theta_r\theta_r}(\boldsymbol{\theta}) - \mathbf{I}_{\theta_r\theta_s}(\boldsymbol{\theta})\mathbf{I}_{\theta_s\theta_s}^{-1}(\boldsymbol{\theta})\mathbf{I}_{\theta_s\theta_r}(\boldsymbol{\theta})\right)^{-1}.$$

b. *Asymptotic performance*
Same as GLRT.

3. Rao test

a. *Test*
Decide $\mathcal{H}_1$ if

$$T_R(\mathbf{x}) = \left.\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}^T \left[\mathbf{I}^{-1}(\tilde{\boldsymbol{\theta}})\right]_{\theta_r\theta_r} \left.\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} > \gamma$$

where $\tilde{\boldsymbol{\theta}} = [\boldsymbol{\theta}_{r_0}^T \, \hat{\boldsymbol{\theta}}_{s_0}^T]^T$ is the MLE of $\boldsymbol{\theta}$ under $\mathcal{H}_0$ (restricted MLE found by maximizing $p(\mathbf{x};\boldsymbol{\theta}_{r_0},\boldsymbol{\theta}_s)$ over $\boldsymbol{\theta}_s$) and $[\mathbf{I}^{-1}(\boldsymbol{\theta})]_{\theta_r\theta_r}$ has been defined in the previous two test descriptions.

b. *Asymptotic performance*
Same as GLRT.

Parameter Tests (One-Sided scalar parameter)

No Nuisance Parameters

This test is used for a one-sided hypothesis test whose PDF is the same under $\mathcal{H}_0$ and $\mathcal{H}_1$ but differs in the scalar parameter value. The PDF is denoted by $p(\mathbf{x};\theta)$. The hypothesis test is

$$\begin{aligned} \mathcal{H}_0 &: \theta = \theta_0 \\ \mathcal{H}_1 &: \theta > \theta_0. \end{aligned}$$

1. Locally most powerful test

a. *Test*
Decide $\mathcal{H}_1$ if

$$T_{\mathrm{LMP}}(\mathbf{x}) = \frac{\left.\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right|_{\theta=\theta_0}}{\sqrt{I(\theta_0)}} > \gamma$$

where $I(\theta)$ is the Fisher information.

b. *Asymptotic performance*

$$T_{\mathrm{LMP}}(\mathbf{x}) \stackrel{a}{\sim} \begin{cases} \mathcal{N}(0,1) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(\sqrt{I(\theta_0)}(\theta_1-\theta_0),1) & \text{under } \mathcal{H}_1 \end{cases}$$

where θ_1 is the value under $\mathcal{H}_1$.

6.3 Composite Hypothesis Testing

The general class of hypothesis testing problems that we will be interested in is the *composite hypothesis test*. As opposed to the simple hypothesis test in which the PDFs under both hypotheses are completely known, the composite hypothesis test must accommodate unknown parameters. The PDF under $\mathcal{H}_0$ or under $\mathcal{H}_1$ or under both hypotheses may not be completely specified. We express our uncertainty by including unknown parameters in the PDFs. For example, if we wish to detect a DC level with *unknown amplitude* A in WGN, then under $\mathcal{H}_1$ the PDF is

$$p(\mathbf{x};A,\mathcal{H}_1) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right]. \tag{6.1}$$

Since the amplitude A is unknown, the PDF is not completely specified. We will highlight this by including the unknown parameter A in the PDF description. Viewed slightly differently, the PDF under $\mathcal{H}_1$ belongs to a family of PDFs, one for each value of A. The PDF is said to be *parameterized* by A. In dealing with unknown parameters, the first step is to design an NP test as if A were known. Then, if possible, the test should be manipulated so that it does not depend on the value of A. The resulting test will be optimal since it is the NP test. An example follows.

Example 6.1 - DC Level in WGN with Unknown Amplitude ($A > 0$)

Consider the DC level in WGN detection problem

$$\begin{aligned} \mathcal{H}_0 &: x[n] = w[n] & n = 0,1,\ldots,N-1 \\ \mathcal{H}_1 &: x[n] = A + w[n] & n = 0,1,\ldots,N-1 \end{aligned} \tag{6.2}$$

where the value of A is unknown, although a priori we know that $A > 0$, and $w[n]$ is WGN with variance σ^2. Then, the NP test is to decide $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x};A,\mathcal{H}_1)}{p(\mathbf{x};\mathcal{H}_0)} = \frac{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right]}{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right]} > \gamma.$$

Taking the logarithm we have

$$-\frac{1}{2\sigma^2}\left(-2A\sum_{n=0}^{N-1}x[n]+NA^2\right)>\ln\gamma$$

or

$$A\sum_{n=0}^{N-1}x[n]>\sigma^2\ln\gamma+\frac{NA^2}{2}.$$

Since it is known that $A>0$, we have

$$\sum_{n=0}^{N-1}x[n]>\frac{\sigma^2}{A}\ln\gamma+\frac{NA}{2}.$$

Finally, scaling by $1/N$ produces the test

$$T(\mathbf{x})=\frac{1}{N}\sum_{n=0}^{N-1}x[n]>\frac{\sigma^2}{NA}\ln\gamma+\frac{A}{2}=\gamma'. \tag{6.3}$$

The key question is whether we can implement this detector without knowledge of the exact value of A. Clearly, the test statistic, which is the sample mean of the data, does not depend on A but it appears that the threshold γ' does. This dependence is only illusory as we will demonstrate. Recall from Chapter 3 that under $\mathcal{H}_0$, $T(\mathbf{x})=\bar{x}\sim\mathcal{N}(0,\sigma^2/N)$. Hence,

$$\begin{aligned}P_{FA}&=\Pr\left\{T(\mathbf{x})>\gamma';\mathcal{H}_0\right\}\\&=Q\left(\frac{\gamma'}{\sqrt{\sigma^2/N}}\right)\end{aligned}$$

so that

$$\gamma'=\sqrt{\frac{\sigma^2}{N}}Q^{-1}(P_{FA})$$

which is independent of A. Since the PDF of $T(\mathbf{x})$ under $\mathcal{H}_0$ does not depend on A, the threshold, which is chosen to maintain a constant P_{FA}, can be found and will not depend on A. Also, since the test is actually the NP detector (see Example 3.2), it is optimal in that it yields the highest P_D for a given P_{FA}. Note, however, that P_D will depend on the value of A. More specifically,

$$P_D=\Pr\left\{T(\mathbf{x})>\gamma';\mathcal{H}_1\right\}.$$

But under $\mathcal{H}_1$, $T(\mathbf{x})=\bar{x}\sim\mathcal{N}(A,\sigma^2/N)$ so that

$$\begin{aligned}P_D&=Q\left(\frac{\gamma'-A}{\sqrt{\sigma^2/N}}\right)\\&=Q\left(Q^{-1}(P_{FA})-\sqrt{\frac{NA^2}{\sigma^2}}\right).\end{aligned}$$

As expected, P_D increases with increasing A. This is illustrated in Figure 6.1. We may say that over all possible detectors that have a given P_{FA} the one that decides $\mathcal{H}_1$ if

$$\frac{1}{N}\sum_{n=0}^{N-1}x[n]>\sqrt{\frac{\sigma^2}{N}}Q^{-1}(P_{FA}) \tag{6.4}$$

yields the highest P_D for any value of A, as long as $A>0$. This type of test, when it

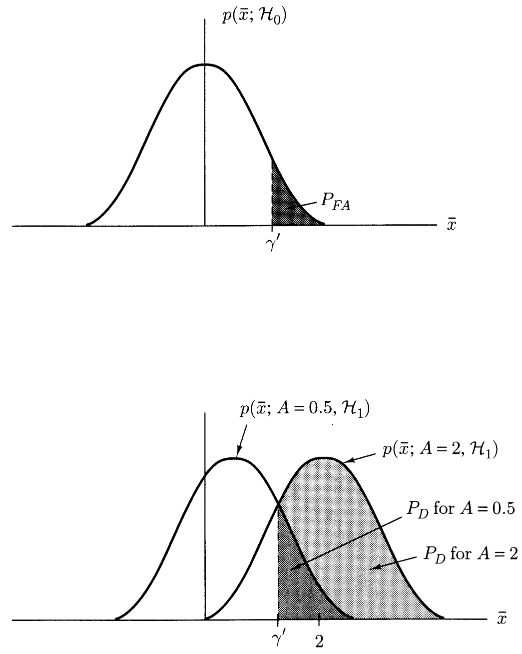


Figure 6.1. Dependence of probability of detection on unknown parameter A.

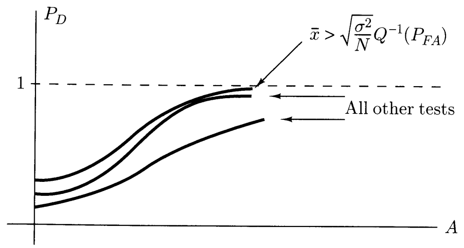


Figure 6.2. Comparison of performance of UMP test to all others with same P_{FA}.

exists, is called a *uniformly most powerful* (UMP) test. Any other test has a poorer performance as shown in Figure 6.2. Unfortunately, UMP tests seldom exist. If, for instance, A could take on any value or $-\infty < A < \infty$, then we would obtain different tests for A positive and A negative. If $A > 0$, we have (6.4) but if $A < 0$, we would decide $\mathcal{H}_1$ if (see Problem 6.1)

$$\frac{1}{N}\sum_{n=0}^{N-1} x[n] < -\sqrt{\frac{\sigma^2}{N}}Q^{-1}(P_{FA}). \tag{6.5}$$

Since the value of A is unknown, the NP approach does not result in a *unique* test.

The hypothesis testing problem of (6.2) can be recast as the *parameter testing problem*

$$\begin{aligned} \mathcal{H}_0 &: A = 0 \\ \mathcal{H}_1 &: A > 0. \end{aligned}$$

Such a test is called a *one-sided* test as opposed to

$$\begin{aligned} \mathcal{H}_0 &: A = 0 \\ \mathcal{H}_1 &: A \neq 0 \end{aligned} \tag{6.6}$$

which is a *two-sided* test. For a UMP test to exist the parameter test must be one-sided [Kendall and Stuart 1979]. Thus, two-sided testing problems never produce UMP tests but the one-sided problem may. See also Problems 6.2 and 6.3.

◇

When a UMP test does not exist, we are forced to implement suboptimal tests. It is of interest to compare the performance of the suboptimal test with that of the optimal NP test. Although the latter will be unrealizable (since we don't know the value of the unknown parameters), its performance can be used as an upper bound. If our suboptimal detector has performance close to that of the NP detector, then there is little motivation in examining other suboptimal detectors. This situation

is analogous to the use of the Cramer-Rao lower bound on estimator variance as a comparison tool on the performance of a suboptimal estimator (see [Kay-I 1993, Chapter 3]). A detector that assumes perfect knowledge of an unknown parameter to design the NP detector is referred to as a *clairvoyant detector*. Its performance is used as an upper bound. We now continue Example 6.1.

Example 6.2 - DC Level in WGN with Unknown Amplitude

When A can take on positive and negative values, the clairvoyant detector decides $\mathcal{H}_1$ if

$$\begin{aligned} \frac{1}{N}\sum_{n=0}^{N-1} x[n] &= \bar{x} > \gamma'_+ \quad \text{for } A > 0 \\ \frac{1}{N}\sum_{n=0}^{N-1} x[n] &= \bar{x} < \gamma'_- \quad \text{for } A < 0. \end{aligned}$$

The detector is clearly unrealizable since it is composed of two different NP tests, the choice of which depends upon the unknown parameter A. It provides an upper bound on performance, which can be found as follows. Under $\mathcal{H}_0$, noting that $\bar{x} \sim \mathcal{N}(0, \sigma^2/N)$, we have for $A > 0$

$$\begin{aligned} P_{FA} &= \Pr\{\bar{x} > \gamma'_+; \mathcal{H}_0\} \\ &= Q\left(\frac{\gamma'_+}{\sqrt{\sigma^2/N}}\right) \end{aligned} \tag{6.7}$$

and for $A < 0$

$$\begin{aligned} P_{FA} &= 1 - Q\left(\frac{\gamma'_-}{\sqrt{\sigma^2/N}}\right) \\ &= Q\left(\frac{-\gamma'_-}{\sqrt{\sigma^2/N}}\right). \end{aligned}$$

For a constant P_{FA} we should choose $\gamma'_- = -\gamma'_+$. Next, under $\mathcal{H}_1$, $\bar{x} \sim \mathcal{N}(A, \sigma^2/N)$ so that for $A > 0$

$$\begin{aligned} P_D &= Q\left(\frac{\gamma'_+ - A}{\sqrt{\sigma^2/N}}\right) \\ &= Q\left(Q^{-1}(P_{FA}) - \sqrt{\frac{NA^2}{\sigma^2}}\right) \end{aligned}$$

which follows from (6.7) and for $A < 0$

$$P_D = 1 - Q\left(\frac{\gamma'_- - A}{\sqrt{\sigma^2/N}}\right)$$

$$
\begin{aligned}
&= Q\left(\frac{-\gamma'_- + A}{\sqrt{\sigma^2/N}}\right) \\
&= Q\left(Q^{-1}(P_{FA}) + \frac{A}{\sqrt{\sigma^2/N}}\right) \\
&= Q\left(Q^{-1}(P_{FA}) - \sqrt{\frac{NA^2}{\sigma^2}}\right)
\end{aligned}
$$

which is the same as for $A > 0$. This is shown in Figure 6.3 for $N = 10$, $\sigma^2 = 1$, and $P_{FA} = 0.1$. The dashed curves indicate the performance if the NP test for $A > 0$ were used but A was negative and similarly for $A < 0$. In this case $P_D < P_{FA}$, which is quite poor, as expected.

◇

As an example of the use of the clairvoyant detector bound, consider the detector that decides $\mathcal{H}_1$ if

$$\left|\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right| > \gamma'' \tag{6.8}$$

which assumes A is unknown and $-\infty < A < \infty$. Large excursions of the sample mean from zero are chosen to be indications of a signal present. It is easily shown

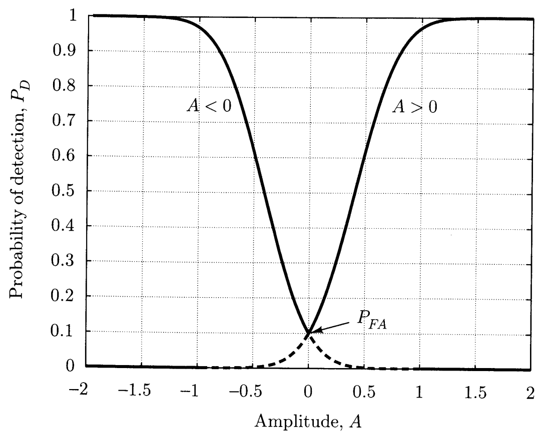


Figure 6.3. Detection performance of clairvoyant NP detector.

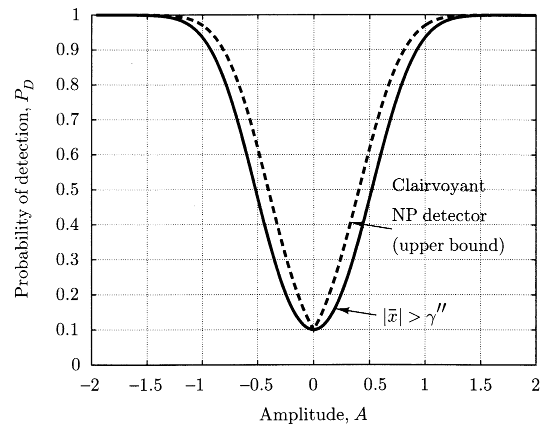


Figure 6.4. Detection performance comparison of clairvoyant NP detector and realizable detector.

that (see Problem 6.5)

$$P_D = Q\left(Q^{-1}\left(\frac{P_{FA}}{2}\right) - \sqrt{\frac{NA^2}{\sigma^2}}\right) + Q\left(Q^{-1}\left(\frac{P_{FA}}{2}\right) + \sqrt{\frac{NA^2}{\sigma^2}}\right). \tag{6.9}$$

For $N = 10$, $\sigma^2 = 1$, $P_{FA} = 0.1$, this is plotted in Figure 6.4, along with the clairvoyant detector bound (shown in Figure 6.3). It is observed that the proposed detector has performance close to the bound. This detector then would appear to be a good compromise in that it is *robust* to the sign of A. In fact, the proposed detector is an example of a more general approach to composite hypothesis testing, the generalized likelihood ratio test, which is described in the next section (see Example 6.4).

6.4 Composite Hypothesis Testing Approaches

There are two major approaches to composite hypothesis testing. The first is to consider the unknown parameters as realizations of random variables and to assign a prior PDF. The second is to estimate the unknown parameters for use in a likelihood ratio test. We will term the first method the Bayesian approach and the second, the generalized likelihood ratio test (GLRT). The Bayesian approach employs the philosophy described in [Kay-I 1993, Chapter 10], where it is applied to parameter estimation. It requires prior knowledge of the unknown parameters whereas the

GLRT does not. In practice, the GLRT appears to be more widely used due to its ease of implementation and less restrictive assumptions. The Bayesian approach requires multidimensional integration, which is usually not possible in closed form. For these reasons our emphasis in succeeding chapters will be on the GLRT.

The general problem is to decide between $\mathcal{H}_0$ and $\mathcal{H}_1$ when the PDFs depend on a set of unknown parameters. These parameters may or may not be the same under each hypothesis. Under $\mathcal{H}_0$ we assume that the vector parameter $\boldsymbol{\theta}_0$ is unknown while under $\mathcal{H}_1$ the vector parameter $\boldsymbol{\theta}_1$ is unknown. We first discuss the Bayesian approach.

6.4.1 Bayesian Approach

The Bayesian approach assigns prior PDFs to $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$. In doing so it models the unknown parameters as *realizations* of a vector random variable. If the prior PDFs are denoted by $p(\boldsymbol{\theta}_0)$ and $p(\boldsymbol{\theta}_1)$, respectively, the PDFs of the data are

$$\begin{aligned} p(\mathbf{x};\mathcal{H}_0) &= \int p(\mathbf{x}|\boldsymbol{\theta}_0;\mathcal{H}_0)p(\boldsymbol{\theta}_0)d\boldsymbol{\theta}_0 \\ p(\mathbf{x};\mathcal{H}_1) &= \int p(\mathbf{x}|\boldsymbol{\theta}_1;\mathcal{H}_1)p(\boldsymbol{\theta}_1)d\boldsymbol{\theta}_1 \end{aligned}$$

where $p(\mathbf{x}|\boldsymbol{\theta}_i;\mathcal{H}_i)$ is the conditional PDF of $\mathbf{x}$, conditioned on $\boldsymbol{\theta}_i$, assuming $\mathcal{H}_i$ is true. The unconditional PDFs $p(\mathbf{x};\mathcal{H}_0)$ and $p(\mathbf{x};\mathcal{H}_1)$ are now completely specified, no longer dependent on the unknown parameters. With the Bayesian approach the optimal NP detector decides $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x};\mathcal{H}_1)}{p(\mathbf{x};\mathcal{H}_0)} = \frac{\int p(\mathbf{x}|\boldsymbol{\theta}_1;\mathcal{H}_1)p(\boldsymbol{\theta}_1)d\boldsymbol{\theta}_1}{\int p(\mathbf{x}|\boldsymbol{\theta}_0;\mathcal{H}_0)p(\boldsymbol{\theta}_0)d\boldsymbol{\theta}_0} > \gamma. \tag{6.10}$$

The required integrations are multidimensional with dimension equal to the unknown parameter dimension. The choice of the prior PDFs can prove difficult. If, indeed, one does have some prior knowledge, then one should use it. If not, then a noninformative prior (see [Kay-I 1993, pg. 332]) should be used. A noninformative prior is a PDF that is as "flat" as possible. As an example, if we wish to detect a sinusoid with unknown phase ϕ and we have no reason to favor any particular value of ϕ over another, then $\phi \sim \mathcal{U}[0, 2\pi]$ would be consistent with our lack of prior knowledge (see also Problem 7.22). A uniform PDF, then, would be a good choice, although the integration may prove difficult. If, however, the parameter takes on values over an infinite interval, say $-\infty < A < \infty$ for the DC level, then we cannot assign a uniform PDF. More commonly, a Gaussian PDF is chosen with $A \sim \mathcal{N}(0, \sigma_A^2)$ and we let $\sigma_A^2 \to \infty$ to reflect our lack of prior knowledge. This approach is illustrated next.

Example 6.3 - DC Level in WGN with Unknown Amplitude - Bayesian Approach

Assume that for the DC level in WGN A is unknown and can take on values $-\infty < A < \infty$. We assign the prior PDF $A \sim \mathcal{N}(0, \sigma_A^2)$, where A is independent of $w[n]$. Note that as $\sigma_A^2 \to \infty$, the PDF becomes a noninformative prior (see also [Kay-I 1993, Problem 10.17]). The conditional PDF under $\mathcal{H}_1$ is assumed to be

$$p(\mathbf{x}|A;\mathcal{H}_1) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right].$$

Under $\mathcal{H}_0$ the PDF is completely known. Hence, according to (6.10) with $\boldsymbol{\theta}_1 = A$, the NP detector decides $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x};\mathcal{H}_1)}{p(\mathbf{x};\mathcal{H}_0)} = \frac{\int_{-\infty}^{\infty} p(\mathbf{x}|A;\mathcal{H}_1)p(A)dA}{p(\mathbf{x};\mathcal{H}_0)} > \gamma.$$

But

$$\begin{aligned} p(\mathbf{x};\mathcal{H}_1) &= \int_{-\infty}^{\infty} p(\mathbf{x}|A;\mathcal{H}_1)p(A)dA \\ &= \int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right] \\ &\quad \cdot \frac{1}{\sqrt{2\pi\sigma_A^2}} \exp\left(-\frac{1}{2\sigma_A^2}A^2\right) dA. \end{aligned}$$

Letting

$$Q(A) = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2 + \frac{A^2}{\sigma_A^2}$$

we have upon completing the square in A

$$\begin{aligned} Q(A) &= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}x^2[n] - \frac{2N}{\sigma^2}\bar{x}A + \frac{N}{\sigma^2}A^2 + \frac{A^2}{\sigma_A^2} \\ &= \underbrace{\left(\frac{N}{\sigma^2} + \frac{1}{\sigma_A^2}\right)}_{1/\sigma^2_{A|x}} A^2 - \frac{2N}{\sigma^2}\bar{x}A + \frac{1}{\sigma^2}\sum_{n=0}^{N-1}x^2[n] \\ &= \frac{A^2}{\sigma^2_{A|x}} - \frac{2N\sigma^2_{A|x}\bar{x}A}{\sigma^2\sigma^2_{A|x}} + \frac{1}{\sigma^2}\sum_{n=0}^{N-1}x^2[n] \end{aligned}$$

$$= \frac{1}{\sigma^2_{A|x}}\left(A - \frac{N\bar{x}\sigma^2_{A|x}}{\sigma^2}\right)^2 - \frac{N^2\bar{x}^2}{\sigma^4}\sigma^2_{A|x} + \frac{1}{\sigma^2}\sum_{n=0}^{N-1} x^2[n]$$

so that

$$\begin{aligned}\frac{p(\mathbf{x};\mathcal{H}_1)}{p(\mathbf{x};\mathcal{H}_0)} &= \frac{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\frac{1}{\sqrt{2\pi\sigma_A^2}}\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}Q(A)\right]dA}{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left(-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right)} \\ &= \frac{1}{\sqrt{2\pi\sigma_A^2}}\sqrt{2\pi\sigma^2_{A|x}}\exp\left(\frac{N^2\bar{x}^2\sigma^2_{A|x}}{2\sigma^4}\right) > \gamma.\end{aligned}$$

Taking logarithms of both sides and retaining only the data dependent terms we decide $\mathcal{H}_1$ if

$$(\bar{x})^2 > \gamma'$$

or

$$|\bar{x}| > \sqrt{\gamma'}. \tag{6.11}$$

This is similar to the detector we proposed in the previous section. In order to set the threshold we do not need knowledge of σ_A^2. However, if we had assumed $A \sim \mathcal{N}(\mu_A, \sigma_A^2)$, then we would need to know μ_A and σ_A^2 (see Section 7.4.2) to implement the Bayesian detector. The performance of the detector of (6.11) is derived in Problem 6.7.

◇

6.4.2 Generalized Likelihood Ratio Test

The GLRT replaces the unknown parameters by their maximum likelihood estimates (MLEs). Although there is no optimality associated with the GLRT, in practice, it appears to work quite well. (Asymptotically, it can be shown that the GLRT is UMP among all tests that are *invariant* [Lehmann 1959].) In general, the GLRT decides $\mathcal{H}_1$ if

$$L_G(\mathbf{x}) = \frac{p(\mathbf{x};\hat{\boldsymbol{\theta}}_1,\mathcal{H}_1)}{p(\mathbf{x};\hat{\boldsymbol{\theta}}_0,\mathcal{H}_0)} > \gamma \tag{6.12}$$

where $\hat{\boldsymbol{\theta}}_1$ is the MLE of $\boldsymbol{\theta}_1$ assuming $\mathcal{H}_1$ is true (maximizes $p(\mathbf{x};\boldsymbol{\theta}_1,\mathcal{H}_1)$), and $\hat{\boldsymbol{\theta}}_0$ is the MLE of $\boldsymbol{\theta}_0$ assuming $\mathcal{H}_0$ is true (maximizes $p(\mathbf{x};\boldsymbol{\theta}_0,\mathcal{H}_0)$). The approach also provides information about the unknown parameters since the first step in determining $L_G(\mathbf{x})$ is to find the MLEs. We now continue the DC level in WGN example.

Example 6.4 - DC Level in WGN with Unknown Amplitude - GLRT

In this case we have $\boldsymbol{\theta}_1 = A$ and there are no unknown parameters under $\mathcal{H}_0$. The hypothesis test becomes

$$\begin{aligned}\mathcal{H}_0 &: A = 0\\ \mathcal{H}_1 &: A \neq 0.\end{aligned}$$

Thus, the GLRT decides $\mathcal{H}_1$ if

$$L_G(\mathbf{x}) = \frac{p(\mathbf{x};\hat{A},\mathcal{H}_1)}{p(\mathbf{x};\mathcal{H}_0)} > \gamma.$$

The MLE of A is found by maximizing

$$p(\mathbf{x};A,\mathcal{H}_1) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right].$$

This was done in [Kay-I 1993, pp. 163–164] with the result that $\hat{A} = \bar{x}$. Thus,

$$L_G(\mathbf{x}) = \frac{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-\bar{x})^2\right]}{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left(-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right)}.$$

Taking logarithms we have

$$\begin{aligned}\ln L_G(\mathbf{x}) &= -\frac{1}{2\sigma^2}\left(\sum_{n=0}^{N-1}x^2[n] - 2\bar{x}\sum_{n=0}^{N-1}x[n] + N\bar{x}^2 - \sum_{n=0}^{N-1}x^2[n]\right)\\ &= -\frac{1}{2\sigma^2}(-2N\bar{x}^2 + N\bar{x}^2)\\ &= \frac{N\bar{x}^2}{2\sigma^2}\end{aligned}$$

or we decide $\mathcal{H}_1$ if

$$|\bar{x}| > \gamma'. \tag{6.13}$$

The performance has already been given by (6.9). When compared to the clairvoyant detector, there is only a slight loss as shown in Figure 6.4.

◇

The GLRT can also be expressed in another form, which is sometimes more convenient. Since $\hat{\boldsymbol{\theta}}_i$ is the MLE under $\mathcal{H}_i$, it maximizes $p(\mathbf{x};\boldsymbol{\theta}_i,\mathcal{H}_i)$ or

$$p(\mathbf{x};\hat{\boldsymbol{\theta}}_i,\mathcal{H}_i) = \max_{\boldsymbol{\theta}_i} p(\mathbf{x};\boldsymbol{\theta}_i,\mathcal{H}_i).$$

Hence, (6.12) can be written as

$$L_G(\mathbf{x}) = \frac{\max_{\boldsymbol{\theta}_1} p(\mathbf{x}; \boldsymbol{\theta}_1, \mathcal{H}_1)}{\max_{\boldsymbol{\theta}_0} p(\mathbf{x}; \boldsymbol{\theta}_0, \mathcal{H}_0)}. \tag{6.14}$$

For the special case where the PDF under $\mathcal{H}_0$ is completely known

$$\begin{aligned} L_G(\mathbf{x}) &= \frac{\max_{\boldsymbol{\theta}_1} p(\mathbf{x}; \boldsymbol{\theta}_1, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} \\ &= \max_{\boldsymbol{\theta}_1} \frac{p(\mathbf{x}; \boldsymbol{\theta}_1, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} \end{aligned}$$

or we maximize the likelihood ratio over $\boldsymbol{\theta}_1$ so that

$$L_G(\mathbf{x}) = \max_{\boldsymbol{\theta}_1} L(\mathbf{x}; \boldsymbol{\theta}_1). \tag{6.15}$$

See also Problem 6.14. We now illustrate the GLRT approach with a slightly more complicated example.

Example 6.5 - DC Level in WGN with Unknown Amplitude and Variance - GLRT

Consider the detection problem

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = A + w[n] & n = 0, 1, \ldots, N-1 \end{array}$$

where A is unknown with $-\infty < A < \infty$ and $w[n]$ is WGN with *unknown variance* σ^2. A UMP test does not exist because the equivalent parameter test is

$$\begin{array}{l} \mathcal{H}_0 : A = 0, \sigma^2 > 0 \\ \mathcal{H}_1 : A \neq 0, \sigma^2 > 0 \end{array} \tag{6.16}$$

which is two-sided. In this example the hypothesis test contains a *nuisance parameter*, which is σ^2. Although we are not directly concerned with σ^2, it enters into the problem since it affects the PDFs under $\mathcal{H}_0$ and $\mathcal{H}_1$. The detector of (6.13) can no longer be implemented since the threshold is dependent upon the PDF under $\mathcal{H}_0$, which in turn depends on σ^2. The variance can take on any value $0 < \sigma^2 < \infty$. The GLRT decides $\mathcal{H}_1$ if

$$L_G(\mathbf{x}) = \frac{p(\mathbf{x}; \hat{A}, \hat{\sigma_1^2}, \mathcal{H}_1)}{p(\mathbf{x}; \hat{\sigma_0^2}, \mathcal{H}_0)} > \gamma$$

where $[\hat{A}\, \hat{\sigma_1^2}]^T$ is the MLE of the vector parameter $\boldsymbol{\theta}_1 = [A\, \sigma^2]^T$ under $\mathcal{H}_1$, and $\hat{\sigma_0^2}$ is the MLE of the parameter $\boldsymbol{\theta}_0 = \sigma^2$ under $\mathcal{H}_0$. Note that we need to estimate the variance under both hypotheses. To find $\hat{A}, \hat{\sigma_1^2}$ we must maximize

$$p(\mathbf{x}; A, \sigma^2, \mathcal{H}_1) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n] - A)^2\right].$$

In [Kay-I 1993, pg. 183] we showed that

$$\begin{aligned} \hat{A} &= \bar{x} \\ \hat{\sigma_1^2} &= \frac{1}{N}\sum_{n=0}^{N-1}(x[n] - \bar{x})^2 \end{aligned}$$

so that

$$p(\mathbf{x}; \hat{A}, \hat{\sigma_1^2}, \mathcal{H}_1) = \frac{1}{(2\pi\hat{\sigma_1^2})^{\frac{N}{2}}} \exp\left(-\frac{N}{2}\right).$$

To find $\hat{\sigma_0^2}$ we must maximize

$$p(\mathbf{x}; \sigma^2, \mathcal{H}_0) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left(-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1} x^2[n]\right).$$

It is easily shown that

$$\hat{\sigma_0^2} = \frac{1}{N}\sum_{n=0}^{N-1} x^2[n]$$

so that

$$p(\mathbf{x}; \hat{\sigma_0^2}, \mathcal{H}_0) = \frac{1}{(2\pi\hat{\sigma_0^2})^{\frac{N}{2}}} \exp\left(-\frac{N}{2}\right).$$

Thus, we have

$$L_G(\mathbf{x}) = \left(\frac{\hat{\sigma_0^2}}{\hat{\sigma_1^2}}\right)^{\frac{N}{2}}$$

or equivalently

$$2\ln L_G(\mathbf{x}) = N \ln \frac{\hat{\sigma_0^2}}{\hat{\sigma_1^2}}. \tag{6.17}$$

In essence, the GLRT decides $\mathcal{H}_1$ if the fit to the data of the signal $\hat{A} = \bar{x}$ produces a much smaller error, as measured by $\hat{\sigma_1^2} = (1/N)\sum_{n=0}^{N-1}(x[n] - \hat{A})^2$ than a fit of no signal or $\hat{\sigma_0^2} = (1/N)\sum_{n=0}^{N-1} x^2[n]$. A slightly more intuitive form can be found as follows. Since

$$\begin{aligned}
\hat{\sigma_1^2} &= \frac{1}{N}\sum_{n=0}^{N-1}(x[n]-\bar{x})^2 \\
&= \frac{1}{N}\sum_{n=0}^{N-1}(x^2[n]-2x[n]\bar{x}+\bar{x}^2) \\
&= \frac{1}{N}\sum_{n=0}^{N-1}x^2[n]-\bar{x}^2 \\
&= \hat{\sigma_0^2}-\bar{x}^2 \qquad (6.18)
\end{aligned}$$

we have upon using this in (6.17)

$$\begin{aligned}
2\ln L_G(\mathbf{x}) &= N\ln\left(\frac{\hat{\sigma_1^2}+\bar{x}^2}{\hat{\sigma_1^2}}\right) \\
&= N\ln\left(1+\frac{\bar{x}^2}{\hat{\sigma_1^2}}\right). \qquad (6.19)
\end{aligned}$$

Since $\ln(1+x)$ is monotonically increasing with increasing x, an equivalent test statistic is

$$T(\mathbf{x}) = \frac{\bar{x}^2}{\hat{\sigma_1^2}}. \qquad (6.20)$$

Comparing this to the known σ^2 case, we observe that the GLRT has *normalized* the statistic by $\hat{\sigma_1^2}$ to allow the threshold to be determined. The threshold is independent of the true value of σ^2 since the PDF of $T(\mathbf{x})$ under $\mathcal{H}_0$ does not depend on σ^2 (see also Example 9.2). This is easily established if we let $w[n] = \sigma u[n]$, where $u[n]$ is WGN with variance *one*. Then, from (6.20) with $x[n] = w[n]$ (under $\mathcal{H}_0$) we have

$$\begin{aligned}
T(\mathbf{x}) &= \frac{\left(\frac{1}{N}\sum_{n=0}^{N-1}w[n]\right)^2}{\frac{1}{N}\sum_{n=0}^{N-1}(w[n]-\bar{w})^2} = \frac{\left(\frac{1}{N}\sum_{n=0}^{N-1}\sigma u[n]\right)^2}{\frac{1}{N}\sum_{n=0}^{N-1}(\sigma u[n]-\sigma\bar{u})^2} \\
&= \frac{\left(\frac{1}{N}\sum_{n=0}^{N-1}u[n]\right)^2}{\frac{1}{N}\sum_{n=0}^{N-1}(u[n]-\bar{u})^2}
\end{aligned}$$

whose PDF does not depend on σ^2. Similarly, $2\ln L_G(\mathbf{x})$ as given by (6.19) has a PDF under $\mathcal{H}_0$ that does not depend on σ^2. The detection performance of the GLRT of (6.19) will be slightly poorer than for the case when σ^2 is known. Surprisingly, if N is large enough, the degradation will be quite small. We explore the large data record performance in the next section.

◇

6.5 Performance of GLRT for Large Data Records

For large data records or *asymptotically* (as $N \to \infty$) the detection performance of the GLRT can be found quite easily. This allows us to set the threshold as well as to determine the probability of detection. The conditions under which the asymptotic expressions hold are:

1. when the data record is large and the signal is weak (see Section 1.5) and

2. when the MLE attains its asymptotic PDF [Kay-I 1993, pg. 183].

The theorem is valid when the composite hypothesis testing problem can be recast as a parameter test of the PDF. Appendices 6A–C contain an outline of the proof.

Consider a PDF $p(\mathbf{x};\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is a $p \times 1$ vector of unknown parameters and partition $\boldsymbol{\theta}$ as

$$\boldsymbol{\theta} = \begin{bmatrix}\boldsymbol{\theta}_r \\ \boldsymbol{\theta}_s\end{bmatrix} = \begin{bmatrix} r\times 1 \\ s\times 1\end{bmatrix}$$

where $p = r+s$. We wish to test if $\boldsymbol{\theta}_r = \boldsymbol{\theta}_{r_0}$ as opposed to $\boldsymbol{\theta}_r \neq \boldsymbol{\theta}_{r_0}$. The parameter vector $\boldsymbol{\theta}_s$ is a set of nuisance parameters, which are unknown *but the same* under either hypothesis. Hence, the parameter test is

$$\begin{array}{l}\mathcal{H}_0 : \boldsymbol{\theta}_r = \boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s \\ \mathcal{H}_1 : \boldsymbol{\theta}_r \neq \boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s.\end{array} \qquad (6.21)$$

The GLRT for this problem is to decide $\mathcal{H}_1$ if

$$L_G(\mathbf{x}) = \frac{p(\mathbf{x};\hat{\boldsymbol{\theta}}_{r_1},\hat{\boldsymbol{\theta}}_{s_1},\mathcal{H}_1)}{p(\mathbf{x};\boldsymbol{\theta}_{r_0},\hat{\boldsymbol{\theta}}_{s_0},\mathcal{H}_0)} > \gamma \qquad (6.22)$$

where $\hat{\boldsymbol{\theta}}_1 = [\hat{\boldsymbol{\theta}}_{r_1}^T\ \hat{\boldsymbol{\theta}}_{s_1}^T]^T$ is the MLE of $\boldsymbol{\theta}$ under $\mathcal{H}_1$, termed the *unrestricted MLE*, and $\hat{\boldsymbol{\theta}}_{s_0}$ is the MLE of $\boldsymbol{\theta}_s$ under $\mathcal{H}_0$ or when $\boldsymbol{\theta}_r = \boldsymbol{\theta}_{r_0}$, termed the *restricted MLE*. Then, as $N \to \infty$ the modified GLRT statistic $2\ln L_G(\mathbf{x})$ has the PDF

$$2\ln L_G(\mathbf{x}) \overset{a}{\sim} \begin{cases} \chi_r^2 & \text{under } \mathcal{H}_0 \\ \chi_r'^2(\lambda) & \text{under } \mathcal{H}_1 \end{cases} \qquad (6.23)$$

where "a" denotes an asymptotic PDF, χ_r^2 denotes a chi-squared PDF with r degrees of freedom, and $\chi_r'^2(\lambda)$ denotes a noncentral chi-squared PDF with r degrees of freedom and noncentrality parameter λ (see Chapter 2). The noncentrality parameter is

$$\begin{aligned}\lambda &= (\boldsymbol{\theta}_{r_1} - \boldsymbol{\theta}_{r_0})^T \Big[\mathbf{I}_{\theta_r\theta_r}(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s) \\ &\quad - \mathbf{I}_{\theta_r\theta_s}(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s)\mathbf{I}^{-1}_{\theta_s\theta_s}(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s)\mathbf{I}_{\theta_s\theta_r}(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s)\Big](\boldsymbol{\theta}_{r_1} - \boldsymbol{\theta}_{r_0}) \qquad (6.24)\end{aligned}$$

where $\boldsymbol{\theta}_{r_1}$ is the true value of $\boldsymbol{\theta}_r$ under $\mathcal{H}_1$ and $\boldsymbol{\theta}_s$ is the true value of the parameter (which is the same under $\mathcal{H}_0$ and $\mathcal{H}_1$). The matrices in (6.24) are partitions of the Fisher information matrix $\mathbf{I}(\boldsymbol{\theta})$ (see [Kay-I 1993, pg. 40]). They are defined by

$$\begin{aligned}\mathbf{I}(\boldsymbol{\theta}) = \mathbf{I}(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s) &= \begin{bmatrix} \mathbf{I}_{\theta_r\theta_r}(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s) & \mathbf{I}_{\theta_r\theta_s}(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s) \\ \mathbf{I}_{\theta_s\theta_r}(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s) & \mathbf{I}_{\theta_s\theta_s}(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s) \end{bmatrix} \qquad (6.25) \\ &= \begin{bmatrix} r\times r & r\times s \\ s\times r & s\times s \end{bmatrix}.\end{aligned}$$

Since the asymptotic PDF under $\mathcal{H}_0$ does not depend on any unknown parameters, the threshold required to maintain a constant P_{FA} can be found. This type of detector is referred to as a *constant false alarm rate* (CFAR) detector. In general, the CFAR property holds only for large data records. For no nuisance parameters $\boldsymbol{\theta}$ is $r \times 1$. The hypothesis test then becomes

$$\begin{aligned}\mathcal{H}_0 &: \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ \mathcal{H}_1 &: \boldsymbol{\theta} \neq \boldsymbol{\theta}_0\end{aligned} \qquad (6.26)$$

and the asymptotic PDFs are given by (6.23) but with

$$\lambda = (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^T\mathbf{I}(\boldsymbol{\theta}_0)(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0) \qquad (6.27)$$

where $\boldsymbol{\theta}_1$ is the true value of $\boldsymbol{\theta}$ under $\mathcal{H}_1$. Note that when nuisance parameters are present the noncentrality parameter is decreased and hence P_D is decreased. This is observed by comparing λ for no nuisance parameters or (6.27) with that for nuisance parameters or (6.24). This is the price paid for having to estimate extra parameters for use in the detector. We now illustrate the asymptotic performance with some examples.

Example 6.6 - DC Level in WGN with Unknown Amplitude

We refer to Example 6.4. There we saw that for an unknown amplitude (no nuisance parameters) we have that $\boldsymbol{\theta} = A$ and $\boldsymbol{\theta}_0 = 0$ so that $r = 1$. The asymptotic PDF of the modified GLRT statistic (see Example 6.4) is from (6.23) and (6.27)

$$2\ln L_G(\mathbf{x}) = \frac{N\bar{x}^2}{\sigma^2} \overset{a}{\sim} \begin{cases} \chi_1^2 & \text{under } \mathcal{H}_0 \\ \chi_1'^2(\lambda) & \text{under } \mathcal{H}_1 \end{cases}$$

where $\lambda = A^2 I(0)$. But $I(A) = N/\sigma^2$ [Kay-I 1993, pp. 31–32] so that

$$\lambda = \frac{NA^2}{\sigma^2}.$$

For this special case the asymptotic PDFs hold *exactly for finite data records* since

$$\bar{x} \sim \begin{cases} \mathcal{N}(0, \frac{\sigma^2}{N}) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(A, \frac{\sigma^2}{N}) & \text{under } \mathcal{H}_1 \end{cases}$$

and therefore

$$\frac{\bar{x}}{\sigma/\sqrt{N}} \sim \begin{cases} \mathcal{N}(0, 1) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(\frac{\sqrt{N}A}{\sigma}, 1) & \text{under } \mathcal{H}_1. \end{cases}$$

Squaring the statistic produces a χ_1^2 random variable under $\mathcal{H}_0$ and a $\chi_1'^2(\lambda)$ random variable under $\mathcal{H}_1$. The result of this example is not purely coincidental. It can be shown that for the case of the classical linear model, for which this example is a special case, the asymptotic statistics are exact for finite data records (see Problems 6.15–6.17).

◇

Example 6.7 - DC Level in WGN with Unknown Amplitude and Variance

We refer to Example 6.5. From (6.16) we see that $\boldsymbol{\theta} = [A\,\sigma^2]^T$ so that $\boldsymbol{\theta}_r = A$, $\boldsymbol{\theta}_{r_0} = 0$, and $\boldsymbol{\theta}_s = \sigma^2$. Also, $p = 2$ and $r = s = 1$. Thus, the asymptotic PDFs are from (6.23) and (6.24)

$$2\ln L_G(\mathbf{x}) = N\ln\left(1 + \frac{\bar{x}^2}{\hat{\sigma}_1^2}\right) \sim \begin{cases} \chi_1^2 & \text{under } \mathcal{H}_0 \\ \chi_1'^2(\lambda) & \text{under } \mathcal{H}_1 \end{cases} \qquad (6.28)$$

where

$$\lambda = A^2\left[I_{AA}(0,\sigma^2) - I_{A\sigma^2}(0,\sigma^2)I^{-1}_{\sigma^2\sigma^2}(0,\sigma^2)I_{\sigma^2 A}(0,\sigma^2)\right].$$

But from [Kay-I 1993, pg. 41]

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{N}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{bmatrix}.$$

As a result, $I_{AA}(0,\sigma^2) = N/\sigma^2$, $I_{A\sigma^2}(0,\sigma^2) = I_{\sigma^2 A}(0,\sigma^2) = 0$, and

$$\lambda = \frac{NA^2}{\sigma^2}$$

as before. Whereas in Example 6.6 the PDFs were exact, now they are only valid for large data records. It is interesting to note that for large N the GLRT performance is the same whether or not σ^2 is known (see Example 6.6). This is due to the diagonal nature of the Fisher information matrix. We explore this result further in Chapter 9. To ascertain how good the large data record approximation is for finite data records, we compare the receiver operating characteristics (ROC) of the GLRT as given by (6.28) to that obtained by a Monte Carlo computer simulation. The ROC based on the asymptotic PDF can be found by observing that a noncentral chi-squared random variable y with one degree of freedom and noncentrality parameter λ is equal to the square of a random variable x with $x \sim \mathcal{N}(\sqrt{\lambda}, 1)$. Thus, under $\mathcal{H}_0$

$$\begin{aligned} P_{FA} &= \Pr\{y > \gamma'; \mathcal{H}_0\} \\ &= \Pr\{x > \sqrt{\gamma'}; \mathcal{H}_0\} + \Pr\{x < -\sqrt{\gamma'}; \mathcal{H}_0\} \\ &= 2Q(\sqrt{\gamma'}). \end{aligned}$$

Similarly, it can be shown that

$$P_D = Q(\sqrt{\gamma'} - \sqrt{\lambda}) + Q(\sqrt{\gamma'} + \sqrt{\lambda})$$

or finally

$$P_D = Q\left(Q^{-1}\left(\frac{P_{FA}}{2}\right) - \sqrt{\lambda}\right) + Q\left(Q^{-1}\left(\frac{P_{FA}}{2}\right) + \sqrt{\lambda}\right) \tag{6.29}$$

where $\lambda = NA^2/\sigma^2$. To compare the asymptotic performance of (6.29) with the true performance as obtained via a Monte Carlo computer simulation we choose $\lambda = 5$, $\sigma^2 = 1$. Then, for $N = 10$ we choose $A = \sqrt{\lambda\sigma^2/N} = \sqrt{5/10}$, and for $N = 30$ we choose $A = \sqrt{\lambda\sigma^2/N} = \sqrt{5/30}$ so that P_D as given by (6.29) will be the same. This allows a direct comparison as N increases. The results are shown in Figure 6.5a for $N = 10$ and Figure 6.5b for $N = 30$. The dashed curves are the theoretical asymptotic ROC (as given by (6.29)), and the solid curves are the computer-generated results. It is seen that the theoretical asymptotic performance adequately summarizes the actual performance for data records as short as $N = 30$ samples.

$\diamond$

6.6 Equivalent Large Data Records Tests

There are two other tests that have the same asymptotic (as $N \to \infty$) detection performance as the GLRT. For finite data records, however, there is no guarantee that the performances will be the same. Their main advantage is that these asymptotically equivalent statistics may be easier to compute. This is especially

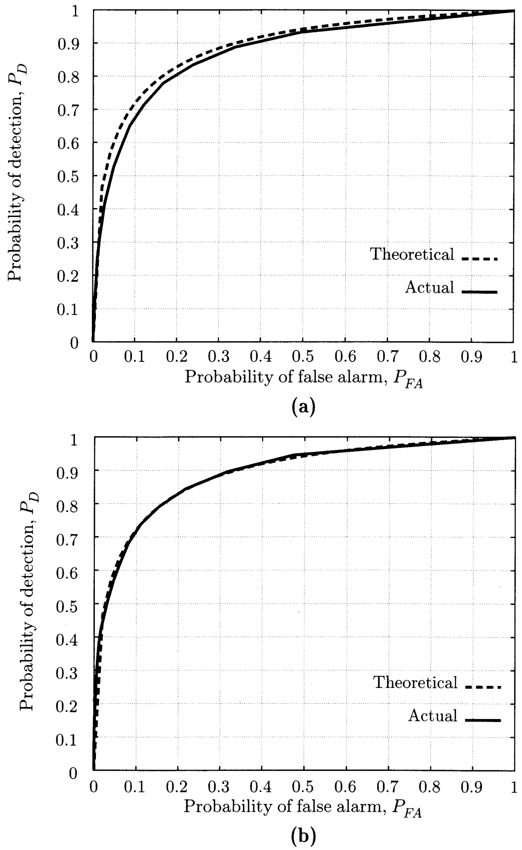


Figure 6.5. Comparison of theoretical asymptotic performance to actual performance for GLRT (a) $N = 10$ (b) $N = 30$.

true of the Rao test for which it is not necessary to determine the MLE under $\mathcal{H}_1$, leaving only the MLE under $\mathcal{H}_0$ to be found. The derivations for these statistics are outlined in Appendices 6A,B. We summarize the results next.

Consider first the case of no nuisance parameters. We wish to test

$$\begin{aligned} \mathcal{H}_0 &: \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ \mathcal{H}_1 &: \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{aligned}$$

where $\boldsymbol{\theta}$ is an $r \times 1$ vector of parameters. The *Wald test* decides $\mathcal{H}_1$ if

$$T_{\mathrm{W}}(\mathbf{x}) = (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0)^T \mathbf{I}(\hat{\boldsymbol{\theta}}_1)(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0) > \gamma \tag{6.30}$$

where $\hat{\boldsymbol{\theta}}_1$ is the MLE of $\boldsymbol{\theta}$ under $\mathcal{H}_1$ or equivalently the MLE of $\boldsymbol{\theta}$ with no restrictions on the parameter space. Note that the computation of $T_{\mathrm{W}}(\mathbf{x})$ may be simpler than the GLRT

$$L_G(\mathbf{x}) = \frac{p(\mathbf{x}; \hat{\boldsymbol{\theta}}_1, \mathcal{H}_1)}{p(\mathbf{x}; \boldsymbol{\theta}_0, \mathcal{H}_0)}$$

which requires an evaluation of the PDFs. The *Rao test* decides $\mathcal{H}_1$ if

$$T_{\mathrm{R}}(\mathbf{x}) = \left.\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}^T \mathbf{I}^{-1}(\boldsymbol{\theta}_0) \left.\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} > \gamma. \tag{6.31}$$

In (6.31) we omit the $\mathcal{H}_i$ descriptor in the PDF since it is assumed that the PDFs under $\mathcal{H}_0$ and $\mathcal{H}_1$ differ only in the value of $\boldsymbol{\theta}$. Alternatively, there is no need to do so when the PDF is parameterized by $\boldsymbol{\theta}$. The MLE under $\mathcal{H}_1$ need not be found for this test. This is clearly advantageous when

$$\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}$$

cannot be easily solved for the MLE of $\boldsymbol{\theta}$ under $\mathcal{H}_1$. Of the three tests the Rao test is the simplest computationally. We illustrate these alternative tests with an example.

Example 6.8 - DC Level in WGN with Unknown Amplitude - Wald and Rao Tests

From Example 6.4 we have the test

$$\begin{aligned} \mathcal{H}_0 &: A = 0 \\ \mathcal{H}_1 &: A \neq 0 \end{aligned}$$

so that $\boldsymbol{\theta} = A$ and $\boldsymbol{\theta}_0 = 0$. The Wald test is found by letting $\hat{\boldsymbol{\theta}}_1 = \hat{A} = \bar{x}$ and $\mathbf{I}(\hat{\boldsymbol{\theta}}_1) = I(\hat{A}) = N/\sigma^2$. Then, from (6.30) the Wald test decides $\mathcal{H}_1$ if

$$T_{\mathrm{W}}(\mathbf{x}) = \frac{N\bar{x}^2}{\sigma^2} > \gamma$$

which is identical to the GLRT or $2 \ln L_G(\mathbf{x})$. The Rao test requires us to find the derivative of the log-likelihood function but not the MLE of A. Since

$$p(\mathbf{x}; A) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2\right]$$

we have that

$$\frac{\partial \ln p(\mathbf{x}; A)}{\partial A} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)$$

and

$$\left.\frac{\partial \ln p(\mathbf{x}; A)}{\partial A}\right|_{A=0} = \frac{N\bar{x}}{\sigma^2}.$$

Also, $\mathbf{I}(\boldsymbol{\theta}_0) = I(A = 0) = N/\sigma^2$. Thus, from (6.31)

$$T_{\mathrm{R}}(\mathbf{x}) = \left(\frac{N\bar{x}}{\sigma^2}\right)^2 \frac{\sigma^2}{N} = \frac{N\bar{x}^2}{\sigma^2} > \gamma$$

which is also identical to the GLRT. In fact, this example is a special case of the linear model in which all three test statistics are identical (see Problem 6.15). In general, these three tests will yield different statistics. The advantage of the alternative tests lies in the hypothesis testing problem for nonlinear signal models [Seber and Wild 1989] and/or nonGaussian noise. The latter situation is explored in the next example.

◇

Example 6.9 - DC Level in NonGaussian Noise - Rao Test

We wish to detect an unknown DC level in independent and identically distributed (IID) *nonGaussian* noise or

$$\begin{aligned} \mathcal{H}_0 &: x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 &: x[n] = A + w[n] & n = 0, 1, \ldots, N-1 \end{aligned}$$

where A is unknown and $-\infty < A < \infty$ and $\{w[0], w[1], \ldots, w[N-1]\}$ are IID with nonGaussian PDF

$$p(w[n]) = \frac{1}{a\sigma\Gamma(\frac{5}{4})2^{\frac{5}{4}}} \exp\left[-\frac{1}{2}\left(\frac{w[n]}{a\sigma}\right)^4\right] \qquad -\infty < w[n] < \infty.$$

The constant a is

$$a = \left(\frac{\Gamma(\frac{1}{4})}{\sqrt{2}\Gamma(\frac{3}{4})}\right)^{\frac{1}{2}} = 1.4464.$$

This PDF is termed a generalized Gaussian or a PDF of the exponential power class [Box and Tiao 1973]. Its mean is zero and its variance is σ^2, which is assumed known. For $\sigma^2 = 1$ it is shown in Figure 6.6 versus a $\mathcal{N}(0, 1)$ PDF. The detection

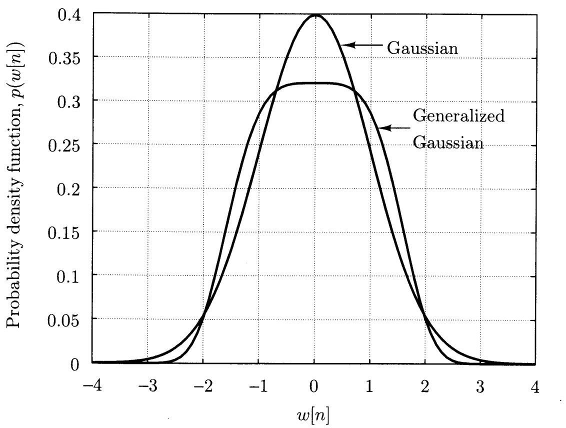


Figure 6.6. Comparison of generalized Gaussian PDF to Gaussian PDF.

problem is equivalent to a parameter test of the PDF

$$\begin{aligned} p(\mathbf{x}; A) &= c^N \prod_{n=0}^{N-1} \exp\left[-\frac{1}{2}\left(\frac{x[n]-A}{a\sigma}\right)^4\right] \\ &= c^N \exp\left[-\frac{1}{2}\sum_{n=0}^{N-1}\left(\frac{x[n]-A}{a\sigma}\right)^4\right] \end{aligned} \tag{6.32}$$

where

$$c = \frac{1}{a\sigma\Gamma(\frac{5}{4})2^{\frac{5}{4}}}$$

for which

$$\begin{aligned} &\mathcal{H}_0 : A = 0 \\ &\mathcal{H}_1 : A \neq 0. \end{aligned}$$

Now, because of the nonGaussian nature of the noise the MLE cannot be easily found. To do so we must minimize

$$J(A) = \sum_{n=0}^{N-1}(x[n]-A)^4$$

which leads to the solution of a cubic equation, followed by a testing of the roots to determine the one that minimizes $J(A)$. The GLRT decides $\mathcal{H}_1$ if

$$L_G(\mathbf{x}) = \frac{p(\mathbf{x}; \hat{A})}{p(\mathbf{x}; 0)} > \gamma$$

or

$$L_G(\mathbf{x}) = \frac{\exp\left[-\frac{1}{2}\sum_{n=0}^{N-1}\left(\frac{x[n]-\hat{A}}{a\sigma}\right)^4\right]}{\exp\left[-\frac{1}{2}\sum_{n=0}^{N-1}\left(\frac{x[n]}{a\sigma}\right)^4\right]}.$$

Taking logarithms produces

$$2\ln L_G(\mathbf{x}) = \sum_{n=0}^{N-1}\left(\frac{x[n]}{a\sigma}\right)^4 - \sum_{n=0}^{N-1}\left(\frac{x[n]-\hat{A}}{a\sigma}\right)^4 > 2\ln\gamma$$

or we decide $\mathcal{H}_1$ if

$$\max_A \left[\sum_{n=0}^{N-1} x^4[n] - \sum_{n=0}^{N-1}(x[n]-A)^4\right] > 2a^4\sigma^4 \ln\gamma = \gamma'.$$

This requires a maximization over A. The Rao test, on the other hand, does not. From (6.31) we have with $\boldsymbol{\theta} = A$ and $\boldsymbol{\theta}_0 = 0$

$$T_{\mathrm{R}}(\mathbf{x}) = \frac{\left(\left.\frac{\partial \ln p(\mathbf{x}; A)}{\partial A}\right|_{A=0}\right)^2}{I(A=0)}.$$

But from (6.32)

$$\begin{aligned} \frac{\partial \ln p(\mathbf{x}; A)}{\partial A} &= -2\sum_{n=0}^{N-1}\left(\frac{x[n]-A}{a\sigma}\right)^3\left(-\frac{1}{a\sigma}\right) \\ &= \frac{2}{a\sigma}\sum_{n=0}^{N-1}\left(\frac{x[n]-A}{a\sigma}\right)^3 \end{aligned} \tag{6.33}$$

and

$$\left.\frac{\partial \ln p(\mathbf{x}; A)}{\partial A}\right|_{A=0} = \frac{2}{a^4\sigma^4}\sum_{n=0}^{N-1} x^3[n].$$

Also [Kay-I 1993, pg. 32]

$$I(A) = -E\left[\frac{\partial^2 \ln p(\mathbf{x}; A)}{\partial A^2}\right]$$

so that from (6.33)

$$
\begin{aligned}
I(A) &= -E\left[\frac{6}{a\sigma}\sum_{n=0}^{N-1}\left(\frac{x[n]-A}{a\sigma}\right)^2\left(-\frac{1}{a\sigma}\right)\right] \\
&= \frac{6}{a^4\sigma^4}E\left[\sum_{n=0}^{N-1}(x[n]-A)^2\right] \\
&= \frac{6N}{a^4\sigma^2}
\end{aligned}
$$

which does not depend on A. Hence,

$$
\begin{aligned}
T_{\mathrm{R}}(\mathbf{x}) &= \frac{\left(\frac{2}{a^4\sigma^4}\sum_{n=0}^{N-1}x^3[n]\right)^2}{\frac{6N}{a^4\sigma^2}} \\
&= \frac{\frac{2}{3}N}{a^4\sigma^6}\left(\frac{1}{N}\sum_{n=0}^{N-1}x^3[n]\right)^2.
\end{aligned}
$$

Note that the statistic is to within a scale factor the square of the third moment. Under $\mathcal{H}_0$ $E(x^3[n]) = 0$ since the PDF is symmetric while under $\mathcal{H}_1$ it will be nonzero.

Lastly, note that the asymptotic PDF of the Rao test will be the same as for the GLRT. It is from (6.23) and (6.27)

$$
T_{\mathrm{R}}(\mathbf{x}) \stackrel{a}{\sim} \begin{cases} \chi_1^2 & \text{under } \mathcal{H}_0 \\ \chi_1'^2(\lambda) & \text{under } \mathcal{H}_1 \end{cases}
$$

where $\lambda = A^2 I(A=0) = 6NA^2/(a^4\sigma^2)$. With respect to Gaussian noise of the same variance (see Example 6.6) the detection performance will be better since $6/a^4 > 1$, and hence the noncentrality parameter is larger. This is because $I(A)$ for $A=0$ can be shown to be minimized when the noise is Gaussian (see Problem 6.20).

$\diamond$

Next we give the corresponding Wald and Rao tests when nuisance parameters are present. Consider the parameter test of (6.21). Then, as shown in Appendix 6B, the Wald test is

$$
T_{\mathrm{W}}(\mathbf{x}) = (\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0})^T\left(\left[\mathbf{I}^{-1}(\hat{\boldsymbol{\theta}}_1)\right]_{\theta_r\theta_r}\right)^{-1}(\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0}) \tag{6.34}
$$

where $\hat{\boldsymbol{\theta}}_1 = [\hat{\boldsymbol{\theta}}_{r_1}^T\ \hat{\boldsymbol{\theta}}_{s_1}^T]^T$ is the MLE of $\boldsymbol{\theta}$ under $\mathcal{H}_1$ (the unrestricted MLE or it maximizes $p(\mathbf{x};\boldsymbol{\theta}) = p(\mathbf{x};\boldsymbol{\theta}_r,\boldsymbol{\theta}_s)$ over $\boldsymbol{\theta}_r,\boldsymbol{\theta}_s$) and $[\mathbf{I}^{-1}(\boldsymbol{\theta})]_{\theta_r\theta_r}$ is the $r\times r$ upper-left

partition of $\mathbf{I}^{-1}(\boldsymbol{\theta})$. Specifically,

$$
\left[\mathbf{I}^{-1}(\boldsymbol{\theta})\right]_{\theta_r\theta_r} = \left(\mathbf{I}_{\theta_r\theta_r}(\boldsymbol{\theta}) - \mathbf{I}_{\theta_r\theta_s}(\boldsymbol{\theta})\mathbf{I}_{\theta_s\theta_s}^{-1}(\boldsymbol{\theta})\mathbf{I}_{\theta_s\theta_r}(\boldsymbol{\theta})\right)^{-1}
$$

where the matrices are defined in (6.25). The Rao test is

$$
T_{\mathrm{R}}(\mathbf{x}) = \left.\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}^T \left[\mathbf{I}^{-1}(\tilde{\boldsymbol{\theta}})\right]_{\theta_r\theta_r} \left.\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} \tag{6.35}
$$

where $\tilde{\boldsymbol{\theta}} = [\boldsymbol{\theta}_{r_0}^T\ \hat{\boldsymbol{\theta}}_{s_0}^T]^T$ is the MLE of $\boldsymbol{\theta}$ under $\mathcal{H}_0$ (the restricted MLE or it maximizes $p(\mathbf{x};\boldsymbol{\theta}_{r_0},\boldsymbol{\theta}_s)$ over $\boldsymbol{\theta}_s$). Each statistic has the same asymptotic PDF as the GLRT or $2\ln L_G(\mathbf{x})$. Clearly, the Rao test, for which only the MLE under $\mathcal{H}_0$ is required, will be the simplest to determine. We illustrate with an example.

Example 6.10 - DC Level in WGN with Unknown Amplitude and Variance - Rao Test

Referring to Example 6.5, we have the parameter test

$$
\begin{aligned}
&\mathcal{H}_0 : A = 0, \sigma^2 > 0 \\
&\mathcal{H}_1 : A \neq 0, \sigma^2 > 0
\end{aligned}
$$

so that $\boldsymbol{\theta} = [A\,\sigma^2]^T$ and $\boldsymbol{\theta}_r = A$, $\boldsymbol{\theta}_s = \sigma^2$. Furthermore, under $\mathcal{H}_0$, $A=0$ so that $\tilde{\boldsymbol{\theta}} = [0\,\hat{\sigma_0^2}]^T$, where $\hat{\sigma_0^2}$ is the MLE of σ^2 under $\mathcal{H}_0$ or when $A=0$. The Rao test is from (6.35)

$$
T_{\mathrm{R}}(\mathbf{x}) = \left.\frac{\partial \ln p(\mathbf{x};A,\sigma^2)}{\partial A}\right|_{A=0,\sigma^2=\hat{\sigma_0^2}}^2 \left[\mathbf{I}^{-1}(\tilde{\boldsymbol{\theta}})\right]_{AA}.
$$

From Example 6.5 we have $\hat{\sigma_0^2} = (1/N)\sum_{n=0}^{N-1}x^2[n]$. Now

$$
p(\mathbf{x};A,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right]
$$

and

$$
\frac{\partial \ln p(\mathbf{x};A,\sigma^2)}{\partial A} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n]-A).
$$

Evaluated at $A=0, \sigma^2 = \hat{\sigma_0^2}$, this becomes

$$
\left.\frac{\partial \ln p(\mathbf{x};A,\sigma^2)}{\partial A}\right|_{A=0,\sigma^2=\hat{\sigma_0^2}} = \frac{N\bar{x}}{\hat{\sigma_0^2}}.
$$

Also,

$$\left[\mathbf{I}^{-1}(\tilde{\boldsymbol{\theta}})\right]_{AA} = \left(I_{AA}(\tilde{\boldsymbol{\theta}}) - I_{A\sigma^2}(\tilde{\boldsymbol{\theta}})I^{-1}_{\sigma^2\sigma^2}(\tilde{\boldsymbol{\theta}})I_{\sigma^2 A}(\tilde{\boldsymbol{\theta}})\right)^{-1}$$

which is easily evaluated using

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{N}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{bmatrix}.$$

Since $I_{A\sigma^2}(\boldsymbol{\theta}) = 0$, we have

$$\left[\mathbf{I}^{-1}(\tilde{\boldsymbol{\theta}})\right]_{AA} = I^{-1}_{AA}(\tilde{\boldsymbol{\theta}}) = \frac{\hat{\sigma_0^2}}{N}.$$

Thus,

$$\begin{aligned} T_{\mathrm{R}}(\mathbf{x}) &= \left(\frac{N\bar{x}}{\hat{\sigma_0^2}}\right)^2 \frac{\hat{\sigma_0^2}}{N} \\ &= \frac{N\bar{x}^2}{\hat{\sigma_0^2}}. \end{aligned}$$

For this example it can easily be shown that the GLRT and Rao test statistics are asymptotically the same. Recall that the GLRT statistic was from (6.19) and (6.18)

$$\begin{aligned} 2\ln L_G(\mathbf{x}) &= N \ln\left(1 + \frac{\bar{x}^2}{\hat{\sigma_1^2}}\right) \\ &= N \ln \frac{1}{1 - \frac{\bar{x}^2}{\hat{\sigma_0^2}}}. \end{aligned}$$

These statistics will be the same if $\bar{x}^2/\hat{\sigma_0^2} \ll 1$ since then $\ln(1/(1-x)) \approx x$ for $x \ll 1$. But for large N, $\bar{x}$ will approach 0 under $\mathcal{H}_0$ and A under $\mathcal{H}_1$, where A is assumed to be small. For finite data records, however, the statistics will be slightly different. The Rao test, of course, is easier to determine. The disadvantage is that it may be somewhat poorer in performance than the GLRT for finite data records. As an example, we find the actual performance of the Rao test for the conditions used to generate Figure 6.5a. Choosing $\lambda = 5$, $\sigma^2 = 1$, $N = 10$, and $A = \sqrt{\lambda\sigma^2/N}$, we have the results shown in Figure 6.7. The theoretical asymptotic performance is given by the dashed curve, the GLRT by the dash-dot curve, and the Rao test by the solid curve. Even for data records as short as $N = 10$ samples, the Rao test yields about the same performance as the GLRT, at least for this example.

$\diamond$

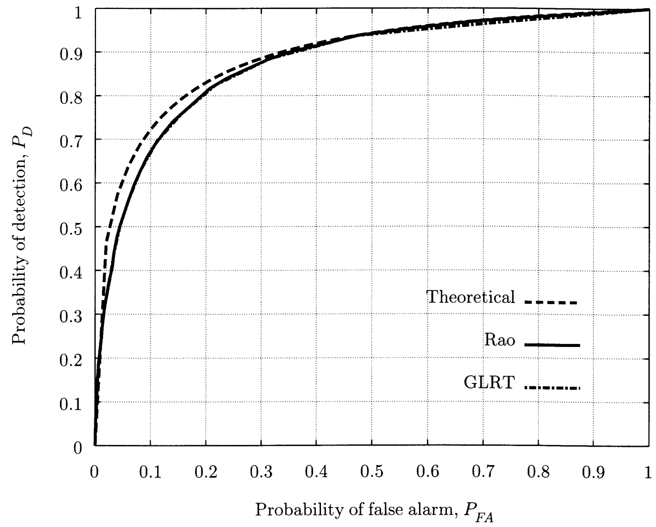


Figure 6.7. Comparison of performance of asymptotically equivalent tests to theoretical performance.

6.7 Locally Most Powerful Detectors

We have noted that a UMP test can only exist if the parameter test is one-sided. Even when a one-sided test does not lead to a UMP test, an asymptotic version of the UMP test may be found. It assumes the parameter to be tested is a scalar, and there are no nuisance parameters. This test is called the *locally most powerful* (LMP) test. To describe its implementation, consider the one-sided parameter test

$$\begin{aligned} &\mathcal{H}_0 : \theta = \theta_0 \\ &\mathcal{H}_1 : \theta > \theta_0 \end{aligned}$$

with no nuisance parameters. The PDF under $\mathcal{H}_0$ and $\mathcal{H}_1$ is parameterized by θ and is given by $p(\mathbf{x};\theta)$. If we wish to test for values of θ that are near θ_0, then an LMP test exists. In the context of signal detection this requirement stipulates that θ is the signal amplitude and it is assumed to be small ($\theta = A$ and $\theta_0 = 0$). The LMP test constrains the P_{FA} and maximizes P_D for all θ close to θ_0 with $\theta > \theta_0$. For large departures of θ from θ_0, however, there is no guarantee of optimality. Hence, a GLRT should also be tried.

To derive the LMP test we first consider the NP test or we decide $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x};\theta)}{p(\mathbf{x};\theta_0)} > \gamma.$$

Taking the logarithm produces

$$\ln p(\mathbf{x};\theta) - \ln p(\mathbf{x};\theta_0) > \ln\gamma.$$

Assuming that $\theta-\theta_0$ is small, we have upon using a first-order Taylor expansion of $\ln p(\mathbf{x};\theta)$ about $\theta=\theta_0$

$$\ln p(\mathbf{x};\theta) = \ln p(\mathbf{x};\theta_0) + \left.\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right|_{\theta=\theta_0}(\theta-\theta_0)$$

so that we decide $\mathcal{H}_1$ if

$$\left.\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right|_{\theta=\theta_0}(\theta-\theta_0) > \ln\gamma$$

or since $\theta>\theta_0$, we decide $\mathcal{H}_1$ if

$$\left.\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right|_{\theta=\theta_0} > \frac{\ln\gamma}{\theta-\theta_0} = \gamma'.$$

The scaled statistic

$$T_{\mathrm{LMP}}(\mathbf{x}) = \frac{\left.\dfrac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right|_{\theta=\theta_0}}{\sqrt{I(\theta_0)}} \tag{6.36}$$

is the LMP test. That it maximizes P_D follows from its equivalence to the NP test under weak signal conditions. Alternatively, it is shown in Appendix 6E to maximize the slope of the P_D curve with θ at $\theta=\theta_0$. For large data records it is shown in Appendix 6D that

$$T_{\mathrm{LMP}}(\mathbf{x}) \overset{a}{\sim} \begin{cases} \mathcal{N}(0,1) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(\sqrt{I(\theta_0)}(\theta_1-\theta_0),1) & \text{under } \mathcal{H}_1 \end{cases} \tag{6.37}$$

where θ_1 is the true value of θ under $\mathcal{H}_1$ and $I(\theta_0)$ is the Fisher information evaluated at θ_0. Note that

$$T_{\mathrm{LMP}}(\mathbf{x}) = \left.\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right|_{\theta=\theta_0}\sqrt{I^{-1}(\theta_0)}$$

may be thought of as the one-sided Rao test for a scalar parameter with no nuisance parameters (see (6.31)). The usual Rao test is for a two-sided test, which necessitates the squaring operation. Finally, from (6.37) we see that the LMP test has the

usual detection performance of a mean-shifted Gauss-Gauss detector (see Chapter 3) whose deflection coefficient is

$$d^2 = (\theta_1-\theta_0)^2 I(\theta_0).$$

An example follows. See also Example 8.6.

Example 6.11 - Correlation Testing

Assume we observe IID Gaussian vectors $\{\mathbf{x}[0],\mathbf{x}[1],\ldots,\mathbf{x}[N-1]\}$, where each $\mathbf{x}[n]$ is 2×1. The PDF is $\mathbf{x}[n]\sim\mathcal{N}(\mathbf{0},\mathbf{C})$, where

$$\mathbf{C} = \sigma^2\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

or $\mathbf{x}[n]$ is bivariate Gaussian. We assume that σ^2 is known and inquire if $\rho=0$ or $\rho>0$. The hypothesis test is

$$\begin{aligned} \mathcal{H}_0 &: \rho = 0 \\ \mathcal{H}_1 &: \rho > 0. \end{aligned}$$

A GLRT would require us to find the MLE of ρ. In [Kay-I 1993, Problem 9.3] we showed that this necessitated the solution of a cubic equation. Furthermore, even if we implemented the GLRT, the asymptotic PDF is not given by (6.23) due to the one-sided nature of the parameter test (see Problem 6.23). The choice of a threshold becomes difficult. Instead, we appeal to the LMP test, assuming that ρ under $\mathcal{H}_1$ is positive and close to zero. To do so we have

$$p(\mathbf{x};\rho) = \prod_{n=0}^{N-1}\frac{1}{2\pi\det^{\frac{1}{2}}(\mathbf{C})}\exp\left(-\frac{1}{2}\mathbf{x}^T[n]\mathbf{C}^{-1}\mathbf{x}[n]\right)$$

and since

$$\begin{aligned} \det(\mathbf{C}) &= \sigma^4(1-\rho^2) \\ \mathbf{C}^{-1} &= \frac{1}{\sigma^2(1-\rho^2)}\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \end{aligned}$$

we have

$$\ln p(\mathbf{x};\rho) = -N\ln 2\pi - \frac{N}{2}\ln\sigma^4(1-\rho^2) - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\mathbf{x}^T[n]\mathbf{C}_0^{-1}\mathbf{x}[n]$$

where

$$\mathbf{C}_0^{-1} = \begin{bmatrix} \dfrac{1}{1-\rho^2} & \dfrac{-\rho}{1-\rho^2} \\ \dfrac{-\rho}{1-\rho^2} & \dfrac{1}{1-\rho^2} \end{bmatrix}.$$

Differentiating produces

$$\frac{\partial \ln p(\mathbf{x};\rho)}{\partial \rho} = \frac{N\rho}{1-\rho^2} - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\mathbf{x}^T[n]\frac{\partial \mathbf{C}_0^{-1}}{\partial \rho}\mathbf{x}[n]$$

where

$$\frac{\partial \mathbf{C}_0^{-1}}{\partial \rho} = \begin{bmatrix} \frac{2\rho}{(1-\rho^2)^2} & -\frac{1+\rho^2}{(1-\rho^2)^2} \\ -\frac{1+\rho^2}{(1-\rho^2)^2} & \frac{2\rho}{(1-\rho^2)^2} \end{bmatrix}$$

so that

$$\left.\frac{\partial \ln p(\mathbf{x};\rho)}{\partial \rho}\right|_{\rho=0} = -\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\mathbf{x}^T[n]\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}\mathbf{x}[n].$$

Letting $\mathbf{x}[n] = [x_1[n]\, x_2[n]]^T$, we have

$$\left.\frac{\partial \ln p(\mathbf{x};\rho)}{\partial \rho}\right|_{\rho=0} = \frac{\sum_{n=0}^{N-1} x_1[n]x_2[n]}{\sigma^2}.$$

It can be shown that [Kay-I 1993, Problem 3.15]

$$I(\rho) = \frac{N(1+\rho^2)}{(1-\rho^2)^2}$$

so that $I(0) = N$, and finally from (6.36) we decide $\mathcal{H}_1$ if

$$T_{\mathrm{LMP}}(\mathbf{x}) = \sqrt{N}\hat{\rho} > \gamma' \tag{6.38}$$

where

$$\hat{\rho} = \frac{\frac{1}{N}\sum_{n=0}^{N-1} x_1[n]x_2[n]}{\sigma^2}$$

is an estimate of ρ (although not the MLE). The asymptotic detection performance is easily found from (6.37) as

$$T_{\mathrm{LMP}}(\mathbf{x}) \overset{a}{\sim} \begin{cases} \mathcal{N}(0,1) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(\sqrt{N}\rho, 1) & \text{under } \mathcal{H}_1. \end{cases}$$

Note that the deflection coefficient is $d^2 = N\rho^2$.

◇

6.8 Multiple Hypothesis Testing

The problem of *multiple composite hypothesis testing* is much more difficult than for the previous binary composite case. The optimal Bayesian approach discussed in Chapter 3 no longer applies. There we saw that to minimize the probability of error P_e for a set of equally likely hypotheses, we should implement the maximum likelihood rule. This rule chose the hypothesis for which $p(\mathbf{x}|\mathcal{H}_i)$ is maximum. It was assumed that the hypotheses could be assigned prior probabilities. When there are unknown PDF parameters $\boldsymbol{\theta}_i$, this approach cannot be implemented. We denote the PDF in this case as $p(\mathbf{x};\boldsymbol{\theta}_i|\mathcal{H}_i)$. It is clear that due to our lack of knowledge of $\boldsymbol{\theta}_i$ the PDF cannot be evaluated. One way to circumvent this problem is to use the Bayesian approach. We consider $\boldsymbol{\theta}_i$ to be the realization of a random vector and assign to it a prior PDF $p(\boldsymbol{\theta}_i)$. In the Bayesian paradigm we will denote the data PDF conditioned on the unknown parameter outcome as $p(\mathbf{x}|\boldsymbol{\theta}_i,\mathcal{H}_i)$. Then, the unconditional data PDF is found as

$$p(\mathbf{x}|\mathcal{H}_i) = \int p(\mathbf{x}|\boldsymbol{\theta}_i,\mathcal{H}_i)p(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i$$

allowing us to implement the maximum likelihood rule. Due to the difficulty of performing the integration and the need for prior knowledge, this approach does not appear to be used much in practice. However, an approximation to it is widely used and will be discussed shortly. Before doing so the question arises as to whether the GLRT philosophy can be extended to accommodate multiple hypotheses. Unfortunately, this does not appear to be possible as is now illustrated.

We consider the problem of detecting a signal that is modeled as a DC level or a line in WGN. The hypothesis test is

$$\begin{aligned} \mathcal{H}_0 &: x[n] = w[n] \\ \mathcal{H}_1 &: x[n] = A + w[n] \\ \mathcal{H}_2 &: x[n] = A + Bn + w[n] \end{aligned}$$

for $n = 0, 1, \ldots, N-1$, and where A, B, σ^2 are unknown nonrandom parameters. As formulated, the PDF conditioned on $\mathcal{H}_2$ is

$$p(\mathbf{x};A,B,\sigma^2|\mathcal{H}_2) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A-Bn)^2\right]$$

while those conditioned on $\mathcal{H}_1$ and $\mathcal{H}_0$ are respectively $p(\mathbf{x};A,B=0,\sigma^2|\mathcal{H}_2)$ and $p(\mathbf{x};A=0,B=0,\sigma^2|\mathcal{H}_2)$. The unknown parameters for the PDFs conditioned on $\mathcal{H}_0$ and $\mathcal{H}_1$ are a subset of those for the PDFs conditioned on $\mathcal{H}_2$. Specifically, we can write

$$\begin{aligned} \boldsymbol{\theta}_0 &= \sigma^2 \\ \boldsymbol{\theta}_1 &= \begin{bmatrix} \sigma^2 \\ A \end{bmatrix} = \begin{bmatrix} \boldsymbol{\theta}_0 \\ A \end{bmatrix} \end{aligned}$$

$$\boldsymbol{\theta}_2 = \begin{bmatrix} \sigma^2 \\ A \\ B \end{bmatrix} = \begin{bmatrix} \boldsymbol{\theta}_1 \\ B \end{bmatrix}.$$

As such, the parameter spaces are *nested.* As a reasonable extension to the maximum likelihood rule when there are unknown parameters, we might decide $\mathcal{H}_k$ if

$$\max_{\boldsymbol{\theta}_i} p(\mathbf{x}; \boldsymbol{\theta}_i | \mathcal{H}_i) \tag{6.39}$$

is maximum for $i = k$, where $i = 0, 1, 2$. This approach is similar in spirit to the GLRT. But because of the nesting, $p(\mathbf{x}; \hat{\boldsymbol{\theta}}_2|\mathcal{H}_2)$ will *always* be maximum and hence $\mathcal{H}_2$ will always be chosen. This is because $\hat{\boldsymbol{\theta}}_2$ is obtained by maximizing $p(\mathbf{x}; A, B, \sigma^2|\mathcal{H}_2)$ over all three parameters while $\hat{\boldsymbol{\theta}}_1$, $\hat{\boldsymbol{\theta}}_0$ are *constrained* MLEs. In fact, we have

$$\hat{\sigma_0^2} = \frac{1}{N}\sum_{n=0}^{N-1} x^2[n]$$

so that

$$p(\mathbf{x}; \hat{\sigma_0^2}|\mathcal{H}_0) = \frac{1}{(2\pi\hat{\sigma_0^2})^{\frac{N}{2}}} \exp\left(-\frac{N}{2}\right).$$

Similarly,

$$\begin{aligned} \hat{\sigma_1^2} &= \frac{1}{N}\sum_{n=0}^{N-1}(x[n] - \hat{A}_1)^2 \\ \hat{\sigma_2^2} &= \frac{1}{N}\sum_{n=0}^{N-1}(x[n] - \hat{A}_2 - \hat{B}_2 n)^2 \end{aligned}$$

where $\hat{A}_1$ is found by maximizing $p(\mathbf{x}; A, B = 0, \sigma^2|\mathcal{H}_2) = p(\mathbf{x}; A, \sigma^2|\mathcal{H}_1)$ and $\hat{A}_2, \hat{B}_2$ are found by maximizing $p(\mathbf{x}; A, B, \sigma^2|\mathcal{H}_2)$. We have then that

$$\begin{aligned} p(\mathbf{x}; \hat{A}_1, \hat{\sigma_1^2}|\mathcal{H}_1) &= \frac{1}{(2\pi\hat{\sigma_1^2})^{\frac{N}{2}}} \exp\left(-\frac{N}{2}\right) \\ p(\mathbf{x}; \hat{A}_2, \hat{B}_2, \hat{\sigma_2^2}|\mathcal{H}_2) &= \frac{1}{(2\pi\hat{\sigma_2^2})^{\frac{N}{2}}} \exp\left(-\frac{N}{2}\right). \end{aligned}$$

To maximize the PDF according to (6.39) we choose the one that has the minimum $\hat{\sigma_i^2}$. But $\hat{\sigma_i^2}$ is just the minimum least squares error or the minimum of

$$\begin{aligned} \hat{\sigma_0^2} &= \frac{1}{N}\sum_{n=0}^{N-1} x^2[n] \\ \hat{\sigma_1^2} &= \min_A \frac{1}{N}\sum_{n=0}^{N-1}(x[n] - A)^2 \\ \hat{\sigma_2^2} &= \min_{A,B} \frac{1}{N}\sum_{n=0}^{N-1}(x[n] - A - Bn)^2. \end{aligned}$$

As such, $\hat{\sigma_2^2}$ will always be minimum and hence $p(\mathbf{x}; \hat{A}_2, \hat{B}_2, \hat{\sigma_2^2})$ will always be maximum. The reader may wish to review [Kay-I 1993, Section 8.6]. In other words, the *modeling error* must decrease as we add more parameters to the model. To offset this tendency we would like to decide $\mathcal{H}_k$ for which the model (or signal) explains the data (via a least squares fit), but is as simple as possible in that it has the *fewest number of parameters.* One such criterion is to decide $\mathcal{H}_k$ if

$$\xi_i = \ln p(\mathbf{x}; \hat{\boldsymbol{\theta}}_i|\mathcal{H}_i) - \frac{1}{2}\ln\det\left(\mathbf{I}(\hat{\boldsymbol{\theta}}_i)\right) \tag{6.40}$$

is maximized for $i = k$. The second term is a penalty term that becomes more negative as i increases. It attempts to counterbalance the decrease in $\hat{\sigma_i^2}$ and hence the increase in $p(\mathbf{x}; \hat{\boldsymbol{\theta}}_i|\mathcal{H}_i)$. It has been assumed that the hypotheses are equally likely or $P(\mathcal{H}_i) = 1/M$. This criterion is termed the *generalized ML rule* and is derived in Appendix 6F. We illustrate its use to solve the previous problem.

Example 6.12 - DC Level or Line in WGN - Unknown Parameters

The MLEs of A, B have been derived in [Kay-I 1993, Example 8.6]. To simplify the subsequent calculations, however, we change the observation interval to be $n = -M, \ldots, 0, \ldots, M$. This makes $\mathbf{H}^T\mathbf{H}$ a diagonal matrix, which is easily inverted. Then, it follows in a similar fashion to [Kay-I 1993, Example 8.6] that

$$\begin{aligned} \hat{A}_1 &= \bar{x} \\ \hat{A}_2 &= \bar{x} \\ \hat{B}_2 &= \frac{\sum_{n=-M}^{M} n x[n]}{\sum_{n=-M}^{M} n^2}. \end{aligned}$$

The Fisher information matrices are also easily shown to be [Kay-I 1993, Example

3.7]

$$
\begin{aligned}
\mathbf{I}(\boldsymbol{\theta}_2) &= \begin{bmatrix} \frac{N}{2\sigma^4} & 0 & 0 \\ 0 & \frac{N}{\sigma^2} & 0 \\ 0 & 0 & \frac{1}{\sigma^2}\sum_{n=-M}^{M} n^2 \end{bmatrix} \\
\mathbf{I}(\boldsymbol{\theta}_1) &= \begin{bmatrix} \frac{N}{2\sigma^4} & 0 \\ 0 & \frac{N}{\sigma^2} \end{bmatrix} \\
\mathbf{I}(\boldsymbol{\theta}_0) &= \frac{N}{2\sigma^4}
\end{aligned}
$$

where $N = 2M+1$ and $\sum_{n=-M}^{M} n^2 = N(N^2-1)/12 \approx N^3/12$. Hence, from (6.40)

$$
\begin{aligned}
\xi_i &= \ln\left[\frac{1}{(2\pi\hat{\sigma_i^2})^{\frac{N}{2}}}\exp\left(-\frac{N}{2}\right)\right] - \frac{1}{2}\ln\det(\mathbf{I}(\hat{\boldsymbol{\theta}}_i)) \\
&= -\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln\hat{\sigma_i^2} - \frac{N}{2} - \frac{1}{2}\ln\det(\mathbf{I}(\hat{\boldsymbol{\theta}}_i)).
\end{aligned}
$$

Ignoring the constant terms and scaling by -1, we have

$$
\xi_i' = \frac{N}{2}\ln\hat{\sigma_i^2} + \frac{1}{2}\ln\det(\mathbf{I}(\hat{\boldsymbol{\theta}}_i)).
$$

We now need to *minimize* ξ_i' over i. But

$$
\begin{aligned}
\xi_0' &= \frac{N}{2}\ln\hat{\sigma_0^2} + \frac{1}{2}\ln N - \frac{1}{2}\ln 2\hat{\sigma}_0^4 \\
\xi_1' &= \frac{N}{2}\ln\hat{\sigma_1^2} + \ln N - \frac{1}{2}\ln 2\hat{\sigma}_1^6 \\
\xi_2' &= \frac{N}{2}\ln\hat{\sigma_2^2} + \frac{5}{2}\ln N - \frac{1}{2}\ln 12 - \frac{1}{2}\ln 2\hat{\sigma}_2^8
\end{aligned}
$$

and ignoring the terms that do not depend on N since their contribution is small as $N \to \infty$, we have

$$
\begin{aligned}
\xi_0'' &= \frac{N}{2}\ln\hat{\sigma_0^2} + \frac{1}{2}\ln N \\
\xi_1'' &= \frac{N}{2}\ln\hat{\sigma_1^2} + \ln N \\
\xi_2'' &= \frac{N}{2}\ln\hat{\sigma_2^2} + \frac{5}{2}\ln N.
\end{aligned}
$$

Note that the penalty term increases with the number of parameters while the fitting term decreases. The chosen model is a compromise between these two conflicting requirements. Underfitting leads to a large bias of the parameters (which inflates $E(\hat{\sigma_i^2})$) while overfitting increases the variance due to the estimation of an excessive number of parameters. This criterion can be shown to be consistent in that as $N \to \infty$, the correct hypothesis will be chosen [Bozdogan 1987]. However, for finite data records there is no optimality.

$\diamond$

An approximation to (6.40) has been proposed based on information theoretic coding considerations. Termed the minimum description length (MDL), it chooses the hypothesis that minimizes [Rissanen 1978]

$$
\mathrm{MDL}(i) = -\ln p(\mathbf{x};\hat{\boldsymbol{\theta}}_i|\mathcal{H}_i) + \frac{n_i}{2}\ln N \tag{6.41}
$$

where n_i is the number of estimated parameters or equivalently the dimensionality of $\boldsymbol{\theta}_i$. To show that it approximates the generalized maximum likelihood rule, we first note that from (6.40)

$$
-\xi_i = -\ln p(\mathbf{x};\hat{\boldsymbol{\theta}}_i|\mathcal{H}_i) + \frac{1}{2}\ln\det(\mathbf{I}(\hat{\boldsymbol{\theta}}_i)).
$$

As an approximation to the determinant term we let $\det(\mathbf{I}(\hat{\boldsymbol{\theta}}_i)) = cN^{i+1}$ for some constant c. In the previous example, we had $\det(\mathbf{I}(\hat{\boldsymbol{\theta}}_0)) = N/(2\hat{\sigma}_0^4)$, $\det(\mathbf{I}(\hat{\boldsymbol{\theta}}_1)) = N^2/(2\hat{\sigma}_1^6)$, and $\det(\mathbf{I}(\hat{\boldsymbol{\theta}}_2)) \approx N^5/(24\hat{\sigma}_2^8)$. Clearly, this approximation is just that and its accuracy depends upon the Fisher information matrix. Then, we have

$$
-\xi_i = -\ln p(\mathbf{x};\hat{\boldsymbol{\theta}}_i|\mathcal{H}_i) + \frac{1}{2}\ln c + \frac{i+1}{2}\ln N
$$

which, after discarding the constant term, is just the MDL, since $n_i = i+1$ is the number of unknown parameters.

References

Box, G.E.P., G.C. Tiao, *Bayesian Inference in Statistical Analysis*, Addison- Wesley, Reading, Mass., 1973.

Bozdogan, H., "Model Selection and Akaike's Information Criterion (AIC): The General Theory and Analytical Extensions," *Psychometrika*, Vol. 512, pp. 345–370, 1987.

Chernoff, H., "On the Distribution of the Likelihood Ratio," *Ann. Math. Statist.*, Vol. 25, pp. 573–578, 1954.

Kendall, Sir M., A. Stuart, *The Advanced Theory of Statistics*, Vol. 2, Macmillan, New York, 1979.

Lehmann, E.L., *Testing Statistical Hypotheses*, J. Wiley, New York, 1959.

Rissanen, J., "Modeling by Shortest Data Description," *Automatica*, Vol. 14, pp. 465-471, 1978.

Seber, G.A.F., C.J. Wild, *Nonlinear Regression*, J. Wiley, New York, 1989.

Zacks, S., *Parametric Statistical Inference*, Pergamon, New York, 1981.

Problems

6.1 Derive the NP test for a DC level in WGN with A unknown and $A < 0$ as given by (6.5).

6.2 We observe two samples $x[n]$ for $n = 0, 1$ from the exponential PDF

$$p(x[n]) = \begin{cases} \lambda \exp(-\lambda x[n]) & \text{for } x[n] > 0 \\ 0 & \text{for } x[n] < 0 \end{cases}$$

where λ is unknown and $\lambda > 0$. The samples are assumed to be independent and identically distributed (IID). For the hypothesis testing problem

$$\begin{aligned} \mathcal{H}_0 &: \lambda = \lambda_0 \\ \mathcal{H}_1 &: \lambda > \lambda_0 \end{aligned}$$

determine if a UMP test exists. If so, find $T(\mathbf{x})$ and P_{FA} as a function of the threshold.

6.3 Repeat Problem 6.2 but now consider the two-sided test

$$\begin{aligned} \mathcal{H}_0 &: \lambda = \lambda_0 \\ \mathcal{H}_1 &: \lambda \neq \lambda_0. \end{aligned}$$

6.4 In Example 6.1 we assume that $A > 0$ and use (6.3) as our test. If, in fact, $A < 0$, find P_D as a function of A.

6.5 Derive the detection probability given by (6.9).

6.6 We wish to detect a DC level A in WGN with variance σ^2 based on the samples $x[n]$ for $n = 0, 1, \ldots, N-1$. The amplitude of the DC level A is known but the variance σ^2 of the noise is unknown. Use a Bayesian approach with the prior PDF

$$p(\sigma^2) = \begin{cases} \dfrac{\lambda \exp(-\lambda/\sigma^2)}{\sigma^4} & \sigma^2 > 0 \\ 0 & \sigma^2 < 0 \end{cases}$$

where $\lambda > 0$. Find the detector that maximizes P_D for a fixed P_{FA}. Do not evaluate the threshold. Explain what happens as $\lambda \to 0$. Hint: You will need the Gamma integral

$$\int_0^\infty x^{b-1} \exp(-ax)dx = a^{-b}\Gamma(b).$$

6.7 Find P_D and P_{FA} for the detector of (6.11). What happens as $\sigma_A^2 \to \infty$?

6.8 We wish to detect a damped exponential $s[n] = Ar^n$, where A is unknown and r is known $(0 < r < 1)$, in WGN with known variance σ^2. Based on $x[n]$ for $n = 0, 1, \ldots, N-1$ show that the GLRT decides $\mathcal{H}_1$ if $\hat{A}^2 > \gamma'$, where $\hat{A}$ is the MLE of A.

6.9 For Problem 6.8 we now assume that the noise variance is also unknown. Find $2\ln L_G(\mathbf{x})$ and its asymptotic PDF. The Fisher information matrix can be shown to be

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} \sum_{n=0}^{N-1} \dfrac{r^{2n}}{\sigma^2} & 0 \\ 0 & \dfrac{N}{2\sigma^4} \end{bmatrix}.$$

Hint: See Examples 6.5 and 6.7.

6.10 We observe the IID samples $x[n]$ for $n = 0, 1, \ldots, N-1$, which consist of noise only with an unknown variance, and wish to determine if the noise if Gaussian or not. We model the nonGaussian noise as Laplacian so that we have the hypothesis testing problem

$$\begin{aligned} \mathcal{H}_0 &: p(x[n]; \mathcal{H}_0) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}x^2[n]\right) \quad n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 &: p(x[n]; \mathcal{H}_1) = \frac{1}{\sqrt{2\sigma^2}} \exp\left(-\sqrt{\frac{2}{\sigma^2}}|x[n]|\right) \quad n = 0, 1, \ldots, N-1 \end{aligned}$$

where σ^2 is the unknown variance. Find the GLRT statistic $L_G(\mathbf{x})$. Hint: The MLE of σ^2 under $\mathcal{H}_1$ is $\hat{\sigma_1^2} = [(\sqrt{2}/N)\sum_{n=0}^{N-1}|x[n]|]^2$.

6.11 Consider the parameter test

$$\begin{aligned} \mathcal{H}_0 &: \theta = \theta_0 \\ \mathcal{H}_1 &: \theta \neq \theta_0. \end{aligned}$$

We observe the IID samples $x[n]$ for $n = 0, 1, \ldots, N-1$. If the PDF can be factored as

$$p(\mathbf{x}; \theta) = g(T(\mathbf{x}), \theta)h(\mathbf{x})$$

then $T(\mathbf{x})$ is a sufficient statistic for θ (see [Kay-I 1993, pp. 104–105]). Assume that $T(\mathbf{x})$ is a sufficient statistic for θ and find the GLRT statistic for this problem to show that it is a function of only the sufficient statistic. Hint: Recall that $h(\mathbf{x}) \geq 0$ for all $\mathbf{x}$.

6.12 Apply the results of Problem 6.11 to Example 6.4 by first performing the factorization to show that $T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]$. Then find the GLRT statistic.

6.13 Consider the detection of a DC level in WGN when A is known and $A > 0$ but σ^2 is not. Show that the clairvoyant detector decides $\mathcal{H}_1$ if

$$\frac{\bar{x}}{\sqrt{\sigma^2}} > \frac{1}{\sqrt{N}} Q^{-1}(P_{FA}).$$

Then replace the assumed known value of σ^2 by the estimator

$$\hat{\sigma_0^2} = (1/N) \sum_{n=0}^{N-1} x^2[n]$$

in the detector. This approach is called an *estimate and plug* detector. Is the threshold correct?

6.14 We wish to detect a DC level in IID Laplacian noise with known variance (see Problem 6.10 for definition of noise PDF). If the amplitude A is unknown, use (6.15) to find the GLRT. Do not perform the maximization.

6.15 The classical linear model was defined by $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$, where $\mathbf{H}$ is a known $N \times p$ matrix, $\boldsymbol{\theta}$ is an unknown $p \times 1$ parameter vector, and $\mathbf{w}$ is a random noise vector with PDF $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ (see [Kay-I 1993, Chapter 4]). Assuming that σ^2 is known, show that for the parameter test

$$\begin{aligned} \mathcal{H}_0 &: \boldsymbol{\theta} = \mathbf{0} \\ \mathcal{H}_1 &: \boldsymbol{\theta} \neq \mathbf{0} \end{aligned}$$

the GLRT, Rao, and Wald tests decide $\mathcal{H}_1$ if

$$2 \ln L_G(\mathbf{x}) = T_{\mathrm{W}}(\mathbf{x}) = T_{\mathrm{R}}(\mathbf{x}) = \frac{\hat{\boldsymbol{\theta}}_1^T \mathbf{H}^T \mathbf{H} \hat{\boldsymbol{\theta}}_1}{\sigma^2} > \gamma.$$

The estimator $\hat{\boldsymbol{\theta}}_1 = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$ is the MLE of $\boldsymbol{\theta}$ under $\mathcal{H}_1$ or equivalently the unrestricted MLE and $\mathbf{I}(\boldsymbol{\theta}) = \mathbf{H}^T\mathbf{H}/\sigma^2$ is the Fisher information matrix. Hint: For the Rao test show that

$$\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{I}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta})$$

by using $\partial \mathbf{b}^T\boldsymbol{\theta}/\partial\boldsymbol{\theta} = \mathbf{b}$ and $\partial \boldsymbol{\theta}^T\mathbf{A}\boldsymbol{\theta}/\partial\boldsymbol{\theta} = 2\mathbf{A}\boldsymbol{\theta}$ for $\mathbf{A}$ a symmetric matrix.

6.16 Apply the results of Problem 6.15 to the problem of detecting a DC level of unknown amplitude in WGN with known variance. Compare the GLRT statistic to that obtained in Example 6.4.

6.17 For the classical linear model described in Problem 6.15 show that the exact PDF (for finite data records) is

$$T(\mathbf{x}) = 2 \ln L_G(\mathbf{x}) \sim \begin{cases} \chi_p^2 & \text{under } \mathcal{H}_0 \\ \chi_p'^2(\lambda) & \text{under } \mathcal{H}_1 \end{cases}$$

where

$$\lambda = \frac{\boldsymbol{\theta}_1^T \mathbf{H}^T \mathbf{H} \boldsymbol{\theta}_1}{\sigma^2}$$

and $\boldsymbol{\theta}_1$ is the true value of $\boldsymbol{\theta}$ under $\mathcal{H}_1$. Hint: Use the result that if $\mathbf{x}$ is a $p \times 1$ vector random variable with $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$, then $\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x} \sim \chi_p'^2(\lambda)$ where $\lambda = \boldsymbol{\mu}^T\mathbf{C}^{-1}\boldsymbol{\mu}$ (see Section 2.3).

6.18 For the problem discussed in Example 6.7 perform a Monte Carlo computer simulation to verify the results of Figure 6.5a. To do so let $N = 10$, $A = 1/\sqrt{2}$, and $\sigma^2 = 1$.

6.19 We wish to detect a sinusoid $s[n] = a \cos 2\pi f_0 n + b \sin 2\pi f_0 n$ for $n = 0, 1, \ldots, N-1$ embedded in WGN with known variance σ^2. The frequency is $f_0 = k/N$ for $k \in \{1, 2, \ldots, N/2-1\}$. If a and b are unknown, determine the GLRT, Rao, and Wald tests. Also, determine the detection performance and interpret the noncentrality parameter. Hint: Note that the data can be put in the form of the linear model so that the results of Problems 6.15 and 6.17 can be applied. Also, the columns of $\mathbf{H}$ are orthogonal.

6.20 The Fisher information for the amplitude of a DC level based on a single sample of a DC level of unknown amplitude in noise with PDF $p(w[n])$ can be shown to be

$$i(A) = \int_{-\infty}^{\infty} \frac{\left(\frac{dp(u)}{du}\right)^2}{p(u)} du$$

[Kay-I 1993, Problem 3.2]. For N IID samples the Fisher information is multiplied by N. Show that of all the PDFs with mean zero and variance one, $i(A)$ is minimum for a Gaussian PDF. To do so use the Cauchy-Schwarz inequality

$$\left(\int_{-\infty}^{\infty} \frac{d \ln p(u)}{du} u p(u) du\right)^2 \leq \int_{-\infty}^{\infty} \left(\frac{d \ln p(u)}{du}\right)^2 p(u) du \int_{-\infty}^{\infty} u^2 p(u) du$$

which holds with equality if and only if

$$\frac{d \ln p(u)}{du} = cu$$

for c a constant.

6.21 Find the LMP test for the parameter testing problem

$$\begin{array}{l} \mathcal{H}_0 : \sigma^2 = \sigma_0^2 \\ \mathcal{H}_1 : \sigma^2 > \sigma_0^2 \end{array}$$

where the IID samples $x[n] \sim \mathcal{N}(0, \sigma^2)$ for $n = 0, 1, \ldots, N-1$ are observed. Does a UMP test exist?

6.22 In Example 6.11 assume that we wish to test if $\rho = 0$ versus $\rho \neq 0$. Find the Rao test and its asymptotic performance.

6.23 For the DC level with unknown amplitude A (but $A > 0$) in WGN with known variance σ^2 (as in Example 6.1), derive the GLRT as follows. First show that the MLE is

$$\hat{A} = \max(0, \bar{x}).$$

Then, find the GLRT test statistic to show that

$$\begin{aligned} 2 \ln L_G(\mathbf{x}) &= 2 \ln \frac{p(\mathbf{x}; \hat{A}, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} \\ &= \frac{N}{\sigma^2}(2\hat{A}\bar{x} - \hat{A}^2) \\ &= \begin{cases} \frac{N}{\sigma^2}\bar{x}^2 & \bar{x} > 0 \\ 0 & \bar{x} \leq 0. \end{cases} \end{aligned}$$

Next show that the PDF of $y = 2 \ln L_G(\mathbf{x})$ under $\mathcal{H}_0$ is

$$p(y) = \frac{1}{2}\delta(y) + \frac{1}{2} p_{\chi_1^2}(y)$$

where $\delta(y)$ is a Dirac delta function and $p_{\chi_1^2}$ is the PDF of a χ_1^2 random variable [Chernoff 1954]. This example serves to illustrate that the asymptotic GLRT statistics only apply when the parameter to be tested is an *interior* point of its domain.

6.24 For the linear model described in Problem 6.15, assume that we wish to decide among multiple linear models having different dimensions, where the ith model is characterized by an $N \times i$ observation matrix. The models have an equal

prior probability of occurrence. The unknown parameter vector $\boldsymbol{\theta}_i$ is $i \times 1$ and the models are nested. That is to say we wish to decide between the hypotheses that $\boldsymbol{\theta}$ is 1×1, 2×1, 3×1, etc. We let the MLE of $\boldsymbol{\theta}_i$ be denoted by $\hat{\boldsymbol{\theta}}_i$. Also, let the $N \times i$ observation matrix be denoted by $\mathbf{H}_i$. Use the generalized ML rule to decide among the hypotheses by using the results

$$\begin{aligned} \hat{\boldsymbol{\theta}}_i &= (\mathbf{H}_i^T \mathbf{H}_i)^{-1} \mathbf{H}_i^T \mathbf{x} \\ \mathbf{I}(\boldsymbol{\theta}_i) &= \frac{\mathbf{H}_i^T \mathbf{H}_i}{\sigma^2} \end{aligned}$$

in (6.40). You should be able to show that

$$\begin{aligned} \xi_i = & -\frac{N}{2} \ln 2\pi\sigma^2 + \frac{i}{2} \ln \sigma^2 - \frac{1}{2} \ln \det(\mathbf{H}_i^T \mathbf{H}_i) \\ & -\frac{1}{2\sigma^2} \mathbf{x}^T (\mathbf{I} - \mathbf{H}_i (\mathbf{H}_i^T \mathbf{H}_i)^{-1} \mathbf{H}_i^T) \mathbf{x}. \end{aligned}$$

If the columns of $\mathbf{H}_i$ are orthogonal so that $\mathbf{H}_i^T \mathbf{H}_i = (N/2)\mathbf{I}_i$, show that we should choose the i that maximizes

$$\xi_i' = \sum_{k=1}^{i} \frac{[\hat{\boldsymbol{\theta}}_i]_k^2}{2\sigma^2/N} - i \ln \frac{N}{2\sigma^2}.$$

Explain the meaning of this result. For an application to Fourier analysis see [Kay-I 1993, Example 4.2].

Appendix 6A

Asymptotically Equivalent Tests - No Nuisance Parameters

We first consider the problem of composite hypothesis testing with no nuisance parameters. If $\boldsymbol{\theta}$ is a $r \times 1$ vector of unknown parameters, we wish to test

$$\begin{array}{l}\mathcal{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ \mathcal{H}_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0\end{array}$$

where under $\mathcal{H}_1$, $\boldsymbol{\theta}$ can take on values near $\boldsymbol{\theta}_0$. (Actually, we assume that under $\mathcal{H}_1$, $||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| = c/\sqrt{N}$ for some constant $c > 0$.) If $p(\mathbf{x}; \boldsymbol{\theta})$ is the PDF, which is parameterized by $\boldsymbol{\theta}$, then a GLRT decides $\mathcal{H}_1$ if

$$L_G(\mathbf{x}) = \frac{p(\mathbf{x}; \hat{\boldsymbol{\theta}}_1)}{p(\mathbf{x}; \boldsymbol{\theta}_0)} > \gamma$$

where $\hat{\boldsymbol{\theta}}_1$ is the MLE of $\boldsymbol{\theta}$ under $\mathcal{H}_1$. Note that in finding $\hat{\boldsymbol{\theta}}_1$, we do not have to impose the constraint $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, but can maximize $p(\mathbf{x}; \boldsymbol{\theta})$ over the entire parameter space. This is because the probability of the MLE yielding the particular value $\boldsymbol{\theta}_0$ is zero. Hence, $\hat{\boldsymbol{\theta}}_1$ is the unrestricted MLE, which we denote as $\hat{\boldsymbol{\theta}}$. Assuming the asymptotic PDF of the MLE is attained, we know that $\hat{\boldsymbol{\theta}}$ attains the Cramer-Rao lower bound. That is to say, it satisfies [Kay-I 1993, pg. 41]

$$\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{I}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \tag{6A.1}$$

where $\boldsymbol{\theta}$ is the true value. The matrix $\mathbf{I}(\boldsymbol{\theta})$ is the Fisher information matrix [Kay-I 1993, pg. 40]. Since the MLE is consistent, as $N \to \infty$, $\hat{\boldsymbol{\theta}} \to \boldsymbol{\theta}$, and we can use a first-order Taylor expansion

$$[\mathbf{I}(\boldsymbol{\theta})]_{ij} = \left[\mathbf{I}(\hat{\boldsymbol{\theta}})\right]_{ij} + \left.\frac{\partial [\mathbf{I}(\boldsymbol{\theta})]_{ij}}{\partial \boldsymbol{\theta}}\right|^T_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$$

so that from (6A.1)

$$\begin{aligned}\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_i} &= \sum_{j=1}^{r} [\mathbf{I}(\boldsymbol{\theta})]_{ij} (\hat{\theta}_j - \theta_j) \\ &= \sum_{j=1}^{r} [\mathbf{I}(\hat{\boldsymbol{\theta}})]_{ij} (\hat{\theta}_j - \theta_j) + \sum_{j=1}^{r} \left.\frac{\partial [\mathbf{I}(\boldsymbol{\theta})]_{ij}}{\partial \boldsymbol{\theta}}\right|^T_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\hat{\theta}_j - \theta_j).\end{aligned}$$

But the last term may be neglected as $N \to \infty$ since it is of higher order than the first. As a result, we have

$$\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{I}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}).$$

Now integrating with respect to $\boldsymbol{\theta}$ produces

$$\ln p(\mathbf{x}; \boldsymbol{\theta}) = -\frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \mathbf{I}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + c(\hat{\boldsymbol{\theta}})$$

or

$$p(\mathbf{x}; \boldsymbol{\theta}) = p(\mathbf{x}; \hat{\boldsymbol{\theta}}) \exp\left[-\frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \mathbf{I}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\right] \tag{6A.2}$$

since the constant of integration must be $c(\hat{\boldsymbol{\theta}}) = \ln p(\mathbf{x}; \hat{\boldsymbol{\theta}})$. This is the asymptotic form of the PDF. The GLRT becomes

$$L_G(\mathbf{x}) = \frac{p(\mathbf{x}; \hat{\boldsymbol{\theta}}_1)}{p(\mathbf{x}; \boldsymbol{\theta}_0)} = \frac{\max_{\boldsymbol{\theta}} p(\mathbf{x}; \boldsymbol{\theta})}{p(\mathbf{x}; \boldsymbol{\theta}_0)}.$$

But from (6A.2), $p(\mathbf{x}; \boldsymbol{\theta})$ is maximized over $\boldsymbol{\theta}$ for $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ since $\mathbf{I}(\hat{\boldsymbol{\theta}})$ is assumed to be positive definite. Recalling that $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_1$, we have

$$L_G(\mathbf{x}) = \frac{p(\mathbf{x}; \hat{\boldsymbol{\theta}}_1)}{p(\mathbf{x}; \hat{\boldsymbol{\theta}}_1) \exp\left[-\frac{1}{2}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0)^T \mathbf{I}(\hat{\boldsymbol{\theta}}_1)(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0)\right]}$$

or

$$2 \ln L_G(\mathbf{x}) = (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0)^T \mathbf{I}(\hat{\boldsymbol{\theta}}_1)(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0). \tag{6A.3}$$

This test statistic is the Wald test, which we have denoted by $T_W(\mathbf{x})$. It is seen to be asymptotically equivalent to $2 \ln L_G(\mathbf{x})$. To derive the Rao test we use (6A.1) with $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_1$

$$\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta} = \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

and assume that the true value of $\boldsymbol{\theta}$ is near $\boldsymbol{\theta}_0$. Then

$$\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0 = \mathbf{I}^{-1}(\boldsymbol{\theta}_0) \left.\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

Substituting this in (6A.3) produces

$$2 \ln L_G(\mathbf{x}) = \left.\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|^T_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \mathbf{I}^{-1}(\boldsymbol{\theta}_0)\mathbf{I}(\hat{\boldsymbol{\theta}}_1)\mathbf{I}^{-1}(\boldsymbol{\theta}_0) \left.\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

Again as $N \to \infty$, $\hat{\boldsymbol{\theta}}_1$ approaches the true value of $\boldsymbol{\theta}$, which is either $\boldsymbol{\theta}_0$ under $\mathcal{H}_0$ or *near* $\boldsymbol{\theta}_0$ under $\mathcal{H}_1$. Thus

$$\mathbf{I}^{-1}(\boldsymbol{\theta}_0)\mathbf{I}(\hat{\boldsymbol{\theta}}_1)\mathbf{I}^{-1}(\boldsymbol{\theta}_0) \to \mathbf{I}^{-1}(\boldsymbol{\theta}_0)$$

and

$$2 \ln L_G(\mathbf{x}) = \left.\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|^T_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \mathbf{I}^{-1}(\boldsymbol{\theta}_0) \left.\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

This is the Rao test statistic, which we have denoted by $T_{\mathrm{R}}(\mathbf{x})$.

Appendix 6B

Asymptotically Equivalent Tests - Nuisance Parameters

Now consider the composite hypothesis testing problem of the $p \times 1$ parameter vector $\boldsymbol{\theta} = [\boldsymbol{\theta}_r^T \, \boldsymbol{\theta}_s^T]^T$, where $\boldsymbol{\theta}_r$ is $r \times 1$, $\boldsymbol{\theta}_s$ is $s \times 1$, and $p = r + s$. The PDF is parameterized by $\boldsymbol{\theta}$ and is denoted by $p(\mathbf{x}; \boldsymbol{\theta}) = p(\mathbf{x}; \boldsymbol{\theta}_r, \boldsymbol{\theta}_s)$. We wish to test

$$\begin{array}{l} \mathcal{H}_0 : \boldsymbol{\theta}_r = \boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s \\ \mathcal{H}_1 : \boldsymbol{\theta}_r \neq \boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s \end{array}$$

where $\boldsymbol{\theta}_s$ is a vector of nuisance parameters. The GLRT decides $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x}; \hat{\boldsymbol{\theta}}_{r_1}, \hat{\boldsymbol{\theta}}_{s_1})}{p(\mathbf{x}; \boldsymbol{\theta}_{r_0}, \hat{\boldsymbol{\theta}}_{s_0})} > \gamma$$

where $\hat{\boldsymbol{\theta}}_{r_1}, \hat{\boldsymbol{\theta}}_{s_1}$ is the MLE of $\boldsymbol{\theta}_r, \boldsymbol{\theta}_s$ under $\mathcal{H}_1$, which is equivalent to the unrestricted MLE $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\theta}}_{s_0}$ is the MLE of $\boldsymbol{\theta}_s$ under $\mathcal{H}_0$ or the MLE of $\boldsymbol{\theta}_s$ when subject to the constraint $\boldsymbol{\theta}_r = \boldsymbol{\theta}_{r_0}$. Using (6A.2), which is

$$p(\mathbf{x}; \boldsymbol{\theta}) = p(\mathbf{x}; \hat{\boldsymbol{\theta}}) \exp\left[-\frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \mathbf{I}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\right]$$

we have

$$\begin{aligned} p(\mathbf{x}; \hat{\boldsymbol{\theta}}_{r_1}, \hat{\boldsymbol{\theta}}_{s_1}) &= \max_{\boldsymbol{\theta}_r, \boldsymbol{\theta}_s} p(\mathbf{x}; \boldsymbol{\theta}_r, \boldsymbol{\theta}_s) \\ &= \max_{\boldsymbol{\theta}} p(\mathbf{x}; \boldsymbol{\theta}) \\ &= p(\mathbf{x}; \hat{\boldsymbol{\theta}}) \end{aligned}$$

since $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ maximizes $p(\mathbf{x}; \boldsymbol{\theta})$ when there are no restrictions on $\boldsymbol{\theta}$. In other words, the MLE of $\boldsymbol{\theta}_r, \boldsymbol{\theta}_s$ under $\mathcal{H}_1$ is just $\hat{\boldsymbol{\theta}}$ so that $\hat{\boldsymbol{\theta}}_1 = [\hat{\boldsymbol{\theta}}_{r_1}^T \, \hat{\boldsymbol{\theta}}_{s_1}^T]^T = \hat{\boldsymbol{\theta}}$. Under $\mathcal{H}_0$, however, we must maximize $p(\mathbf{x}; \boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s)$ over $\boldsymbol{\theta}_s$, which yields the restricted MLE $\hat{\boldsymbol{\theta}}_{s_0}$. To do so we need to minimize over $\boldsymbol{\theta}_s$

$$J(\boldsymbol{\theta}_s) = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \mathbf{I}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$$

where $\boldsymbol{\theta} = [\boldsymbol{\theta}_{r_0}^T \, \boldsymbol{\theta}_s^T]^T$. We partition the Fisher information matrix as

$$\mathbf{I}(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} \mathbf{I}_{\theta_r\theta_r}(\hat{\boldsymbol{\theta}}) & \mathbf{I}_{\theta_r\theta_s}(\hat{\boldsymbol{\theta}}) \\ \mathbf{I}_{\theta_s\theta_r}(\hat{\boldsymbol{\theta}}) & \mathbf{I}_{\theta_s\theta_s}(\hat{\boldsymbol{\theta}}) \end{bmatrix} \tag{6B.1}$$

so that

$$\begin{aligned} J(\boldsymbol{\theta}_s) &= (\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0})^T \mathbf{I}_{\theta_r\theta_r}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0}) + (\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0})^T \mathbf{I}_{\theta_r\theta_s}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{s_1} - \boldsymbol{\theta}_s) \\ &\quad + (\hat{\boldsymbol{\theta}}_{s_1} - \boldsymbol{\theta}_s)^T \mathbf{I}_{\theta_s\theta_r}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0}) + (\hat{\boldsymbol{\theta}}_{s_1} - \boldsymbol{\theta}_s)^T \mathbf{I}_{\theta_s\theta_s}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{s_1} - \boldsymbol{\theta}_s). \end{aligned}$$

Taking the gradient with respect to $\boldsymbol{\theta}_s$ yields

$$\begin{aligned} \frac{\partial J(\boldsymbol{\theta}_s)}{\partial \boldsymbol{\theta}_s} &= -\mathbf{I}_{\theta_r\theta_s}^T(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0}) - \mathbf{I}_{\theta_s\theta_r}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0}) - 2\mathbf{I}_{\theta_s\theta_s}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{s_1} - \boldsymbol{\theta}_s) \\ &= -2\mathbf{I}_{\theta_s\theta_r}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0}) - 2\mathbf{I}_{\theta_s\theta_s}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{s_1} - \boldsymbol{\theta}_s). \end{aligned}$$

Setting this equal to zero and solving yields $\boldsymbol{\theta}_s = \hat{\boldsymbol{\theta}}_{s_0}$ where

$$\hat{\boldsymbol{\theta}}_{s_0} = \hat{\boldsymbol{\theta}}_{s_1} + \mathbf{I}_{\theta_s\theta_s}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{I}_{\theta_s\theta_r}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0}). \tag{6B.2}$$

Substituting into $J(\boldsymbol{\theta}_s)$ produces

$$\begin{aligned} J(\hat{\boldsymbol{\theta}}_{s_0}) &= (\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0})^T \mathbf{I}_{\theta_r\theta_r}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0}) \\ &\quad -2\,(\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0})^T \mathbf{I}_{\theta_r\theta_s}(\hat{\boldsymbol{\theta}})\mathbf{I}_{\theta_s\theta_s}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{I}_{\theta_s\theta_r}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0}) \\ &\quad + (\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0})^T \mathbf{I}_{\theta_s\theta_r}^T(\hat{\boldsymbol{\theta}})\mathbf{I}_{\theta_s\theta_s}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{I}_{\theta_s\theta_s}(\hat{\boldsymbol{\theta}})\mathbf{I}_{\theta_s\theta_s}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{I}_{\theta_s\theta_r}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0}) \\ &= (\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0})^T \left[\mathbf{I}_{\theta_r\theta_r}(\hat{\boldsymbol{\theta}}) - \mathbf{I}_{\theta_r\theta_s}(\hat{\boldsymbol{\theta}})\mathbf{I}_{\theta_s\theta_s}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{I}_{\theta_s\theta_r}(\hat{\boldsymbol{\theta}})\right](\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0}). \end{aligned} \tag{6B.3}$$

Hence

$$p(\mathbf{x}; \boldsymbol{\theta}_{r_0}, \hat{\boldsymbol{\theta}}_{s_0}) = p(\mathbf{x}; \hat{\boldsymbol{\theta}}_1)\exp\left[-\frac{1}{2}J(\hat{\boldsymbol{\theta}}_{s_0})\right]$$

and the GLRT is

$$\begin{aligned} L_G(\mathbf{x}) &= \frac{p(\mathbf{x}; \hat{\boldsymbol{\theta}}_{r_1}, \hat{\boldsymbol{\theta}}_{s_1})}{p(\mathbf{x}; \boldsymbol{\theta}_{r_0}, \hat{\boldsymbol{\theta}}_{s_0})} \\ &= \frac{p(\mathbf{x}; \hat{\boldsymbol{\theta}}_1)}{p(\mathbf{x}; \hat{\boldsymbol{\theta}}_1)\exp\left[-\frac{1}{2}J(\hat{\boldsymbol{\theta}}_{s_0})\right]} \end{aligned}$$

or

$$2\ln L_G(\mathbf{x}) = J(\hat{\boldsymbol{\theta}}_{s_0})$$

which from (6B.3) is

$$2\ln L_G(\mathbf{x}) = (\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0})^T \left(\left[\mathbf{I}^{-1}(\hat{\boldsymbol{\theta}}_1)\right]_{\theta_r\theta_r}\right)^{-1}(\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0}) \tag{6B.4}$$

where $[\mathbf{I}^{-1}(\hat{\boldsymbol{\theta}}_1)]_{\theta_r\theta_r}$ is the $r \times r$ upper-left partition of $\mathbf{I}^{-1}(\hat{\boldsymbol{\theta}}_1)$. This follows from the formula for the inverse of a partitioned matrix since if

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} = \begin{bmatrix} r \times r & r \times s \\ s \times r & s \times s \end{bmatrix}$$

then

$$\mathbf{A}^{-1} = \begin{bmatrix} (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} & -(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1} \end{bmatrix}.$$

This is the Wald test statistic or $T_W(\mathbf{x})$. To derive the Rao test we use (6A.1), which in partitioned form is

$$\begin{bmatrix} \dfrac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r} \\ \dfrac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_s} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{\theta_r\theta_r}(\boldsymbol{\theta}) & \mathbf{I}_{\theta_r\theta_s}(\boldsymbol{\theta}) \\ \mathbf{I}_{\theta_s\theta_r}(\boldsymbol{\theta}) & \mathbf{I}_{\theta_s\theta_s}(\boldsymbol{\theta}) \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\theta}}_r - \boldsymbol{\theta}_r \\ \hat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_s \end{bmatrix}$$

where $\hat{\boldsymbol{\theta}}_r, \hat{\boldsymbol{\theta}}_s$ is the unrestricted MLE and $\boldsymbol{\theta}_r, \boldsymbol{\theta}_s$ is the true value. The upper partition yields

$$\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r} = \mathbf{I}_{\theta_r\theta_r}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_r - \boldsymbol{\theta}_r) + \mathbf{I}_{\theta_r\theta_s}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_s).$$

Evaluating this at

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_r \\ \boldsymbol{\theta}_s \end{bmatrix} = \begin{bmatrix} \boldsymbol{\theta}_{r_0} \\ \hat{\boldsymbol{\theta}}_{s_0} \end{bmatrix} = \tilde{\boldsymbol{\theta}}$$

which is close to the true value of $\boldsymbol{\theta}$ and noting that $\hat{\boldsymbol{\theta}}_r = \hat{\boldsymbol{\theta}}_{r_1}, \hat{\boldsymbol{\theta}}_s = \hat{\boldsymbol{\theta}}_{s_1}$, we have

$$\left.\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} = \mathbf{I}_{\theta_r\theta_r}(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0}) + \mathbf{I}_{\theta_r\theta_s}(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{s_1} - \hat{\boldsymbol{\theta}}_{s_0}).$$

But from (6B.2)

$$\hat{\boldsymbol{\theta}}_{s_1} - \hat{\boldsymbol{\theta}}_{s_0} = -\mathbf{I}^{-1}_{\theta_s\theta_s}(\hat{\boldsymbol{\theta}})\mathbf{I}_{\theta_s\theta_r}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0})$$

so that

$$\left.\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} = \left[\mathbf{I}_{\theta_r\theta_r}(\tilde{\boldsymbol{\theta}}) - \mathbf{I}_{\theta_r\theta_s}(\tilde{\boldsymbol{\theta}})\mathbf{I}^{-1}_{\theta_s\theta_s}(\hat{\boldsymbol{\theta}})\mathbf{I}_{\theta_s\theta_r}(\hat{\boldsymbol{\theta}})\right](\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0}). \tag{6B.5}$$

The difference between evaluation of $\mathbf{I}(\boldsymbol{\theta})$ at $\tilde{\boldsymbol{\theta}}$ or $\hat{\boldsymbol{\theta}}$ is negligible as $N \to \infty$ so that from (6B.4)

$$2\ln L_G(\mathbf{x}) = (\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0})^T \left(\left[\mathbf{I}^{-1}(\tilde{\boldsymbol{\theta}})\right]_{\theta_r\theta_r}\right)^{-1} (\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0}) \tag{6B.6}$$

and from (6B.5)

$$\begin{aligned}\left.\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} &= \left[\mathbf{I}_{\theta_r\theta_r}(\tilde{\boldsymbol{\theta}}) - \mathbf{I}_{\theta_r\theta_s}(\tilde{\boldsymbol{\theta}})\mathbf{I}^{-1}_{\theta_s\theta_s}(\tilde{\boldsymbol{\theta}})\mathbf{I}_{\theta_s\theta_r}(\tilde{\boldsymbol{\theta}})\right](\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0}) \\ &= \left(\left[\mathbf{I}^{-1}(\tilde{\boldsymbol{\theta}})\right]_{\theta_r\theta_r}\right)^{-1}(\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0})\end{aligned}$$

or

$$\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0} = \left[\mathbf{I}^{-1}(\tilde{\boldsymbol{\theta}})\right]_{\theta_r\theta_r} \left.\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}$$

and thus from (6B.6)

$$2\ln L_G(\mathbf{x}) = \left.\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r}\right|^T_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} \left[\mathbf{I}^{-1}(\tilde{\boldsymbol{\theta}})\right]_{\theta_r\theta_r} \left.\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}$$

which is the Rao test statistic $T_{\mathrm{R}}(\mathbf{x})$.

Appendix 6C

Asymptotic PDF of GLRT

We first consider the case of no nuisance parameters so that $\boldsymbol{\theta}$ is $r \times 1$. In determining the asymptotic PDF of $2\ln L_G(\mathbf{x})$, we begin with the asymptotically equivalent Wald test statistic (see (6A.3))

$$2\ln L_G(\mathbf{x}) = (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0)^T \mathbf{I}(\hat{\boldsymbol{\theta}}_1)(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0).$$

Since $\hat{\boldsymbol{\theta}}_1$ is the unrestricted MLE of $\boldsymbol{\theta}$ or $\hat{\boldsymbol{\theta}}$, we have as $N \to \infty$

$$\hat{\boldsymbol{\theta}}_1 \sim \begin{cases} \mathcal{N}(\boldsymbol{\theta}_0, \mathbf{I}^{-1}(\boldsymbol{\theta}_0)) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(\boldsymbol{\theta}_1, \mathbf{I}^{-1}(\boldsymbol{\theta}_1)) & \text{under } \mathcal{H}_1. \end{cases}$$

But as $N \to \infty$

$$\mathbf{I}(\hat{\boldsymbol{\theta}}_1)(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0) = \mathbf{I}(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0) = \mathbf{I}(\boldsymbol{\theta}_1)(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0) \tag{6C.1}$$

so that under $\mathcal{H}_0$ as $N \to \infty$

$$2\ln L_G(\mathbf{x}) = (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0)^T \mathbf{I}(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0) \sim \chi^2_r$$

and under $\mathcal{H}_1$

$$2\ln L_G(\mathbf{x}) = (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0)^T \mathbf{I}(\boldsymbol{\theta}_1)(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0) \sim \chi'^2_r(\lambda)$$

where

$$\lambda = (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^T \mathbf{I}(\boldsymbol{\theta}_1)(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)$$

or equivalently as $N \to \infty$

$$\lambda = (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^T \mathbf{I}(\boldsymbol{\theta}_0)(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0).$$

We have used the result that if $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$, then $\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x} \sim \chi_r'^2(\lambda)$, where $\lambda = \boldsymbol{\mu}^T\mathbf{C}^{-1}\boldsymbol{\mu}$ (see Section 2.3). Next for the nuisance parameter case we have from (6B.4)

$$2\ln L_G(\mathbf{x}) = (\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0})^T \left(\left[\mathbf{I}^{-1}(\hat{\boldsymbol{\theta}}_1)\right]_{\theta_r\theta_r}\right)^{-1} (\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0})$$

where $\hat{\boldsymbol{\theta}}_1 = [\hat{\boldsymbol{\theta}}_{r_1}^T\ \hat{\boldsymbol{\theta}}_{s_1}^T]^T$. Since $\hat{\boldsymbol{\theta}}_{r_1}$ is the unrestricted MLE of $\boldsymbol{\theta}_r$, we have as $N \to \infty$

$$\hat{\boldsymbol{\theta}}_{r_1} \sim \begin{cases} \mathcal{N}(\boldsymbol{\theta}_{r_0}, [\mathbf{I}^{-1}(\boldsymbol{\theta}_0)]_{\theta_r\theta_r}) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(\boldsymbol{\theta}_{r_1}, [\mathbf{I}^{-1}(\boldsymbol{\theta}_1)]_{\theta_r\theta_r}) & \text{under } \mathcal{H}_1 \end{cases} \tag{6C.2}$$

where $\boldsymbol{\theta}_0 = [\boldsymbol{\theta}_{r_0}^T\ \boldsymbol{\theta}_s^T]^T$ and $\boldsymbol{\theta}_1 = [\boldsymbol{\theta}_{r_1}^T\ \boldsymbol{\theta}_s^T]^T$. But using the same argument as in (6C.1) we have under $\mathcal{H}_0$

$$2\ln L_G(\mathbf{x}) \sim \chi_r^2$$

and under $\mathcal{H}_1$

$$2\ln L_G(\mathbf{x}) \sim \chi_r'^2(\lambda)$$

where

$$\lambda = (\boldsymbol{\theta}_{r_1} - \boldsymbol{\theta}_{r_0})^T \left(\left[\mathbf{I}^{-1}(\boldsymbol{\theta}_{r_1}, \boldsymbol{\theta}_s)\right]_{\theta_r\theta_r}\right)^{-1} (\boldsymbol{\theta}_{r_1} - \boldsymbol{\theta}_{r_0})$$

which as $N \to \infty$ is equivalent to

$$\lambda = (\boldsymbol{\theta}_{r_1} - \boldsymbol{\theta}_{r_0})^T \left(\left[\mathbf{I}^{-1}(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s)\right]_{\theta_r\theta_r}\right)^{-1} (\boldsymbol{\theta}_{r_1} - \boldsymbol{\theta}_{r_0}).$$

Appendix 6D

Asymptotic Detection Performance of LMP Test

The LMP test statistic was defined to be

$$T_{\mathrm{LMP}}(\mathbf{x}) = \frac{\left.\dfrac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right|_{\theta=\theta_0}}{\sqrt{I(\theta_0)}}.$$

Under $\mathcal{H}_0$ for which $\theta = \theta_0$ (recall the regularity conditions [Kay-I 1993, pg. 67])

$$E\left(\left.\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right|_{\theta=\theta_0}\right) = 0$$

and (see [Kay-I 1993, pg. 67])

$$E\left[\left(\left.\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right|_{\theta=\theta_0}\right)^2\right] = I(\theta_0)$$

which is the definition of the Fisher information. Furthermore, if the $x[n]$'s are IID so that

$$\ln p(\mathbf{x};\theta) = \sum_{n=0}^{N-1} \ln p(x[n];\theta)$$

then

$$T_{\mathrm{LMP}}(\mathbf{x}) = \frac{1}{\sqrt{I(\theta_0)}} \sum_{n=0}^{N-1} \left.\frac{\partial \ln p(x[n];\theta)}{\partial\theta}\right|_{\theta=\theta_0}$$

which by the central limit theorem becomes Gaussian. Hence, under $\mathcal{H}_0$

$$T_{\mathrm{LMP}}(\mathbf{x}) \sim \mathcal{N}(0,1).$$

Under $\mathcal{H}_1$ and for θ_0 near θ_1 a first-order Taylor expansion yields

$$\left.\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right|_{\theta=\theta_0} = \left.\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right|_{\theta=\theta_1} + \left.\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2}\right|_{\theta=\theta_1}(\theta_0-\theta_1)$$

so that due to the regularity conditions

$$\begin{aligned} E\left(\left.\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right|_{\theta=\theta_0}\right) &= 0 + E\left(\left.\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2}\right|_{\theta=\theta_1}\right)(\theta_0-\theta_1) \\ &= -I(\theta_1)(\theta_0-\theta_1) \\ &\to I(\theta_0)(\theta_1-\theta_0) \end{aligned}$$

as $N\to\infty$. Also, under $\mathcal{H}_1$

$$\begin{aligned} &\mathrm{var}\left(\left.\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right|_{\theta=\theta_0}\right) \\ &= E\left\{\left[\left.\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right|_{\theta=\theta_0} - E\left(\left.\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right|_{\theta=\theta_0}\right)\right]^2\right\} \\ &= E\left\{\left[\left.\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right|_{\theta=\theta_1} + \left(\left.\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2}\right|_{\theta=\theta_1} + I(\theta_1)\right)(\theta_0-\theta_1)\right]^2\right\}. \end{aligned}$$

Ignoring the $(\theta_0-\theta_1)$ and $(\theta_0-\theta_1)^2$ terms we have

$$\mathrm{var}\left(\left.\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right|_{\theta=\theta_0}\right) = E\left[\left(\left.\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right|_{\theta=\theta_1}\right)^2\right] = I(\theta_1)$$

or equivalently as $N\to\infty$

$$\mathrm{var}\left(\left.\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right|_{\theta=\theta_0}\right) = I(\theta_0).$$

Thus

$$\left.\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right|_{\theta=\theta_0} \sim \mathcal{N}(I(\theta_0)(\theta_1-\theta_0), I(\theta_0))$$

and under $\mathcal{H}_1$

$$T_{\mathrm{LMP}}(\mathbf{x}) \sim \mathcal{N}(\sqrt{I(\theta_0)}(\theta_1-\theta_0), 1)$$

since as before the central limit theorem applies under the IID assumption.

Appendix 6E

Alternate Derivation of Locally Most Powerful Test

Let the region in R^N for which we choose $\mathcal{H}_1$ be denoted by R_1. Then

$$\begin{aligned} P_{FA} &= \int_{R_1} p(\mathbf{x};\mathcal{H}_0)d\mathbf{x} \\ P_D &= \int_{R_1} p(\mathbf{x};\mathcal{H}_1)d\mathbf{x}. \end{aligned}$$

But the PDF is parameterized by θ so that we write $p(\mathbf{x};\mathcal{H}_0)$ as $p(\mathbf{x};\theta_0)$ and $p(\mathbf{x};\mathcal{H}_1)$ as $p(\mathbf{x};\theta)$, resulting in

$$\begin{aligned} P_{FA} &= \int_{R_1} p(\mathbf{x};\theta_0)d\mathbf{x} \\ P_D(\theta) &= \int_{R_1} p(\mathbf{x};\theta)d\mathbf{x}. \end{aligned}$$

The probability of detection is explicitly shown to depend on θ. For $\theta>\theta_0$ and $\theta-\theta_0$ small we can expand $P_D(\theta)$ in a first-order Taylor expansion about $\theta=\theta_0$ to yield

$$\begin{aligned} P_D(\theta) &= P_D(\theta_0) + \left.\frac{dP_D(\theta)}{d\theta}\right|_{\theta=\theta_0}(\theta-\theta_0) \\ &= P_{FA} + \left.\frac{dP_D(\theta)}{d\theta}\right|_{\theta=\theta_0}(\theta-\theta_0). \end{aligned}$$

Since P_{FA} is fixed and only the derivative of $P_D(\theta)$ depends on R_1, we maximize $P_D(\theta)$ for any θ by maximizing the slope of $P_D(\theta)$ at $\theta=\theta_0$. Now

$$\frac{dP_D(\theta)}{d\theta} = \frac{d}{d\theta}\int_{R_1} p(\mathbf{x};\theta)d\mathbf{x}$$

$$
\begin{aligned}
&= \int_{R_1} \frac{dp(\mathbf{x};\theta)}{d\theta} d\mathbf{x} \\
&= \int_{R_1} \frac{d\ln p(\mathbf{x};\theta)}{d\theta} p(\mathbf{x};\theta) d\mathbf{x}.
\end{aligned}
$$

To maximize the slope subject to the P_{FA} constraint we use Lagrangian multipliers to form

$$F = \left.\frac{dP_D(\theta)}{d\theta}\right|_{\theta=\theta_0} + \lambda(P_{FA} - \alpha).$$

But

$$
\begin{aligned}
F &= \int_{R_1} \left[\left.\frac{d\ln p(\mathbf{x};\theta)}{d\theta}\right|_{\theta=\theta_0} p(\mathbf{x};\theta_0) + \lambda p(\mathbf{x};\theta_0) \right] d\mathbf{x} - \lambda\alpha \\
&= \int_{R_1} \left(\left.\frac{d\ln p(\mathbf{x};\theta)}{d\theta}\right|_{\theta=\theta_0} + \lambda \right) p(\mathbf{x};\theta_0) d\mathbf{x} - \lambda\alpha.
\end{aligned}
$$

To maximize F we should include $\mathbf{x}$ in R_1 if

$$\left.\frac{d\ln p(\mathbf{x};\theta)}{d\theta}\right|_{\theta=\theta_0} + \lambda > 0$$

or

$$\left.\frac{d\ln p(\mathbf{x};\theta)}{d\theta}\right|_{\theta=\theta_0} > -\lambda = \gamma.$$

See also Appendix 3A.

Appendix 6F

Derivation of Generalized ML Rule

For large data records as shown in Appendix 6A (see (6A.2))

$$p(\mathbf{x};\boldsymbol{\theta}_i) = p(\mathbf{x};\hat{\boldsymbol{\theta}}_i) \exp\left[-\frac{1}{2}(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)^T \mathbf{I}(\hat{\boldsymbol{\theta}}_i)(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)\right] \tag{6F.1}$$

where $\hat{\boldsymbol{\theta}}_i$ is the unrestricted MLE of $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_i$ has dimension $n_i \times 1$. If we now employ the Bayesian philosophy and consider $\boldsymbol{\theta}_i$ as a random variable (as well as the hypothesis chosen), the conditional data PDF becomes

$$p(\mathbf{x}|\boldsymbol{\theta}_i,\mathcal{H}_i) = p(\mathbf{x}|\hat{\boldsymbol{\theta}}_i,\mathcal{H}_i) \exp\left[-\frac{1}{2}(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)^T \mathbf{I}(\hat{\boldsymbol{\theta}}_i)(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)\right]. \tag{6F.2}$$

Note that we must include the $\mathcal{H}_i$ descriptor in the PDF since the PDF is different under each hypothesis. By assigning the prior PDF $p(\boldsymbol{\theta}_i)$ to the unknown parameter vector, we have

$$p(\mathbf{x}|\mathcal{H}_i) = \int p(\mathbf{x}|\boldsymbol{\theta}_i,\mathcal{H}_i) p(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i. \tag{6F.3}$$

In order not to bias our results by the prior, we choose a "wide" prior or one that is flat over the values of $p(\mathbf{x}|\boldsymbol{\theta}_i,\mathcal{H}_i)$ for which it is significantly positive. Such a prior is the Gaussian prior

$$p(\boldsymbol{\theta}_i) = \frac{1}{(2\pi)^{\frac{n_i}{2}} \det^{\frac{1}{2}}(\mathbf{C}_{\theta_i})} \exp\left[-\frac{1}{2}(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)^T \mathbf{C}_{\theta_i}^{-1}(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)\right] \tag{6F.4}$$

which is centered about $\hat{\boldsymbol{\theta}}_i$ and is wider than $p(\mathbf{x}|\boldsymbol{\theta}_i,\mathcal{H}_i)$ if we choose $\mathbf{C}_{\theta_i} \gg \mathbf{I}^{-1}(\hat{\boldsymbol{\theta}}_i)$. Furthermore, to avoid imparting more information from one prior to another (since the dimensionality of $\boldsymbol{\theta}_i$ depends on i), we choose $\mathbf{C}_{\theta_i}$ to ensure that the information is the same for all i. The rationale is that the information in the prior can be quantified by the negative of the entropy [Zacks 1981] which is

$$E(\ln p(\boldsymbol{\theta}_i)) = \int p(\boldsymbol{\theta}_i)\ln p(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i.$$

But for the Gaussian prior chosen

$$\begin{aligned} E(\ln p(\boldsymbol{\theta}_i)) &= -\frac{n_i}{2}\ln 2\pi - \frac{1}{2}\ln\det(\mathbf{C}_{\theta_i}) - \frac{1}{2}E\left[(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)^T\mathbf{C}_{\theta_i}^{-1}(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)\right] \\ &= -\frac{n_i}{2}\ln 2\pi - \frac{1}{2}\ln\det(\mathbf{C}_{\theta_i}) - \frac{1}{2}n_i \\ &= -\frac{n_i}{2}\ln 2\pi e - \frac{1}{2}\ln\det(\mathbf{C}_{\theta_i}) = c \end{aligned} \tag{6F.5}$$

for c a constant. We have used the result that if $\mathbf{x}$ is $n_i \times 1$ and $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$, then

$$\begin{aligned} E[(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})] &= E[\mathrm{tr}((\mathbf{x}-\boldsymbol{\mu})^T\mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu}))] \\ &= E[\mathrm{tr}((\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{C}^{-1})] \\ &= \mathrm{tr}[E((\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T)\mathbf{C}^{-1}] \\ &= \mathrm{tr}(\mathbf{I}) = n_i. \end{aligned}$$

Hence, we should choose $\mathbf{C}_{\theta_i}$ to satisfy the constraint of (6F.5). Substituting (6F.2) and (6F.4) into (6F.3) produces

$$\begin{aligned} p(\mathbf{x}|\mathcal{H}_i) &= \frac{p(\mathbf{x}|\hat{\boldsymbol{\theta}}_i, \mathcal{H}_i)}{(2\pi)^{\frac{n_i}{2}}\det^{\frac{1}{2}}(\mathbf{C}_{\theta_i})}\int \exp\left[-\frac{1}{2}(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)^T\left(\mathbf{I}(\hat{\boldsymbol{\theta}}_i) + \mathbf{C}_{\theta_i}^{-1}\right)(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)\right]d\boldsymbol{\theta}_i \\ &= \frac{p(\mathbf{x}|\hat{\boldsymbol{\theta}}_i, \mathcal{H}_i)(2\pi)^{\frac{n_i}{2}}\det^{\frac{1}{2}}\left[\left(\mathbf{I}(\hat{\boldsymbol{\theta}}_i) + \mathbf{C}_{\theta_i}^{-1}\right)^{-1}\right]}{(2\pi)^{\frac{n_i}{2}}\det^{\frac{1}{2}}(\mathbf{C}_{\theta_i})} \\ &= \frac{p(\mathbf{x}|\hat{\boldsymbol{\theta}}_i, \mathcal{H}_i)}{\det^{\frac{1}{2}}\left[\mathbf{C}_{\theta_i}\left(\mathbf{I}(\hat{\boldsymbol{\theta}}_i) + \mathbf{C}_{\theta_i}^{-1}\right)\right]} \\ &\approx \frac{p(\mathbf{x}|\hat{\boldsymbol{\theta}}_i, \mathcal{H}_i)}{\det^{\frac{1}{2}}(\mathbf{C}_{\theta_i})\det^{\frac{1}{2}}\left(\mathbf{I}(\hat{\boldsymbol{\theta}}_i)\right)}. \end{aligned}$$

Taking logarithms yields

$$\ln p(\mathbf{x}|\mathcal{H}_i) = \ln p(\mathbf{x}|\hat{\boldsymbol{\theta}}_i, \mathcal{H}_i) - \frac{1}{2}\ln\det(\mathbf{I}(\hat{\boldsymbol{\theta}}_i)) - \frac{1}{2}\ln\det(\mathbf{C}_{\theta_i})$$

which from (6F.5) becomes

$$\ln p(\mathbf{x}|\mathcal{H}_i) = \ln p(\mathbf{x}|\hat{\boldsymbol{\theta}}_i, \mathcal{H}_i) - \frac{1}{2}\ln\det(\mathbf{I}(\hat{\boldsymbol{\theta}}_i)) + c + \frac{n_i}{2}\ln 2\pi e.$$

Ignoring the constant c and also $(n_i/2)\ln 2\pi e$, which is small relative to the other terms which grow with N, we have

$$\xi_i = \ln p(\mathbf{x}|\mathcal{H}_i) = \ln p(\mathbf{x}|\hat{\boldsymbol{\theta}}_i, \mathcal{H}_i) - \frac{1}{2}\ln\det(\mathbf{I}(\hat{\boldsymbol{\theta}}_i)).$$

Even though the generalized ML rule is derived under the Bayesian assumption for $\boldsymbol{\theta}_i$ it is applied in the classical case so that

$$\xi_i = \ln p(\mathbf{x};\hat{\boldsymbol{\theta}}_i|\mathcal{H}_i) - \frac{1}{2}\ln\det(\mathbf{I}(\hat{\boldsymbol{\theta}}_i))$$

where $\hat{\boldsymbol{\theta}}_i$ is the MLE or the value that maximizes $p(\mathbf{x};\boldsymbol{\theta}_i|\mathcal{H}_i)$.

Chapter 7

Deterministic Signals with Unknown Parameters

7.1 Introduction

A problem of great practical importance is the detection of signals that are not completely known. The lack of signal knowledge may be due to uncertain propagation effects as in mobile radio or to unknown signal generation mechanisms responsible for stock market trends, for example. Being able to design effective detectors requires the advanced theory of composite hypothesis testing that was described in Chapter 6. In this and succeeding chapters we apply that theory to many detection problems of practical interest. The present chapter examines the detection of *deterministic* signals with unknown parameters while the next chapter studies detection of *random* signals with unknown parameters.

7.2 Summary

The lack of signal knowledge results in a degradation of detection performance when compared with the optimal matched filter. An illustration of this effect is the energy detector described in Section 7.3. The detector for the unknown amplitude signal is discussed in Section 7.4 with the GLRT being given by (7.14) and its performance by (7.16). A Bayesian detector yields (7.18) as the test statistic and requires knowledge of a prior Gaussian PDF. In Section 7.5 the detection of a signal whose arrival time is unknown leads to the GLRT of (7.21) or equivalently (7.22). The important case of sinusoidal detection is explored in Section 7.6. For an unknown amplitude and phase the GLRT reduces to a computation of the periodogram at the known frequency or (7.25). The detection performance is given by (7.28). If the frequency is also unknown, the GLRT "peak picks" the periodogram or (7.30). Its performance is given by (7.31). Finally, for an unknown arrival time as well, the GLRT "peak

picks" the spectrogram which is given by (7.33). Its performance is analyzed in Section 7.8. In Section 7.7 the classical linear model is discussed. The general results are summarized in Theorem 7.1. The GLRT is given as the simple closed-form expression (7.35), as is its detection performance. The theorem has wide applicability to many practical detection problems. In Section 7.8 some signal processing applications are described. The performance of the usual narrowband detector for an active sonar or radar is determined to be (7.38). Also, the GLRT for detection in the presence of a sinusoidal interference is given by (7.39).

7.3 Signal Modeling and Detection Performance

The general model we shall employ for a deterministic signal with uncertainty is one whose form is completely known except for a few parameters. For example, if $s[n] = A\cos(2\pi f_0 n + \phi)$, then we may not know the values of the set of parameters $\{A, f_0, \phi\}$ or those of some subset of these parameters. The two principal approaches to designing a good detector for this composite hypothesis testing problem were described in Chapter 6. They are the generalized likelihood ratio test (GLRT) if the unknown parameters are modeled as deterministic and the Bayesian approach if they are modeled as realizations of random variables. These two methods employ fundamentally different philosophies, and so a direct comparison is not possible.

In the classical case of unknown deterministic signal parameters, an optimal detector, i.e., a uniformly most powerful (UMP) test (one that produces the highest P_D for all values of the unknown parameters and for a given P_{FA}) will usually not exist. The GLRT, a suboptimal detector, will however, usually produce good detection performance. For large data records it can be shown to be UMP within the class of *invariant* detectors [Lehmann 1959]. The detection loss incurred by using a GLRT can be bounded by comparing its performance to that of the clairvoyant detector as discussed in Chapter 6. The latter assumes perfect knowledge of the unknown parameters in implementing the NP detector. Therefore, the performance of the unrealizable but optimal NP detector will be an upper bound on the performance of any realizable detector. Usually the GLRT loss will be quite small. An example was given in Chapter 6 for a DC level with unknown amplitude in WGN.

If the Bayesian philosophy is adopted, then the resulting detector can be said to be optimal in the NP sense. In effect, the troublesome unknown parameters are "integrated out," leaving a simple hypothesis testing problem. The latter is then easily solved by application of the NP theorem. The difficulty, of course, is in specifying the prior PDF and in practice, of actually carrying out the integration. Also, the resulting detector cannot be claimed to be optimal if the unknown parameters are indeed deterministic or if they are random but with a different prior PDF than that assumed.

In this chapter we will have occasion to use both signal models. We choose the one that most nearly models the problem at hand and that is analytically tractable. From the standpoint of practicality the GLRT is usually easier to implement as it requires a maximization rather than an integration. This is because numerical maximization, as a last resort, is generally simpler.

Before discussing specific detectors it is instructive to study the importance of signal information. If a completely known signal is to be detected in WGN with variance σ^2, then a matched filter is optimal and its detection performance is given by (4.14) as

$$P_D = Q\left(Q^{-1}(P_{FA}) - \sqrt{d^2}\right) \tag{7.1}$$

where $d^2 = \mathcal{E}/\sigma^2$ is the deflection coefficient and $\mathcal{E}$ is the signal energy. When unknown signal parameters are present we can expect the performance of any detector to be degraded. Clearly, (7.1) provides an upper bound on detection performance (since it is the clairvoyant detector). The importance of signal information can be illustrated by comparing the matched filter to a detector which assumes no knowledge of the signal. Consider the detection problem

$$\begin{array}{ll}\mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = s[n] + w[n] & n = 0, 1, \ldots, N-1\end{array}$$

where $s[n]$ is deterministic and *completely unknown* and $w[n]$ is WGN with variance σ^2. A GLRT would decide $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x}; \hat{s}[0], \ldots, \hat{s}[N-1], \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > \gamma \tag{7.2}$$

where $\hat{s}[n]$ for $n = 0, 1, \ldots, N-1$ is the MLE under $\mathcal{H}_1$ (see also Problem 7.6 for a Bayesian approach). To determine the MLE we maximize the likelihood function

$$p(\mathbf{x}; s[0], \ldots, s[N-1], \mathcal{H}_1) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - s[n])^2\right]$$

over the signal samples. Clearly, the MLE is $\hat{s}[n] = x[n]$. Thus, from (7.2)

$$\frac{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}}{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left(-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n]\right)} > \gamma.$$

Taking logarithms produces

$$\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n] > \ln \gamma$$

or we decide $\mathcal{H}_1$ if

$$\sum_{n=0}^{N-1} x^2[n] > \gamma'. \tag{7.3}$$

This is just an energy detector, which also resulted from the modeling of the signal as a white Gaussian random process (see Example 5.1). As in Example 5.1, it has the form of an estimator-correlator since we decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} x^2[n] = \sum_{n=0}^{N-1} x[n]\hat{s}[n] > \gamma' \tag{7.4}$$

where $\hat{s}[n] = x[n]$. The detection performance will differ from that in Example 5.1 because of the different modeling assumptions (see also Problem 7.7). It is shown in Appendix 7A that for large N the detection performance of the energy detector (ED) or of (7.4) is given by (7.1) but with the deflection coefficient

$$d^2_{\mathrm{ED}} = \frac{\left(\frac{\mathcal{E}}{\sigma^2}\right)^2}{2N}. \tag{7.5}$$

Large N means that $\mathcal{E}/\sigma^2 \ll N$ or $(\mathcal{E}/N)/\sigma^2 \ll 1$, where the latter can be viewed as the input SNR. Recall that for a matched filter (MF) (the clairvoyant detector) the performance is also given by (7.1) but with

$$d^2_{\mathrm{MF}} = \frac{\mathcal{E}}{\sigma^2}. \tag{7.6}$$

The loss in performance can be quantified by comparing the deflection coefficients since P_D is monotonically increasing with d^2. This leads to a loss of

$$10\log_{10}\frac{d^2_{\mathrm{MF}}}{d^2_{\mathrm{ED}}} = 10\log_{10}\frac{2N}{\frac{\mathcal{E}}{\sigma^2}} \quad \mathrm{dB}.$$

As an example, for a DC level in WGN or $s[n] = A$ the loss is

$$10\log_{10}\frac{2\sigma^2}{A^2} = 3 - 10\log_{10}\frac{A^2}{\sigma^2} \quad \mathrm{dB} \tag{7.7}$$

since $\mathcal{E} = NA^2$. It is valid if $\mathcal{E}/\sigma^2 = NA^2/\sigma^2 \ll N$ or if the input SNR satisfies $A^2/\sigma^2 \ll 1$. As N becomes larger, the loss between the matched filter and energy detector increases. This follows from (7.7) since as N increases, the input SNR decreases for the same (P_{FA}, P_D) operating point. In essence, the matched filter *coherently* combines the data ($T_{\mathrm{MF}}(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]$) while the energy detector *incoherently* combines the data ($T_{\mathrm{ED}}(\mathbf{x}) = \sum_{n=0}^{N-1} x^2[n]$). A coherent combination reduces the noise more rapidly (see Problem 7.8).

To illustrate this effect we can determine the necessary input SNR required for a given d^2 and hence a given detection performance. It is easily shown that for the DC level in WGN the required input SNR or $\eta = A^2/\sigma^2$ for a given deflection coefficient d^2 and hence a given P_D is from (7.5) and (7.6)

$$\begin{aligned} 10\log_{10}\eta_{\mathrm{MF}} &= 10\log_{10}d^2 - 10\log_{10}N \quad \mathrm{dB} \\ 10\log_{10}\eta_{\mathrm{ED}} &= 5\log_{10}d^2 + 1.5 - 5\log_{10}N \quad \mathrm{dB} \end{aligned}$$

an example of which is shown in Figure 7.1. For instance, the loss for $N = 1000$ is 11.5 dB. Note that the required input SNR decreases by $10\log_{10}N$ for the matched filter but only by $5\log_{10}N$ for the energy detector. The loss is due to the incoherent averaging employed by the energy detector. We may interpret these results by saying that the processing gain of the matched filter is $10\log_{10}N$ while that of the energy detector is only $5\log_{10}N$. Therefore, a substantial penalty is paid for the lack of signal knowledge. For signals which are partially known, i.e., signals with only a few unknown parameters we can expect a detection performance somewhere between the matched filter (clairvoyant detector) and the energy detector. In the next sections we examine some common detection problems for signals with unknown parameters.

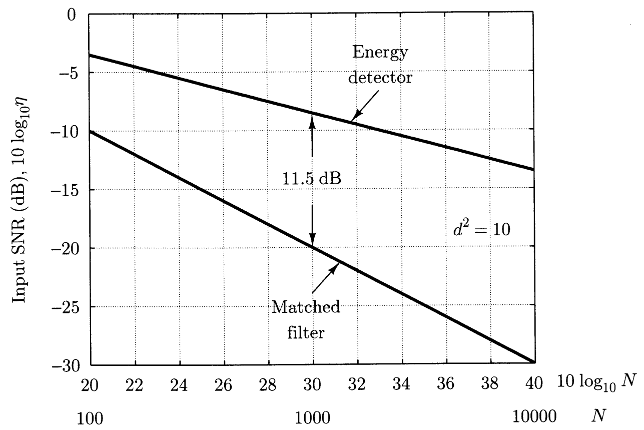


Figure 7.1. Required input SNR for given detection performance.

7.4 Unknown Amplitude

We now consider the problem of detecting a deterministic signal known except for amplitude in WGN. Specifically,

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = w[n] & n = 0,1,\ldots,N-1 \\ \mathcal{H}_1 : x[n] = As[n] + w[n] & n = 0,1,\ldots,N-1 \end{array}$$

where $s[n]$ is known, the amplitude A is *unknown* and $w[n]$ is WGN with variance σ^2. This problem is exactly of the form described in Chapter 6 if $s[n] = 1$ (DC level of unknown amplitude in WGN). There it was shown that if the sign of A is known, then a UMP test exists. If the sign is not known, then a UMP test does not exist and we must employ either a GLRT or a Bayesian approach. To determine if a UMP test exists for this case we assume that A is known and then construct the NP test. If the test statistic and its threshold can be found without knowledge of A, then the test is UMP. The LRT decides $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x};\mathcal{H}_1)}{p(\mathbf{x};\mathcal{H}_0)} > \gamma$$

or

$$\frac{\dfrac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left[-\dfrac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-As[n])^2\right]}{\dfrac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left[-\dfrac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right]} > \gamma.$$

Taking logarithms and simplifying yields

$$-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(-2As[n]x[n] + A^2s^2[n]\right) > \ln\gamma \tag{7.8}$$

or

$$A\sum_{n=0}^{N-1}x[n]s[n] > \sigma^2\ln\gamma + \frac{A^2}{2}\sum_{n=0}^{N-1}s^2[n] = \gamma'.$$

If $A > 0$, then the NP test is to decide $\mathcal{H}_1$ if

$$\sum_{n=0}^{N-1}x[n]s[n] > \frac{\gamma'}{A} = \gamma'' \tag{7.9}$$

while if $A < 0$, we decide $\mathcal{H}_1$ if

$$\sum_{n=0}^{N-1}x[n]s[n] < \frac{\gamma'}{A} = \gamma''. \tag{7.10}$$

In either case the test is UMP and reduces to the usual correlator structure. If, however, the sign of A is unknown, we cannot construct a *unique* test. For example, if we assume $A > 0$ and use (7.9), then poor results would be expected if A is actually negative. This is because for $A < 0$

$$E(T(\mathbf{x})) = E\left(\sum_{n=0}^{N-1} x[n]s[n]\right) = A\sum_{n=0}^{N-1} s^2[n] < 0.$$

Note that when a UMP test does exist so that the test statistic and threshold do not depend on the value of A, the *performance* of the detector will depend on the *magnitude* of A. The performance of the UMP test for either $A > 0$ or $A < 0$ is easily shown to be (see Problem 7.10)

$$P_D = Q\left(Q^{-1}(P_{FA}) - \sqrt{d^2}\right)$$

where

$$d^2 = \frac{A^2\sum_{n=0}^{N-1} s^2[n]}{\sigma^2} = \frac{\mathcal{E}}{\sigma^2}.$$

When the sign of the amplitude is unknown, we must resort to either a GLRT or a Bayesian approach, as described next.

7.4.1 GLRT

The GLRT decides $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x}; \hat{A}, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > \gamma$$

where $\hat{A}$ is the MLE of A under $\mathcal{H}_1$. The latter can be shown to be (see Problem 7.11)

$$\hat{A} = \frac{\sum_{n=0}^{N-1} x[n]s[n]}{\sum_{n=0}^{N-1} s^2[n]} \tag{7.11}$$

so that from (7.8)

$$-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(-2\hat{A}s[n]x[n] + \hat{A}^2 s^2[n]\right) > \ln\gamma$$

and using (7.11) we have

$$-\frac{1}{2\sigma^2}\left(-2\hat{A}\hat{A}\sum_{n=0}^{N-1} s^2[n] + \hat{A}^2\sum_{n=0}^{N-1} s^2[n]\right) > \ln\gamma.$$

Finally, we decide $\mathcal{H}_1$ if

$$\hat{A}^2 > \frac{2\sigma^2\ln\gamma}{\sum_{n=0}^{N-1} s^2[n]} \tag{7.12}$$

or equivalently if

$$|\hat{A}| > \sqrt{\frac{2\sigma^2\ln\gamma}{\sum_{n=0}^{N-1} s^2[n]}}.$$

For noise only we expect $\hat{A} \approx 0$ (since $E(\hat{A}) = 0$), and when a signal is present $|\hat{A}|$ should depart from zero. It is therefore not unreasonable that the detector is based on the magnitude of the amplitude estimate. Alternatively, we have from (7.12)

$$T(\mathbf{x}) = \left(\sum_{n=0}^{N-1} x[n]s[n]\right)^2 > 2\sigma^2\ln\gamma\sum_{n=0}^{N-1} s^2[n] = \gamma' \tag{7.13}$$

or

$$\left|\sum_{n=0}^{N-1} x[n]s[n]\right| > \sqrt{2\sigma^2\ln\gamma\sum_{n=0}^{N-1} s^2[n]} = \sqrt{\gamma'}. \tag{7.14}$$

The detector is just a correlator that accounts for the unknown sign of A by taking the absolute value. The detector of (7.13) is shown in Figure 7.2.

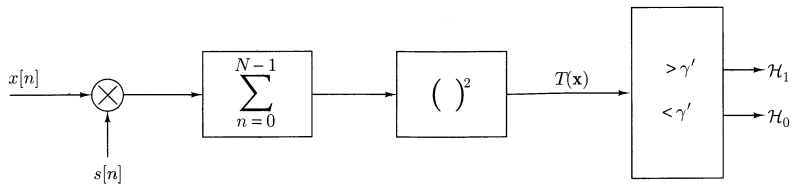


Figure 7.2. GLRT for unknown amplitude signal.

The lack of amplitude knowledge will degrade the detection performance, but only slightly so from that of the correlator. To determine the detection performance we first note that (see the derivation of (4.11))

$$u(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]s[n] \sim \begin{cases} \mathcal{N}\left(0, \sigma^2 \sum_{n=0}^{N-1} s^2[n]\right) & \text{under } \mathcal{H}_0 \\ \mathcal{N}\left(A\sum_{n=0}^{N-1} s^2[n], \sigma^2 \sum_{n=0}^{N-1} s^2[n]\right) & \text{under } \mathcal{H}_1. \end{cases}$$

Thus,

$$\begin{aligned} P_{FA} &= \Pr\left\{|u(\mathbf{x})| > \sqrt{\gamma'}; \mathcal{H}_0\right\} \\ &= \Pr\left\{u(\mathbf{x}) > \sqrt{\gamma'}; \mathcal{H}_0\right\} + \Pr\left\{u(\mathbf{x}) < -\sqrt{\gamma'}; \mathcal{H}_0\right\} \\ &= Q\left(\frac{\sqrt{\gamma'}}{\sqrt{\sigma^2 \sum_{n=0}^{N-1} s^2[n]}}\right) + 1 - Q\left(\frac{-\sqrt{\gamma'}}{\sqrt{\sigma^2 \sum_{n=0}^{N-1} s^2[n]}}\right) \\ &= 2Q\left(\frac{\sqrt{\gamma'}}{\sqrt{\sigma^2 \sum_{n=0}^{N-1} s^2[n]}}\right). \end{aligned} \tag{7.15}$$

Similarly,

$$\begin{aligned} P_D &= \Pr\left\{|u(\mathbf{x})| > \gamma'; \mathcal{H}_1\right\} \\ &= Q\left(\frac{\sqrt{\gamma'} - A\sum_{n=0}^{N-1} s^2[n]}{\sqrt{\sigma^2 \sum_{n=0}^{N-1} s^2[n]}}\right) + 1 - Q\left(\frac{-\sqrt{\gamma'} - A\sum_{n=0}^{N-1} s^2[n]}{\sqrt{\sigma^2 \sum_{n=0}^{N-1} s^2[n]}}\right) \\ &= Q\left(\frac{\sqrt{\gamma'} - A\sum_{n=0}^{N-1} s^2[n]}{\sqrt{\sigma^2 \sum_{n=0}^{N-1} s^2[n]}}\right) + Q\left(\frac{\sqrt{\gamma'} + A\sum_{n=0}^{N-1} s^2[n]}{\sqrt{\sigma^2 \sum_{n=0}^{N-1} s^2[n]}}\right). \end{aligned}$$

But from (7.15) we have

$$\frac{\sqrt{\gamma'}}{\sqrt{\sigma^2 \sum_{n=0}^{N-1} s^2[n]}} = Q^{-1}\left(\frac{P_{FA}}{2}\right)$$

so that

$$P_D = Q\left(Q^{-1}(P_{FA}/2) - \sqrt{\frac{\mathcal{E}}{\sigma^2}}\right) + Q\left(Q^{-1}(P_{FA}/2) + \sqrt{\frac{\mathcal{E}}{\sigma^2}}\right)$$

or finally

$$P_D = Q\left(Q^{-1}(P_{FA}/2) - \sqrt{d^2}\right) + Q\left(Q^{-1}(P_{FA}/2) + \sqrt{d^2}\right) \tag{7.16}$$

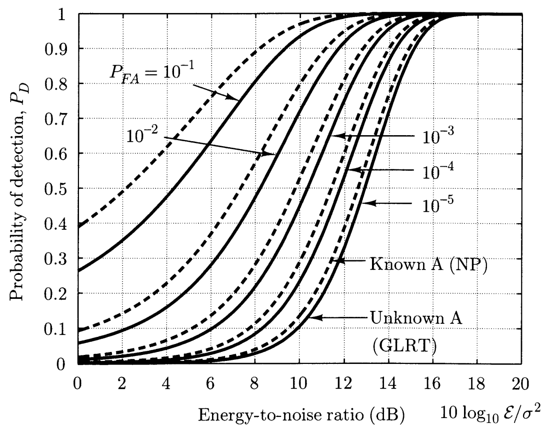


Figure 7.3. Detection performance for GLRT and clairvoyant detector.

where $d^2 = (A^2 \sum_{n=0}^{N-1} s^2[n])/\sigma^2 = \mathcal{E}/\sigma^2$ is the deflection coefficient. This performance is plotted in Figure 7.3 versus the case of known A. The degradation is seen to be only about 0.5 dB in ENR for low P_{FA}'s. These results could also be obtained more simply by employing the theory of the classical linear model as described in Section 7.7.

7.4.2 Bayesian Approach

We now assume that A is a random variable with PDF $\mathcal{N}(\mu_A, \sigma_A^2)$ and A is *independent* of $w[n]$. With these assumptions we have that under $\mathcal{H}_1$

$$\mathbf{x} = \mathbf{s}A + \mathbf{w}$$

where $\mathbf{s} = [s[0]\, s[1] \ldots s[N-1]]^T$. This is just the Bayesian linear model with $\mathbf{H} = \mathbf{s}$ and $\boldsymbol{\theta} = A$ [Kay-I 1993, pp. 325–326]. From the results in Section 5.6 (see (5.30)) it is easily shown that the NP test is to decide $\mathcal{H}_1$ if

$$\begin{aligned} T'(\mathbf{x}) &= \mathbf{x}^T\left(\mathbf{H}\mathbf{C}_\theta\mathbf{H}^T + \mathbf{C}_w\right)^{-1}\mathbf{H}\boldsymbol{\mu}_\theta \\ &\quad + \frac{1}{2}\mathbf{x}^T\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_\theta\mathbf{H}^T\left(\mathbf{H}\mathbf{C}_\theta\mathbf{H}^T + \mathbf{C}_w\right)^{-1}\mathbf{x} > \gamma'. \end{aligned} \tag{7.17}$$

Here we have $\mathbf{H} = \mathbf{s}$, $\boldsymbol{\theta} = A$, $\boldsymbol{\mu}_\theta = \mu_A$, $\mathbf{C}_\theta = \sigma_A^2$, and $\mathbf{C}_w = \sigma^2\mathbf{I}$ so that

$$T'(\mathbf{x}) = \mathbf{x}^T\left(\sigma_A^2\mathbf{s}\mathbf{s}^T + \sigma^2\mathbf{I}\right)^{-1}\mathbf{s}\mu_A$$

$$
\begin{aligned}
&+\frac{1}{2\sigma^2}\mathbf{x}^T\sigma_A^2\mathbf{s}\mathbf{s}^T\left(\sigma_A^2\mathbf{s}\mathbf{s}^T+\sigma^2\mathbf{I}\right)^{-1}\mathbf{x}\\
=\;& \mathbf{x}^T\left(\sigma_A^2\mathbf{s}\mathbf{s}^T+\sigma^2\mathbf{I}\right)^{-1}\mathbf{s}\mu_A\\
&+\frac{\sigma_A^2}{2\sigma^2}\mathbf{x}^T\mathbf{s}\mathbf{x}^T\left(\sigma_A^2\mathbf{s}\mathbf{s}^T+\sigma^2\mathbf{I}\right)^{-1}\mathbf{s}.
\end{aligned}
$$

Using Woodbury's identity (see Appendix 1)

$$
\begin{aligned}
\left(\sigma_A^2\mathbf{s}\mathbf{s}^T+\sigma^2\mathbf{I}\right)^{-1}\mathbf{s} &= \left(\frac{1}{\sigma^2}\mathbf{I}-\frac{\sigma_A^2}{\sigma^4}\frac{\mathbf{s}\mathbf{s}^T}{1+\frac{\sigma_A^2\mathbf{s}^T\mathbf{s}}{\sigma^2}}\right)\mathbf{s}\\
&= \frac{1}{\sigma^2}\left(\mathbf{s}-\frac{\sigma_A^2\mathbf{s}^T\mathbf{s}\mathbf{s}}{\sigma^2+\sigma_A^2\mathbf{s}^T\mathbf{s}}\right)\\
&= \frac{1}{\sigma^2+\sigma_A^2\mathbf{s}^T\mathbf{s}}\mathbf{s}.
\end{aligned}
$$

Thus, we have the NP test statistic

$$
T'(\mathbf{x}) = \frac{\mu_A}{\sigma^2+\sigma_A^2\mathbf{s}^T\mathbf{s}}\mathbf{x}^T\mathbf{s}+\frac{\sigma_A^2}{2\sigma^2(\sigma^2+\sigma_A^2\mathbf{s}^T\mathbf{s})}(\mathbf{x}^T\mathbf{s})^2 \tag{7.18}
$$

which is a combination of a correlator ($\mathbf{x}^T\mathbf{s}$) and a squared-correlator (see also Problem 5.14). Note that if $\sigma_A^2=0$ we have $T'(\mathbf{x})=\frac{1}{\sigma^2}\mathbf{x}^T\mu_A\mathbf{s}$, which is the known amplitude case (with $\mu_A=A$). Also, if $\sigma_A^2\to\infty$, we have $T'(\mathbf{x})=\frac{1}{2\sigma^2\mathbf{s}^T\mathbf{s}}(\mathbf{x}^T\mathbf{s})^2$, which is the Bayesian detector with no prior amplitude knowledge. In order to implement the Bayesian detector we require knowledge of μ_A and σ_A^2. When this information is available, the resultant detector is optimal in the NP sense (highest P_D for a given P_{FA}), unlike the GLRT, which is suboptimal. In Problem 7.12 we determine the performance of the detector when $\mu_A=0$.

7.5 Unknown Arrival Time

In many situations it is desired to detect a signal whose arrival time, or equivalently its delay, is unknown. An example might be a radar system in which a detection indicates an aircraft, and its round trip delay is used to estimate range [Kay-I 1993, pp. 53–56]. Hence, a GLRT might be employed as a detector/estimator. We now consider the detection problem

$$
\begin{array}{ll}
\mathcal{H}_0 : x[n] = w[n] & n=0,1,\ldots,N-1\\
\mathcal{H}_1 : x[n] = s[n-n_0]+w[n] & n=0,1,\ldots,N-1
\end{array}
$$

where $s[n]$ is a known deterministic signal that is nonzero over the interval $[0, M-1]$, n_0 is the unknown delay, and $w[n]$ is WGN with variance σ^2. Clearly, the observation interval $[0, N-1]$ should include the signal for all possible delays. A GLRT would decide $\mathcal{H}_1$ if

$$
\frac{p(\mathbf{x};\hat{n}_0,\mathcal{H}_1)}{p(\mathbf{x};\mathcal{H}_0)} > \gamma
$$

where $\hat{n}_0$ is the MLE of n_0. It has been shown [Kay-I 1993, pg. 192] that $\hat{n}_0$ is found by maximizing

$$
\sum_{n=n_0}^{n_0+M-1} x[n]s[n-n_0] \tag{7.19}
$$

over all possible n_0. Hence, we correlate the data with all possible delayed signals and choose $\hat{n}_0$ as the value that maximizes the correlation of (7.19). To evaluate the GLRT note that

$$
\begin{aligned}
p(\mathbf{x};n_0,\mathcal{H}_1) &= \prod_{n=0}^{n_0-1}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2\sigma^2}x^2[n]\right]\\
&\cdot\prod_{n=n_0}^{n_0+M-1}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2\sigma^2}(x[n]-s[n-n_0])^2\right]\\
&\cdot\prod_{n=n_0+M}^{N-1}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2\sigma^2}x^2[n]\right]\\
&= \prod_{n=0}^{N-1}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2\sigma^2}x^2[n]\right]\\
&\cdot\prod_{n=n_0}^{n_0+M-1}\exp\left[-\frac{1}{2\sigma^2}(-2x[n]s[n-n_0]+s^2[n-n_0])\right]
\end{aligned}
$$

where $0\le n_0\le N-M$. Thus, we have

$$
\frac{p(\mathbf{x};\hat{n}_0,\mathcal{H}_1)}{p(\mathbf{x};\mathcal{H}_0)} = \prod_{n=\hat{n}_0}^{\hat{n}_0+M-1}\exp\left[-\frac{1}{2\sigma^2}(-2x[n]s[n-\hat{n}_0]+s^2[n-\hat{n}_0])\right].
$$

Taking logarithms we decide $\mathcal{H}_1$ if

$$
-\frac{1}{2\sigma^2}\sum_{n=\hat{n}_0}^{\hat{n}_0+M-1}\left(-2x[n]s[n-\hat{n}_0]+s^2[n-\hat{n}_0]\right) > \ln\gamma.
$$

But

$$
\sum_{n=\hat{n}_0}^{\hat{n}_0+M-1} s^2[n-\hat{n}_0] = \sum_{n=0}^{M-1} s^2[n] = \mathcal{E}
$$

so that we decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \sum_{n=\hat{n}_0}^{\hat{n}_0+M-1} x[n]s[n-\hat{n}_0] > \frac{\mathcal{E}}{2} + \sigma^2 \ln \gamma = \gamma'. \tag{7.20}$$

This says that the GLRT implements a correlation of $x[n]$ with $s[n-n_0]$ and compares the maximum value, i.e., that obtained when $n_0 = \hat{n}_0$, to a threshold γ'. If the threshold is exceeded, a signal is declared to be present and its delay is estimated as $\hat{n}_0$; otherwise noise only is declared. Note that the test statistic may also be written as

$$T(\mathbf{x}) = \max_{n_0 \in [0, N-M]} \sum_{n=n_0}^{n_0+M-1} x[n]s[n-n_0] \tag{7.21}$$

with this implementation shown in Figure 7.4.

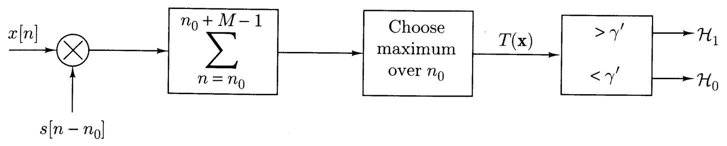


Figure 7.4. GLRT for signal with unknown arrival time.

The determination of the detection performance of the GLRT is difficult. From (7.21) we need to determine the PDF of the maximum of $N-M+1$ *correlated* Gaussian random variables. The correlation arises because the same noise samples enter into the sum for successive values of n_0. We do not pursue this matter further but refer the reader to Example 7.5 for an approximation. (Also see Problem 7.13 for a further discussion.)

The use of (7.21) may be undesirable in situations where the delay must be determined to be less than the sampling interval. In such a case the test statistic can be implemented in the frequency domain as (see Problem 7.14)

$$T(\mathbf{x}) = \max_{n_0 \in [0, N-M]} \int_{-\frac{1}{2}}^{\frac{1}{2}} X(f)\,[S(f)\exp(-j2\pi f n_0)]^*\,df \tag{7.22}$$

where $X(f) = \sum_{n=0}^{N-1} x[n]\exp(-j2\pi f n)$ and $S(f) = \sum_{n=0}^{M-1} s[n]\exp(-j2\pi f n)$ are the Fourier transforms of $x[n]$ and $s[n]$, respectively. Now noninteger values of delay can be accommodated.

The case of unknown arrival time and unknown amplitude can also be addressed. In Problem 7.15 it is shown that we decide $\mathcal{H}_1$ if

$$\max_{n_0 \in [0, N-M]} \left| \sum_{n=n_0}^{n_0+M-1} x[n]s[n-n_0] \right| > \gamma'. \tag{7.23}$$

7.6 Sinusoidal Detection

The detection of a sinusoid in WGN is a common problem in many fields. Because of its wide practical utility, we examine in some detail the detector structures as well as their performance. The results form the basis for many practical sonar, radar, and communication systems. The general detection problem is

$$\mathcal{H}_0 : x[n] = w[n] \qquad n = 0, 1, \ldots, N-1$$

$$\mathcal{H}_1 : x[n] = \begin{cases} w[n] & n = 0, 1, \ldots, n_0 - 1, n_0 + M, \ldots, N-1 \\ A\cos(2\pi f_0 n + \phi) + w[n] & n = n_0, n_0 + 1, \ldots, n_0 + M - 1 \end{cases}$$

where $w[n]$ is WGN with known variance σ^2 and any subset of the parameter set $\{A, f_0, \phi\}$ may be unknown. The sinusoid is assumed to be nonzero over the interval $[n_0, n_0 + M - 1]$ with M denoting the signal length and n_0 the delay time. Initially, we will assume that n_0 is known and that $n_0 = 0$. The observation interval is then just the signal interval or $[0, N-1] = [0, M-1]$. Later we will allow for unknown time delays. We now consider

$$\begin{aligned} \mathcal{H}_0 &: x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 &: x[n] = A\cos(2\pi f_0 n + \phi) + w[n] & n = 0, 1, \ldots, N-1 \end{aligned}$$

where the unknown parameters are deterministic. Recall that in Example 5.5 we assumed that A and ϕ were unknown but modeled as realizations of random variables. This Bayesian approach led to the Rayleigh fading sinusoidal model and the detector as a quadrature matched filter. Our approach now will be to use the GLRT for the following cases:

1. A unknown
2. A, ϕ unknown
3. A, ϕ, f_0 unknown
4. A, ϕ, f_0, n_0 unknown.

7.6.1 Amplitude Unknown

We have that the signal is $As[n]$, where $s[n] = \cos(2\pi f_0 n + \phi)$ and $s[n]$ is known. Thus, this is exactly the case studied in Section 7.4. Typically, the sign of the amplitude is not known. Thus, from (7.13) we decide $\mathcal{H}_1$ if

$$\left(\sum_{n=0}^{N-1} x[n]\cos(2\pi f_0 n + \phi)\right)^2 > \gamma' \tag{7.24}$$

with the detection performance given by (7.16). The MLE of the amplitude is from (7.11)

$$
\begin{aligned}
\hat{A} &= \frac{\sum_{n=0}^{N-1} x[n]\cos(2\pi f_0 n + \phi)}{\sum_{n=0}^{N-1} \cos^2(2\pi f_0 n + \phi)} \\
&\approx \frac{2}{N}\sum_{n=0}^{N-1} x[n]\cos(2\pi f_0 n + \phi)
\end{aligned}
$$

for f_0 not near 0 or 1/2. The detector is shown in Figure 7.5a and its performance is given in Figure 7.6a. Here, we have that $\mathcal{E}/\sigma^2 \approx NA^2/(2\sigma^2)$.

7.6.2 Amplitude and Phase Unknown

When A and ϕ are unknown, we must assume that $A > 0$. Otherwise, two different sets of A, ϕ will yield the same signal, and thus the parameters will not be *identifiable*. The reader should consider what happens if $A = 1, \phi = 0$ and $A = -1, \phi = \pi$. The GLRT decides $\mathcal{H}_1$ if

$$
\frac{p(\mathbf{x}; \hat{A}, \hat{\phi}, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > \gamma
$$

where $\hat{A}, \hat{\phi}$ is the MLE and f_0 is not near 0 or 1/2. Hence, we decide $\mathcal{H}_1$ if

$$
L_G(\mathbf{x}) = \frac{\dfrac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left[-\dfrac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n] - \hat{A}\cos(2\pi f_0 n + \hat{\phi})\right)^2\right]}{\dfrac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left[-\dfrac{1}{2\sigma^2}\sum_{n=0}^{N-1} x^2[n]\right]} > \gamma.
$$

The MLE for large N can be shown to be approximately [Kay-I 1993, pp. 193–195]

$$
\begin{aligned}
\hat{A} &= \sqrt{\hat{\alpha}_1^2 + \hat{\alpha}_2^2} \\
\hat{\phi} &= \arctan\left(\frac{-\hat{\alpha}_2}{\hat{\alpha}_1}\right)
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{\alpha}_1 &= \frac{2}{N}\sum_{n=0}^{N-1} x[n]\cos 2\pi f_0 n \\
\hat{\alpha}_2 &= \frac{2}{N}\sum_{n=0}^{N-1} x[n]\sin 2\pi f_0 n.
\end{aligned}
$$

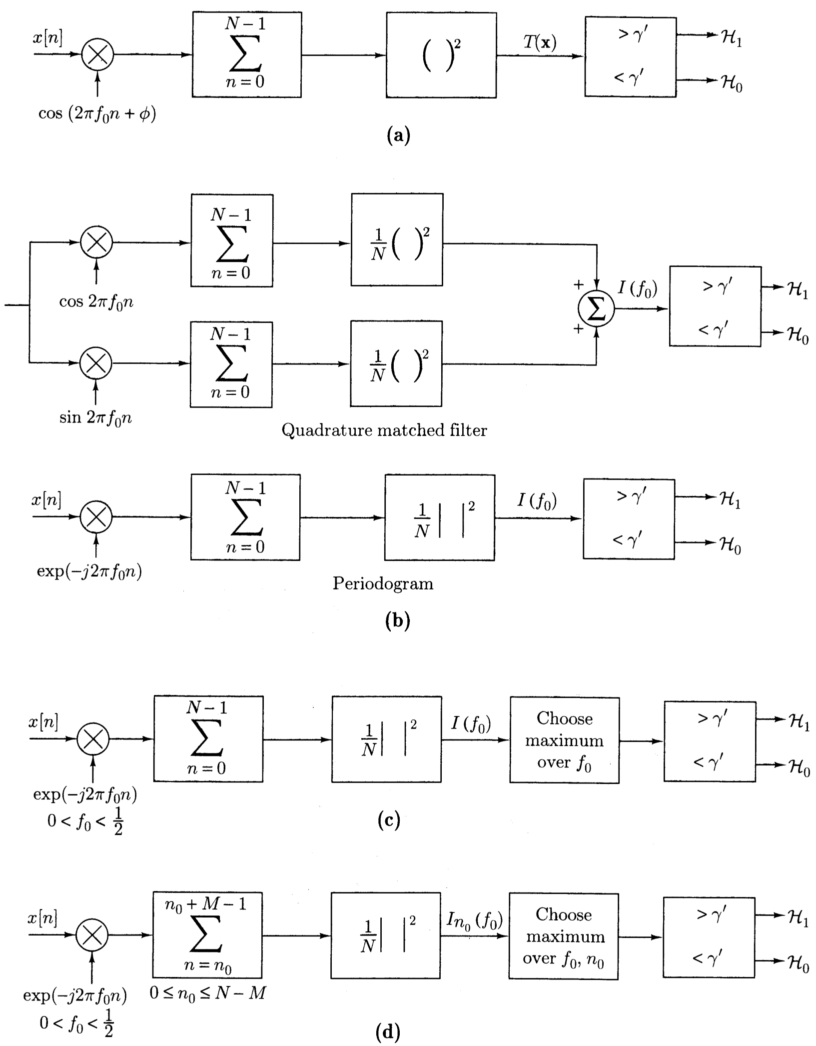
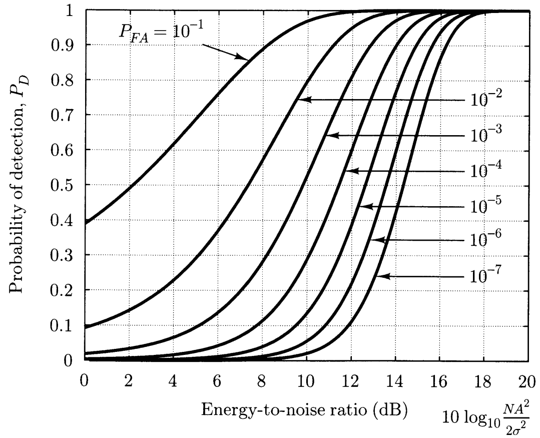
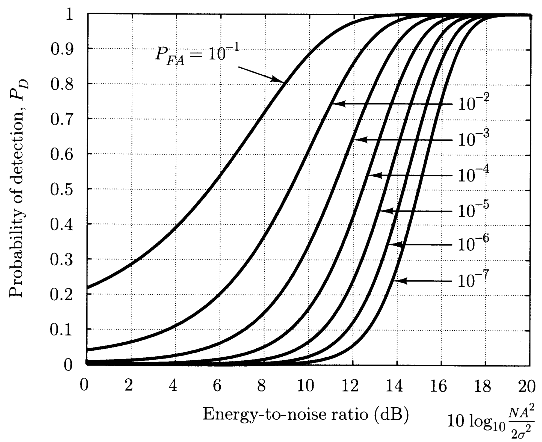


Figure 7.5. GLRT for sinusoidal detection (a) Unknown amplitude (b) Unknown amplitude and phase (c) Unknown amplitude, phase, frequency (d) Unknown amplitude, phase, frequency, arrival time.

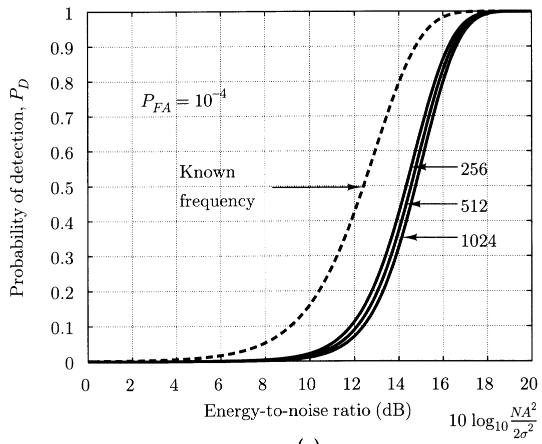


(a)



(b)

Figure 7.6. GLRT detection performance for sinusoid in WGN (a) Unknown amplitude (b) Unknown amplitude, phase.



(c)

Figure 7.6. Continued. GLRT detection performance for sinusoid in WGN (c) Unknown amplitude, phase, frequency.

In deriving the MLE the parameter transformation $\alpha_1 = A\cos\phi$, $\alpha_2 = -A\sin\phi$ has been employed. Now

$$\ln L_G(\mathbf{x}) = -\frac{1}{2\sigma^2}\left[\sum_{n=0}^{N-1} -2x[n]\hat{A}\cos(2\pi f_0 n + \hat{\phi}) + \sum_{n=0}^{N-1} \hat{A}^2\cos^2(2\pi f_0 n + \hat{\phi})\right].$$

Using the parameter transformation we have $\hat{\alpha}_1 = \hat{A}\cos\hat{\phi}$, $\hat{\alpha}_2 = -\hat{A}\sin\hat{\phi}$ so that

$$\begin{aligned}
&\sum_{n=0}^{N-1} x[n]\hat{A}\cos(2\pi f_0 n + \hat{\phi}) \\
&= \sum_{n=0}^{N-1} x[n]\cos(2\pi f_0 n)\hat{A}\cos\hat{\phi} - \sum_{n=0}^{N-1} x[n]\sin(2\pi f_0 n)\hat{A}\sin\hat{\phi} \\
&= \frac{N}{2}(\hat{\alpha}_1^2 + \hat{\alpha}_2^2).
\end{aligned}$$

Also, making use of

$$\sum_{n=0}^{N-1}\cos^2(2\pi f_0 n + \hat{\phi}) \approx \frac{N}{2}$$

results in

$$\begin{aligned}\ln L_G(\mathbf{x}) &= -\frac{1}{2\sigma^2}\left[-2\frac{N}{2}(\hat{\alpha}_1^2+\hat{\alpha}_2^2)+\frac{N}{2}\hat{A}^2\right] \\ &= -\frac{1}{2\sigma^2}\left[-\frac{N}{2}(\hat{\alpha}_1^2+\hat{\alpha}_2^2)\right] \\ &= \frac{N}{4\sigma^2}(\hat{\alpha}_1^2+\hat{\alpha}_2^2)\end{aligned}$$

or we decide $\mathcal{H}_1$ if

$$\frac{N}{4\sigma^2}(\hat{\alpha}_1^2+\hat{\alpha}_2^2) > \ln\gamma.$$

But

$$\begin{aligned}\hat{\alpha}_1^2+\hat{\alpha}_2^2 &= \left(\frac{2}{N}\right)^2\left[\left(\sum_{n=0}^{N-1}x[n]\cos 2\pi f_0 n\right)^2+\left(\sum_{n=0}^{N-1}x[n]\sin 2\pi f_0 n\right)^2\right] \\ &= \frac{4}{N}\frac{1}{N}\left|\sum_{n=0}^{N-1}x[n]\exp(-j2\pi f_0 n)\right|^2 \\ &= \frac{4}{N}I(f_0)\end{aligned}$$

where $I(f_0)$ is the periodogram [Kay 1988] evaluated at $f=f_0$. Finally, we decide $\mathcal{H}_1$ if

$$I(f_0) > \sigma^2\ln\gamma = \gamma'. \tag{7.25}$$

The form of the detector is identical to that for the Rayleigh fading model. Of course, the detection performance will be different. This detector is also called an *incoherent or quadrature matched filter* and is shown in Figure 7.5b in its two equivalent forms.

The detection performance can easily be found from Theorem 7.1 since this is a special case of the classical linear model. This approach is left to the reader as an exercise (see Problem 7.17). A "first principles" approach first rewrites the periodogram as

$$I(f_0)=\xi_1^2+\xi_2^2$$

where

$$\begin{aligned}\xi_1 &= \frac{1}{\sqrt{N}}\sum_{n=0}^{N-1}x[n]\cos 2\pi f_0 n \\ \xi_2 &= \frac{1}{\sqrt{N}}\sum_{n=0}^{N-1}x[n]\sin 2\pi f_0 n.\end{aligned}$$

Then, since ξ_1,ξ_2 are linear transformations of $\mathbf{x}$, they are jointly Gaussian. For f_0 not near 0 or 1/2 it can be shown that if $\boldsymbol{\xi}=[\xi_1\,\xi_2]^T$

$$\begin{aligned}&\boldsymbol{\xi}\sim\mathcal{N}\left(\mathbf{0},\tfrac{\sigma^2}{2}\mathbf{I}\right) && \text{under } \mathcal{H}_0 \\ &\boldsymbol{\xi}\sim\mathcal{N}\left(\begin{bmatrix}\frac{\sqrt{N}}{2}A\cos\phi \\ -\frac{\sqrt{N}}{2}A\sin\phi\end{bmatrix},\tfrac{\sigma^2}{2}\mathbf{I}\right) && \text{under } \mathcal{H}_1\end{aligned}$$

as outlined in Problem 7.18. Since under either hypothesis the random variables are independent, the PDF is related to the central χ^2 under $\mathcal{H}_0$ and the noncentral χ^2 under $\mathcal{H}_1$. Specifically, consider the normalized periodogram $I(f_0)/(\sigma^2/2)$. The latter is distributed as a χ_2^2 PDF under $\mathcal{H}_0$ and as a $\chi_2'^2(\lambda)$ PDF under $\mathcal{H}_1$, where

$$\lambda=\left(\frac{\sqrt{N}\frac{A}{2}\cos\phi}{\sigma/\sqrt{2}}\right)^2+\left(\frac{-\sqrt{N}\frac{A}{2}\sin\phi}{\sigma/\sqrt{2}}\right)^2=\frac{NA^2}{2\sigma^2}$$

is the noncentrality parameter. As a result

$$\begin{aligned}P_{FA} &= \Pr\{I(f_0)>\gamma';\mathcal{H}_0\} \\ &= \Pr\left\{\frac{I(f_0)}{\sigma^2/2}>\frac{\gamma'}{\sigma^2/2};\mathcal{H}_0\right\} \\ &= Q_{\chi_2^2}\left(\frac{2\gamma'}{\sigma^2}\right) \\ &= \exp\left(-\frac{\gamma'}{\sigma^2}\right).\end{aligned} \tag{7.26}$$

Also

$$\begin{aligned}P_D &= \Pr\{I(f_0)>\gamma';\mathcal{H}_1\} \\ &= \Pr\left\{\frac{I(f_0)}{\sigma^2/2}>\frac{\gamma'}{\sigma^2/2};\mathcal{H}_1\right\} \\ &= Q_{\chi_2'^2(\lambda)}\left(\frac{2\gamma'}{\sigma^2}\right).\end{aligned} \tag{7.27}$$

In summary, the detection performance may be written using (7.26) as

$$P_D=Q_{\chi_2'^2(\lambda)}\left(2\ln\frac{1}{P_{FA}}\right) \tag{7.28}$$

where $\lambda=NA^2/(2\sigma^2)$. Using the MATLAB program Qchipr2.m in Appendix 2D some detection curves are plotted in Figure 7.6b. As expected, there is a slight degradation as compared to the previous case of unknown amplitude only. However, it is less than 1 dB for small P_{FA}'s as can be seen by comparing Figure 7.6b to 7.6a. See also Problem 7.20, in which P_D is expressed in terms of the Rician PDF, leading to the Marcum Q function.

Before concluding our discussion, we should note that it is possible to use a Bayesian approach to detect a sinusoid of unknown amplitude and phase. To do so one usually chooses a noninformative or uniform PDF for the phase. Then as described in Problem 7.22, it is possible to obtain the NP detector, which is optimal under the stated assumptions. The test statistic is again the periodogram evaluated at $f = f_0$.

7.6.3 Amplitude, Phase, and Frequency Unknown

When the frequency is also unknown, the GLRT decides $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x}; \hat{A}, \hat{\phi}, \hat{f}_0, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > \gamma$$

or

$$\frac{\max_{f_0} p(\mathbf{x}; \hat{A}, \hat{\phi}, f_0, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > \gamma.$$

Since the PDF under $\mathcal{H}_0$ does not depend on f_0 and is nonnegative, we have

$$\max_{f_0} \frac{p(\mathbf{x}; \hat{A}, \hat{\phi}, f_0, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > \gamma.$$

Also, because the logarithm is a monotonic function we have the equivalent test

$$\ln \max_{f_0} \frac{p(\mathbf{x}; \hat{A}, \hat{\phi}, f_0, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > \ln \gamma$$

and again because of the monotonicity (see Problem 7.21)

$$\max_{f_0} \ln \frac{p(\mathbf{x}; \hat{A}, \hat{\phi}, f_0, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > \ln \gamma.$$

But from (7.25)

$$\ln \frac{p(\mathbf{x}; \hat{A}, \hat{\phi}, f_0, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} = \frac{I(f_0)}{\sigma^2} \tag{7.29}$$

so that finally we decide $\mathcal{H}_1$ if

$$\max_{f_0} I(f_0) > \sigma^2 \ln \gamma = \gamma'. \tag{7.30}$$

The detector decides that a sinusoid is present if the peak value of the periodogram exceeds a threshold, and if so, the frequency location of that peak is the MLE of the frequency. This explains why the periodogram or its fast Fourier transform (FFT) implementation is a basic component in nearly all narrowband detection systems. The detector is illustrated in Figure 7.5c. Its performance can be found in a similar fashion to that of the previous case of amplitude and phase unknown. The only difference is that the P_{FA} increases by the number of frequencies searched over (see (7.37)). Assuming an N-point FFT is used to evaluate $I(f)$ and that the maximum is found over the frequencies $f_k = k/N$ for $k = 1, 2, \ldots, N/2 - 1$, we have that

$$P_D = Q_{\chi_2'^2\left(\frac{NA^2}{2\sigma^2}\right)}\left(2 \ln \frac{N/2 - 1}{P_{FA}}\right). \tag{7.31}$$

See Figure 7.6c for some detection curves.

7.6.4 Amplitude, Phase, Frequency, and Arrival Time Unknown

In a similar fashion as before the GLRT decides $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x}; \hat{A}, \hat{\phi}, \hat{f}_0, \hat{n}_0, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > \gamma.$$

The MLE of A, ϕ, f_0 is the same as for the case of a known arrival time n_0 except now the data interval is modified to coincide with the signal interval $[n_0, n_0+M-1]$. Hence, for *known* n_0

$$\begin{aligned} \hat{A} &= \sqrt{\hat{\alpha}_1^2 + \hat{\alpha}_2^2} \\ \hat{\phi} &= \arctan\left(\frac{-\hat{\alpha}_2}{\hat{\alpha}_1}\right) \end{aligned}$$

where

$$\begin{aligned} \hat{\alpha}_1 &= \frac{2}{M} \sum_{n=n_0}^{n_0+M-1} x[n] \cos[2\pi \hat{f}_0 (n - n_0)] \\ \hat{\alpha}_2 &= \frac{2}{M} \sum_{n=n_0}^{n_0+M-1} x[n] \sin[2\pi \hat{f}_0 (n - n_0)] \end{aligned}$$

and $\hat{f}_0$ is the frequency at which the periodogram attains its maximum. Substituting into (7.29) we have, as in the previous case,

$$\ln \frac{p(\mathbf{x}; \hat{A}, \hat{\phi}, \hat{f}_0, n_0, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} = \frac{I_{n_0}(\hat{f}_0)}{\sigma^2} \tag{7.32}$$

where

$$I_{n_0}(\hat{f}_0) = \frac{1}{M}\left|\sum_{n=n_0}^{n_0+M-1} x[n]\exp(-j2\pi \hat{f}_0 n)\right|^2.$$

Finally, to find the MLE of n_0 we need to maximize $p(\mathbf{x}; \hat{A}, \hat{\phi}, \hat{f}_0, n_0, \mathcal{H}_1)$ over n_0 or equivalently just (7.32). Hence, we decide $\mathcal{H}_1$ if

$$\max_{n_0} \frac{I_{n_0}(\hat{f}_0)}{\sigma^2} > \ln\gamma$$

or

$$\max_{n_0, f_0} \frac{I_{n_0}(f_0)}{\sigma^2} > \ln\gamma$$

or finally if

$$\max_{n_0, f_0} I_{n_0}(f_0) > \gamma' \tag{7.33}$$

where

$$I_{n_0}(f_0) = \frac{1}{M}\left|\sum_{n=n_0}^{n_0+M-1} x[n]\exp(-j2\pi f_0 n)\right|^2.$$

The $I_{n_0}(f_0)$ statistic is called the *short-time periodogram* or *spectrogram*. The GLRT thus computes the periodogram for all delays and then compares the maximum to a threshold. If the threshold is exceeded, the MLE of the delay and frequency is the location of the maximum. This detector is a standard one in active sonar and radar. Its performance is described in more detail in Section 7.8. An example of the spectrogram is shown in Figure 7.7, for which the parameters are $A = 1$, $f_0 = 0.25$, $\phi = 0$, $M = 128$, $n_0 = 128$, $N = 512$, and $\sigma^2 = 0.5$. The SNR is $A^2/2\sigma^2 = 1$ or 0 dB. A realization of $x[n]$ under $\mathcal{H}_1$ is shown in Figure 7.7a. The signal is present in the interval $[n_0, n_0 + M - 1] = [128, 255]$, although not clearly visible because of the low SNR. The spectrogram is shown in Figure 7.7b. The maximum clearly indicates the presence of the signal and occurs at about the correct frequency and delay. A gray-scale representation of the spectrogram is shown in Figure 7.7c. As a side note, the frequency estimate $\hat{f}_0 = 0.25$ is more accurate than the delay estimate $\hat{n}_0 = 141$ because of the shape of the spectrogram peak. This result can be shown to be true in general and is predicted by the CRLB. See [Van Trees 1971] for a discussion of the CRLB and its dependence on the signal ambiguity function.

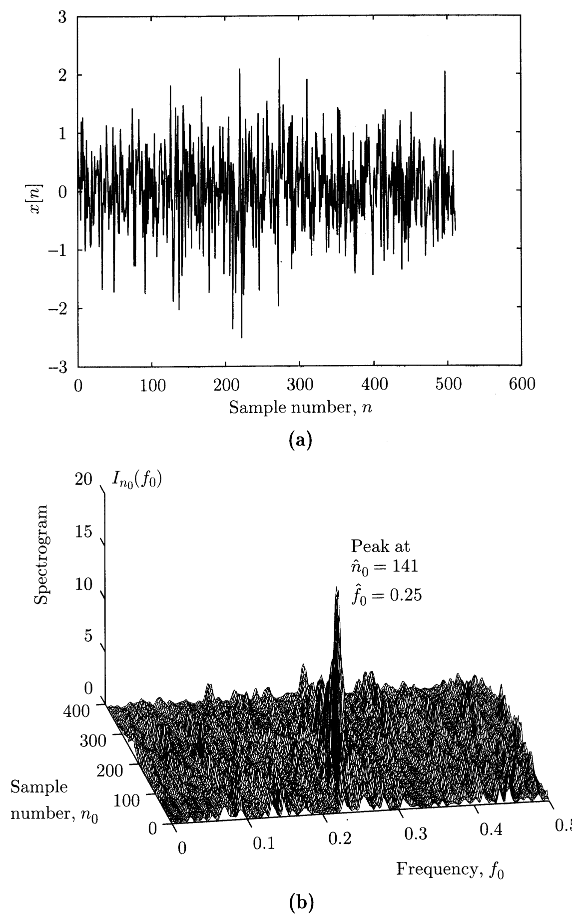


Figure 7.7. GLRT detection of unknown amplitude, phase, frequency, arrival time sinusoid in WGN (a) Time series data (b) Spectrogram of data.

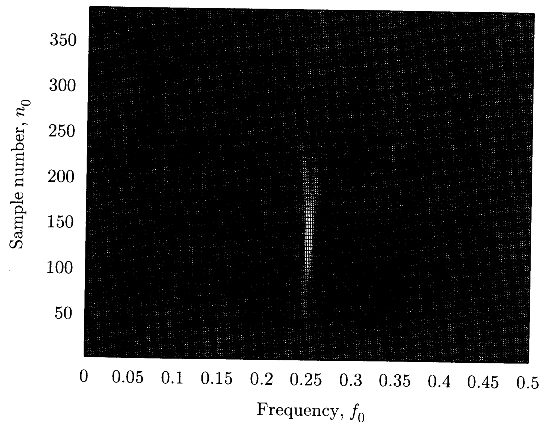


Figure 7.7. Continued. GLRT detection of unknown amplitude, phase, frequency, arrival time sinusoid in WGN (c) Grayscale spectrogram of data.

7.7 Classical Linear Model

A large class of detection problems can be easily solved by applying the classical or Bayesian linear model. In using the Bayesian linear model as already discussed in Chapter 5, we can effectively reduce the detection problem with unknown signal parameters to a special case of the general Gaussian problem (see Section 5.6). The NP detector follows immediately. For the classical linear model (parameters assumed deterministic), however, no UMP test exists. Thus, the GLRT is the standard approach. An example of this has been given in Section 7.4.1 in which under $\mathcal{H}_1$

$$x[n] = As[n] + w[n] \qquad n = 0, 1, \ldots, N-1$$

or

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

where $\mathbf{H} = [s[0]\, s[1] \ldots s[N-1]]^T$, $\boldsymbol{\theta} = A$, and $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. We now generalize the results to obtain the GLRT for a wide class of detection problems. Specifically, under $\mathcal{H}_1$ we assume [Kay-I 1993, Chapter 4]

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

where $\mathbf{H}$ is a known $N \times p$ matrix, $\boldsymbol{\theta}$ is a $p \times 1$ vector of parameters (a subset of which is unknown), and $\mathbf{w}$ is an $N \times 1$ random noise vector with PDF $\mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$

with σ^2 known. We wish to test if $\boldsymbol{\theta}$ satisfies the linear equations $\mathbf{A}\boldsymbol{\theta} = \mathbf{b}$, where $\mathbf{A}$ is a known $r \times p$ (with $r \le p$) matrix and $\mathbf{b}$ is a known $r \times 1$ vector. The linear equations are assumed to be consistent or to possess at least one solution. The hypothesis testing problem is

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{A}\boldsymbol{\theta} = \mathbf{b} \\ \mathcal{H}_1 &: \mathbf{A}\boldsymbol{\theta} \neq \mathbf{b}. \end{aligned} \tag{7.34}$$

To illustrate the modeling we consider the following examples.

Example 7.1 - Unknown Amplitude Signal in WGN

In Section 7.4 we considered the detection of a sinusoid with unknown amplitude in WGN or we assumed that $x[n] = As[n] + w[n]$ under $\mathcal{H}_1$ with A unknown and $x[n] = w[n]$ under $\mathcal{H}_0$. We can formulate the detection problem as a special case of the linear model. To do so note that

$$\begin{aligned} \mathcal{H}_0 &: x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 &: x[n] = As[n] + w[n] & n = 0, 1, \ldots, N-1 \end{aligned}$$

can equivalently be written as

$$\begin{aligned} \mathcal{H}_0 &: x[n] = As[n] + w[n], \ A = 0 & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 &: x[n] = As[n] + w[n], \ A \neq 0 & n = 0, 1, \ldots, N-1 \end{aligned}$$

or in vector-matrix form, we have $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$ with $\mathbf{H} = [s[0]\, s[1] \ldots s[N-1]]^T$, $\boldsymbol{\theta} = A$, and

$$\begin{aligned} \mathcal{H}_0 &: \boldsymbol{\theta} = \mathbf{0} \\ \mathcal{H}_1 &: \boldsymbol{\theta} \neq \mathbf{0}. \end{aligned}$$

This is of the form of (7.34) with $\mathbf{A} = \mathbf{I}$ and $\mathbf{b} = \mathbf{0}$.

◇

Example 7.2 - Sinusoid with Unknown Amplitude and Phase in WGN

As in Section 7.6.2, we assume that $x[n] = A\cos(2\pi f_0 n + \phi) + w[n]$ under $\mathcal{H}_1$ and $x[n] = w[n]$ under $\mathcal{H}_0$. As a result we can write under $\mathcal{H}_1$

$$\begin{aligned} x[n] &= A\cos\phi\cos 2\pi f_0 n - A\sin\phi\sin 2\pi f_0 n + w[n] \\ &= \alpha_1 \cos 2\pi f_0 n + \alpha_2 \sin 2\pi f_0 n + w[n]. \end{aligned}$$

Clearly, $A = 0$ if and only if $\alpha_1 = \alpha_2 = 0$ since $A = \sqrt{\alpha_1^2 + \alpha_2^2}$. Thus, we have that the detection problem

$$\begin{aligned} \mathcal{H}_0 &: x[n] = A\cos(2\pi f_0 n + \phi) + w[n], \ A = 0 \\ \mathcal{H}_1 &: x[n] = A\cos(2\pi f_0 n + \phi) + w[n], \ A \neq 0 \end{aligned}$$

is equivalent to

$$\begin{array}{l} \mathcal{H}_0 : x[n] = \alpha_1 \cos 2\pi f_0 n + \alpha_2 \sin 2\pi f_0 n + w[n], \ \alpha_1 = \alpha_2 = 0 \\ \mathcal{H}_1 : x[n] = \alpha_1 \cos 2\pi f_0 n + \alpha_2 \sin 2\pi f_0 n + w[n], \ \alpha_1^2 + \alpha_2^2 \neq 0. \end{array}$$

In terms of the linear model we have $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$ where

$$\begin{aligned} \mathbf{H} &= \begin{bmatrix} 1 & 0 \\ \cos 2\pi f_0 & \sin 2\pi f_0 \\ \vdots & \vdots \\ \cos 2\pi f_0(N-1) & \sin 2\pi f_0(N-1) \end{bmatrix} \\ \boldsymbol{\theta} &= \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \end{aligned}$$

and the hypothesis testing problem is

$$\begin{array}{l} \mathcal{H}_0 : \boldsymbol{\theta} = \mathbf{0} \\ \mathcal{H}_1 : \boldsymbol{\theta} \neq \mathbf{0}. \end{array}$$

Note that under $\mathcal{H}_1$ we allow $\boldsymbol{\theta}$ or equivalently A and ϕ to take on any values in the range $0 < A < \infty, 0 \leq \phi \leq 2\pi$. This is the assertion that the parameters are unknown under $\mathcal{H}_1$.

◇

The reader may also wish to refer to Chapter 12 in which the linear model is applied to detect changes in parameters.

The GLRT for the classical linear model is now summarized as a theorem. The proof is given in Appendix 7B.

Theorem 7.1 (GLRT for Classical Linear Model) *Assume the data have the form $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$, where $\mathbf{H}$ is a known $N \times p$ $(N > p)$ observation matrix of rank p, $\boldsymbol{\theta}$ is a $p \times 1$ vector of parameters, and $\mathbf{w}$ is an $N \times 1$ noise vector with PDF $\mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. The GLRT for the hypothesis testing problem*

$$\begin{array}{l} \mathcal{H}_0 : \mathbf{A}\boldsymbol{\theta} = \mathbf{b} \\ \mathcal{H}_1 : \mathbf{A}\boldsymbol{\theta} \neq \mathbf{b} \end{array}$$

where $\mathbf{A}$ is an $r \times p$ matrix $(r \leq p)$ of rank r, $\mathbf{b}$ is an $r \times 1$ vector, and $\mathbf{A}\boldsymbol{\theta} = \mathbf{b}$ is a consistent set of linear equations, is to decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = 2 \ln L_G(\mathbf{x}) = \frac{(\mathbf{A}\hat{\boldsymbol{\theta}}_1 - \mathbf{b})^T \left[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\right]^{-1} (\mathbf{A}\hat{\boldsymbol{\theta}}_1 - \mathbf{b})}{\sigma^2} > \gamma' \qquad (7.35)$$

where

$$\hat{\boldsymbol{\theta}}_1 = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$$

is the MLE of $\boldsymbol{\theta}$ under $\mathcal{H}_1$. The exact detection performance is given by

$$\begin{aligned} P_{FA} &= Q_{\chi_r^2}(\gamma') \\ P_D &= Q_{\chi_r'^2(\lambda)}(\gamma') \end{aligned}$$

where the noncentrality parameter is

$$\lambda = \frac{(\mathbf{A}\boldsymbol{\theta}_1 - \mathbf{b})^T \left[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\right]^{-1} (\mathbf{A}\boldsymbol{\theta}_1 - \mathbf{b})}{\sigma^2}.$$

The extension of this theorem for σ^2 unknown is given in Chapter 9. It is interesting to observe that for the linear model the asymptotic GLRT performance (see Chapter 6) is exact for finite data records as given by the theorem. An example of the application of this important theorem follows.

Example 7.3 - Unknown Amplitude Signal in WGN (continued)

Referring to Example 7.1, we have the classical linear model with

$$\begin{aligned} \mathbf{H} &= [\ s[0] \ \ s[1] \ \ \ldots \ \ s[N-1]\ ]^T \\ \boldsymbol{\theta} &= A \end{aligned}$$

and the test

$$\begin{array}{l} \mathcal{H}_0 : A = 0 \\ \mathcal{H}_1 : A \neq 0 \end{array}$$

so that $r = p = 1$, $\mathbf{A} = 1$ and $\mathbf{b} = 0$. Applying the theorem we have

$$T(\mathbf{x}) = \frac{\hat{\theta}_1(\mathbf{H}^T\mathbf{H})\hat{\theta}_1}{\sigma^2} = \frac{\mathbf{H}^T\mathbf{H}\hat{\theta}_1^2}{\sigma^2}$$

where

$$\begin{aligned} \hat{\theta}_1 = \hat{A} &= (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x} \\ &= \frac{\sum_{n=0}^{N-1} x[n]s[n]}{\sum_{n=0}^{N-1} s^2[n]}. \end{aligned}$$

We decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \frac{\left(\sum_{n=0}^{N-1} x[n]s[n]\right)^2}{\sigma^2 \sum_{n=0}^{N-1} s^2[n]} > \gamma'.$$

This is equivalent to (7.13) with γ' as used above equal to $2\ln\gamma$. The detection performance is

$$\begin{aligned} P_{FA} &= Q_{\chi_1^2}(\gamma') \\ P_D &= Q_{\chi_1'^2(\lambda)}(\gamma') \end{aligned}$$

where

$$\lambda = \frac{\theta_1^2(\mathbf{H}^T\mathbf{H})}{\sigma^2} = \frac{A^2\sum_{n=0}^{N-1} s^2[n]}{\sigma^2} = \frac{\mathcal{E}}{\sigma^2}.$$

But a noncentral χ^2 random variable with one degree of freedom is the same as the square of a unit variance Gaussian random variable or

$$\chi_1'^2(\lambda) = z^2$$

where $z \sim \mathcal{N}(\sqrt{\lambda}, 1)$. As a result, we have

$$\begin{aligned} P_D &= \Pr\left\{\chi_1'^2(\lambda) > \gamma'\right\} \\ &= \Pr\left\{z^2 > \gamma'\right\} \\ &= \Pr\left\{z > \sqrt{\gamma'}\right\} + \Pr\left\{z < -\sqrt{\gamma'}\right\} \\ &= Q\left(\sqrt{\gamma'} - \sqrt{\lambda}\right) + 1 - Q\left(-\sqrt{\gamma'} - \sqrt{\lambda}\right) \\ &= Q\left(\sqrt{\gamma'} - \sqrt{\lambda}\right) + Q\left(\sqrt{\gamma'} + \sqrt{\lambda}\right) \end{aligned}$$

and letting $\lambda = 0$, it follows that $P_{FA} = 2Q(\sqrt{\gamma'})$ so that $\sqrt{\gamma'} = Q^{-1}(P_{FA}/2)$ and

$$P_D = Q\left(Q^{-1}\left(\frac{P_{FA}}{2}\right) - \sqrt{\frac{\mathcal{E}}{\sigma^2}}\right) + Q\left(Q^{-1}\left(\frac{P_{FA}}{2}\right) + \sqrt{\frac{\mathcal{E}}{\sigma^2}}\right)$$

which agrees with (7.16).

◇

Example 7.4 - Compensation for DC Offset

In this example we illustrate how "nuisance parameters" are easily accommodated with the linear model. These parameters are of no interest to us but because they are unknown, they must also be estimated. The hypothesis testing problem then is

to test a *subset* of the unknown parameters. We now modify the previous example to include a DC offset under either hypothesis. This condition can occur due to amplifier biases or may be inherent in the data as in image processing for which each pixel value is nonnegative or in financial data, which is also nonnegative. Consider then the detection problem

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = \mu + w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = As[n] + \mu + w[n] & n = 0, 1, \ldots, N-1 \end{array}$$

where $s[n]$ is known but the amplitude A is not, μ is the unknown DC offset (nuisance parameter), and $w[n]$ is WGN with variance σ^2. Then, we have the classical linear model

$$\mathbf{x} = \underbrace{\begin{bmatrix} s[0] & 1 \\ s[1] & 1 \\ \vdots & \vdots \\ s[N-1] & 1 \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} A \\ \mu \end{bmatrix}}_{\boldsymbol{\theta}} + \mathbf{w}.$$

The detection problem is

$$\begin{aligned} \mathcal{H}_0 &: A = 0 \\ \mathcal{H}_1 &: A \neq 0 \end{aligned}$$

or equivalently

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{A}\boldsymbol{\theta} = \mathbf{0} \\ \mathcal{H}_1 &: \mathbf{A}\boldsymbol{\theta} \neq \mathbf{0} \end{aligned}$$

where $\mathbf{A} = [1\ 0]$. Thus, the classical linear model applies and the GLRT is from (7.35)

$$T(\mathbf{x}) = \frac{\hat{\boldsymbol{\theta}}_1^T \mathbf{A}^T \left[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\right]^{-1} \mathbf{A}\hat{\boldsymbol{\theta}}_1}{\sigma^2}.$$

But $\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T$ is a scalar so that

$$T(\mathbf{x}) = \frac{\hat{\boldsymbol{\theta}}_1^T \mathbf{A}^T \mathbf{A}\hat{\boldsymbol{\theta}}_1}{\sigma^2 \mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T}$$

and also $\hat{\boldsymbol{\theta}}_1^T \mathbf{A}^T \mathbf{A}\hat{\boldsymbol{\theta}}_1 = [\hat{\boldsymbol{\theta}}_1]_1^2$. Now the MLE of $\boldsymbol{\theta}_1$ is $\hat{\boldsymbol{\theta}}_1 = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$, which is easily found using

$$\mathbf{H}^T\mathbf{H} = \begin{bmatrix} \sum_{n=0}^{N-1} s^2[n] & \sum_{n=0}^{N-1} s[n] \\ \sum_{n=0}^{N-1} s[n] & N \end{bmatrix}$$

$$
\begin{aligned}
(\mathbf{H}^T\mathbf{H})^{-1} &= \frac{\begin{bmatrix} N & -\sum_{n=0}^{N-1} s[n] \\ -\sum_{n=0}^{N-1} s[n] & \sum_{n=0}^{N-1} s^2[n] \end{bmatrix}}{N\sum_{n=0}^{N-1} s^2[n] - \left(\sum_{n=0}^{N-1} s[n]\right)^2} \\
\mathbf{H}^T\mathbf{x} &= \begin{bmatrix} \sum_{n=0}^{N-1} x[n]s[n] \\ \sum_{n=0}^{N-1} x[n] \end{bmatrix}.
\end{aligned}
$$

We have that

$$
\begin{aligned}
[\hat{\boldsymbol{\theta}}_1]_1^2 &= \left[\frac{N\sum_{n=0}^{N-1} x[n]s[n] - \sum_{n=0}^{N-1} s[n]\sum_{n=0}^{N-1} x[n]}{N\sum_{n=0}^{N-1} s^2[n] - \left(\sum_{n=0}^{N-1} s[n]\right)^2}\right]^2 \\
\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T &= \left[(\mathbf{H}^T\mathbf{H})^{-1}\right]_{11} \\
&= \frac{N}{N\sum_{n=0}^{N-1} s^2[n] - \left(\sum_{n=0}^{N-1} s[n]\right)^2}.
\end{aligned}
$$

The test statistic then becomes

$$
\begin{aligned}
T(\mathbf{x}) &= \frac{\left(N\sum_{n=0}^{N-1} x[n]s[n] - \sum_{n=0}^{N-1} s[n]\sum_{n=0}^{N-1} x[n]\right)^2}{N\sigma^2\left(N\sum_{n=0}^{N-1} s^2[n] - \left(\sum_{n=0}^{N-1} s[n]\right)^2\right)} \\
&= \frac{\left(\sum_{n=0}^{N-1} x[n]s[n] - \bar{x}\sum_{n=0}^{N-1} s[n]\right)^2}{\sigma^2\left(\sum_{n=0}^{N-1} s^2[n] - \frac{1}{N}\left(\sum_{n=0}^{N-1} s[n]\right)^2\right)}
\end{aligned}
$$

$$
= \frac{\left(\sum_{n=0}^{N-1} (x[n]-\bar{x})s[n]\right)^2}{\sigma^2\left(\sum_{n=0}^{N-1} s^2[n] - \frac{1}{N}\left(\sum_{n=0}^{N-1} s[n]\right)^2\right)}
$$

or we decide $\mathcal{H}_1$ if

$$
\left(\sum_{n=0}^{N-1} (x[n]-\bar{x})s[n]\right)^2 > \gamma'.
$$

We obtain the intuitive result that we should subtract out the sample mean of the data before correlating and squaring. A similar result is given in the next section for sinusoidal interference suppression.

$\diamond$

In many problems we wish to test $\boldsymbol{\theta} = \mathbf{0}$ under $\mathcal{H}_0$ versus $\boldsymbol{\theta} \neq \mathbf{0}$ under $\mathcal{H}_1$. Such was the case in Example 7.3. This amounts to testing whether the signal $\mathbf{s} = \mathbf{H}\boldsymbol{\theta}$ is zero or not. Thus, with $\mathbf{A} = \mathbf{I}, \mathbf{b} = \mathbf{0}$ the GLRT reduces to

$$
T(\mathbf{x}) = \frac{\hat{\boldsymbol{\theta}}_1^T\mathbf{H}^T\mathbf{H}\hat{\boldsymbol{\theta}}_1}{\sigma^2} = \frac{\hat{\mathbf{s}}^T\hat{\mathbf{s}}}{\sigma^2}
$$

where $\hat{\mathbf{s}} = \mathbf{H}\hat{\boldsymbol{\theta}}_1$ is the MLE of the signal under $\mathcal{H}_1$. We compare the energy of the estimated signal to a threshold. Also, we can interpret the test statistic as

$$
T(\mathbf{x}) = \hat{\boldsymbol{\theta}}_1^T\mathbf{C}_{\hat{\theta}_1}^{-1}\hat{\boldsymbol{\theta}}_1
$$

since $\mathbf{C}_{\hat{\theta}_1} = \sigma^2(\mathbf{H}^T\mathbf{H})^{-1}$. This is similar to the known signal case except that $\boldsymbol{\theta}_1$ is replaced by its MLE (see Section 4.6). Still another interpretation is as an estimator-correlator (see Problem 7.24). The case of nuisance parameters is described in Problem 7.25.

7.8 Signal Processing Examples

Example 7.5 - Active Sonar/Radar Detection

In Section 7.6.4 we showed that the GLRT for the detection of a sinusoid of unknown amplitude, phase, frequency, and time of arrival in WGN was to decide $\mathcal{H}_1$ if

$$
\max_{n_0, f_0} \frac{1}{M}\left|\sum_{n=n_0}^{n_0+M-1} x[n]\exp(-j2\pi f_0 n)\right|^2 > \gamma'.
$$

In essence we "pick the peak" of the spectrogram and compare it to a threshold. We continue this example to determine the detection performance of this common processor [Knight, Pridham, and Kay 1981]

In actual practice, an FFT is used to evaluate the Fourier transform at different frequencies. Hence, if the signal length is now represented by N (to coincide with the usual notation for FFTs) we decide $\mathcal{H}_1$ if

$$\max_{n_0, k \in \{1,2,\ldots,N/2-1\}} \frac{1}{N} \left| \sum_{n=n_0}^{n_0+N-1} x[n] \exp\left(-j2\pi \frac{k}{N} n\right) \right|^2 > \gamma'$$

for N even. We omit the $k = 0$ and $k = N/2$ frequency samples since it is assumed that $0 < f_0 < 1/2$. Clearly, there may be some loss in performance if f_0 is not near a multiple of $1/N$ (called the "scalloping loss" [Harris 1978]). Similarly, in computing the spectrogram, we do not usually determine its value for all integer n_0 but instead may use nonoverlapping blocks to reduce the computation. Again, some loss occurs if the signal is not positioned entirely within the processing block. Using nonoverlapping blocks allows us to easily analyze the detection performance since then the blocks are independent. Recall that the noise is WGN, which is independent from block to block. For the more usual case of 50% overlap our results will be a reasonable approximation. We will assume for the purposes of this analysis that the signal is indeed wholely contained within one of the blocks and also that the frequency is a multiple of $1/N$. The sinusoidal frequency is then said to lie at the FFT *bin center.* With these assumptions the detection problem is

$$\mathcal{H}_0 : x_i[n] = w_i[n] \qquad \begin{array}{l} n = 0,1,\ldots,N-1 \\ i = 0,1,\ldots,I-1 \end{array}$$

$$\mathcal{H}_1 : x_i[n] = \begin{cases} w_i[n] & \begin{array}{l} n = 0,1,\ldots,N-1 \\ i = 0,1,\ldots,i_0-1 \end{array} \\ A\cos(2\pi f_0 n + \phi) + w_i[n] & \begin{array}{l} n = 0,1,\ldots,N-1 \\ i = i_0 \end{array} \\ w_i[n] & \begin{array}{l} n = 0,1,\ldots,N-1 \\ i = i_0+1, i_0+2, \ldots, I-1 \end{array} \end{cases}$$

where $x_i[n]$ denotes the ith block of N data points, $f_0 = k_0/N$ for $k_0 \in \{1, 2, \ldots, N/2-1\}$, and the $w_i[n]$'s are WGN processes that are independent for $i \neq j$. The number of blocks is I so that the total data record length is NI points. The GLRT can be written as

$$T(\mathbf{x}) = \max_{\substack{0 \le i \le I-1 \\ 1 \le k \le N/2-1}} \frac{1}{N} \left| \sum_{n=0}^{N-1} x_i[n] \exp\left(-j2\pi \frac{k}{N} n\right) \right|^2 > \gamma'$$

or

$$T(\mathbf{x}) = \max_{\substack{0 \le i \le I-1 \\ 1 \le k \le N/2-1}} \frac{1}{N} |X_i[k]|^2 > \gamma'$$

where $X_i[k]$ is the kth DFT coefficient for the ith block of data. Note that i is referred to as the ith *range bin* since range to target depends on the delay time and k is the kth *Doppler bin* since the frequency f_0 is related to the Doppler shift. The spectrogram is then referred to as a *range-Doppler map.* The overall processor is shown in Figure 7.8.

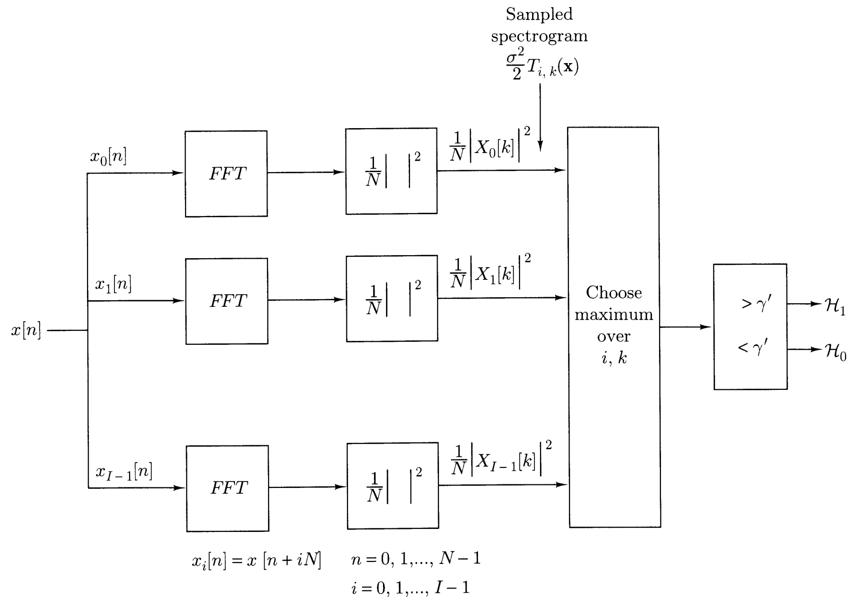


Figure 7.8. Typical active sonar/radar detector.

We first determine P_{FA}. Under $\mathcal{H}_0$

$$\begin{aligned} \sum_{n=0}^{N-1} x_i[n] \exp\left(-j2\pi \frac{k}{N} n\right) &= \sum_{n=0}^{N-1} x_i[n] \cos\left(j2\pi \frac{k}{N} n\right) \\ &\quad - j \sum_{n=0}^{N-1} x_i[n] \sin\left(j2\pi \frac{k}{N} n\right) \\ &= U_i\left(\frac{k}{N}\right) + jV_i\left(\frac{k}{N}\right). \end{aligned}$$

It has been shown, however, that for a given i the random variables $U_i(k/N)$, $V_i(k/N)$ for $k = 1, 2, \ldots, N/2-1$ are independent and Gaussian [Kay-I 1993, Example 15.3]. Also, because of the nonoverlapping block assumption the U_i's and V_i's are independent from block to block. Thus, the set of random variables $\{U_i(k/N), V_i(k/N)\}$ for $i = 0, 1, \ldots, I-1; k = 1, 2, \ldots, N/2-1$ are all independent and Gaussian. Furthermore, they are identically distributed, each with mean zero and variance $N\sigma^2/2$ [Kay-I 1993, pp. 509–511]. Consequently,

$$\begin{aligned} T_{i,k}(\mathbf{x}) &= \frac{\left|\sum_{n=0}^{N-1} x_i[n] \exp\left(-j2\pi \frac{k}{N} n\right)\right|^2}{\frac{N\sigma^2}{2}} \\ &= \left(\frac{U_i\left(\frac{k}{N}\right)}{\sqrt{N\sigma^2/2}}\right)^2 + \left(\frac{V_i\left(\frac{k}{N}\right)}{\sqrt{N\sigma^2/2}}\right)^2 \\ &\sim \chi_2^2 \end{aligned}$$

for all i and k and the $T_{i,k}(\mathbf{x})$ are all independent. The probability of false alarm follows as

$$\begin{aligned} P_{FA} &= \Pr\left\{\max_{i,k} \frac{\sigma^2}{2} T_{i,k}(\mathbf{x}) > \gamma'; \mathcal{H}_0\right\} \\ &= \Pr\left\{\max_{i,k} T_{i,k}(\mathbf{x}) > \frac{2\gamma'}{\sigma^2}; \mathcal{H}_0\right\} \\ &= 1 - \Pr\left\{\max_{i,k} T_{i,k}(\mathbf{x}) < \frac{2\gamma'}{\sigma^2}; \mathcal{H}_0\right\} \\ &= 1 - \Pr\left\{\bigcap_{i,k} T_{i,k}(\mathbf{x}) < \frac{2\gamma'}{\sigma^2}; \mathcal{H}_0\right\} \\ &= 1 - \prod_{i,k} \Pr\left\{T_{i,k}(\mathbf{x}) < \frac{2\gamma'}{\sigma^2}; \mathcal{H}_0\right\} \end{aligned}$$

since all the $T_{i,k}(\mathbf{x})$'s are independent. But

$$\begin{aligned} \Pr\left\{T_{i,k}(\mathbf{x}) < \frac{2\gamma'}{\sigma^2}; \mathcal{H}_0\right\} &= \int_0^{\frac{2\gamma'}{\sigma^2}} \frac{1}{2} \exp(-u/2) du \\ &= 1 - \exp(-\gamma'/\sigma^2). \end{aligned}$$

Hence

$$P_{FA} = 1 - \prod_{i,k} \left(1 - \exp(-\gamma'/\sigma^2)\right)$$

or finally we have

$$P_{FA} = 1 - \left(1 - \exp(-\gamma'/\sigma^2)\right)^L \tag{7.36}$$

where $L = I(N/2-1)$ is the number of Doppler and range bins examined. Note that for a small P_{FA} we must have $\exp(-\gamma'/\sigma^2) \ll 1$ so that using $(1-x)^L \approx 1 - Lx$ for $x \ll 1$

$$\begin{aligned} P_{FA} &\approx 1 - \left(1 - L\exp(-\gamma'/\sigma^2)\right) \\ &= L\exp(-\gamma'/\sigma^2) \\ &= LP_{FA}(\text{bin}) \end{aligned} \tag{7.37}$$

where $P_{FA}(\text{bin})$ is the probability of false alarm if we examine one bin. Hence, P_{FA} increases approximately linearly with the number of bins examined.

To find the probability of detection we first define a detection as a threshold crossing in the *correct* range-Doppler bin. This bin corresponds to the actual delay and frequency of the signal. Hence, P_D is defined as the probability that the maximum of the spectrogram occurs in the correct bin, i.e., at $i = i_0$, $k = k_0$, when a signal is present. With this definition we have

$$\begin{aligned} P_D &= \Pr\left\{\frac{\sigma^2}{2} T_{i_0,k_0}(\mathbf{x}) > \gamma'; \mathcal{H}_1\right\} \\ &= \Pr\left\{T_{i_0,k_0}(\mathbf{x}) > \frac{2\gamma'}{\sigma^2}; \mathcal{H}_1\right\}. \end{aligned}$$

We need to determine the PDF of $T_{i_0,k_0}(\mathbf{x})$ under $\mathcal{H}_1$. But for a given delay i_0 and Doppler bin k_0, this is just the periodogram statistic for a sinusoid of unknown amplitude and phase. We have already evaluated P_D for this case which is given by (7.27) as

$$P_D = Q_{\chi_2'^2(\lambda)}\left(\frac{2\gamma'}{\sigma^2}\right)$$

where $\lambda = NA^2/(2\sigma^2)$. Making use of (7.37) we have finally that

$$P_D = Q_{\chi_2'^2\left(\frac{NA^2}{2\sigma^2}\right)}\left(2\ln\frac{L}{P_{FA}}\right). \tag{7.38}$$

As an example, for an active sonar assume that it is desired to attain $P_{FA} = 10^{-4}$, $P_D = 0.5$ for an input SNR (or $A^2/2\sigma^2$) which may be as low as -10 dB. In practice, the input SNR will decrease as the target's range from the receiver increases. Hence, the minimum SNR corresponds to a maximum anticipated target range. For shorter ranges the increased SNR will yield a higher P_D. Assume that the maximum detection range is to be 5000 yds. Since the speed of sound in the ocean is about 5000 ft/sec, the maximum round trip delay is 6 sec. For a sampling rate of 2000 samples/sec, as might be used for a low-pass bandwidth of 1000 Hz, the data record length is $NI = 6(2000) = 12,000$ samples. We wish to determine the length of a sinusoidal transmit pulse required to attain the given detection performance. Hence, from (7.38) with $L = I(N/2-1) \approx NI/2 = 6000$ and $(A^2/2)/\sigma^2 = 0.1$ we have

$$P_D = Q_{\chi_2'^2(0.1N)}\left(2\ln\frac{6000}{10^{-4}}\right).$$

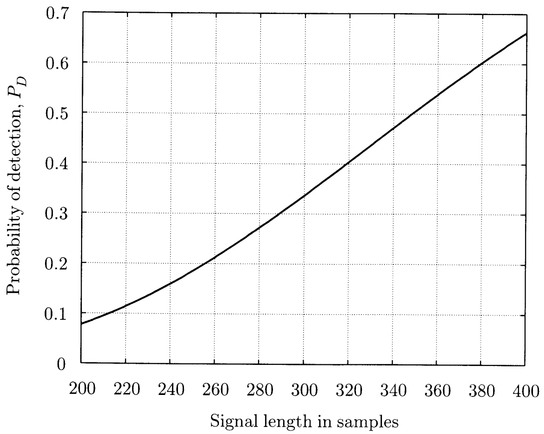


Figure 7.9. Required signal length for design example.

This is plotted in Figure 7.9 versus N. The required signal length is about $N = 350$ for a pulse length of $350/2000 = 175$ msec.

$\diamond$

Example 7.6 - Detection in Interference

In some military situations it is desired to detect a signal in the presence of an interference or jammer. A typical jammer is a narrowband or sinusoidal interference. Hence the detection problem becomes

$$\begin{aligned} \mathcal{H}_0 &: x[n] = B\cos(2\pi f_i n + \phi) + w[n] & n = 0,1,\ldots,N-1 \\ \mathcal{H}_1 &: x[n] = As[n] + B\cos(2\pi f_i n + \phi) + w[n] & n = 0,1,\ldots,N-1 \end{aligned}$$

where $w[n]$ is WGN with variance σ^2. We assume the signal $s[n]$ is known except for its amplitude A, and that the amplitude and phase of the interference are unknown

but that the frequency f_i is known. Then, using the alternative representation $B\cos(2\pi f_i n + \phi) = \alpha_1 \cos 2\pi f_i n + \alpha_2 \sin 2\pi f_i n$ as in Section 7.6.2 we have the linear model

$$\mathbf{x} = \underbrace{\begin{bmatrix} s[0] & 1 & 0 \\ s[1] & \cos 2\pi f_i & \sin 2\pi f_i \\ \vdots & \vdots & \vdots \\ s[N-1] & \cos[2\pi f_i(N-1)] & \sin[2\pi f_i(N-1)] \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} A \\ \alpha_1 \\ \alpha_2 \end{bmatrix}}_{\boldsymbol{\theta}} + \mathbf{w}.$$

The detection problem is now recast as

$$\begin{aligned} \mathcal{H}_0 &: A = 0 \\ \mathcal{H}_1 &: A \neq 0 \end{aligned}$$

or

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{A}\boldsymbol{\theta} = \mathbf{0} \\ \mathcal{H}_1 &: \mathbf{A}\boldsymbol{\theta} \neq \mathbf{0} \end{aligned}$$

where $\mathbf{A} = [1\,0\,0]$. The GLRT follows immediately from (7.35) as

$$T(\mathbf{x}) = \frac{\hat{\boldsymbol{\theta}}_1^T \mathbf{A}^T \left[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\right]^{-1} \mathbf{A}\hat{\boldsymbol{\theta}}_1}{\sigma^2} > \gamma'$$

where $\hat{\boldsymbol{\theta}}_1 = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$. But for f_i not near 0 or 1/2

$$\mathbf{H}^T\mathbf{H} \approx \begin{bmatrix} \sum_{n=0}^{N-1} s^2[n] & \sum_{n=0}^{N-1} s[n]\cos 2\pi f_i n & \sum_{n=0}^{N-1} s[n]\sin 2\pi f_i n \\ \sum_{n=0}^{N-1} s[n]\cos 2\pi f_i n & \frac{N}{2} & 0 \\ \sum_{n=0}^{N-1} s[n]\sin 2\pi f_i n & 0 & \frac{N}{2} \end{bmatrix}$$

and letting $S_c = \sum_{n=0}^{N-1} s[n]\cos 2\pi f_i n$ and $S_s = \sum_{n=0}^{N-1} s[n]\sin 2\pi f_i n$

$$(\mathbf{H}^T\mathbf{H})^{-1} = \frac{\begin{bmatrix} -\frac{N}{2} & S_c & S_s \\ S_c & -\sum_{n=0}^{N-1} s^2[n] + \frac{2}{N}S_s^2 & -\frac{2}{N}S_sS_c \\ S_s & -\frac{2}{N}S_sS_c & -\sum_{n=0}^{N-1} s^2[n] + \frac{2}{N}S_c^2 \end{bmatrix}}{S_s^2 + S_c^2 - \frac{N}{2}\sum_{n=0}^{N-1} s^2[n]}$$

so that

$$\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1} = \frac{\left[\begin{array}{ccc} -\frac{N}{2} & S_c & S_s \end{array}\right]}{S_s^2 + S_c^2 - \frac{N}{2}\sum_{n=0}^{N-1} s^2[n]}$$

$$\mathbf{H}^T\mathbf{x} = \left[\begin{array}{c} \sum_{n=0}^{N-1} x[n]s[n] \\ \sum_{n=0}^{N-1} x[n]\cos 2\pi f_i n \\ \sum_{n=0}^{N-1} x[n]\sin 2\pi f_i n \end{array}\right].$$

It follows that

$$\mathbf{A}\hat{\boldsymbol{\theta}}_1 = \mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x} =$$

$$\frac{-\frac{N}{2}\sum_{n=0}^{N-1} x[n]s[n] + S_c\left(\sum_{n=0}^{N-1} x[n]\cos 2\pi f_i n\right) + S_s\left(\sum_{n=0}^{N-1} x[n]\sin 2\pi f_i n\right)}{S_s^2 + S_c^2 - \frac{N}{2}\sum_{n=0}^{N-1} s^2[n]}$$

and also that

$$\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T = \frac{-\frac{N}{2}}{S_s^2 + S_c^2 - \frac{N}{2}\sum_{n=0}^{N-1} s^2[n]}.$$

Now we define the Fourier transforms

$$S(f_i) = \sum_{n=0}^{N-1} s[n]\exp(-j2\pi f_i n) = S_c - jS_s$$

$$X(f_i) = \sum_{n=0}^{N-1} x[n]\exp(-j2\pi f_i n)$$

which are evaluated at $f = f_i$. Then,

$$\mathbf{A}\hat{\boldsymbol{\theta}}_1 = \frac{\sum_{n=0}^{N-1} x[n]s[n] - \frac{2}{N}\mathrm{Re}(X(f_i)S^*(f_i))}{-\frac{2}{N}|S(f_i)|^2 + \sum_{n=0}^{N-1} s^2[n]}$$

and

$$\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T = \frac{-\frac{N}{2}}{|S(f_i)|^2 - \frac{N}{2}\sum_{n=0}^{N-1} s^2[n]}.$$

We have then that

$$T(\mathbf{x}) = \frac{\left[\sum_{n=0}^{N-1} x[n]s[n] - \frac{2}{N}\mathrm{Re}(X(f_i)S^*(f_i))\right]^2}{\sigma^2\left[\sum_{n=0}^{N-1} s^2[n] - \frac{2}{N}|S(f_i)|^2\right]}.$$

A better intuitive grasp of the test statistic is provided by expressing it in terms of the DFT coefficient. Assume first that $f_i = l/N$, i.e., the interference frequency, lies at the lth DFT bin center. Then, since [Oppenheim and Schafer 1975]

$$\sum_{n=0}^{N-1} x[n]s[n] = \frac{1}{N}\sum_{k=0}^{N-1} X[k]S^*[k]$$

where $X[k], S[k]$ are the DFT coefficients, we have

$$T(\mathbf{x}) = \frac{\left[\frac{1}{N}\sum_{n=0}^{N-1} X[k]S^*[k] - \frac{1}{N}X[l]S^*[l] - \frac{1}{N}X[N-l]S^*[N-l]\right]^2}{\sigma^2\left[\frac{1}{N}\sum_{k=0}^{N-1}|S[k]|^2 - \frac{1}{N}|S[l]|^2 - \frac{1}{N}|S[N-l]|^2\right]}$$

since $X[N-l] = X^*[l]$ and $S[N-l] = S^*[l]$. Finally we have the result that

$$T(\mathbf{x}) = \frac{\left(\frac{1}{N}\sum_{\substack{k=0 \\ k\neq l,N-l}}^{N-1} X[k]S^*[k]\right)^2}{\sigma^2\frac{1}{N}\sum_{\substack{k=0 \\ k\neq l,N-l}}^{N-1} |S[k]|^2} > \gamma' \tag{7.39}$$

or we decide $\mathcal{H}_1$ if

$$\left(\frac{1}{N}\sum_{\substack{k=0 \\ k\neq l,N-l}}^{N-1} X[k]S^*[k]\right)^2 > \frac{\sigma^2}{N}\gamma'\sum_{\substack{k=0 \\ k\neq l,N-l}}^{N-1} |S[k]|^2 = \gamma'' \tag{7.40}$$

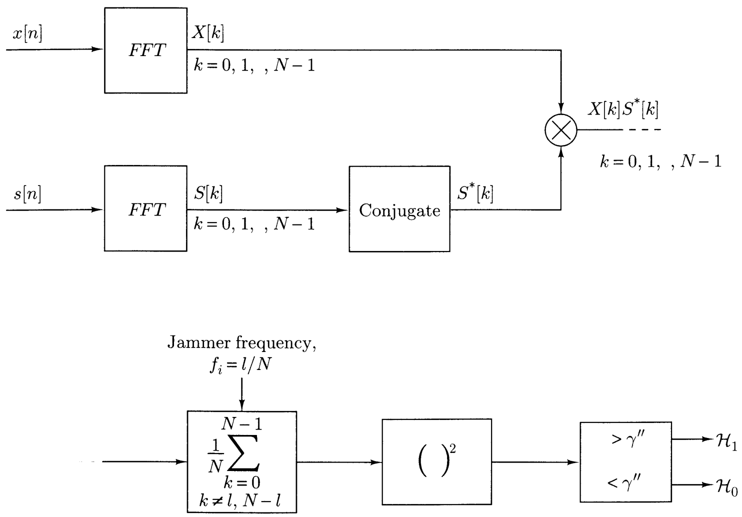


Figure 7.10. GLRT for detection of unknown amplitude signal in presence of sinusoidal jammer.

as shown in Figure 7.10. To interpret our results recall that from (7.13) for an unknown amplitude signal the GLRT is

$$T(\mathbf{x}) = \frac{\left(\sum_{n=0}^{N-1} x[n]s[n]\right)^2}{\sigma^2 \sum_{n=0}^{N-1} s^2[n]} > 2\ln\gamma.$$

This can also be written as

$$T(\mathbf{x}) = \frac{\left(\frac{1}{N}\sum_{k=0}^{N-1} X[k]S^*[k]\right)^2}{\sigma^2 \frac{1}{N}\sum_{k=0}^{N-1} |S[k]|^2} > 2\ln\gamma.$$

In the presence of a sinusoidal interference the GLRT implements the same detector as for the unknown amplitude signal case except that it *"blanks out" the DFT bins that contain the interference.* Also, because of the orthogonality of the DFT sinusoids, the interference will not be present in the remaining bins (see Problem 7.26). In practice, because of leakage, data windowing is used to reduce the spillage of the interference into adjacent bins [Kay 1988].

◇

References

Harris, F.J., "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform," *Proc. IEEE*, Vol. 66, pp. 51–83, Jan. 1978.

Kay, S.M., *Modern Spectral Estimation: Theory and Application*, Prentice-Hall, Englewood Cliffs, N.J., 1988.

Knight, W.S., R.G. Pridham, S.M. Kay, "Digital Signal Processing for Sonar," *Proc. IEEE*, Vol. 69, pp. 1451–1506, Nov. 1981.

Lehmann, E.L., *Testing Statistical Hypotheses*, J. Wiley, New York, 1959.

Oppenheim, A.V., R.W. Schafer, *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, N.J., 1975.

Van Trees, H.L., *Detection, Estimation, and Modulation Theory*, Vol. 3, McGraw-Hill, New York, 1971.

Problems

7.1 Consider the detection problem

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = A + w[n] & n = 0, 1, \ldots, N-1 \end{array}$$

where $w[n]$ is WGN with variance σ^2. The DC level is either $A = 1$ or $A = -1$ and is to be modeled as an unknown deterministic constant. Does a UMP test exist? If not, find the GLRT statistic. Hint: You should be able to show that $\hat{A} = \operatorname{sgn}(\bar{x})$, where $\operatorname{sgn}(x) = 1$ for $x > 0$ and $\operatorname{sgn}(x) = -1$ for $x < 0$.

7.2 For the detection problem described in Problem 7.1 use a Bayesian approach. Assume that A is a random variable with equal probabilities of being either 1 or -1 and is independent of $w[n]$. Derive the optimal NP test statistic. Hint: The inequality $\exp(u) + \exp(-u) > \gamma$ is equivalent to $|u| > \gamma'$.

7.3 Consider the detection problem

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = r^n + w[n] & n = 0, 1, \ldots, N-1 \end{array}$$

where $0 < r < 1$ but is otherwise unknown and $w[n]$ is WGN with variance σ^2. Find the GLRT statistic. Hint: The MLE of r cannot be found in closed form [Kay-I 1993, pg. 178].

7.4 For the detection problem described in Problem 7.1, now assume that A is an unknown deterministic constant that satisfies $-A_0 \le A \le A_0$. The MLE of A can be shown to be

$$\hat{A} = \begin{cases} -A_0 & \text{if } \bar{x} < -A_0 \\ \bar{x} & \text{if } -A_0 \le \bar{x} \le A_0 \\ A_0 & \text{if } \bar{x} > A_0. \end{cases}$$

Find the GLRT statistic. What happens as $A_0 \to \infty$?

7.5 For the detection problem described in Problem 7.1 now assume that A is random with $A \sim \mathcal{U}[-A_0, A_0]$ and is independent of $w[n]$. Find the optimal NP test statistic. What happens as $A_0 \to \infty$? Hint: The integral needed to find $p(\mathbf{x}; \mathcal{H}_1)$ can only be found in closed-form as $A_0 \to \infty$.

7.6 Consider the detection problem

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = s[n] + w[n] & n = 0, 1, \ldots, N-1 \end{array}$$

where $w[n]$ is WGN with variance σ^2. For a completely unknown deterministic signal $s[n]$, we use a Bayesian approach and assume that the prior PDF is approximately uniform. The approximate uniform PDF assumption can be obtained by assuming $\mathbf{s} \sim \mathcal{N}(\mathbf{0}, \sigma_s^2\mathbf{I})$ if σ_s^2 is large. Also, assume $s[n]$ is independent of $w[n]$. Find the NP test statistic and interpret the results. Hint: See Example 5.1.

7.7 For the detection problem

$$\begin{array}{l} \mathcal{H}_0 : x[n] = w[0] \\ \mathcal{H}_1 : x[n] = A + w[0] \end{array}$$

where $w[0] \sim \mathcal{N}(0, \sigma^2)$. Assume that A is deterministic and unknown and show that we decide $\mathcal{H}_1$ if $|x[0]| > \gamma'$ using a GLRT. Then, assume that A is a random variable with $A \sim \mathcal{N}(0, \sigma_A^2)$ and independent of $w[0]$, and show that the NP detector is identical (with the same threshold). Finally, if P_D for the GLRT is denoted by $P_D(A)$ and that for the NP detector denoted by P_D, explain why

$$P_D = \int_{-\infty}^{\infty} P_D(A) \frac{1}{\sqrt{2\pi\sigma_A^2}} \exp\left(-\frac{1}{2\sigma_A^2} A^2\right) dA$$

for the NP detector.

7.8 For a DC level in WGN or $x[n] = A + w[n]$ for $n = 0, 1, \ldots, N-1$, consider the following averagers

$$\begin{array}{lll} T_1(\mathbf{x}) = \left(\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right)^2 & & \text{coherent} \\ T_2(\mathbf{x}) = \frac{1}{N}\sum_{n=0}^{N-1} x^2[n] & & \text{incoherent.} \end{array}$$

Each one attempts to determine A^2, the signal power, by reducing the noise effects as much as possible. Show that the coherent averager is more accurate by computing the mean square error. This is defined as

$$\text{mse} = E\left[\left(T(\mathbf{x}) - A^2\right)^2\right].$$

Hint: The following results will be useful:

$$\text{mse} = \text{var}(T(\mathbf{x})) + \left[E(T(\mathbf{x})) - A^2\right]^2$$

[Kay-I 1993, pg. 19]. If $\xi \sim \mathcal{N}(\mu, \sigma^2)$, then $\text{var}(\xi^2) = 4\mu^2\sigma^2 + 2\sigma^4$.

7.9 For a DC level in WGN with $N = 100$, compute the required input SNRs for a matched filter detector (η_{MF}) and an energy detector (η_{ED}) to attain $P_{FA} = 10^{-4}$ and $P_D = 0.99$. What is the loss in dB of the energy detector?

7.10 For the detection problem

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = As[n] + w[n] & n = 0, 1, \ldots, N-1 \end{array}$$

where $s[n]$ is known, A is unknown with $A > 0$, and $w[n]$ is WGN with variance σ^2, show that a UMP test exists. Find the detection performance.

7.11 If we observe the data $x[n] = As[n] + w[n]$ for $n = 0, 1, \ldots, N-1$, where $s[n]$ is known and $w[n]$ is WGN with variance σ^2, verify that the MLE of A is given by (7.11).

7.12 In the test statistic given by (7.18) assume that $\mu_A = 0$. To find the detection performance first note that

$$\begin{aligned} P_D &= \Pr\{T'(\mathbf{x}) > \gamma'; \mathcal{H}_1\} \\ &= \Pr\left\{|\mathbf{x}^T\mathbf{s}| > \gamma''; \mathcal{H}_1\right\} \end{aligned}$$

and that $\mathbf{x}^T\mathbf{s}$ is a Gaussian random variable. Determine P_D and then in a similar fashion determine P_{FA}.

7.13 Define a random variable as $z = \max(x, y)$, where $[x\, y]^T \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$. Find the CDF of z if $\mathbf{C} = \mathbf{I}$ and hence the PDF. What happens if $\mathbf{C}$ is an arbitrary covariance matrix?

7.14 Show that (7.21) can be written as (7.22) by using Parseval's theorem.

7.15 Verify (7.23) by extending the result of (7.14), which assumed that n_0 was known and that $n_0 = 0$.

7.16 Consider the detection problem

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = As[n; \boldsymbol{\theta}] + w[n] & n = 0, 1, \ldots, N-1 \end{array}$$

where A is unknown, the signal $s[n; \boldsymbol{\theta}]$ depends upon a $p \times 1$ vector of unknown parameters, and $w[n]$ is WGN with variance σ^2. Find the GLRT statistic if $\sum_{n=0}^{N-1} s^2[n; \boldsymbol{\theta}]$ does not depend on $\boldsymbol{\theta}$. As an example of the latter, let $s[n; f_0] = A \cos 2\pi f_0 n$ for $0 < f_0 < 1/2$ and large N.

7.17 Use Theorem 7.1 to find the detection performance of $I(f_0)$ for the case of a sinusoid of unknown amplitude and phase in WGN as discussed in Section 7.6.2. Assume that $\mathbf{H}^T\mathbf{H} \approx (N/2)\mathbf{I}$. The result is given by (7.28).

7.18 Find the means, variances, and covariance for the random variables

$$\begin{aligned} \xi_1 &= \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n \\ \xi_2 &= \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x[n] \sin 2\pi f_0 n. \end{aligned}$$

The data are $x[n] = A\cos(2\pi f_0 n + \phi) + w[n]$, where $w[n]$ is WGN with variance σ^2. Assume that f_0 is not near 0 or 1/2 so that any "double-frequency" term can be approximated as zero.

7.19 Define a random variable as $z = \sqrt{x^2 + y^2}$, where $x \sim \mathcal{N}(\mu_x, \sigma^2)$, $y \sim \mathcal{N}(\mu_y, \sigma^2)$, and x and y are independent. Show that the PDF of z is the *Rician* PDF given by

$$p(z) = \begin{cases} \frac{z}{\sigma^2} \exp\left[-\frac{1}{2\sigma^2}(z^2 + \alpha^2)\right] I_0\left(\frac{\alpha z}{\sigma^2}\right) & z > 0 \\ 0 & z < 0 \end{cases}$$

where $\alpha^2 = \mu_x^2 + \mu_y^2$ and $I_0(x)$ is the *modified Bessel function of order zero.* This function is defined as

$$I_0(x) = \int_0^{2\pi} \exp(x \cos\theta) \frac{d\theta}{2\pi}.$$

What PDF results if $\mu_x = \mu_y = 0$? Hint: Transform (x, y) into polar coordinates (r, θ) and use the identity

$$\mu_x \cos\theta + \mu_y \sin\theta = \sqrt{\mu_x^2 + \mu_y^2} \cos(\theta - \psi)$$

where $\psi = \arctan(\mu_y/\mu_x)$. Then, find the PDF of $r = z$ by integrating the joint PDF over θ.

7.20 In this problem we express (7.27) in terms of the Marcum Q function, which is the right-tail probability for a Rician PDF with $\sigma^2 = 1$ (see Problem 7.19). The Marcum Q function is defined as

$$Q_M(\alpha, \gamma) = \int_\gamma^\infty z \exp\left[-\frac{1}{2}(z^2 + \alpha^2)\right] I_0(\alpha z)\, dz.$$

To do so note that from (7.27)

$$\begin{aligned} P_D &= \Pr\left\{\chi_2'^2(\lambda) > \gamma''\right\} \\ &= \Pr\left\{\sqrt{\chi_2'^2(\lambda)} > \sqrt{\gamma''}\right\} \end{aligned}$$

where $\gamma'' = 2\gamma'/\sigma^2$. Then find the PDF of $\sqrt{\chi_2'^2(\lambda)}$ by recalling that if $x \sim \mathcal{N}(\mu_x, 1)$, $y \sim \mathcal{N}(\mu_y, 1)$ with x, y independent, then $x^2 + y^2 \sim \chi_2'^2(\lambda)$, where $\lambda = \mu_x^2 + \mu_y^2$ and use the results of Problem 7.19. You should be able to show that

$$P_D = Q_M\left(\sqrt{\frac{NA^2}{2\sigma^2}}, \sqrt{\frac{2\gamma'}{\sigma^2}}\right).$$

7.21 Prove that the usual GLRT definition for a signal with unknown parameter vector $\boldsymbol{\theta}$ in WGN, which is to decide $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x}; \hat{\boldsymbol{\theta}}, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > \gamma$$

is equivalent to

$$\max_{\boldsymbol{\theta}} \ln L(\mathbf{x}; \boldsymbol{\theta}) > \ln \gamma$$

where

$$L(\mathbf{x}; \boldsymbol{\theta}) = \frac{p(\mathbf{x}; \boldsymbol{\theta}, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)}$$

is the LR.

7.22 In this problem we derive the NP detector for a *randomly phased* sinusoid in WGN. The detection problem is defined as

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = A\cos(2\pi f_0 n + \phi) + w[n] & n = 0, 1, \ldots, N-1 \end{array}$$

where A is an unknown but deterministic amplitude (with $A > 0$), f_0 is known, ϕ is unknown and is modeled as a uniformly distributed random variable or $\phi \sim \mathcal{U}[0, 2\pi]$, and $w[n]$ is WGN with variance σ^2 and is independent of ϕ. The resultant detector is UMP with respect to the unknown A. The steps in the derivation are as follows. First show that the LR is

$$L(\mathbf{x}) =$$

$$\int_0^{2\pi} \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{n=0}^{N-1} -2Ax[n]\cos(2\pi f_0 n+\phi) + \sum_{n=0}^{N-1} A^2\cos^2(2\pi f_0 n+\phi)\right)\right]\frac{d\phi}{2\pi}.$$

Next use the double-frequency approximation to yield

$$L(\mathbf{x}) = \exp\left(-\frac{NA^2}{4\sigma^2}\right)\int_0^{2\pi}\exp\left[\frac{A}{\sigma^2}\sum_{n=0}^{N-1} x[n]\cos(2\pi f_0 n + \phi)\right]\frac{d\phi}{2\pi}.$$

Then letting $\beta_1 = \sum_{n=0}^{N-1} x[n]\cos 2\pi f_0 n$, $\beta_2 = -\sum_{n=0}^{N-1} x[n]\sin 2\pi f_0 n$, show that

$$\begin{aligned} \sum_{n=0}^{N-1} x[n]\cos(2\pi f_0 n + \phi) &= \beta_1\cos\phi + \beta_2\sin\phi \\ &= \sqrt{\beta_1^2 + \beta_2^2}\cos(\phi - \psi) \end{aligned}$$

where $\psi = \arctan(\beta_2/\beta_1)$ so that

$$\begin{aligned} L(\mathbf{x}) &= \exp\left(-\frac{NA^2}{4\sigma^2}\right)\int_0^{2\pi}\exp\left[\sqrt{\frac{NA^2}{\sigma^4}I(f_0)}\cos(\phi-\psi)\right]\frac{d\phi}{2\pi} \\ &= \exp\left(-\frac{NA^2}{4\sigma^2}\right) I_0\left(\sqrt{\frac{NA^2}{\sigma^4}I(f_0)}\right). \end{aligned}$$

The function $I_0(x)$ is defined in Problem 7.19. In the last step, note that the integrand is periodic in ϕ. Finally, show that $I_0(x)$ is monotonically increasing with x so that the test $L(\mathbf{x}) > \gamma$ is equivalent to $I(f_0) > \gamma'$. As asserted, the test statistic does not depend on A and the threshold, which is only dependent on the PDF of $I(f_0)$ under $\mathcal{H}_0$, is also independent of A.

7.23 It is desired to detect a trend in stock market data. To do so we assume that the data are modeled as

$$x[n] = A + Bn + w[n] \qquad n = 0, 1, \ldots, N-1$$

where $w[n]$ is WGN with variance σ^2. The average stock price A is unknown but is of no interest to us. More importantly, we wish to test whether $B = 0$ or $B \neq 0$, i.e., that a trend is present. Find the GLRT statistic for this problem. What is $T(\mathbf{x})$ if $w[n] = 0$? Hint: Use Theorem 7.1.

7.24 Show that to test $\boldsymbol{\theta} = \mathbf{0}$ versus $\boldsymbol{\theta} \neq \mathbf{0}$ for the classical linear model the GLRT can be written as

$$T(\mathbf{x}) = \frac{\mathbf{x}^T\hat{\mathbf{s}}}{\sigma^2} > \gamma'$$

where $\hat{\mathbf{s}} = \mathbf{H}\hat{\boldsymbol{\theta}}_1$.

7.25 In the classical linear model we let $\boldsymbol{\theta} = [\boldsymbol{\theta}_r^T\ \boldsymbol{\theta}_s^T]^T$, where $\boldsymbol{\theta}_r$ is $r \times 1$ and $\boldsymbol{\theta}_s$ is $s \times 1$. Show that the GLRT for $\boldsymbol{\theta}_r = \mathbf{0}$ versus $\boldsymbol{\theta}_r \neq \mathbf{0}$ ($\boldsymbol{\theta}_s$ are nuisance parameters) is

$$T(\mathbf{x}) = \hat{\boldsymbol{\theta}}_r^T\mathbf{C}_{\hat{\theta}_r}^{-1}\hat{\boldsymbol{\theta}}_r > \gamma'$$

where

$$\hat{\boldsymbol{\theta}}_1 = \begin{bmatrix} \hat{\boldsymbol{\theta}}_r \\ \hat{\boldsymbol{\theta}}_s \end{bmatrix} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$$

and

$$\mathbf{C}_{\hat{\theta}_r} = \sigma^2\left[(\mathbf{H}^T\mathbf{H})^{-1}\right]_{rr}$$

by using Theorem 7.1. Note that $\mathbf{C}_{\hat{\theta}_r}$ is the covariance matrix of $\hat{\boldsymbol{\theta}}_r$ (since $\mathbf{C}_{\hat{\theta}} = \sigma^2(\mathbf{H}^T\mathbf{H})^{-1}$ [Kay-I 1993, pg. 86]) which is obtained by the partitioning

$$(\mathbf{H}^T\mathbf{H})^{-1} = \begin{bmatrix} [(\mathbf{H}^T\mathbf{H})^{-1}]_{rr} & [(\mathbf{H}^T\mathbf{H})^{-1}]_{rs} \\ [(\mathbf{H}^T\mathbf{H})^{-1}]_{sr} & [(\mathbf{H}^T\mathbf{H})^{-1}]_{ss} \end{bmatrix} = \begin{bmatrix} r \times r & r \times s \\ s \times r & s \times s \end{bmatrix}.$$

Finally, show that

$$\lambda = \frac{\boldsymbol{\theta}_r^T\left([(\mathbf{H}^T\mathbf{H})^{-1}]_{rr}\right)^{-1}\boldsymbol{\theta}_r}{\sigma^2}.$$

7.26 For the sinusoidal interference $i[n] = B\cos(2\pi f_i n + \phi)$, where $f_i = l/N$, show that

$$I[k] = 0 \qquad \text{for all } k \neq l, N-l$$

where $I[k]$ is the N-point DFT of the interference. Hence, there is no "leakage" of the interference into the other DFT bins. Hint: Use the identity

$$\sum_{n=0}^{N-1} \exp\left(j2\pi \frac{k}{N} n\right) = 0 \qquad \text{for } k \neq rN$$

where r is an integer [Oppenheim and Schafer 1975].

Appendix 7A

Asymptotic Performance of the Energy Detector

The energy detector decides $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} x^2[n] > \gamma'$$

where $x[n] = w[n]$ under $\mathcal{H}_0$, $x[n] = s[n] + w[n]$ under $\mathcal{H}_1$, and $w[n]$ is WGN. If N is large, then $T(\mathbf{x})$ can be approximated by a Gaussian random variable since it is the sum of N independent, although not identically distributed random variables (unless $s[n] = A$). Thus, we need only find the first two moments to characterize the detection performance. To do so note that (see Chapter 2)

$$T'(\mathbf{x}) = \frac{T(\mathbf{x})}{\sigma^2} = \begin{cases} \chi_N^2 & \text{under} \mathcal{H}_0 \\ \chi_N'^2(\lambda) & \text{under } \mathcal{H}_1 \end{cases}$$

where $\lambda = \sum_{n=0}^{N-1} s^2[n]/\sigma^2 = \mathcal{E}/\sigma^2$. This is because under $\mathcal{H}_1$

$$T'(\mathbf{x}) = \sum_{n=0}^{N-1} \left(\frac{s[n] + w[n]}{\sigma}\right)^2$$

and hence the mean of $x[n]/\sigma$ is $s[n]/\sigma$. Using the properties of chi-squared random variables we have

$$\begin{aligned} E(T'(\mathbf{x}); \mathcal{H}_0) &= N \\ E(T'(\mathbf{x}); \mathcal{H}_1) &= \lambda + N \\ \text{var}(T'(\mathbf{x}); \mathcal{H}_0) &= 2N \\ \text{var}(T'(\mathbf{x}); \mathcal{H}_1) &= 4\lambda + 2N. \end{aligned}$$

Now

$$
\begin{aligned}
P_{FA} &= Q\left(\frac{\gamma'/\sigma^2 - N}{\sqrt{2N}}\right) \\
P_D &= Q\left(\frac{\gamma'/\sigma^2 - \lambda - N}{\sqrt{4\lambda + 2N}}\right)
\end{aligned}
$$

so that

$$
\begin{aligned}
P_D &= Q\left(\frac{\sqrt{2N}Q^{-1}(P_{FA}) - \lambda}{\sqrt{4\lambda + 2N}}\right) \\
&= Q\left(\frac{Q^{-1}(P_{FA}) - \sqrt{\frac{N}{2}\frac{\lambda}{N}}}{\sqrt{1 + 2\frac{\lambda}{N}}}\right).
\end{aligned}
$$

If $\lambda/N \ll 1$, then by expanding the argument $g(x)$ of Q about $x = \lambda/N = 0$ in a first-order Taylor expansion, we have

$$
\begin{aligned}
g(x) &= \frac{Q^{-1}(P_{FA}) - \sqrt{\frac{N}{2}}x}{\sqrt{1+2x}} \\
&\approx Q^{-1}(P_{FA}) - \left(\sqrt{\frac{N}{2}} + Q^{-1}(P_{FA})\right)x \\
&\approx Q^{-1}(P_{FA}) - \sqrt{\frac{N}{2}}x.
\end{aligned}
$$

The last approximation is valid for large N. Finally, we have

$$
\begin{aligned}
P_D &\approx Q\left(Q^{-1}(P_{FA}) - \sqrt{\frac{N}{2}\frac{\lambda}{N}}\right) \\
&= Q\left(Q^{-1}(P_{FA}) - \sqrt{\frac{\lambda^2}{2N}}\right)
\end{aligned}
$$

which the reader will recognize as the performance of the NP detector for the mean-shifted Gauss-Gauss problem with deflection coefficient

$$d^2 = \frac{\lambda^2}{2N}$$

or since $\lambda = \mathcal{E}/\sigma^2$, we have

$$d^2 = \frac{(\mathcal{E}/\sigma^2)^2}{2N}.$$

Appendix 7B

Derivation of GLRT for Classical Linear Model

The GLRT decides $\mathcal{H}_1$ if

$$L_G(\mathbf{x}) = \frac{p(\mathbf{x}; \hat{\boldsymbol{\theta}}_1)}{p(\mathbf{x}; \hat{\boldsymbol{\theta}}_0)} > \gamma$$

where

$$p(\mathbf{x}; \boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})\right]$$

and $\hat{\boldsymbol{\theta}}_i$ is the MLE of $\boldsymbol{\theta}$ under $\mathcal{H}_i$. Under $\mathcal{H}_1$ there are no constraints on $\boldsymbol{\theta}$ other than that we must exclude the set of $\boldsymbol{\theta}$ for which $\mathbf{A}\boldsymbol{\theta} = \mathbf{b}$. However, the unconstrained MLE of $\boldsymbol{\theta}$ or $\hat{\boldsymbol{\theta}}$ (where we allow $\boldsymbol{\theta}$ to take on any value in R^p) has zero probability of satisfying $\mathbf{A}\hat{\boldsymbol{\theta}} = \mathbf{b}$. The reader may wish to consider for the unknown signal amplitude case the probability that $\hat{A} = \sum_{n=0}^{N-1} x[n]s[n] / \sum_{n=0}^{N-1} s^2[n]$ satisfies $\hat{A} = 0$. Hence, $\hat{\boldsymbol{\theta}}_1$ is equivalent to the *unconstrained MLE* $\hat{\boldsymbol{\theta}}$, which for the linear model is

$$\hat{\boldsymbol{\theta}}_1 = \hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}.$$

To find $\hat{\boldsymbol{\theta}}_0$ we must find the *constrained MLE* or the MLE of $\boldsymbol{\theta}$ for which $\mathbf{A}\boldsymbol{\theta} = \mathbf{b}$. This is equivalent to the constrained LS estimator (since $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$), which has been shown to be [Kay-I 1993, pp. 251–252]

$$\hat{\boldsymbol{\theta}}_0 = \hat{\boldsymbol{\theta}}_1 - \underbrace{(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\left[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\right]^{-1}(\mathbf{A}\hat{\boldsymbol{\theta}}_1 - \mathbf{b})}_{\mathbf{d}}.$$

Note that $\mathbf{A}\hat{\boldsymbol{\theta}}_0 = \mathbf{b}$ as required. Thus,

$$\ln L_G(\mathbf{x})$$

$$
\begin{aligned}
&= -\frac{1}{2\sigma^2}\left[(\mathbf{x}-\mathbf{H}\hat{\boldsymbol{\theta}}_1)^T(\mathbf{x}-\mathbf{H}\hat{\boldsymbol{\theta}}_1) - (\mathbf{x}-\mathbf{H}\hat{\boldsymbol{\theta}}_0)^T(\mathbf{x}-\mathbf{H}\hat{\boldsymbol{\theta}}_0)\right] \\
&= -\frac{1}{2\sigma^2}\left[(\mathbf{x}-\mathbf{H}\hat{\boldsymbol{\theta}}_1)^T(\mathbf{x}-\mathbf{H}\hat{\boldsymbol{\theta}}_1) - [(\mathbf{x}-\mathbf{H}\hat{\boldsymbol{\theta}}_1)+\mathbf{H}\mathbf{d}]^T[(\mathbf{x}-\mathbf{H}\hat{\boldsymbol{\theta}}_1)+\mathbf{H}\mathbf{d}]\right] \\
&= -\frac{1}{2\sigma^2}\left[-(\mathbf{x}-\mathbf{H}\hat{\boldsymbol{\theta}}_1)^T\mathbf{H}\mathbf{d} - \mathbf{d}^T\mathbf{H}^T(\mathbf{x}-\mathbf{H}\hat{\boldsymbol{\theta}}_1) - \mathbf{d}^T\mathbf{H}^T\mathbf{H}\mathbf{d}\right].
\end{aligned}
$$

But $(\mathbf{x}-\mathbf{H}\hat{\boldsymbol{\theta}}_1)^T\mathbf{H} = \mathbf{0}$ so that

$$
\begin{aligned}
\ln L_G(\mathbf{x}) &= \frac{1}{2\sigma^2}\mathbf{d}^T\mathbf{H}^T\mathbf{H}\mathbf{d} \\
&= \frac{1}{2\sigma^2}(\mathbf{A}\hat{\boldsymbol{\theta}}_1-\mathbf{b})^T\left[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\right]^{-1}\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{H} \\
&\quad\cdot(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\left[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\right]^{-1}(\mathbf{A}\hat{\boldsymbol{\theta}}_1-\mathbf{b}) \\
&= \frac{(\mathbf{A}\hat{\boldsymbol{\theta}}_1-\mathbf{b})^T\left[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\right]^{-1}(\mathbf{A}\hat{\boldsymbol{\theta}}_1-\mathbf{b})}{2\sigma^2}
\end{aligned}
$$

or

$$
2\ln L_G(\mathbf{x}) = \frac{(\mathbf{A}\hat{\boldsymbol{\theta}}_1-\mathbf{b})^T\left[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\right]^{-1}(\mathbf{A}\hat{\boldsymbol{\theta}}_1-\mathbf{b})}{\sigma^2}
$$

which is (7.35). To determine the detection performance we first note that [Kay-I 1993, pp. 85–86]

$$
\hat{\boldsymbol{\theta}}_1 \sim \mathcal{N}(\boldsymbol{\theta}, \sigma^2(\mathbf{H}^T\mathbf{H})^{-1})
$$

so that

$$
\begin{aligned}
\mathbf{A}\hat{\boldsymbol{\theta}}_1-\mathbf{b} &\sim \mathcal{N}(\mathbf{A}\boldsymbol{\theta}-\mathbf{b}, \sigma^2\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T) \\
&\sim \begin{cases} \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(\mathbf{A}\boldsymbol{\theta}_1-\mathbf{b}, \sigma^2\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T) & \text{under } \mathcal{H}_1. \end{cases}
\end{aligned}
$$

It has been shown (see Chapter 2) that if $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$, where $\mathbf{x}$ is $r \times 1$, then

$$
\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x} \sim \chi_r'^2(\lambda)
$$

where $\lambda = \boldsymbol{\mu}^T\mathbf{C}^{-1}\boldsymbol{\mu}$ or

$$
\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x} \sim \begin{cases} \chi_r^2 & \text{for } \boldsymbol{\mu} = \mathbf{0} \\ \chi_r'^2(\boldsymbol{\mu}^T\mathbf{C}^{-1}\boldsymbol{\mu}) & \text{for } \boldsymbol{\mu} \neq \mathbf{0}. \end{cases}
$$

As a result, we have upon letting $\mathbf{x} = \mathbf{A}\hat{\boldsymbol{\theta}}_1 - \mathbf{b}$, $\boldsymbol{\mu} = \mathbf{0}$ under $\mathcal{H}_0$, $\boldsymbol{\mu} = \mathbf{A}\boldsymbol{\theta}_1 - \mathbf{b}$ under $\mathcal{H}_1$, and $\mathbf{C} = \sigma^2\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T$

$$
2\ln L_G(\mathbf{x}) \sim \begin{cases} \chi_r^2 & \text{under } \mathcal{H}_0 \\ \chi_r'^2(\lambda) & \text{under } \mathcal{H}_1 \end{cases}
$$

where

$$
\lambda = \frac{(\mathbf{A}\boldsymbol{\theta}_1-\mathbf{b})^T\left[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\right]^{-1}(\mathbf{A}\boldsymbol{\theta}_1-\mathbf{b})}{\sigma^2}.
$$

Hence

$$
\begin{aligned}
P_{FA} &= Q_{\chi_r^2}(\gamma') \\
P_D &= Q_{\chi_r'^2(\lambda)}(\gamma')
\end{aligned}
$$

which completes the proof of Theorem 7.1.

Chapter 8

Random Signals with Unknown Parameters

8.1 Introduction

In this chapter we discuss the detection of a Gaussian random signal with unknown parameters in white Gaussian noise. Although we restrict our discussion to *white noise*, the colored noise problem is a simple extension. One need only prewhiten the data, assuming the noise covariance matrix is known, to reduce the problem to that of a Gaussian random signal in *white* Gaussian noise. Since we generally assume the signal is zero mean, it is the *signal covariance matrix* that is not completely known. As we will see, this creates mathematical difficulties in determining the MLE, which is required by the GLRT. We illustrate this and then proceed to analyze some cases of practical interest that are mathematically tractable. The use of a large data record approximation to the PDF for WSS random signals is shown to reduce the complexity of the GLRT implementation. Finally, the locally most powerful detector is applied to the detection of a weak signal of unknown power.

8.2 Summary

We first attempt to determine the GLRT for a signal with a covariance matrix known except for scale. The MLE is found by minimizing (8.5), which unfortunately cannot be done in closed form. For the special case of a white signal, the MLE is given by (8.7) and the GLRT by (8.9). Another case of interest is the rank one signal covariance matrix for which the MLE is given by (8.12) and the GLRT by (8.13). In general, the GLRT statistic is given by (8.14). For large data records and WSS signals, some approximations can be made to simplify the MLE determination. The function to be minimized is (8.16). Some examples for the detection of a signal with unknown power and a signal with unknown center frequency are given in Section

8.4. If the signal can be assumed to be weak, which is usually the case of interest, then the locally most powerful detector can be employed. For a signal with unknown power the detector is given by (8.25). Finally, a practical example of the detection of periodic random signals with unknown PSD is discussed in Section 8.6. The comb filter detector (8.31) is derived and its performance illustrated. When the period of the signal is unknown, some modifications to the GLRT are required and are shown to alleviate the inconsistency of the GLRT.

8.3 Incompletely Known Signal Covariance

In general, the Bayesian approach to this problem proves to be mathematically intractable due to the required integration. The GLRT, on the other hand, is more analytically manageable. As a last resort, the required MLE for the GLRT can be found numerically. Therefore, we will concentrate on the GLRT approach. The general problem is to detect a Gaussian random signal $s[n]$ in WGN or

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = w[n] & n = 0,1,\ldots,N-1 \\ \mathcal{H}_1 : x[n] = s[n] + w[n] & n = 0,1,\ldots,N-1 \end{array} \tag{8.1}$$

where $s[n]$ is a Gaussian random process with zero mean and covariance matrix $\mathbf{C}_s$ and $w[n]$ is WGN with variance σ^2. The signal covariance matrix depends upon some parameters which are unknown. As an example, we might not know the power of the signal. If $s[n]$ is WSS, then this is equivalent to not knowing $r_{ss}[0] = E(s^2[n])$. For $N = 2$ the signal covariance matrix is

$$\begin{aligned} \mathbf{C}_s &= \begin{bmatrix} r_{ss}[0] & r_{ss}[1] \\ r_{ss}[1] & r_{ss}[0] \end{bmatrix} \\ &= r_{ss}[0] \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \end{aligned} \tag{8.2}$$

where $\rho = r_{ss}[1]/r_{ss}[0]$ is the correlation coefficient and is assumed known. Hence, the PDF under $\mathcal{H}_1$ is

$$p(\mathbf{x}; r_{ss}[0], \mathcal{H}_1) = \frac{1}{2\pi \det^{\frac{1}{2}}(\mathbf{C}_s + \sigma^2\mathbf{I})} \exp\left[-\frac{1}{2}\mathbf{x}^T(\mathbf{C}_s + \sigma^2\mathbf{I})^{-1}\mathbf{x}\right].$$

To implement a GLRT we must maximize $p(\mathbf{x}; r_{ss}[0], \mathcal{H}_1)$ over $r_{ss}[0]$ to find the MLE (see Problem 8.1). Because of the need to analytically evaluate the determinant and inverse of a covariance matrix, this may be difficult. In general, when the signal covariance matrix depends upon several unknown parameters, the task is even more demanding. We generalize this example next.

Example 8.1 - Unknown Signal Power

Consider the detection problem of (8.1) where the Gaussian random signal is zero mean and has covariance $\mathbf{C}_s = P_0\mathbf{C}$ with P_0 unknown and $\mathbf{C}$ known. The previous example is a special case, in which $P_0 = r_{ss}[0]$ and

$$\mathbf{C} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

Hence, we know the signal covariance matrix to within a scale factor. The PDF under $\mathcal{H}_1$ is

$$p(\mathbf{x}; P_0, \mathcal{H}_1) = \frac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}(P_0\mathbf{C} + \sigma^2\mathbf{I})} \exp\left[-\frac{1}{2}\mathbf{x}^T(P_0\mathbf{C} + \sigma^2\mathbf{I})^{-1}\mathbf{x}\right]. \tag{8.3}$$

To implement a GLRT requires us to determine the MLE of P_0. The determinant and inverse of the covariance matrix is easily found if $P_0\mathbf{C}+\sigma^2\mathbf{I}$ is first diagonalized. One way to accomplish this is to utilize an eigendecomposition or $\mathbf{V}^T\mathbf{C}\mathbf{V} = \mathbf{\Lambda}$, where $\mathbf{V} = [\mathbf{v}_1\, \mathbf{v}_2 \ldots \mathbf{v}_N]$, with $\mathbf{v}_i$ the ith eigenvector, is the modal matrix of $\mathbf{C}$ and $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N)$, with λ_i the corresponding ith eigenvalue of $\mathbf{C}$. Since $\mathbf{C}$ is symmetric and positive definite, we have that $\mathbf{V}^T = \mathbf{V}^{-1}$ ($\mathbf{V}$ is an orthogonal matrix) and λ_i is real with $\lambda_i > 0$. Hence, $\mathbf{C} = \mathbf{V\Lambda V}^{-1}$ and

$$\begin{aligned}
\det(P_0\mathbf{C} + \sigma^2\mathbf{I}) &= \det(P_0\mathbf{V\Lambda V}^{-1} + \sigma^2\mathbf{I}) \\
&= \det[\mathbf{V}(P_0\mathbf{\Lambda} + \sigma^2\mathbf{I})\mathbf{V}^{-1}] \\
&= \det(P_0\mathbf{\Lambda} + \sigma^2\mathbf{I}) \\
&= \prod_{i=1}^{N}(P_0\lambda_i + \sigma^2).
\end{aligned}$$

Also, we have

$$\begin{aligned}
(P_0\mathbf{C} + \sigma^2\mathbf{I})^{-1} &= (P_0\mathbf{V\Lambda V}^{-1} + \sigma^2\mathbf{I})^{-1} \\
&= [\mathbf{V}(P_0\mathbf{\Lambda} + \sigma^2\mathbf{I})\mathbf{V}^{-1}]^{-1} \\
&= \mathbf{V}(P_0\mathbf{\Lambda} + \sigma^2\mathbf{I})^{-1}\mathbf{V}^T.
\end{aligned}$$

Thus, from (8.3)

$$\begin{aligned}
&\ln p(\mathbf{x}; P_0, \mathcal{H}_1) \\
&= -\frac{N}{2}\ln 2\pi - \frac{1}{2}\sum_{i=1}^{N}\ln(P_0\lambda_i + \sigma^2) - \frac{1}{2}\mathbf{x}^T\mathbf{V}(P_0\mathbf{\Lambda} + \sigma^2\mathbf{I})^{-1}\mathbf{V}^T\mathbf{x} \\
&= -\frac{N}{2}\ln 2\pi - \frac{1}{2}\sum_{i=1}^{N}\ln(P_0\lambda_i + \sigma^2) - \frac{1}{2}\sum_{i=1}^{N}\frac{(\mathbf{v}_i^T\mathbf{x})^2}{P_0\lambda_i + \sigma^2} \qquad (8.4)
\end{aligned}$$

since

$$\mathbf{V}^T\mathbf{x} = \begin{bmatrix} \mathbf{v}_1^T\mathbf{x} \\ \mathbf{v}_2^T\mathbf{x} \\ \vdots \\ \mathbf{v}_N^T\mathbf{x} \end{bmatrix}$$

and $P_0\mathbf{\Lambda} + \sigma^2\mathbf{I}$ is a diagonal matrix. To find the MLE of P_0 we must minimize

$$J(P_0) = \sum_{i=1}^{N}\left[\ln(P_0\lambda_i + \sigma^2) + \frac{(\mathbf{v}_i^T\mathbf{x})^2}{P_0\lambda_i + \sigma^2}\right]. \tag{8.5}$$

Differentiating J leads to a very nonlinear equation in P_0, the general solution of which is unknown. For some special cases, however, we can find $\hat{P}_0$. The following examples describe how this is done.

Example 8.2 - Unknown Signal Power (White Signal)

If all the λ_i's are equal and say $\lambda_i = \lambda$, then

$$J(P_0) = N\ln(P_0\lambda + \sigma^2) + \frac{1}{P_0\lambda + \sigma^2}\sum_{i=1}^{N}(\mathbf{v}_i^T\mathbf{x})^2.$$

But

$$\begin{aligned}
\sum_{i=1}^{N}(\mathbf{v}_i^T\mathbf{x})^2 &= \sum_{i=1}^{N}\mathbf{x}^T\mathbf{v}_i\mathbf{v}_i^T\mathbf{x} \\
&= \mathbf{x}^T\sum_{i=1}^{N}\mathbf{v}_i\mathbf{v}_i^T\mathbf{x} \\
&= \mathbf{x}^T\mathbf{V}\mathbf{V}^T\mathbf{x} \\
&= \mathbf{x}^T\mathbf{x}
\end{aligned}$$

so that

$$J(P_0) = N\ln(P_0\lambda + \sigma^2) + \frac{\mathbf{x}^T\mathbf{x}}{P_0\lambda + \sigma^2}. \tag{8.6}$$

This case corresponds to $\mathbf{C} = \mathbf{V\Lambda V}^{-1} = \lambda\mathbf{I}$ or the random signal is *white*. We now have upon differentiation

$$\frac{N\lambda}{P_0\lambda + \sigma^2} - \frac{\lambda\mathbf{x}^T\mathbf{x}}{(P_0\lambda + \sigma^2)^2} = 0$$

or

$$\hat{P}_0 = \frac{\frac{1}{N}\sum_{n=0}^{N-1} x^2[n] - \sigma^2}{\lambda}$$

as long as $\hat{P}_0 > 0$. If, however, $\hat{P}_0 \le 0$, then the MLE is $\hat{P}_0 = 0$ (see Problem 8.2). This is in accordance with the parameter constraint. Hence, the MLE is

$$\hat{P}_0 = \max\left(0, \left(\frac{1}{N}\sum_{n=0}^{N-1} x^2[n] - \sigma^2\right) \Big/ \lambda\right). \tag{8.7}$$

This is reasonable since $E((1/N)\sum_{n=0}^{N-1} x^2[n]) = P_0\lambda + \sigma^2$. We will let

$$\hat{P}_0^+ = \frac{\frac{1}{N}\sum_{n=0}^{N-1} x^2[n] - \sigma^2}{\lambda}$$

where the "plus" superscript is meant to remind us that $\hat{P}_0^+$ is the MLE if it is positive. The GLRT decides $\mathcal{H}_1$ if

$$\ln L_G(\mathbf{x}) = \ln \frac{p(\mathbf{x}; \hat{P}_0, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > \ln \gamma$$

which becomes from (8.4) and (8.5)

$$\ln L_G(\mathbf{x}) = -\frac{1}{2}J(\hat{P}_0) + \frac{N}{2}\ln\sigma^2 + \frac{1}{2\sigma^2}\sum_{n=0}^{N-1} x^2[n]. \tag{8.8}$$

Using (8.6) this is

$$\begin{aligned}
\ln L_G(\mathbf{x}) &= -\frac{N}{2}\ln(\hat{P}_0\lambda + \sigma^2) - \frac{1}{2}\frac{\sum_{n=0}^{N-1} x^2[n]}{\hat{P}_0\lambda + \sigma^2} \\
&\quad + \frac{N}{2}\ln\sigma^2 + \frac{1}{2\sigma^2}\sum_{n=0}^{N-1} x^2[n] \\
&= -\frac{N}{2}\ln\left(\frac{\hat{P}_0\lambda}{\sigma^2} + 1\right) + \frac{N}{2}\left[\left(\frac{1}{N}\sum_{n=0}^{N-1} x^2[n] \Big/ \sigma^2\right)\left(1 - \frac{\sigma^2}{\hat{P}_0\lambda + \sigma^2}\right)\right]
\end{aligned}$$

$$\begin{aligned}
&= -\frac{N}{2}\ln\left(\frac{\hat{P}_0\lambda}{\sigma^2} + 1\right) + \frac{N}{2}\left[\frac{\hat{P}_0^+\lambda + \sigma^2}{\sigma^2}\frac{\hat{P}_0\lambda}{\hat{P}_0\lambda + \sigma^2}\right] \\
&= \frac{N}{2}\left[\left(\frac{\hat{P}_0^+\lambda + \sigma^2}{\hat{P}_0\lambda + \sigma^2}\frac{\hat{P}_0\lambda}{\sigma^2} + 1\right) - \ln\left(\frac{\hat{P}_0\lambda}{\sigma^2} + 1\right) - 1\right].
\end{aligned}$$

Now if $\hat{P}_0 = 0$, $\ln L_G(\mathbf{x}) = 0$, and since $\ln\gamma > 0$ (see Problem 8.3), we choose $\mathcal{H}_0$, consistent with our intuition. On the other hand, if $\hat{P}_0 > 0$ so that $\hat{P}_0 = \hat{P}_0^+$,

$$\ln L_G(\mathbf{x}) = \frac{N}{2}\left[\left(\frac{\hat{P}_0\lambda}{\sigma^2} + 1\right) - \ln\left(\frac{\hat{P}_0\lambda}{\sigma^2} + 1\right) - 1\right].$$

But the function $g(x) = x - \ln x - 1$ is a monotonically increasing function of x for $x > 1$ since $dg/dx = 1 - 1/x > 0$. A plot of this function, which is defined for $x > 1$, is shown in Figure 8.1. Thus, its inverse g^{-1} exists for $x > 1$, and upon letting $x = \frac{\hat{P}_0\lambda}{\sigma^2} + 1 > 1$ (since $\hat{P}_0 > 0$), we decide $\mathcal{H}_1$ if

$$\frac{N}{2}g\left(\frac{\hat{P}_0\lambda}{\sigma^2} + 1\right) > \ln\gamma$$

or

$$\hat{P}_0 > \frac{\sigma^2}{\lambda}\left[g^{-1}\left(\frac{2}{N}\ln\gamma\right) - 1\right] = \gamma'.$$

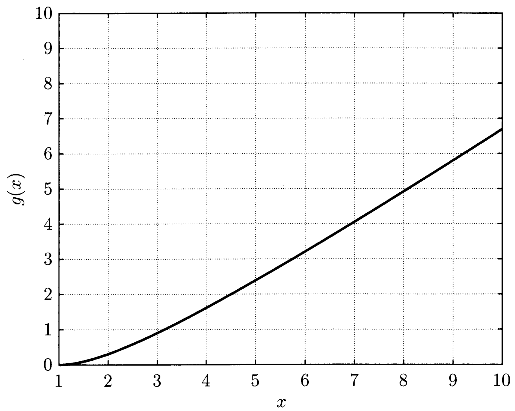


Figure 8.1. Function required for GLRT of white signal of unknown power.

But $\gamma' > 0$ since $(2/N)\ln\gamma > 0$ and thus $g^{-1}((2/N)\ln\gamma) > 1$. Hence, we decide $\mathcal{H}_1$ if $\hat{P}_0 > \gamma'$, where $\gamma' > 0$, and also $\hat{P}_0 > 0$ as originally assumed. The assumption that $\hat{P}_0 > 0$ is, however, subsumed by the condition $\hat{P}_0 > \gamma'$. Finally, since we decide $\mathcal{H}_1$ if $\hat{P}_0 > 0$ we have from (8.7) that we decide $\mathcal{H}_1$ if

$$\sum_{n=0}^{N-1} x^2[n] > \gamma''. \tag{8.9}$$

It is interesting to note in this example that the asymptotic GLRT statistics do not hold [Chernoff 1954]. Under $\mathcal{H}_0$, for example, the statistics of the MLE of P_0 given by (8.7) are that of a random variable that is zero half of the time and Gaussian for the other half (see Problem 8.4). This is in contrast to the usual asymptotic statistics of the MLE, which are Gaussian. This type of result is a consequence of the implicit one-sided parameter testing problem $\mathcal{H}_0 : P_0 = 0$ versus $\mathcal{H}_1 : P_0 > 0$ (see also Problem 6.23).

◇

Example 8.3 - Unknown Signal Power (Low-Rank Covariance Matrix)

Another special case of interest occurs when the signal covariance matrix is low-rank as already discussed in Problems 5.13, 5.14. If, for example, $\lambda_1 \neq 0$, $\lambda_i = 0$ for $i = 2, 3, \ldots, N$, then $\mathbf{C}_s = P_0\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ reduces to

$$\mathbf{C}_s = P_0\lambda_1\mathbf{v}_1\mathbf{v}_1^T$$

which is a rank one matrix. This arises for a random DC level (see [Kay-I 1993, pg. 49]) in which $s[n] = A$ where $A \sim \mathcal{N}(0, \sigma_A^2)$. In this case, $\mathbf{C}_s = \sigma_A^2\mathbf{1}\mathbf{1}^T$ so that $\mathbf{v}_1 = \mathbf{1}/\sqrt{N}$ and $\lambda_1 = N\sigma_A^2$. The remaining eigenvalues are zero. Also, sinusoids can be modeled by low-rank covariance matrices as shown in Problems 8.5 and 8.6. For the rank one covariance matrix the MLE of P_0 is found from (8.5) by minimizing

$$J(P_0) = \ln(P_0\lambda_1 + \sigma^2) + \frac{(\mathbf{v}_1^T\mathbf{x})^2}{P_0\lambda_1 + \sigma^2} + (N-1)\ln\sigma^2 + \frac{1}{\sigma^2}\sum_{i=2}^{N}(\mathbf{v}_i^T\mathbf{x})^2. \tag{8.10}$$

Upon differentiation and setting equal to zero we have

$$\hat{P}_0 = \frac{(\mathbf{v}_1^T\mathbf{x})^2 - \sigma^2}{\lambda_1} \tag{8.11}$$

as long as $\hat{P}_0 > 0$. Hence, the MLE is given as

$$\hat{P}_0 = \max\left(0, \frac{(\mathbf{v}_1^T\mathbf{x})^2 - \sigma^2}{\lambda_1}\right). \tag{8.12}$$

As before we let

$$\hat{P}_0^+ = \frac{(\mathbf{v}_1^T\mathbf{x})^2 - \sigma^2}{\lambda_1}.$$

The GLRT becomes

$$L_G(\mathbf{x}) = \frac{\dfrac{1}{(2\pi)^{\frac{N}{2}}\det^{\frac{1}{2}}(\hat{P}_0\mathbf{C} + \sigma^2\mathbf{I})}\exp\left[-\frac{1}{2}\mathbf{x}^T(\hat{P}_0\mathbf{C} + \sigma^2\mathbf{I})^{-1}\mathbf{x}\right]}{\dfrac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left(-\frac{1}{2\sigma^2}\mathbf{x}^T\mathbf{x}\right)}.$$

Taking the logarithm and using (8.10) we have

$$\begin{aligned}
\ln L_G(\mathbf{x}) &= -\frac{1}{2}J(\hat{P}_0) + \frac{N}{2}\ln\sigma^2 + \frac{1}{2\sigma^2}\mathbf{x}^T\mathbf{x} \\
&= -\frac{1}{2}\ln(\hat{P}_0\lambda_1 + \sigma^2) - \frac{1}{2}\frac{(\mathbf{v}_1^T\mathbf{x})^2}{\hat{P}_0\lambda_1 + \sigma^2} \\
&\quad -\frac{N-1}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=2}^{N}(\mathbf{v}_i^T\mathbf{x})^2 \\
&\quad +\frac{N}{2}\ln\sigma^2 + \frac{1}{2\sigma^2}\mathbf{x}^T\mathbf{x} \\
&= -\frac{1}{2}\ln\left(\frac{\hat{P}_0\lambda_1}{\sigma^2} + 1\right) - \frac{1}{2}\frac{\hat{P}_0^+\lambda_1 + \sigma^2}{\hat{P}_0\lambda_1 + \sigma^2} \\
&\quad +\frac{1}{2\sigma^2}\left(\mathbf{x}^T\mathbf{x} - \sum_{i=2}^{N}\mathbf{x}^T\mathbf{v}_i\mathbf{v}_i^T\mathbf{x}\right).
\end{aligned}$$

But $\mathbf{V}\mathbf{V}^T = \sum_{i=1}^{N}\mathbf{v}_i\mathbf{v}_i^T = \mathbf{I}$ so that $\mathbf{I} - \sum_{i=2}^{N}\mathbf{v}_i\mathbf{v}_i^T = \mathbf{v}_1\mathbf{v}_1^T$. Using this we have

$$\begin{aligned}
\ln L_G(\mathbf{x}) &= -\frac{1}{2}\ln\left(\frac{\hat{P}_0\lambda_1}{\sigma^2} + 1\right) - \frac{1}{2}\frac{\hat{P}_0^+\lambda_1 + \sigma^2}{\hat{P}_0\lambda_1 + \sigma^2} + \frac{1}{2\sigma^2}(\mathbf{v}_1^T\mathbf{x})^2 \\
&= -\frac{1}{2}\ln\left(\frac{\hat{P}_0\lambda_1}{\sigma^2} + 1\right) - \frac{1}{2}\frac{\hat{P}_0^+\lambda_1 + \sigma^2}{\hat{P}_0\lambda_1 + \sigma^2} + \frac{1}{2}\frac{\hat{P}_0^+\lambda_1 + \sigma^2}{\sigma^2} \\
&= \frac{1}{2}\left[(\hat{P}_0^+\lambda_1 + \sigma^2)\left(\frac{1}{\sigma^2} - \frac{1}{\hat{P}_0\lambda_1 + \sigma^2}\right) - \ln\left(\frac{\hat{P}_0\lambda_1}{\sigma^2} + 1\right)\right] \\
&= \frac{1}{2}\left[\left(\frac{\hat{P}_0^+\lambda_1 + \sigma^2}{\hat{P}_0\lambda_1 + \sigma^2}\frac{\hat{P}_0\lambda_1}{\sigma^2} + 1\right) - \ln\left(\frac{\hat{P}_0\lambda_1}{\sigma^2} + 1\right) - 1\right].
\end{aligned}$$

By the same arguments as before we decide $\mathcal{H}_1$ if $\hat{P}_0 > \gamma'$ or, equivalently, if $\hat{P}_0^+ > \gamma'$, or if

$$T(\mathbf{x}) = (\mathbf{v}_1^T\mathbf{x})^2 > \gamma''. \tag{8.13}$$

As an example, for the random DC level in WGN we have $\mathbf{v}_1 = \mathbf{1}/\sqrt{N}$ so that $T(\mathbf{x}) = N\bar{x}^2$ or finally we decide $\mathcal{H}_1$ if

$$\bar{x}^2 > \frac{\gamma''}{N}.$$

Note that if P_0 were known we would obtain the same detector (see Problem 5.14). Thus, the GLRT is UMP in this example. The reader is asked to determine the performance of this detector in Problem 8.8.

◇

An extension of the low-rank covariance matrix to periodic signals is described in Section 8.6. The analytical difficulties encountered in these examples is typical. Numerical determination of the MLE is often required. By using the asymptotic form of the likelihood function as described in the next section, we can reduce the required computation.

Before proceeding it is worthwhile to summarize the GLRT for the general problem discussed here. If the signal covariance matrix is denoted by $\mathbf{C}_s(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of unknown parameters, the GLRT decides $\mathcal{H}_1$ if

$$\begin{aligned} L_G(\mathbf{x}) &= \frac{p(\mathbf{x};\hat{\boldsymbol{\theta}},\mathcal{H}_1)}{p(\mathbf{x};\mathcal{H}_0)} \\ &= \frac{\dfrac{1}{(2\pi)^{\frac{N}{2}}\det^{\frac{1}{2}}(\mathbf{C}_s(\hat{\boldsymbol{\theta}})+\sigma^2\mathbf{I})}\exp\left[-\dfrac{1}{2}\mathbf{x}^T(\mathbf{C}_s(\hat{\boldsymbol{\theta}})+\sigma^2\mathbf{I})^{-1}\mathbf{x}\right]}{\dfrac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left(-\dfrac{1}{2\sigma^2}\mathbf{x}^T\mathbf{x}\right)} > \gamma \end{aligned}$$

or by taking logarithms and scaling by two

$$2\ln L_G(\mathbf{x}) = -\ln\det(\mathbf{C}_s(\hat{\boldsymbol{\theta}})+\sigma^2\mathbf{I}) + N\ln\sigma^2 - \mathbf{x}^T(\mathbf{C}_s(\hat{\boldsymbol{\theta}})+\sigma^2\mathbf{I})^{-1}\mathbf{x} + \frac{\mathbf{x}^T\mathbf{x}}{\sigma^2}.$$

Using (5.4) we have

$$\begin{aligned} \frac{\mathbf{x}^T\mathbf{x}}{\sigma^2} - \mathbf{x}^T(\mathbf{C}_s(\hat{\boldsymbol{\theta}})+\sigma^2\mathbf{I})^{-1}\mathbf{x} &= \frac{\mathbf{x}^T\mathbf{x}}{\sigma^2} - \mathbf{x}^T\left[\frac{1}{\sigma^2}\mathbf{I} - \frac{1}{\sigma^4}\left(\frac{1}{\sigma^2}\mathbf{I}+\mathbf{C}_s^{-1}(\hat{\boldsymbol{\theta}})\right)^{-1}\right]\mathbf{x} \\ &= \frac{1}{\sigma^4}\mathbf{x}^T\left(\frac{1}{\sigma^2}\mathbf{I}+\mathbf{C}_s^{-1}(\hat{\boldsymbol{\theta}})\right)^{-1}\mathbf{x} \\ &= \frac{1}{\sigma^2}\mathbf{x}^T\mathbf{C}_s(\hat{\boldsymbol{\theta}})(\mathbf{C}_s(\hat{\boldsymbol{\theta}})+\sigma^2\mathbf{I})^{-1}\mathbf{x} \end{aligned}$$

as in Section 5.3. Hence,

$$2\ln L_G(\mathbf{x}) = \frac{1}{\sigma^2}\mathbf{x}^T\mathbf{C}_s(\hat{\boldsymbol{\theta}})(\mathbf{C}_s(\hat{\boldsymbol{\theta}})+\sigma^2\mathbf{I})^{-1}\mathbf{x} - \ln\det\left(\frac{\mathbf{C}_s(\hat{\boldsymbol{\theta}})}{\sigma^2}+\mathbf{I}\right) \tag{8.14}$$

where $\hat{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}$.

8.4 Large Data Record Approximations

If the signal random process is WSS, then for large data records it has been shown that the log-likelihood ratio is (see Section 5.5)

$$\begin{aligned} l(\mathbf{x}) &= \ln\frac{p(\mathbf{x};\mathcal{H}_1)}{p(\mathbf{x};\mathcal{H}_0)} \\ &= -\frac{N}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}}\left[\ln\left(\frac{P_{ss}(f)}{\sigma^2}+1\right) - \frac{P_{ss}(f)}{P_{ss}(f)+\sigma^2}\frac{I(f)}{\sigma^2}\right]df \end{aligned} \tag{8.15}$$

where $P_{ss}(f)$ is the PSD of $s[n]$ and $I(f)$ is the periodogram. Since only the PSD of the signal is assumed to have unknown parameters, the MLE can be found by maximizing $l(\mathbf{x})$. Note that $p(\mathbf{x};\mathcal{H}_0)$ does not depend on the signal parameters. Hence, the MLE is found by minimizing

$$J(\boldsymbol{\theta}) = \int_{-\frac{1}{2}}^{\frac{1}{2}}\left[\ln\left(\frac{P_{ss}(f;\boldsymbol{\theta})}{\sigma^2}+1\right) - \frac{P_{ss}(f;\boldsymbol{\theta})}{P_{ss}(f;\boldsymbol{\theta})+\sigma^2}\frac{I(f)}{\sigma^2}\right]df \tag{8.16}$$

where we now denote the dependence of $P_{ss}(f)$ on $\boldsymbol{\theta}$. Once the MLE is found the GLRT decides $\mathcal{H}_1$ if

$$\begin{aligned} \ln L_G(\mathbf{x}) &= \ln\frac{p(\mathbf{x};\hat{\boldsymbol{\theta}},\mathcal{H}_1)}{p(\mathbf{x};\mathcal{H}_0)} \\ &= -\frac{N}{2}J(\hat{\boldsymbol{\theta}}) > \ln\gamma \end{aligned}$$

or

$$-J(\hat{\boldsymbol{\theta}}) > \gamma'. \tag{8.17}$$

In the special case in which the logarithmic term in the integrand of (8.16) does not depend on $\boldsymbol{\theta}$, the GLRT decides $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \int_{-\frac{1}{2}}^{\frac{1}{2}} H(f;\hat{\boldsymbol{\theta}})I(f)df > \gamma'' \tag{8.18}$$

or since $\hat{\boldsymbol{\theta}}$ is found by maximizing $-J(\boldsymbol{\theta})$

$$\max_{\boldsymbol{\theta}} \int_{-\frac{1}{2}}^{\frac{1}{2}} H(f;\boldsymbol{\theta})I(f)df > \gamma'' \tag{8.19}$$

where $H(f;\boldsymbol{\theta})$ is the Wiener filter or

$$H(f;\boldsymbol{\theta}) = \frac{P_{ss}(f;\boldsymbol{\theta})}{P_{ss}(f;\boldsymbol{\theta}) + \sigma^2}.$$

Note that if numerical minimization of $J(\boldsymbol{\theta})$ is required, this may be done by approximating the integrals using FFTs. No matrix inversions or determinant evaluations are necessary. Some examples follow.

Example 8.4 - Unknown Signal Power

This is the large data record version of Example 8.1. Here we assume $P_{ss}(f;P_0) = P_0 Q(f)$ where $\int_{-\frac{1}{2}}^{\frac{1}{2}} Q(f)df = 1$ so that P_0 is the total power in $s[n]$. The reader may also wish to refer to [Kay-I 1993, Problem 3.16] for the CRLB. From (8.16) the MLE is found by minimizing

$$J(P_0) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\ln\left(\frac{P_0 Q(f)}{\sigma^2} + 1\right) - \frac{P_0 Q(f)}{P_0 Q(f) + \sigma^2}\frac{I(f)}{\sigma^2}\right] df. \tag{8.20}$$

Differentiating produces

$$\frac{d\,J(P_0)}{d\,P_0} = \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\frac{Q(f)}{P_0 Q(f) + \sigma^2} - \frac{Q(f)}{(P_0 Q(f) + \sigma^2)^2} I(f)\right] df$$

and setting this equal to zero yields

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{Q(f)(P_0 Q(f) + \sigma^2) - Q(f)I(f)}{(P_0 Q(f) + \sigma^2)^2} df = 0. \tag{8.21}$$

Unfortunately, this cannot be solved for P_0. As an approximation if we assume a low SNR or $P_0 Q(f) \ll \sigma^2$, we have

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{Q(f)(P_0 Q(f) + \sigma^2) - Q(f)I(f)}{\sigma^4} df \approx 0$$

which yields

$$\hat{P}_0 \approx \frac{\displaystyle\int_{-\frac{1}{2}}^{\frac{1}{2}} Q(f)(I(f) - \sigma^2)df}{\displaystyle\int_{-\frac{1}{2}}^{\frac{1}{2}} Q^2(f)df} \tag{8.22}$$

assuming $\hat{P}_0 > 0$. As before, if $\hat{P}_0 \le 0$, we set $\hat{P}_0 = 0$. The GLRT becomes from (8.17) and (8.20) to decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[-\ln\left(\frac{\hat{P}_0 Q(f)}{\sigma^2} + 1\right) + \frac{\hat{P}_0 Q(f)}{\hat{P}_0 Q(f) + \sigma^2}\frac{I(f)}{\sigma^2}\right] df > \gamma' \tag{8.23}$$

where $\hat{P}_0$ is given as the solution of (8.21) or by the low SNR approximation of (8.22). In either case, if $\hat{P}_0 \le 0$, we set $\hat{P}_0$ equal to zero and hence from (8.23), $T(\mathbf{x}) = 0$ and we decide $\mathcal{H}_0$.

◇

Example 8.5 - Unknown Center Frequency

This example was discussed in [Kay-I 1993, Example 3.12] where we determined the CRLB. We assume that

$$P_{ss}(f;f_c) = Q(f - f_c) + Q(-f - f_c)$$

where $Q(f)$ is a low-pass PSD which is zero for $f < -f_1$ and $f > f_2$. The center frequency is f_c. We further assume that for all possible center frequencies the PSD for $f \ge 0$ is given by $Q(f - f_c)$ and will lie entirely in the $[0, 1/2]$ interval. As a result

$$\begin{aligned}\int_{-\frac{1}{2}}^{\frac{1}{2}} \ln\left(\frac{P_{ss}(f;f_c)}{\sigma^2} + 1\right) df &= 2\int_0^{\frac{1}{2}} \ln\left(\frac{P_{ss}(f;f_c)}{\sigma^2} + 1\right) df \\ &= 2\int_0^{\frac{1}{2}} \ln\left(\frac{Q(f - f_c)}{\sigma^2} + 1\right) df\end{aligned}$$

does not depend on f_c and hence can be omitted from (8.16) in determining the MLE. The GLRT becomes from (8.19) to decide $\mathcal{H}_1$ if

$$\begin{aligned}T(\mathbf{x}) &= \max_{f_c} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{P_{ss}(f;f_c)}{P_{ss}(f;f_c) + \sigma^2} I(f)df \\ &= 2\max_{f_c} \int_0^{\frac{1}{2}} \frac{Q(f - f_c)}{Q(f - f_c) + \sigma^2} I(f)df\end{aligned}$$

and could be implemented as a grid search. For low SNR or $Q(f - f_c) \ll \sigma^2$ this reduces to

$$T(\mathbf{x}) \approx \frac{2}{\sigma^2} \max_{f_c} \int_0^{\frac{1}{2}} Q(f - f_c)I(f)df$$

which is a *spectral correlation* of the estimated PSD, i.e., the periodogram $I(f)$, with the signal PSD $Q(f - f_c)$ shifted to all possible center frequencies. In accordance with the results of [Kay-I 1993, Example 3.12] we would expect better estimation/detection performance when $Q(f)$ is narrowband.

◇

Some further examples may be found in [Levin 1965] and in Problem 8.10.

8.5 Weak Signal Detection

The detection problem discussed in Example 8.1 can be recast as the parameter test

$$\begin{aligned}\mathcal{H}_0 &: P_0 = 0\\ \mathcal{H}_1 &: P_0 > 0.\end{aligned}$$

This is a one-sided hypothesis test which was studied in Chapter 6. If the signal is assumed to be weak or P_0 small, then the detection problem reduces to a test of a *small* departure of P_0 from zero. In such a case, the locally most powerful (LMP) detector can be employed. Omitting the normalizing factor $\sqrt{I(P_0)}$ (see (6.36)), this detector decides $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \left.\frac{\partial \ln p(\mathbf{x}; P_0, \mathcal{H}_1)}{\partial P_0}\right|_{P_0=0} > \gamma. \tag{8.24}$$

The advantage of this detector is that only a partial derivative need be evaluated. Hence, *MLE evaluation is not required.* We now explicitly determine the LMP detector.

Example 8.6 - Locally Most Powerful Detector for Unknown Power Signal

The log-PDF under $\mathcal{H}_1$ is given by (8.3) as

$$\ln p(\mathbf{x}; P_0, \mathcal{H}_1) = -\frac{N}{2}\ln 2\pi - \frac{1}{2}\ln\det(P_0\mathbf{C} + \sigma^2\mathbf{I}) - \frac{1}{2}\mathbf{x}^T(P_0\mathbf{C} + \sigma^2\mathbf{I})^{-1}\mathbf{x}.$$

To evaluate the derivative we use the following formulas [Kay-I 1993, pg. 73]

$$\begin{aligned}\frac{\partial \ln\det\mathbf{C}(\theta)}{\partial\theta} &= \operatorname{tr}\left(\mathbf{C}^{-1}(\theta)\frac{\partial\mathbf{C}(\theta)}{\partial\theta}\right)\\ \frac{\partial\mathbf{C}^{-1}(\theta)}{\partial\theta} &= -\mathbf{C}^{-1}(\theta)\frac{\partial\mathbf{C}(\theta)}{\partial\theta}\mathbf{C}^{-1}(\theta)\end{aligned}$$

where the $N \times N$ covariance matrix $\mathbf{C}(\theta)$ is assumed to depend on the parameter θ. The notations $\partial\mathbf{C}(\theta)/\partial\theta$ and $\partial\mathbf{C}^{-1}(\theta)/\partial\theta$ denote the $N \times N$ matrices whose $[i, j]$

element is $\partial[\mathbf{C}(\theta)]_{ij}/\partial\theta$ and $\partial[\mathbf{C}^{-1}(\theta)]_{ij}/\partial\theta$, respectively. We have upon letting $\mathbf{C}(P_0) = P_0\mathbf{C} + \sigma^2\mathbf{I}$

$$\frac{\partial \ln p(\mathbf{x}; P_0, \mathcal{H}_1)}{\partial P_0} = -\frac{1}{2}\operatorname{tr}\left(\mathbf{C}^{-1}(P_0)\frac{\partial\mathbf{C}(P_0)}{\partial P_0}\right) + \frac{1}{2}\mathbf{x}^T\mathbf{C}^{-1}(P_0)\frac{\partial\mathbf{C}(P_0)}{\partial P_0}\mathbf{C}^{-1}(P_0)\mathbf{x}.$$

But $\partial\mathbf{C}(P_0)/\partial P_0 = \mathbf{C}$ so that

$$\frac{\partial \ln p(\mathbf{x}; P_0, \mathcal{H}_1)}{\partial P_0} = -\frac{1}{2}\operatorname{tr}\left((P_0\mathbf{C} + \sigma^2\mathbf{I})^{-1}\mathbf{C}\right) + \frac{1}{2}\mathbf{x}^T(P_0\mathbf{C}+\sigma^2\mathbf{I})^{-1}\mathbf{C}(P_0\mathbf{C}+\sigma^2\mathbf{I})^{-1}\mathbf{x}.$$

Evaluating the derivative at $P_0 = 0$ yields

$$T(\mathbf{x}) = -\frac{1}{2\sigma^2}\operatorname{tr}(\mathbf{C}) + \frac{1}{2}\frac{\mathbf{x}^T\mathbf{C}\mathbf{x}}{\sigma^4}$$

so that we decide $\mathcal{H}_1$ if

$$\mathbf{x}^T\mathbf{C}\mathbf{x} > 2\sigma^4\left(\gamma + \frac{1}{2\sigma^2}\operatorname{tr}(\mathbf{C})\right) = \gamma'. \tag{8.25}$$

See Problem 8.12 for the detection performance. As an example, if $\mathbf{C}$ is the low-rank covariance matrix $\mathbf{C} = \lambda_1\mathbf{v}_1\mathbf{v}_1^T$, then we decide $\mathcal{H}_1$ if

$$\mathbf{x}^T\lambda_1\mathbf{v}_1\mathbf{v}_1^T\mathbf{x} > \gamma'$$

or

$$(\mathbf{v}_1^T\mathbf{x})^2 > \frac{\gamma'}{\lambda_1} = \gamma''$$

in agreement with the results of Example 8.3. Thus, the GLRT is also the LMP detector for this example.

◇

8.6 Signal Processing Example

In this section we extend the low-rank signal covariance detection problem discussed in Example 8.3 to the problem of detecting a WSS Gaussian *periodic* signal of *unknown PSD.* This problem arises in many fields. In speech processing the voiced phonemes are approximately periodic signals [Rabiner and Schafer 1978]. Similarly, many military vehicles have onboard machinery that radiates periodic signals [Urick 1975]. We first assume the period is known and then briefly describe the modifications to the detector if it is unknown.

Example 8.7 - Detection of Periodic Random Signals

The signal that is of interest is a WSS Gaussian random process with PSD $P_{ss}(f)$ and is periodic with known period M. Consequently, the ACF $r_{ss}[k]$ is also periodic with period M or $r_{ss}[k+M] = r_{ss}[k]$ for all k. Since the ACF is the inverse Fourier transform of the PSD, we have

$$r_{ss}[k] = \int_{-\frac{1}{2}}^{\frac{1}{2}} P_{ss}(f) \exp(j2\pi fk) df. \tag{8.26}$$

Due to the periodicity, it follows that

$$r_{ss}[k+M] = \int_{-\frac{1}{2}}^{\frac{1}{2}} P_{ss}(f) \exp[j2\pi f(M+k)] = r_{ss}[k]$$

which can only be true if $\exp(j2\pi fM) = 1$. This is satisfied for the frequencies $f = 0, 1/M, 2/M, \ldots, M/2$ (for M even) in the interval $[0, 1/2]$. Thus, the PSD must be zero at all other frequencies. The PSD is a set of Dirac impulses at these frequencies or

$$P_{ss}(f) = P_0\delta(f) + \sum_{i=1}^{\frac{M}{2}-1} \frac{P_i}{2}\delta\left(f - \frac{i}{M}\right) + P_{M/2}\delta(f - 1/2) \qquad 0 \le f \le \frac{1}{2}$$

where the total power at $f = f_i$ is P_i. The "frequency" component at $f = 0$ is the DC component, that at $1/M$ is termed the *fundamental* and the remaining ones are termed the *harmonics*. Since we usually assume $s[n]$ to be zero mean, we let $P_0 = 0$, and also to avoid power at the Nyquist frequency ($f = 1/2$), we let $P_{M/2} = 0$ as well. Hence,

$$P_{ss}(f) = \sum_{i=1}^{\frac{M}{2}-1} \frac{P_i}{2}\delta\left(f - \frac{i}{M}\right) \qquad 0 \le f \le \frac{1}{2}.$$

Finally, because $P_{ss}(-f) = P_{ss}(f)$ we have

$$P_{ss}(f) = \sum_{i=-(\frac{M}{2}-1)}^{\frac{M}{2}-1} \frac{P_i}{2}\delta\left(f - \frac{i}{M}\right) \qquad -\frac{1}{2} \le f \le \frac{1}{2}$$

where $P_{-i} = P_i$ and $P_0 = 0$. Note that the total power of the signal at f_i (contribution from positive and negative frequencies) is P_i. The corresponding ACF is from (8.26)

$$r_{ss}[k] = \sum_{i=-L}^{L} \frac{P_i}{2} \exp\left(j2\pi \frac{i}{M} k\right) \tag{8.27}$$

or equivalently

$$r_{ss}[k] = \sum_{i=1}^{L} P_i \cos\left(2\pi \frac{i}{M} k\right)$$

where $L = M/2 - 1$. This model also arises if the signal is modeled as a sum of sinusoids with frequencies $f_i = i/M$ or

$$s[n] = \sum_{i=1}^{L} A_i \cos(2\pi f_i n + \phi_i)$$

where each A_i is a Rayleigh random variable with $E(A_i^2/2) = P_i$, each $\phi_i \sim \mathcal{U}[0, 2\pi]$ and all the random variables are independent (see Problem 8.13). Thus, the periodic random signal is a generalization of the Rayleigh fading model described in Example 5.5.

To simplify the derivation we assume that the data record length is an integral number of periods or $N = KM$ for K a positive integer. Since this assumption does not generally hold, the results that we will obtain are approximate. The approximation, however, will be quite accurate for moderately sized data records. The problem then is to detect $s[n]$ whose period M is known but whose powers P_i are *unknown*. The log-PDF under $\mathcal{H}_1$ is derived in Appendix 8A as

$$\ln p(\mathbf{x}; \mathbf{P}, \mathcal{H}_1) = -\frac{N}{2}\ln 2\pi\sigma^2 - \frac{1}{2}\frac{\mathbf{x}^T\mathbf{x}}{\sigma^2} - \sum_{i=1}^{L}\left[\ln\left(\frac{NP_i/2}{\sigma^2} + 1\right) - \frac{NP_i/2}{NP_i/2 + \sigma^2}\frac{I(f_i)}{\sigma^2}\right]$$

where $f_i = i/M$ and $\mathbf{P} = [P_1 \, P_2 \ldots P_L]^T$ are the unknown parameters. The log-likelihood ratio becomes

$$\begin{aligned} l(\mathbf{x}) &= \ln \frac{p(\mathbf{x}; \mathbf{P}, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} \\ &= -\sum_{i=1}^{L}\left[\ln\left(\frac{NP_i/2}{\sigma^2} + 1\right) - \frac{NP_i/2}{NP_i/2 + \sigma^2}\frac{I(f_i)}{\sigma^2}\right]. \end{aligned} \tag{8.28}$$

Note the similarity to (8.15). The GLRT is obtained by replacing the P_i's in (8.28) by their MLE. Because of the decoupling $\hat{P}_k$ is found by minimizing

$$J(P_k) = \ln\left(\frac{NP_k/2}{\sigma^2} + 1\right) - \frac{NP_k/2}{NP_k/2 + \sigma^2}\frac{I(f_k)}{\sigma^2}.$$

It is easily shown that (see Problem 8.14)

$$\hat{P}_k = \max\left(0, \frac{2}{N}(I(f_k) - \sigma^2)\right). \tag{8.29}$$

Note that $I(f_k) - \sigma^2$ is an estimate of the *signal* PSD and the $1/N$ factor is the bandwidth. The result is an estimate of the sum of the signal powers at $f = f_k$ and $f = -f_k$ (recall that the power at $f = f_k$ is $P_k/2$ and that at $f = -f_k$ is also $P_k/2$). For the $\hat{P}_k$'s which are zero there is no contribution to $l(\mathbf{x})$. The log-GLRT becomes from (8.28) and (8.29)

$$\begin{aligned}
\ln \frac{p(\mathbf{x}; \hat{\mathbf{P}}, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} &= \sum_{\substack{i=1 \\ \hat{P}_i>0}}^{L} \left[\frac{N\hat{P}_i}{2\sigma^2} - \ln\left(\frac{N\hat{P}_i}{2\sigma^2} + 1\right)\right] \\
&= \sum_{\substack{i=1 \\ \hat{P}_i>0}}^{L} \left[\left(\frac{I(f_i) - \sigma^2}{\sigma^2}\right) - \ln\left(\frac{I(f_i) - \sigma^2}{\sigma^2} + 1\right)\right] \\
&= \sum_{\substack{i=1 \\ I(f_i)/\sigma^2>1}}^{L} \left(\frac{I(f_i)}{\sigma^2} - \ln\frac{I(f_i)}{\sigma^2} - 1\right) \qquad (8.30)
\end{aligned}$$

or we decide $\mathcal{H}_1$ if

$$\sum_{i=1}^{L} g\left(\frac{I(f_i)}{\sigma^2}\right) > \gamma'$$

where $g(x) = \max(0, x - \ln x - 1)$. Note that for $I(f_i)/\sigma^2 > 1$, $g(I(f_i)/\sigma^2) > 0$ as shown in Figure 8.1. This detector is shown in Figure 8.2. In practice, the nonlinearity is usually omitted to yield

$$\sum_{i=1}^{L} I(f_i) > \gamma''. \qquad (8.31)$$

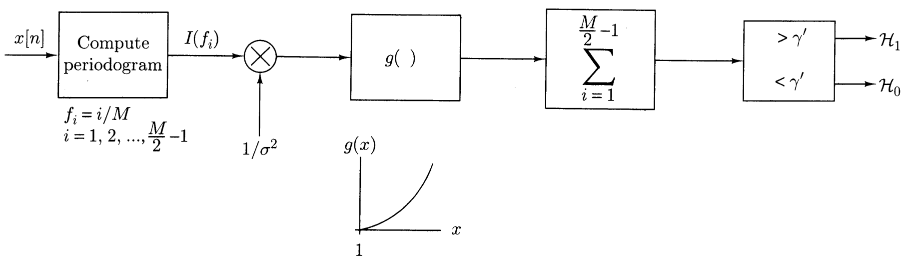


Figure 8.2. GLRT for detection of periodic Gaussian random process in WGN.

This test statistic essentially sums up the power at each signal frequency and compares it to a threshold. Heuristically, this operation can be viewed as the limiting form of (5.27) where $P_{ss}(f) \to \infty$ at the signal frequencies and $P_{ss}(f) = 0$ otherwise. Hence, the Wiener filter only passes the energy at the signal frequencies and zeros out the noise power at all other frequencies. This type of detector is referred to as a *comb filter* since it filters the data with a filter having passbands centered about the signal frequencies (which are equally separated as in a comb), followed by an energy detector in frequency (see also Problem 8.16). To see this we rewrite the periodogram as

$$\begin{aligned}
I(f_i) &= \frac{1}{N}\left|\sum_{k=0}^{N-1} x[k]\exp(-j2\pi f_i k)\right|^2 \\
&= \frac{\left|\sum_{k=0}^{N-1} x[k]h_i[n-k]\right|^2 \Bigg|_{n=0}}{\frac{1}{N}}
\end{aligned}$$

where

$$h_i[n] = \begin{cases} \frac{1}{N}\exp(j2\pi f_i n) & n = -(N-1), -(N-2), \ldots, 0 \\ 0 & \text{otherwise.} \end{cases}$$

This operation is recognized as an FIR filtering of $x[n]$ with a complex filter with impulse response $h_i[n]$, followed by a squaring and sampling to obtain a power estimate, and finally a division by $1/N$, the filter bandwidth, to obtain an estimated PSD. The frequency response of the ith periodogram filter is easily shown to have magnitude

$$|H_i(f)| = \left|\frac{\sin[N\pi(f - f_i)]}{N\sin[\pi(f - f_i)]}\right|.$$

Hence, the ith filter has a passband centered about $f = f_i$ and has a bandwidth of about $1/N$ [Kay 1988]. This interpretation is shown in Figure 8.3.

Other interpretations of the detector of (8.31) are as an estimator-correlator and an averager followed by an energy detector. We now manipulate the test statistic into these two equivalent forms. From (8.31) we have upon scaling by two

$$T(\mathbf{x}) = 2\sum_{i=1}^{\frac{M}{2}-1} I(f_i) = \sum_{\substack{i=-(\frac{M}{2}-1) \\ i\neq 0}}^{\frac{M}{2}-1} I(f_i)$$

since $I(-f_i) = I(f_i)$. But it may be shown that

$$I'(f_i) = \frac{1}{N}\left|\sum_{n=0}^{N-1} (x[n] - \bar{x})\exp(-j2\pi f_i n)\right|^2$$

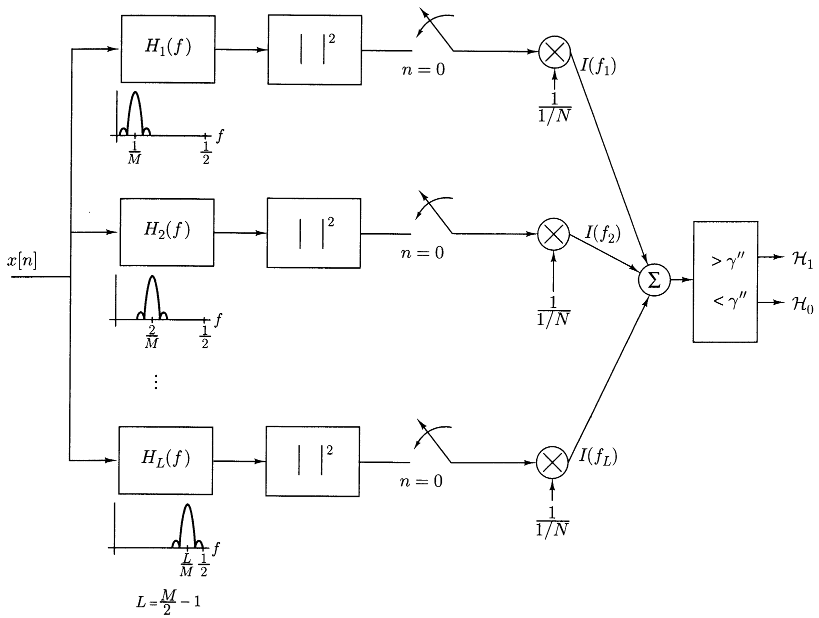


Figure 8.3. Comb filter detector for periodic Gaussian random signal in WGN.

$$= \begin{cases} I(f_i) & i = 1, 2, \ldots, \frac{M}{2} - 1 \\ 0 & i = 0 \end{cases} \tag{8.32}$$

where $\bar{x} = (1/N) \sum_{n=0}^{N-1} x[n]$ (see Problem 8.15). As a result, we can write

$$T(\mathbf{x}) = \sum_{i=-(\frac{M}{2}-1)}^{\frac{M}{2}-1} I'(f_i)$$

where $I'(f_i)$ is the periodogram of the mean compensated data $y[n] = x[n] - \bar{x}$. Also, due to the periodicity of $I'(f)$ we have $I'(f_i) = I'(f_{i+M})$ and assuming that $I'(f_{M/2})$ is zero leads to

$$T(\mathbf{x}) = \sum_{i=0}^{M-1} I'(f_i)$$

$$\begin{aligned}
&= \sum_{i=0}^{M-1} \frac{1}{N} \left| \sum_{n=0}^{N-1} y[n] \exp(-j2\pi f_i n) \right|^2 \\
&= \sum_{i=0}^{M-1} \frac{1}{N} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} y[m] y[n] \exp[j2\pi f_i (m-n)] \\
&= \frac{1}{N} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} y[m] y[n] \sum_{i=0}^{M-1} \exp[j2\pi f_i (m-n)].
\end{aligned}$$

But it may be shown that

$$\sum_{i=0}^{M-1} \exp(j2\pi f_i k) = \begin{cases} M & \text{for } k = 0, \pm M, \pm 2M, \ldots \\ 0 & \text{otherwise} \end{cases}$$

since $f_i = i/M$. As a result, we must have that $m - n = rM$ for r an integer or $m = n + rM$ for a nonzero contribution to the sum over m and n. It follows that

$$\begin{aligned}
T(\mathbf{x}) &= \frac{M}{N} \sum_{n=0}^{N-1} \sum_{r=-\infty}^{\infty} y[n] y[n + rM] \\
&= \sum_{n=0}^{N-1} y[n] \frac{1}{K} \sum_{r=-\infty}^{\infty} y[n + rM]
\end{aligned}$$

where we define $y[n] = 0$ for $n < 0$ and $n > N - 1$. But

$$\hat{s}[n] = \frac{1}{K} \sum_{r=-\infty}^{\infty} y[n + rM]$$

for $n = 0, 1, \ldots, N-1$ is an estimate of $s[n]$ obtained by averaging $y[n]$ over all available periods for the nth sample. For example, if $n = 0$, we have $\hat{s}[0] = (1/K) \sum_{r=0}^{K-1} y[rM]$. Note also that $\hat{s}[n]$ is periodic with period M as required. Hence,

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} y[n] \hat{s}[n] \tag{8.33}$$

which is recognized as an estimator-correlator. This implementation is shown in Figure 8.4.

The averager-energy detector implementation follows by invoking the periodicity of $\hat{s}[n]$. We have from (8.33)

$$T(\mathbf{x}) = \sum_{r=0}^{K-1} \sum_{n=0}^{M-1} y[n + rM] \hat{s}[n + rM]$$

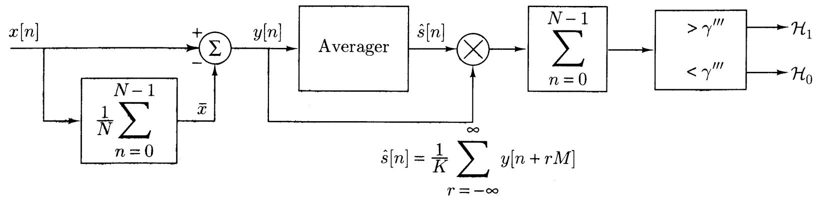


Figure 8.4. Estimator-correlator for detection of periodic Gaussian random signal in WGN.

$$
\begin{aligned}
&= \sum_{r=0}^{K-1}\sum_{n=0}^{M-1} y[n+rM]\hat{s}[n] \\
&= K\sum_{n=0}^{M-1} \hat{s}[n]\frac{1}{K}\sum_{r=0}^{K-1} y[n+rM] \\
&= K\sum_{n=0}^{M-1} \hat{s}^2[n]
\end{aligned}
$$

where

$$
\hat{s}[n] = \frac{1}{K}\sum_{r=0}^{K-1} y[n+rM] \qquad n = 0, 1, \ldots, M-1. \tag{8.34}
$$

Here we estimate the signal over one period by averaging and follow by an energy detector (see also Problem 8.16). This is shown in Figure 8.5.

As an example, consider a WSS Gaussian random process consisting of four harmonically related sinusoids with Rayleigh distributed amplitudes and uniformly

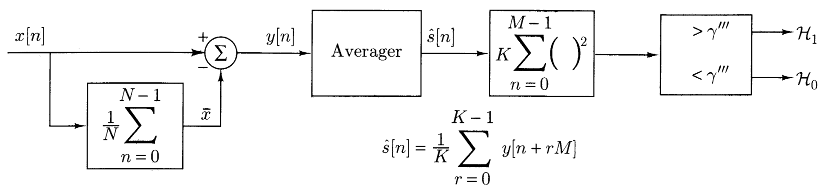


Figure 8.5. Averager-energy detector for periodic Gaussian random signal in WGN.

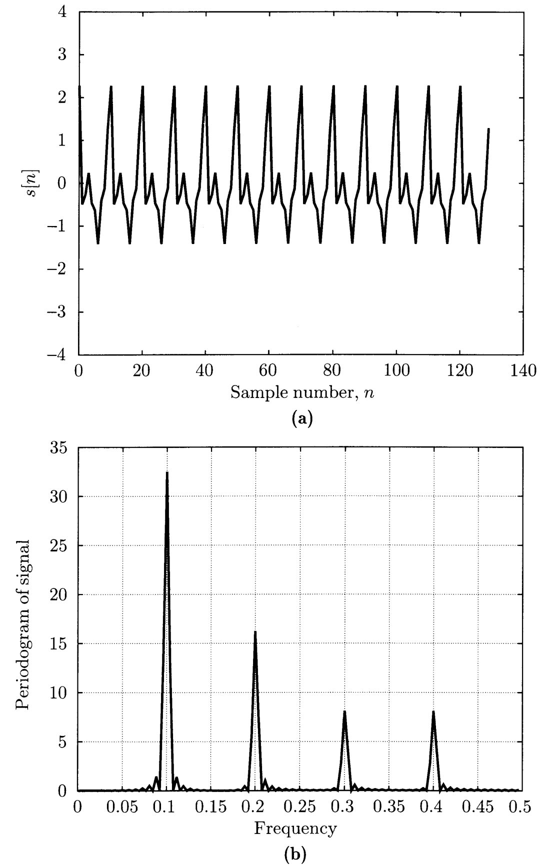


Figure 8.6. Realization of periodic Gaussian random signal (a) Time series (b) Periodogram.

distributed phases, all of which are independent. A realization of the resulting periodic signal is shown in Figure 8.6a. The actual realization is given by

$$
\begin{aligned}
s[n] &= \cos(2\pi f_0 n) + \frac{1}{\sqrt{2}}\cos(2\pi(2f_0)n + \pi/3) \\
&\quad + \frac{1}{2}\cos(2\pi(3f_0)n + \pi/7) + \frac{1}{2}\cos(2\pi(4f_0)n + \pi/9)
\end{aligned}
$$

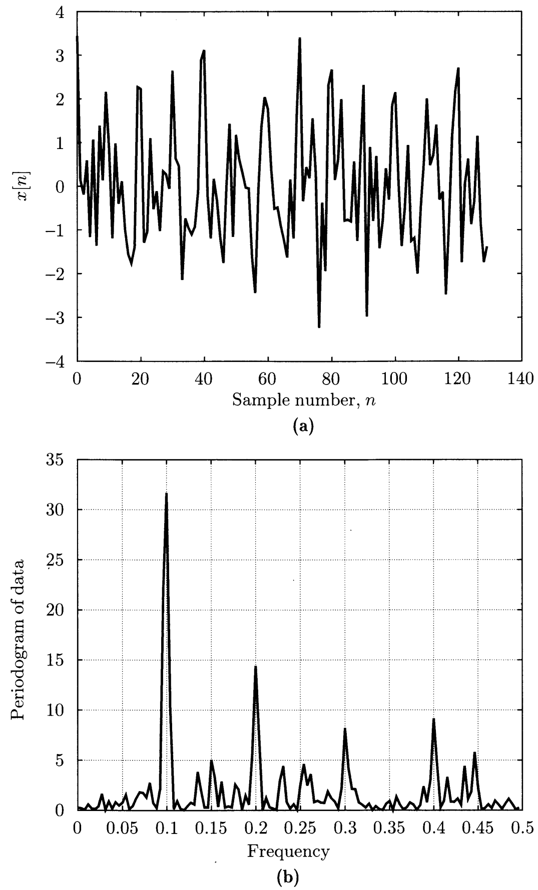


Figure 8.7. Realization of periodic Gaussian random signal in WGN (a) Time series (b) Periodogram.

for $n = 0, 1, \ldots, N-1 = 129$. The fundamental frequency is $f_0 = 1/10$ so that the period is $M = 10$. There are $K = 13$ periods of the signal process. The periodogram of the signal is shown in Figure 8.6b. As expected, there is power at the frequencies $kf_0 = 0.1k$ for $k = 1, 2, 3, 4$. Adding WGN with $\sigma^2 = 1$ to the signal produces the data $x[n]$ shown in Figure 8.7a, with the corresponding periodogram displayed in Figure 8.7b. The signal averager given by (8.34) is used to estimate the signal

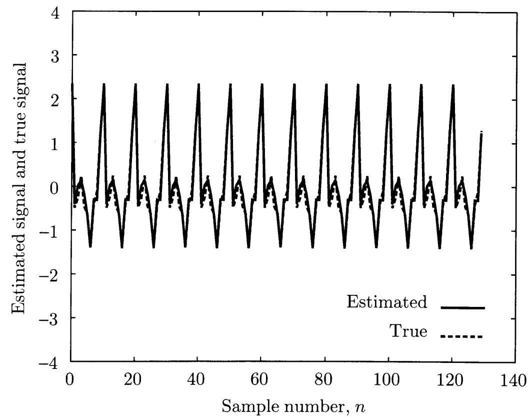


Figure 8.8. Estimated signal and true signal.

over the initial period ($[0, M-1] = [0, 9]$). This estimate is then replicated over the remaining 12 periods to yield the signal estimate shown in Figure 8.8. The solid curve indicates the estimated signal, while the dashed curve is the true signal. Since the SNRs are relatively high in this example, the signal is easily estimated and would be easily detectable.

The question as to what to do if the period M is unknown is of considerable practical interest. It might be supposed that we could implement the test statistic of (8.31) or $T(\mathbf{x}) = \sum_{i=1}^{L} I(f_i)$ for different assumed periods and then choose the maximum as our test statistic. In doing so, we are computing

$$T(\mathbf{x}) = \sum_{i=1}^{\frac{M}{2}-1} I\left(\frac{i}{M}\right) \tag{8.35}$$

for different values of M. For this example the test statistic versus M is shown in Figure 8.9. It is seen that the statistic does have a local maximum at the correct value of $M = 10$ but the maximum value occurs at $M = 40$, a multiple of the true value. This is because the test statistic of (8.35) sums the power at frequencies $1/M, 2/M, \ldots, M/2-1$, so that as M increases the statistic will always be larger at a multiple of M. This is a manifestation of the fact that the MLE is not consistent for this problem. As M increases, we are estimating more parameters (recall the P_i's) and therefore there is no averaging. To remedy this situation we can modify the GLRT with a factor that causes the MLE to be consistent. Based on the minimum

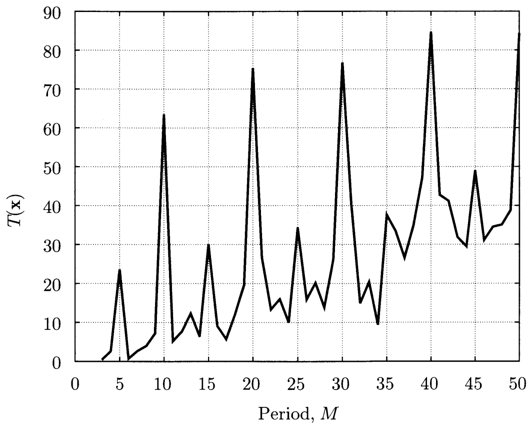


Figure 8.9. GLRT statistic as a function of assumed signal period.

description length idea discussed in Chapter 6 we should decide $\mathcal{H}_1$ if (see (6.41))

$$\max_M \left[\ln \frac{p(\mathbf{x}; \hat{\mathbf{P}}, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} - \frac{\frac{M}{2} - 1}{2} \ln N \right] > \gamma.$$

Note that the number of parameters estimated is $M/2 - 1$ so that the additional term forms a *penalty factor* with the penalty increasing as M increases. Using the GLRT of (8.30) we decide $\mathcal{H}_1$ if

$$T'(\mathbf{x}) = \max_M \left[\sum_{\substack{i=1 \\ I(i/M)/\sigma^2 > 1}}^{\frac{M}{2}-1} \left(\frac{I(\frac{i}{M})}{\sigma^2} - \ln \frac{I(\frac{i}{M})}{\sigma^2} - 1 \right) - \frac{\frac{M}{2} - 1}{2} \ln N \right] > \gamma.$$

In Figure 8.10 the term in brackets is plotted versus M. As desired the effect of the penalty term is to make the MLE consistent. The maximum now occurs at the correct period. The reader may also wish to consult [Wise, Caprio, and Parks, 1976] and [Eddy 1980] for further approaches to this problem.

◇

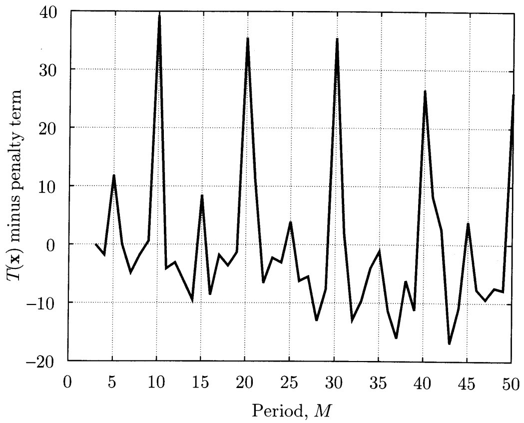


Figure 8.10. Modified GLRT statistic as a function of assumed signal period.

References

Chernoff, H., "On the Distribution of the Likelihood Ratio," *Ann. Math. Statist.*, Vol. 25, pp. 573–578, 1954.

Eddy, T.W., "Maximum Likelihood Detection and Estimation for Harmonic Sets," *J. Acoust. Soc. Am.*, Vol. 68, pp. 149–155, 1980.

Kay, S.M., *Modern Spectral Estimation: Theory and Application*, Prentice-Hall, Englewood Cliffs, N.J., 1988.

Levin, M.J.,"Power Spectrum Parameter Estimation," *IEEE Trans. Inform. Theory*, Vol. IT-11, pp. 100-107, 1965.

Rabiner, L.R., R.W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, Englewood Cliffs, N.J., 1978.

Urick, R.J., *Principles of Underwater Sound*, McGraw-Hill, New York, 1975.

Wise, J.D., J.R. Caprio, T.W.Parks, "Maximum Likelihood Pitch Estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, Vol. ASSP-24, pp.418–423, 1976.

Problems

8.1 If a random Gaussian signal $s[n]$ embedded in WGN with variance σ^2 has the signal covariance matrix given by (8.2), find the equation to be minimized that yields the MLE for $r_{ss}[0]$.

8.2 Show that $J(P_0)$ as given by (8.6) is monotonically increasing for $P_0 > P_0^+$ and monotonically decreasing for $P_0 < P_0^+$. Then, argue that the minimum of $J(P_0)$ for $0 \leq P_0 < \infty$ is given by (8.7).

8.3 Referring to the problem described in Example 8.2, the GLRT decides $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x}; \hat{P}_0, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > \gamma.$$

Argue why we must have $\gamma > 1$ to avoid $P_{FA} = 1$. To do so note that $p(\mathbf{x}; \mathcal{H}_0) = p(\mathbf{x}; P_0 = 0, \mathcal{H}_1)$.

8.4 Consider the MLE of P_0 for Example 8.2 as given by (8.7). First show that under $\mathcal{H}_0$, $\hat{P}_0^+ \sim \mathcal{N}(0, 2\sigma^4/(N\lambda^2))$ for large data records. Then, find the asymptotic PDF of the MLE $\hat{P}_0$ under $\mathcal{H}_0$. Hint: Use the central limit theorem to find the asymptotic PDF of $(1/N)\sum_{n=0}^{N-1} x^2[n]$.

8.5 From Example 5.5 the ACF of a Rayleigh fading sinusoid is $r_{ss}[k] = \sigma_s^2 \cos 2\pi f_0 k$. If $0 < f < 1/2$, show that the 3×3 covariance matrix of the signal has rank two. To do so show that

$$\det\left[r_{ss}[0]\right] \neq 0$$

$$\det\begin{bmatrix} r_{ss}[0] & r_{ss}[1] \\ r_{ss}[1] & r_{ss}[0] \end{bmatrix} \neq 0$$

but

$$\det \mathbf{C}_s = \det\begin{bmatrix} r_{ss}[0] & r_{ss}[1] & r_{ss}[2] \\ r_{ss}[1] & r_{ss}[0] & r_{ss}[1] \\ r_{ss}[2] & r_{ss}[1] & r_{ss}[0] \end{bmatrix} = 0.$$

This can be generalized to show that the $N \times N$ matrix $\mathbf{C}_s$ has rank two. See Problem 8.6.

8.6 For the Rayleigh fading sinusoidal model of Example 5.5 the $N \times N$ signal covariance matrix is

$$[\mathbf{C}_s]_{mn} = \sigma_s^2 \cos[2\pi f_0(m-n)].$$

Assuming that $0 < f_0 < 1/2$ show that

$$\mathbf{C}_s = \sigma_s^2 \mathbf{c}\mathbf{c}^T + \sigma_s^2 \mathbf{s}\mathbf{s}^T$$

where

$$\begin{aligned} \mathbf{c} &= [1 \ \cos 2\pi f_0 \ldots \cos[2\pi f_0(N-1)]]^T \\ \mathbf{s} &= [0 \ \sin 2\pi f_0 \ldots \sin[2\pi f_0(N-1)]]^T. \end{aligned}$$

Now, if $f_0 = k/N$ for $k = 1, 2, \ldots, N/2 - 1$, show that $\mathbf{c}^T\mathbf{s} = 0$ and hence that $\mathbf{v}_1 = \mathbf{c}/(\sqrt{N/2})$, $\mathbf{v}_2 = \mathbf{s}/(\sqrt{N/2})$. Also, find the eigenvalues. Hint: Use

$$\begin{aligned} \sum_{n=0}^{N-1} \sin \alpha n &= \operatorname{Im}\left(\sum_{n=0}^{N-1} \exp(j\alpha n)\right) \\ &= \operatorname{Im}\left(\exp[j(N-1)\alpha/2]\frac{\sin N\frac{\alpha}{2}}{\sin\frac{\alpha}{2}}\right). \end{aligned}$$

8.7 Consider the detection problem of (8.1). If $s[n] = Ar^n$ for $n = 0, 1, \ldots, N-1$, where $A \sim \mathcal{N}(0, \sigma_A^2)$ and σ_A^2 is unknown, find the GLRT statistic. Hint: Apply the results of Example 8.3.

8.8 For Example 8.3 the GLRT for a random DC level in WGN is to decide $\mathcal{H}_1$ if $\bar{x}^2 > \gamma''/N$. Determine P_{FA} and P_D for this detector.

8.9 Consider the detection problem of (8.1). Assume that $s[n]$ is WSS with $r_{ss}[k] = P_0 \cos 2\pi f_0 k$ where P_0 is unknown and $f_0 = k/N$ for $k = 1, 2, \ldots, N/2 - 1$. Find the GLRT statistic as follows. First note that according to Problem 8.6

$$\mathbf{C}_s = P_0 \mathbf{c}\mathbf{c}^T + P_0 \mathbf{s}\mathbf{s}^T$$

so that $\mathbf{C} = \mathbf{c}\mathbf{c}^T + \mathbf{s}\mathbf{s}^T$ and thus $\mathbf{v}_1 = \mathbf{c}/\sqrt{N/2}$, $\mathbf{v}_2 = \mathbf{s}/\sqrt{N/2}$, $\lambda_1 = \lambda_2 = N/2$. Then show that

$$\hat{P}_0 = \max\left(0, \frac{2}{N}(I(f_0) - \sigma^2)\right).$$

As in Example 8.3 it may be shown that the GLRT decides $\mathcal{H}_1$ if $\hat{P}_0 > \gamma'$. This leads to the GLRT, which decides $\mathcal{H}_1$ if $I(f_0) > \gamma''$.

8.10 Using the large data record approximation of Section 8.4, find the GLRT statistic if $s[n]$ is WSS with low-pass PSD

$$P_{ss}(f; \delta) = P((1+\delta)f)$$

where $-\delta_0 < \delta < \delta_0$, $\delta_0 \ll 1$, and δ is unknown. The PSD $P(f)$ is known and assumed to be nonzero only over the frequency range $|f| \leq f_1$. Furthermore, $(1+\delta_0)f_1 < 1/2$. This type of signal is a model for the effects of Doppler on a signal transmitted from a moving platform. The factor δ is referred to as the *dilation factor* and gives rise to a compression ($\delta < 0$) or expansion ($\delta > 0$) of the transmitted waveform. Hint: Show that (8.19) applies, assuming that $\delta_0 \ll 1$.

8.11 Using the large data record approximation of Section 8.4, find the GLRT if $s[n]$ is WSS with

$$P_{ss}(f;P_0) = \begin{cases} 2P_0 & 0 \le f \le \frac{1}{4} \\ 0 & \frac{1}{4} < f \le \frac{1}{2}. \end{cases}$$

P_0 is assumed to be unknown. Explain your results. Hint: Note that $g(x) = x - \ln x - 1$ is monotonically increasing for $x > 1$.

8.12 Find P_{FA} and P_D for the LMP detector of (8.25) by invoking the central limit theorem for large N. Hence, assume that $\mathbf{x}^T\mathbf{C}\mathbf{x}$ is Gaussian. Hint: Use the formulas for $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_x)$ [Kay-I 1993, pg. 76]

$$\begin{aligned} E(\mathbf{x}^T\mathbf{A}\mathbf{x}) &= \operatorname{tr}(\mathbf{A}\mathbf{C}_x) \\ \operatorname{var}(\mathbf{x}^T\mathbf{A}\mathbf{x}) &= 2\operatorname{tr}\left[(\mathbf{A}\mathbf{C}_x)^2\right]. \end{aligned}$$

8.13 If $s[n] = \sum_{i=1}^{L} A_i \cos(2\pi f_i n + \phi_i)$ where each A_i is a Rayleigh random variable with $E(A_i^2/2) = P_i$ and each $\phi_i \sim \mathcal{U}[0, 2\pi]$ and all the random variables are independent, show that $r_{ss}[k] = \sum_{i=1}^{L} P_i \cos 2\pi f_i k$.

8.14 Verify (8.29) by using the results of Problem 8.2.

8.15 Verify (8.32) by noting that $N = KM$ and also that

$$\sum_{l=0}^{N-1} \exp\left(-j2\pi \frac{i}{M} l\right) = \begin{cases} KM & i = 0 \\ 0 & i = 1, 2, \ldots, \frac{M}{2} - 1. \end{cases}$$

Hint: Use the identity that for $N = KM$

$$\sum_{n=0}^{N-1} a_n = \sum_{m=0}^{K-1} \sum_{r=0}^{M-1} a_{r+mM}.$$

8.16 In this problem we show that an averager is equivalent to a comb filter. The impulse response of an FIR averager that averages K samples spaced M samples apart is

$$h[n] = \frac{1}{K} \sum_{r=0}^{K-1} \delta[n - rM].$$

Show that the output of this filter to an input $y[n]$, where $y[n] = 0$ for $n < 0$, at $n = N - 1$ is

$$z[N-1] = \frac{1}{K} \sum_{r=0}^{K-1} y[(M-1) + rM] = \hat{s}[M-1]$$

if $N = KM$. In a similar fashion it can be shown that

$$z[N-1-k] = \frac{1}{K} \sum_{r=0}^{K-1} y[(M-1-k) + rM] = \hat{s}[M-1-k]$$

for $k = 0, 1, \ldots, M-1$. Find the frequency response of this filter and plot its magnitude using the hint in Problem 8.6. Why do you think it is called a *comb filter?*

Appendix 8A

Derivation of PDF for Periodic Gaussian Random Process

From (8.27) the $[m, n]$ element of the $N \times N$ signal covariance matrix can be written as

$$\begin{aligned}
[\mathbf{C}_s]_{mn} &= r_{ss}[m-n] \\
&= \sum_{\substack{i=-L \\ i\neq 0}}^{L} \frac{P_i}{2} \exp\left(j2\pi \frac{i}{M} m\right) \exp\left(-j2\pi \frac{i}{M} n\right) \\
&= \sum_{\substack{i=-L \\ i\neq 0}}^{L} \frac{P_i}{2} [\mathbf{e}_i \mathbf{e}_i^H]_{mn}
\end{aligned}$$

where $\mathbf{e}_i = [1 \ \exp[j(2\pi/M)i] \ldots \exp[j(2\pi/M)i(N-1)]]^T$ and H denotes complex conjugate transpose so that

$$\mathbf{C}_s = \sum_{\substack{i=-L \\ i\neq 0}}^{L} \frac{P_i}{2} \mathbf{e}_i \mathbf{e}_i^H. \tag{8A.1}$$

The vectors $\mathbf{e}_i$ are orthogonal to each other. Hence, they are the eigenvectors (to within a scale factor of $1/\sqrt{N}$) of $\mathbf{C}_s$ with nonzero eigenvalues. This is the critical assumption that allows us to easily find the PDF. To show that this is true consider for $r, s = \pm 1, \pm 2, \ldots, \pm M/2 - 1 = \pm L$

$$\mathbf{e}_r^H \mathbf{e}_s = \sum_{n=0}^{N-1} \exp\left[j\frac{2\pi}{M}(s-r)n\right]$$

$$= \sum_{m=0}^{K-1} \sum_{l=0}^{M-1} \exp\left[j\frac{2\pi}{M}(s-r)(l+mM)\right]$$

since $N = KM$. But

$$\begin{aligned}
\mathbf{e}_r^H \mathbf{e}_s &= \sum_{m=0}^{K-1} \sum_{l=0}^{M-1} \exp\left[j\frac{2\pi}{M}(s-r)l\right] \\
&= K \sum_{l=0}^{M-1} \exp\left[j\frac{2\pi}{M}(s-r)l\right] \\
&= \begin{cases} KM & \text{if } r = s \\ 0 & \text{if } r \neq s. \end{cases}
\end{aligned}$$

This follows from the orthogonality of the DFT sinusoids (see Appendix 1). Hence, if we define

$$\begin{aligned}
\mathbf{E} &= [\ \mathbf{e}_{-L} \ \ \mathbf{e}_{-L+1} \ \ldots \ \mathbf{e}_{-1} \ \ \mathbf{e}_1 \ \ldots \ \mathbf{e}_L\ ] & N \times 2L \\
\mathbf{D} &= \operatorname{diag}\left(\frac{P_{-L}}{2}, \frac{P_{-L+1}}{2}, \ldots, \frac{P_{-1}}{2}, \frac{P_1}{2}, \ldots, \frac{P_L}{2}\right) & 2L \times 2L
\end{aligned}$$

then from (8A.1)

$$\mathbf{C}_s = \mathbf{E}\mathbf{D}\mathbf{E}^H$$

and due to the orthogonality of the columns of $\mathbf{E}$, $\mathbf{E}^H\mathbf{E} = N\mathbf{I}$. The covariance matrix of $x[n]$ is

$$\mathbf{C}_x = \mathbf{C}_s + \sigma^2\mathbf{I} = \mathbf{E}\mathbf{D}\mathbf{E}^H + \sigma^2\mathbf{I}.$$

Using the matrix inversion lemma we have

$$\begin{aligned}
\mathbf{C}_x^{-1} &= (\sigma^2\mathbf{I} + \mathbf{E}\mathbf{D}\mathbf{E}^H)^{-1} \\
&= \frac{1}{\sigma^2}\mathbf{I} - \frac{1}{\sigma^4}\mathbf{E}\left(\mathbf{E}^H\frac{1}{\sigma^2}\mathbf{E} + \mathbf{D}^{-1}\right)^{-1}\mathbf{E}^H \\
&= \frac{1}{\sigma^2}\mathbf{I} - \frac{1}{\sigma^4}\mathbf{E}\left(\frac{N}{\sigma^2}\mathbf{I} + \mathbf{D}^{-1}\right)^{-1}\mathbf{E}^H \\
&= \frac{1}{\sigma^2}\mathbf{I} - \frac{1}{\sigma^4}\sum_{\substack{i=-L \\ i\neq 0}}^{L} \frac{\mathbf{e}_i\mathbf{e}_i^H}{\frac{N}{\sigma^2} + \frac{1}{P_i/2}} \\
&= \frac{1}{\sigma^2}\left(\mathbf{I} - \sum_{\substack{i=-L \\ i\neq 0}}^{L} \frac{P_i/2}{NP_i/2 + \sigma^2}\mathbf{e}_i\mathbf{e}_i^H\right).
\end{aligned}$$

Next we find the determinant.

$$\begin{aligned}\det(\mathbf{C}_x) &= \det(\sigma^2\mathbf{I}+\mathbf{EDE}^H)\\ &= \sigma^{2N}\det(\mathbf{I}+\frac{1}{\sigma^2}\mathbf{EDE}^H).\end{aligned}$$

Using the identity

$$\det(\mathbf{I}_K+\mathbf{A}_{KL}\mathbf{B}_{LK})=\det(\mathbf{I}_L+\mathbf{B}_{LK}\mathbf{A}_{LK})$$

where K, L indicate the dimensions of the matrix, we have

$$\begin{aligned}\det(\mathbf{C}_x) &= \sigma^{2N}\det\left(\mathbf{I}+\frac{1}{\sigma^2}\mathbf{E}^H\mathbf{ED}\right)\\ &= \sigma^{2N}\det\left(\mathbf{I}+\frac{N}{\sigma^2}\mathbf{D}\right)\\ &= \sigma^{2N}\prod_{\substack{i=-L\\ i\neq 0}}^{L}\left(1+\frac{NP_i/2}{\sigma^2}\right).\end{aligned}$$

The inverse and determinant can also be found by applying the formulas of Problem 2.13. With these results the log-PDF under $\mathcal{H}_1$ is

$$\begin{aligned}\ln p(\mathbf{x};\mathbf{P},\mathcal{H}_1) &= -\frac{N}{2}\ln 2\pi-\frac{1}{2}\ln\det(\mathbf{C}_x)-\frac{1}{2}\mathbf{x}^T\mathbf{C}_x^{-1}\mathbf{x}\\ &= -\frac{N}{2}\ln 2\pi-\frac{1}{2}\ln\sigma^{2N}-\frac{1}{2}\sum_{\substack{i=-L\\ i\neq 0}}^{L}\ln\left(\frac{NP_i/2}{\sigma^2}+1\right)\\ &\quad -\frac{1}{2}\frac{\mathbf{x}^T\mathbf{x}}{\sigma^2}+\frac{1}{2\sigma^2}\sum_{\substack{i=-L\\ i\neq 0}}^{L}\frac{P_i/2}{NP_i/2+\sigma^2}|\mathbf{e}_i^H\mathbf{x}|^2\\ &= -\frac{N}{2}\ln(2\pi\sigma^2)-\frac{1}{2}\frac{\mathbf{x}^T\mathbf{x}}{\sigma^2}\\ &\quad -\frac{1}{2}\sum_{\substack{i=-L\\ i\neq 0}}^{L}\left[\ln\left(\frac{NP_i/2}{\sigma^2}+1\right)-\frac{P_i/2}{NP_i/2+\sigma^2}\frac{|\mathbf{e}_i^H\mathbf{x}|^2}{\sigma^2}\right]\end{aligned}$$

or finally

$$\begin{aligned}\ln p(\mathbf{x};\mathbf{P},\mathcal{H}_1) &= -\frac{N}{2}\ln(2\pi\sigma^2)-\frac{1}{2}\frac{\mathbf{x}^T\mathbf{x}}{\sigma^2}\end{aligned}$$

$$\begin{aligned}&\quad -\frac{1}{2}\sum_{\substack{i=-L\\ i\neq 0}}^{L}\left[\ln\left(\frac{NP_i/2}{\sigma^2}+1\right)-\frac{NP_i/2}{NP_i/2+\sigma^2}\frac{I(f_i)}{\sigma^2}\right]\\ &= -\frac{N}{2}\ln(2\pi\sigma^2)-\frac{1}{2}\frac{\mathbf{x}^T\mathbf{x}}{\sigma^2}\\ &\quad -\sum_{i=1}^{L}\left[\ln\left(\frac{NP_i/2}{\sigma^2}+1\right)-\frac{NP_i/2}{NP_i/2+\sigma^2}\frac{I(f_i)}{\sigma^2}\right]\end{aligned}$$

where

$$I(f_i)=\frac{1}{N}\left|\sum_{n=0}^{N-1}x[n]\exp(-j2\pi f_i n)\right|^2.$$

Chapter 9

Unknown Noise Parameters

9.1 Introduction

The problem of detecting a signal in Gaussian noise when the PDF of the noise is known except for a few parameters, is of considerable practical interest. It is very seldom that the statistical characterization of the noise is completely known a priori. Typically, if the noise is WGN, the power may be unknown or if the noise is colored, the PSD may not be completely known. Matters are further complicated when the signal is also uncertain. In either case, the resultant hypothesis test has unknown parameters under *both* $\mathcal{H}_0$ and $\mathcal{H}_1$ (due to the noise parameters). It is termed the *doubly composite* hypothesis testing problem. In this chapter, we explore a number of detectors, as well as their performance, that are applicable to this problem.

9.2 Summary

The case of WGN with unknown variance is first discussed. A constant false alarm rate detector is defined in Section 9.3. The GLRT for a known deterministic signal is derived in Section 9.4.1 and is shown not to possess the constant false alarm rate property. For a random signal with a known PDF, the GLRT is difficult to implement, as shown in Section 9.4.2. For a deterministic signal with unknown parameters in the form of the linear model, the GLRT and its exact performance are given in Theorem 9.1. Colored WSS Gaussian noise with unknown power spectral density parameters is examined in Section 9.5. When the noise is modeled as an autoregressive process of order one, the ACF and PSD are given by (9.18) and (9.19), respectively. For a known signal in autoregressive noise the GLRT produces the detector of (9.22), which is not constant false alarm rate. However, the Rao test for a signal described by the linear model in zero mean Gaussian noise with an unknown covariance matrix produces the test of (9.36). Its asymptotic performance is given in Theorem 9.2 and produces a constant false alarm rate detector. Furthermore, the asymptotic performance is as good as if the noise parameters were known. Finally,

in Section 9.6 the detection of a linear FM signal in autoregressive noise produces the Rao test of (9.40) and an asymptotic performance of (9.39) and (9.41). The detector is shown to outperform a matched filter.

9.3 General Considerations

In designing detectors when the noise PDF is incompletely known, we encounter a new problem than when only signal parameters are unknown. A primary consideration (and the basis for the NP detector) was to be able to constrain the P_{FA}. Recall that the threshold γ' is found from the requirement that $P_{FA} = \alpha$, which is defined as

$$\begin{aligned} P_{FA} &= \Pr\{T(\mathbf{x}) > \gamma'; \mathcal{H}_0\} \\ &= \int_{\gamma'}^{\infty} p(T; \mathcal{H}_0)\, dT \end{aligned}$$

where $p(T; \mathcal{H}_0)$ is the PDF of the test statistic under $\mathcal{H}_0$. If, however, the PDF of $T(\mathbf{x})$ is not completely known, then we cannot determine the threshold γ'. As an example, for the detection of a known DC level (with $A > 0$) in WGN (see Example 3.2) we decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \frac{1}{N}\sum_{n=0}^{N-1} x[n] > \frac{\sigma^2}{NA}\ln\gamma + \frac{A}{2} = \gamma'. \tag{9.1}$$

Under $\mathcal{H}_0$ we have $T(\mathbf{x}) \sim \mathcal{N}(0, \sigma^2/N)$ so that

$$P_{FA} = Q\left(\frac{\gamma'}{\sqrt{\sigma^2/N}}\right) \tag{9.2}$$

and thus $\gamma' = \sqrt{\sigma^2/N}Q^{-1}(P_{FA})$. Clearly, the threshold will depend on the WGN variance σ^2. If it is unknown, then the threshold cannot be determined. One possible approach would be to estimate σ^2, assuming $\mathcal{H}_0$ is true, and then let $\hat{\gamma}' = \sqrt{\hat{\sigma^2}/N}Q^{-1}(P_{FA})$. This approach, which is sometimes referred to as an *estimate and plug* detector, suffers from the property that the estimator

$$\hat{\sigma^2} = \frac{1}{N}\sum_{n=0}^{N-1} x^2[n]$$

will be biased when a signal is present. Under $\mathcal{H}_1$, we have that $E(\hat{\sigma^2}) = \sigma^2 + A^2$, which has the effect of inflating the threshold and thereby reducing P_D. Furthermore, the use of an estimated threshold in (9.2) will cause the true P_{FA} to be different than the computed P_{FA}. To avoid the effect of the signal induced bias, one can use additional or reference data samples, known to consist of noise only, and independent of the signal and noise data to form

$$\hat{\sigma^2} = \frac{1}{N}\sum_{n=0}^{N-1} w_R^2[n]$$

where $w_R[n]$ are the reference noise samples. The reference noise samples are assumed to be samples of WGN with variance σ^2. Then we could use

$$T(\mathbf{x}, \mathbf{w}_R) = \frac{\frac{1}{N}\sum_{n=0}^{N-1} x[n]}{\sqrt{\hat{\sigma^2}/N}}.$$

This test statistic can be shown not to depend on the actual value of σ^2 under $\mathcal{H}_0$. Hence, the threshold can be set. In fact, $T(\mathbf{x}, \mathbf{w}_R)$ has a Student's t PDF as shown in Problem 9.1. The denominator $\sqrt{\hat{\sigma^2}/N}$ is sometimes called a *normalization* factor. It yields a detector whose P_{FA} can be constrained by the choice of an appropriate threshold. This is because the PDF of the statistic under $\mathcal{H}_0$ does not depend on σ^2 (see also Problem 9.3). As a result, independent of the noise power, the threshold can be chosen to yield a detector with a constant false alarm probability. Such a detector is called a *constant false alarm rate* (CFAR) detector. It is not clear, however, if $T(\mathbf{x}, \mathbf{w}_R)$ is an optimal test statistic, in that it maximizes P_D.

Returning to the standard hypothesis testing problem of a known DC level in WGN we have

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = A + w[n] & n = 0, 1, \ldots, N-1 \end{array} \tag{9.3}$$

where A is known and $w[n]$ is WGN with *unknown* variance σ^2. The PDFs under $\mathcal{H}_0$ and $\mathcal{H}_1$ are denoted by $p(\mathbf{x}; \sigma^2, \mathcal{H}_0)$ and $p(\mathbf{x}; \sigma^2, \mathcal{H}_1)$, respectively. This is a composite hypothesis test for which there is no optimal solution. If we are willing to consider σ^2 to be a random variable, and assign to it a prior PDF, then the resulting Bayes detector can be said to be optimal (see Section 6.4.1). Otherwise, a GLRT will usually produce good results. We now illustrate the GLRT approach for this example. The Bayesian approach was considered in Problem 6.6. We will also apply the Rao test (see Section 6.6), which has the advantage of not requiring the MLE of the signal parameters, when they too are unknown.

Example 9.1 - GLRT for DC Level in WGN with Unknown Variance

The detection problem of (9.3) is examined when A is known and $A > 0$. Then, the GLRT decides $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x}; \hat{\sigma_1^2}, \mathcal{H}_1)}{p(\mathbf{x}; \hat{\sigma_0^2}, \mathcal{H}_0)} > \gamma$$

where $\hat{\sigma_i^2}$ is the MLE of σ^2 under $\mathcal{H}_i$. Note that in contrast to the estimate and plug detector *the GLRT estimates σ^2 under both hypotheses.* Also, because of this, the resultant detector will in general be different than the estimate and plug detector. Now the MLEs can be shown to be [Kay-I 1993, pp. 176–177]

$$\begin{aligned} \hat{\sigma_0^2} &= \frac{1}{N}\sum_{n=0}^{N-1} x^2[n] \\ \hat{\sigma_1^2} &= \frac{1}{N}\sum_{n=0}^{N-1} (x[n]-A)^2 \end{aligned}$$

so that

$$\begin{aligned} L_G(\mathbf{x}) &= \frac{p(\mathbf{x}; \hat{\sigma_1^2}, \mathcal{H}_1)}{p(\mathbf{x}; \hat{\sigma_0^2}, \mathcal{H}_0)} \\ &= \frac{\frac{1}{(2\pi\hat{\sigma_1^2})^{\frac{N}{2}}} \exp\left[-\frac{1}{2\hat{\sigma_1^2}}\sum_{n=0}^{N-1}(x[n]-A)^2\right]}{\frac{1}{(2\pi\hat{\sigma_0^2})^{\frac{N}{2}}} \exp\left[-\frac{1}{2\hat{\sigma_0^2}}\sum_{n=0}^{N-1}x^2[n]\right]} \\ &= \frac{\frac{1}{(2\pi\hat{\sigma_1^2})^{\frac{N}{2}}} \exp\left(-\frac{N}{2}\right)}{\frac{1}{(2\pi\hat{\sigma_0^2})^{\frac{N}{2}}} \exp\left(-\frac{N}{2}\right)} \\ &= \left(\frac{\hat{\sigma_0^2}}{\hat{\sigma_1^2}}\right)^{\frac{N}{2}}. \end{aligned}$$

Hence, we decide $\mathcal{H}_1$ if

$$\frac{\hat{\sigma_0^2}}{\hat{\sigma_1^2}} > \gamma^{\frac{2}{N}}.$$

An equivalent test statistic that is slightly more intuitive is

$$\begin{aligned} T(\mathbf{x}) &= \frac{1}{2A}\left(\frac{\hat{\sigma_0^2}}{\hat{\sigma_1^2}} - 1\right) \\ &= \frac{\bar{x} - A/2}{\frac{1}{N}\sum_{n=0}^{N-1}(x[n]-A)^2} \end{aligned} \tag{9.4}$$

where we have used $\hat{\sigma_1^2} = \hat{\sigma_0^2} - 2A\bar{x} + A^2$. Note that the GLRT automatically introduces a normalizing factor (although it is correct only under $\mathcal{H}_1$). It can be shown (see Problem 9.5) that for weak signals or $A \to 0$ and hence $N \to \infty$ that

$$T(\mathbf{x}) \overset{a}{\sim} \begin{cases} \mathcal{N}\left(-\frac{A}{2\sigma^2}, \frac{1}{N\sigma^2}\right) & \text{under } \mathcal{H}_0 \\ \mathcal{N}\left(\frac{A}{2\sigma^2}, \frac{1}{N\sigma^2}\right) & \text{under } \mathcal{H}_1. \end{cases} \tag{9.5}$$

Unfortunately, under $\mathcal{H}_0$ the PDF of $T(\mathbf{x})$ depends on σ^2, and therefore, the threshold cannot be set. This example illustrates that in the doubly composite hypothesis tesing problem the GLRT may not be a CFAR detector, even asymptotically. (Note that in this example the conditions necessary for the asymptotic PDF of the GLRT to hold, and thus the PDF under $\mathcal{H}_0$ not to depend on σ^2, are violated (see Problem 9.6)).

◇

A more common detection problem is when, in addition to σ^2 being unknown, the DC level A is also unknown. In this case the GLRT is an asymptotically CFAR detector. In fact, we have already addressed this problem in Example 6.5. There we saw that the GLRT statistic was

$$T(\mathbf{x}) = 2\ln L_G(\mathbf{x}) = N \ln\left(1 + \frac{\bar{x}^2}{\hat{\sigma_1^2}}\right)$$

where $\hat{\sigma_1^2} = (1/N)\sum_{n=0}^{N-1}(x[n] - \bar{x})^2$. Under $\mathcal{H}_0$ the asymptotic PDF of $2\ln L_G(\mathbf{x})$ was shown to be independent of σ^2. Furthermore, in Example 6.7 it was explicitly shown that

$$T(\mathbf{x}) \overset{a}{\sim} \begin{cases} \chi_1^2 & \text{under } \mathcal{H}_0 \\ \chi_1'^2(\lambda) & \text{under } \mathcal{H}_1 \end{cases}$$

where $\lambda = NA^2/\sigma^2$, verifying the CFAR property. (The asymptotically equivalent Rao test for this problem was studied in Example 6.10). In general, if the detection problem can be recast as the parameter testing problem of Section 6.5 or

$$\begin{aligned} \mathcal{H}_0 &: \boldsymbol{\theta}_r = \boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s \\ \mathcal{H}_1 &: \boldsymbol{\theta}_r \neq \boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s \end{aligned}$$

where $\boldsymbol{\theta}_r$ has dimension $r \times 1$ ($r \geq 1$) and $\boldsymbol{\theta}_s$ is an $s \times 1$ ($s \geq 0$) vector of nuisance parameters, then the GLRT will be an asymptotically CFAR detector with $2\ln L_G(\mathbf{x}) \overset{a}{\sim} \chi_r^2$ under $\mathcal{H}_0$.

A point worth mentioning is that the GLRT for the doubly composite hypothesis testing problem is not found by maximizing the LRT (in contrast to (6.15)). This is because the PDF must be maximized under *both* $\mathcal{H}_0$ and $\mathcal{H}_1$ when there are unknown noise parameters.

In the next few sections we study the detection of signals in Gaussian noise with unknown parameters. A hierarchy of problems is summarized in Figure 9.1. We will build upon our results in Chapters 7 and 8 to now accommodate an incompletely known *noise* PDF. In Section 9.4 we examine WGN, while in Section 9.5 we extend the discussion to colored WSS Gaussian noise.

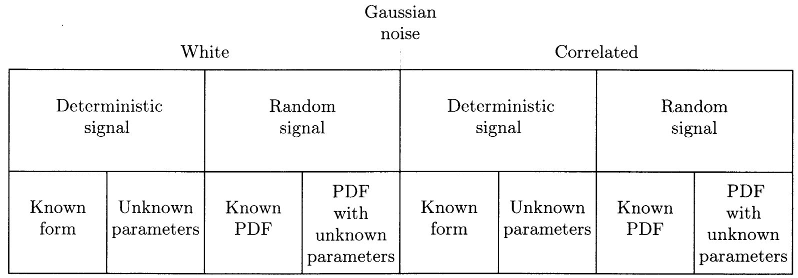


Figure 9.1. Hierarchy of detection problems in presence of unknown parameters.

9.4 White Gaussian Noise

If the noise is WGN, then the only possible unknown parameter is the noise variance σ^2. Thus, under $\mathcal{H}_0$ we need only estimate σ^2. We have already seen that the MLE is

$$\hat{\sigma_0^2} = \frac{1}{N}\sum_{n=0}^{N-1} x^2[n]$$

and thus

$$p(\mathbf{x}; \hat{\sigma_0^2}, \mathcal{H}_0) = \frac{1}{(2\pi\hat{\sigma_0^2})^{\frac{N}{2}}} \exp\left(-\frac{N}{2}\right).$$

Under $\mathcal{H}_1$, however, the MLE of σ^2 will depend upon the signal assumptions. We next determine the GLRT for the different signal possibilities described in Figure 9.1.

9.4.1 Known Deterministic Signal

The detection problem is

$$\begin{aligned} \mathcal{H}_0 &: x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 &: x[n] = s[n] + w[n] & n = 0, 1, \ldots, N-1 \end{aligned}$$

where $s[n]$ is a known deterministic signal. In this case there are no unknown signal parameters so that the GLRT decides $\mathcal{H}_1$ if

$$L_G(\mathbf{x}) = \frac{p(\mathbf{x}; \hat{\sigma_1^2}, \mathcal{H}_1)}{p(\mathbf{x}; \hat{\sigma_0^2}, \mathcal{H}_0)} > \gamma$$

where $\hat{\sigma_i^2}$ is the MLE of σ^2 under $\mathcal{H}_i$. Having already found the MLE under $\mathcal{H}_0$, we need only find $\hat{\sigma_1^2}$. But this is easily shown to be

$$\hat{\sigma_1^2} = \frac{1}{N}\sum_{n=0}^{N-1}(x[n] - s[n])^2$$

so that in a similar fashion to Example 9.1 we find that

$$L_G(\mathbf{x}) = \left(\frac{\hat{\sigma_0^2}}{\hat{\sigma_1^2}}\right)^{\frac{N}{2}}.$$

This generalizes Example 9.1 for which $s[n] = A$. Alternatively, if we consider the equivalent test statistic

$$\begin{aligned} T(\mathbf{x}) &= \frac{N}{2}\left(L_G(\mathbf{x})^{\frac{2}{N}} - 1\right) \\ &= \frac{N}{2}\frac{\hat{\sigma_0^2} - \hat{\sigma_1^2}}{\hat{\sigma_1^2}} \end{aligned}$$

we have

$$\begin{aligned} T(\mathbf{x}) &= \frac{\sum_{n=0}^{N-1} x^2[n] - \sum_{n=0}^{N-1}(x[n] - s[n])^2}{2\hat{\sigma_1^2}} \\ &= \frac{\sum_{n=0}^{N-1} x[n]s[n] - \frac{1}{2}\sum_{n=0}^{N-1} s^2[n]}{\hat{\sigma_1^2}}. \end{aligned}$$

When σ^2 is known, the NP test is

$$\frac{\sum_{n=0}^{N-1} x[n]s[n] - \frac{1}{2}\sum_{n=0}^{N-1} s^2[n]}{\sigma^2} > \ln\gamma$$

as shown in Section 4.3.1. The only difference is in the normalization factor. It can further be shown (in a similar fashion to Problem 9.5) that as $N \to \infty$

$$T(\mathbf{x}) \overset{a}{\sim} \begin{cases} \mathcal{N}\left(-\frac{\mathcal{E}}{2\sigma^2}, \frac{\mathcal{E}}{\sigma^2}\right) & \text{under } \mathcal{H}_0 \\ \mathcal{N}\left(\frac{\mathcal{E}}{2\sigma^2}, \frac{\mathcal{E}}{\sigma^2}\right) & \text{under } \mathcal{H}_1 \end{cases}$$

where $\mathcal{E} = \sum_{n=0}^{N-1} s^2[n]$ is the signal energy. Hence, the detector is not CFAR. One way to remedy this is to consider the statistic

$$T'(\mathbf{x}) = \frac{\sum_{n=0}^{N-1} x[n]s[n]}{\sqrt{\hat{\sigma_0^2}}}$$

as described in Problem 9.9.

9.4.2 Random Signal with Known PDF

We now assume that $s[n]$ is a Gaussian random process with known PDF. The detection problem then becomes

$$\begin{aligned} &\mathcal{H}_0 : x[n] = w[n] && n = 0, 1, \ldots, N-1 \\ &\mathcal{H}_1 : x[n] = s[n] + w[n] && n = 0, 1, \ldots, N-1 \end{aligned}$$

where $s[n]$ is a Gaussian random process with zero mean and *known* covariance matrix $\mathbf{C}_s$. The noise, as usual, is WGN with unknown variance σ^2. The PDF under $\mathcal{H}_1$ is

$$p(\mathbf{x}; \sigma^2, \mathcal{H}_1) = \frac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}(\mathbf{C}_s + \sigma^2\mathbf{I})} \exp\left[-\frac{1}{2}\mathbf{x}^T(\mathbf{C}_s + \sigma^2\mathbf{I})^{-1}\mathbf{x}\right].$$

In order to implement the GLRT we must find the MLE of σ^2 under $\mathcal{H}_1$. To do so we use the equivalent expression for the PDF based on an eigendecomposition of the covariance matrix (see Example 8.1). It follows that

$$\ln p(\mathbf{x}; \sigma^2, \mathcal{H}_1) = -\frac{N}{2}\ln 2\pi - \frac{1}{2}\sum_{i=1}^{N}\left[\ln(\lambda_{s_i} + \sigma^2) + \frac{(\mathbf{v}_i^T\mathbf{x})^2}{\lambda_{s_i} + \sigma^2}\right] \tag{9.6}$$

where $\mathbf{C}_s$ has eigenvectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N\}$ and corresponding eigenvalues $\{\lambda_{s_1}, \lambda_{s_2}, \ldots, \lambda_{s_N}\}$. The MLE of σ^2 is found by minimizing

$$J(\sigma^2) = \sum_{i=1}^{N}\left[\ln(\lambda_{s_i} + \sigma^2) + \frac{(\mathbf{v}_i^T\mathbf{x})^2}{\lambda_{s_i} + \sigma^2}\right]. \tag{9.7}$$

Differentiating and setting equal to zero produces

$$\sum_{i=1}^{N} \frac{\lambda_{s_i} + \sigma^2 - (\mathbf{v}_i^T \mathbf{x})^2}{(\lambda_{s_i} + \sigma^2)^2} = 0. \tag{9.8}$$

Unfortunately, this cannot be solved analytically. A special case occurs when the signal is weak or $\lambda_{s_i} \ll \sigma^2$ for all i, which allows the approximation to (9.8)

$$\sum_{i=1}^{N} \left(\lambda_{s_i} + \sigma^2 - (\mathbf{v}_i^T \mathbf{x})^2\right) = 0.$$

Solving this produces

$$\hat{\sigma_1^2} = \frac{1}{N} \sum_{i=1}^{N} \left[(\mathbf{v}_i^T \mathbf{x})^2 - \lambda_{s_i}\right]$$

which is equivalent to (see Problem 9.10)

$$\hat{\sigma_1^2} = \frac{1}{N} \sum_{n=0}^{N-1} x^2[n] - \frac{1}{N} \mathrm{tr}(\mathbf{C}_s). \tag{9.9}$$

Of course, since σ^2 must be nonnegative, the MLE becomes

$$\hat{\sigma_1^2} = \max\left(0, \frac{1}{N} \sum_{n=0}^{N-1} x^2[n] - \frac{1}{N} \mathrm{tr}(\mathbf{C}_s)\right). \tag{9.10}$$

(Recall that we are maximizing the PDF over $0 < \sigma^2 < \infty$.) The GLRT is from (9.6)

$$\begin{aligned}
2 \ln L_G(\mathbf{x}) &= 2 \ln p(\mathbf{x}; \hat{\sigma_1^2}, \mathcal{H}_1) - 2 \ln p(\mathbf{x}; \hat{\sigma_0^2}, \mathcal{H}_0) \\
&= -N \ln 2\pi - \sum_{i=1}^{N} \left[\ln(\lambda_{s_i} + \hat{\sigma_1^2}) + \frac{(\mathbf{v}_i^T \mathbf{x})^2}{\lambda_{s_i} + \hat{\sigma_1^2}}\right] \\
&\quad + N \ln 2\pi + N \ln \hat{\sigma_0^2} + N \\
&= \sum_{i=1}^{N} \left[\ln \frac{\hat{\sigma_0^2}}{\lambda_{s_i} + \hat{\sigma_1^2}} - \frac{(\mathbf{v}_i^T \mathbf{x})^2}{\lambda_{s_i} + \hat{\sigma_1^2}} + 1\right]
\end{aligned}$$

or finally we decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \sum_{i=1}^{N} \left[\ln \frac{\hat{\sigma_0^2}}{\lambda_{s_i} + \hat{\sigma_1^2}} - \frac{(\mathbf{v}_i^T \mathbf{x})^2}{\lambda_{s_i} + \hat{\sigma_1^2}} + 1\right] > \gamma' \tag{9.11}$$

where $\hat{\sigma_0^2} = (1/N) \sum_{n=0}^{N-1} x^2[n]$ and $\hat{\sigma_1^2}$ is given by the solution of (9.8). The asymptotic PDF will *not* be given by the standard one of Section 6.5.

9.4.3 Deterministic Signal with Unknown Parameters

A wide class of detection problems can be easily solved when σ^2 is unknown if the signal conforms to the linear model. Recall from Section 7.7 that the classical linear model assumes that $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$, where $\mathbf{x}$ is the $N \times 1$ data vector, $\mathbf{H}$ is an $N \times p$ known observation matrix, $\boldsymbol{\theta}$ is a $p \times 1$ vector of *unknown* signal parameters, and $\mathbf{w}$ is a random Gaussian vector with $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. In contrast to the linear model discussed in Section 7.7, *we now assume that σ^2 is also unknown.* Hence, the unknown parameter vector is $[\boldsymbol{\theta}^T \, \sigma^2]^T$. Note that the somewhat inconsistent use of $\boldsymbol{\theta}$ for the unknown signal parameters only, as opposed to the entire vector of unknown signal and noise parameters, is done to preserve the usual linear model notation. Hopefully, the meaning of $\boldsymbol{\theta}$ will be clear from the context. The extension of Theorem 7.1 to accommodate an unknown noise variance is now given. See Appendix 9A for the derivation.

Theorem 9.1 (GLRT for Classical Linear Model – σ^2 Unknown)
Assume the data have the form $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$, where $\mathbf{H}$ is a known $N \times p$ $(N > p)$ observation matrix of rank p, $\boldsymbol{\theta}$ is a $p \times 1$ vector of parameters, and $\mathbf{w}$ is an $N \times 1$ noise vector with PDF $\mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. The GLRT for the hypothesis testing problem

$$\begin{aligned}
\mathcal{H}_0 &: \mathbf{A}\boldsymbol{\theta} = \mathbf{b}, \sigma^2 > 0 \\
\mathcal{H}_1 &: \mathbf{A}\boldsymbol{\theta} \neq \mathbf{b}, \sigma^2 > 0
\end{aligned} \tag{9.12}$$

where $\mathbf{A}$ is an $r \times p$ matrix $(r \leq p)$ of rank r, $\mathbf{b}$ is an $r \times 1$ vector, and $\mathbf{A}\boldsymbol{\theta} = \mathbf{b}$ is a consistent set of linear equations, is to decide $\mathcal{H}_1$ if

$$\begin{aligned}
T(\mathbf{x}) &= \frac{N-p}{r} \left(L_G(\mathbf{x})^{\frac{2}{N}} - 1\right) & (9.13) \\
&= \frac{N-p}{r} \frac{(\mathbf{A}\hat{\boldsymbol{\theta}}_1 - \mathbf{b})^T [\mathbf{A}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{A}^T]^{-1} (\mathbf{A}\hat{\boldsymbol{\theta}}_1 - \mathbf{b})}{\mathbf{x}^T (\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) \mathbf{x}} > \gamma' & (9.14)
\end{aligned}$$

where $\hat{\boldsymbol{\theta}}_1 = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}$ is the MLE of $\boldsymbol{\theta}$ under $\mathcal{H}_1$ or the unrestricted MLE. The exact detection performance (holds for finite data records) is given by

$$\begin{aligned}
P_{FA} &= Q_{F_{r,N-p}}(\gamma') \\
P_D &= Q_{F'_{r,N-p}(\lambda)}(\gamma')
\end{aligned} \tag{9.15}$$

where $F_{r,N-p}$ denotes an F distribution with r numerator degrees of freedom and $N-p$ denominator degrees of freedom, and $F'_{r,N-p}(\lambda)$ denotes a noncentral F distribution with r numerator degrees of freedom, $N-p$ denominator degrees of freedom, and noncentrality parameter λ. The noncentrality parameter is given by

$$\lambda = \frac{(\mathbf{A}\boldsymbol{\theta}_1 - \mathbf{b})^T [\mathbf{A}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{A}^T]^{-1} (\mathbf{A}\boldsymbol{\theta}_1 - \mathbf{b})}{\sigma^2}$$

where $\boldsymbol{\theta}_1$ is the true value of $\boldsymbol{\theta}$ under $\mathcal{H}_1$.

See Chapter 2 for a discussion of the F distribution. The test leads to a CFAR detector since P_{FA} does not depend on σ^2. Note that the test statistic is nearly identical to that when σ^2 is known (see Theorem 7.1). The principal difference (apart from a scale factor) is that the denominator σ^2 has been replaced by its unbiased estimator

$$\hat{\sigma_1^2} = \frac{1}{N-p}\mathbf{x}^T(\mathbf{I} - \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T)\mathbf{x}.$$

Also, the PDF as given by (9.15) is *exact in that it holds for finite data records.* As expected, though, as $N \to \infty$, the PDFs converge to the usual chi-squared statistics for the GLRT (see Problem 9.14). We now illustrate the use of this important theorem with some examples.

Example 9.2 - DC Level in WGN with Unknown Amplitude and Variance

We have already studied this problem in Examples 6.5 and 6.7. Here we apply Theorem 9.1 to obtain an equivalent test statistic as well as the *exact PDF.* In Example 6.7 only the asymptotic detection performance (as $N \to \infty$) was given. This was because we employed the asymptotic properties of the GLRT. Hence, we have the detection problem

$$\begin{array}{ll}\mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = A + w[n] & n = 0, 1, \ldots, N-1\end{array}$$

where A is an unknown DC level with $-\infty < A < \infty$, and $w[n]$ is WGN with unknown variance σ^2. In terms of the classical linear model we have $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$, where $\mathbf{H} = [1\,1\ldots 1]^T$, and $\boldsymbol{\theta} = A$. The hypothesis test expressed as a parameter test is

$$\begin{array}{l}\mathcal{H}_0 : A = 0, \sigma^2 > 0 \\ \mathcal{H}_1 : A \neq 0, \sigma^2 > 0.\end{array}$$

In terms of the classical linear model parameter test of (9.12), this is equivalent to $\mathbf{A} = 1, \boldsymbol{\theta} = A, \mathbf{b} = 0$ so that $r = p = 1$. Hence, the test is given by (9.14). Now $\mathbf{H}^T\mathbf{H} = N$ and $\mathbf{H}^T\mathbf{x} = N\bar{x}$ so that $\hat{\boldsymbol{\theta}}_1 = \bar{x}$, $\mathbf{A}\hat{\boldsymbol{\theta}}_1 - \mathbf{b} = \bar{x}$, and $[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1} = N$. From (9.14) we have

$$\begin{aligned}T(\mathbf{x}) &= (N-1)\frac{N\bar{x}^2}{\mathbf{x}^T\mathbf{x} - N\bar{x}^2} \\ &= (N-1)\frac{\bar{x}^2}{\frac{1}{N}\sum_{n=0}^{N-1}(x[n]-\bar{x})^2} \\ &= (N-1)\frac{\bar{x}^2}{\hat{\sigma_1^2}}\end{aligned}$$

which agrees with (6.19). This is because for $p = r = 1$, upon using (9.13)

$$\begin{aligned}L_G(\mathbf{x}) &= \left(\frac{T(\mathbf{x})}{N-1} + 1\right)^{\frac{N}{2}} \\ &= \left(\frac{\bar{x}^2}{\hat{\sigma_1^2}} + 1\right)^{\frac{N}{2}}\end{aligned}$$

or

$$2\ln L_G(\mathbf{x}) = N \ln\left(1 + \frac{\bar{x}^2}{\hat{\sigma_1^2}}\right).$$

The detection performance is from (9.15)

$$\begin{aligned}P_{FA} &= Q_{F_{1,N-1}}(\gamma') \\ P_D &= Q_{F'_{1,N-1}(\lambda)}(\gamma')\end{aligned}$$

where $\lambda = NA^2/\sigma^2$. It is shown in Problem 9.14 that as $N \to \infty$, we obtain the usual GLRT statistics.

◇

Example 9.3 - Signal with Unknown Linear Parameters in WGN with Unknown Variance

We now extend the results of the previous example to the case where the signal is $\mathbf{s} = \mathbf{H}\boldsymbol{\theta}$ and we wish to test if $\mathbf{s} = \mathbf{0}$ versus $\mathbf{s} \neq \mathbf{0}$. Since $\mathbf{H}$ is assumed to be full-rank, the equivalent problem is to test $\boldsymbol{\theta} = \mathbf{0}$ versus $\boldsymbol{\theta} \neq \mathbf{0}$. Using Theorem 9.1 with $\mathbf{A} = \mathbf{I}$, $\mathbf{b} = \mathbf{0}$ and $r = p$, we decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \frac{N-p}{p}\frac{\hat{\boldsymbol{\theta}}_1^T\mathbf{H}^T\mathbf{H}\hat{\boldsymbol{\theta}}_1}{\mathbf{x}^T(\mathbf{I} - \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T)\mathbf{x}} > \gamma'$$

where $\hat{\boldsymbol{\theta}}_1 = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$ is the MLE of $\boldsymbol{\theta}$ under $\mathcal{H}_1$. See also Section 7.7 for some alternative forms of the test statistic. One interesting interpretation of $T(\mathbf{x})$ follows as

$$\begin{aligned}T(\mathbf{x}) &= \frac{N-p}{p}\frac{[(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}]^T\mathbf{H}^T\mathbf{H}[(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}]}{\mathbf{x}^T(\mathbf{I} - \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}} \\ &= \frac{N-p}{p}\frac{\mathbf{x}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}}{\mathbf{x}^T(\mathbf{I} - \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T)\mathbf{x}} \\ &= \frac{N-p}{p}\frac{\mathbf{x}^T\mathbf{P}_H\mathbf{x}}{\mathbf{x}^T\mathbf{P}_H^\perp\mathbf{x}}\end{aligned}$$

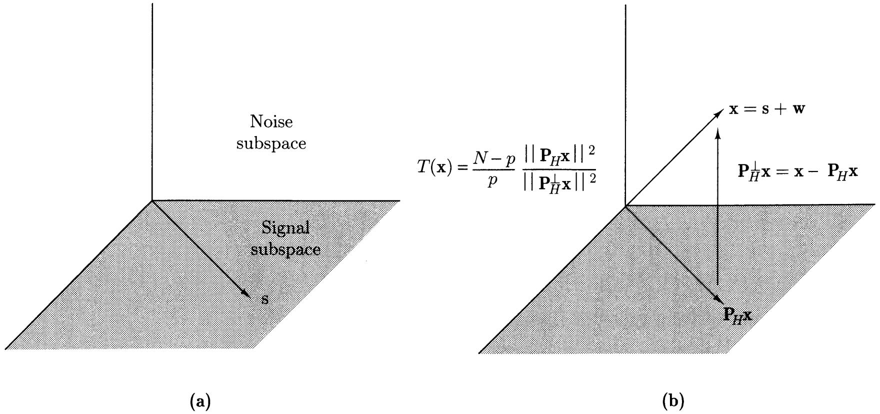


Figure 9.2. Vector space interpretation of detector for linear model.

$$= \frac{N-p}{p} \frac{||\mathbf{P}_H\mathbf{x}||^2}{||\mathbf{P}_H^{\perp}\mathbf{x}||^2} \tag{9.16}$$

where $\mathbf{P}_H = \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$ is the orthogonal projection matrix that projects a vector onto the columns of $\mathbf{H}$, and $\mathbf{P}_H^{\perp} = \mathbf{I} - \mathbf{P}_H$ is the orthogonal projection matrix that projects a vector onto the space orthogonal to that spanned by the columns of $\mathbf{H}$ (see [Kay-I 1993, pp. 231–232]). Also, $||\boldsymbol{\xi}||$ denotes the Euclidean norm of the vector $\boldsymbol{\xi}$. If we think of the subspace spanned by the columns of $\mathbf{H}$ as a "signal subspace" and the orthogonal subspace as the "noise subspace," then the test statistic is an estimated SNR as illustrated in Figure 9.2. As shown there, we decompose $\mathbf{x}$ into its two orthogonal components, $\mathbf{P}_H\mathbf{x}$ and $\mathbf{P}_H^{\perp}\mathbf{x}$, and compute the ratio of the squared-length of the two vectors. The reader may also wish to consult [Scharf and Friedlander 1994] for additional interpretations. Finally, note that $\mathbf{P}_H\mathbf{x}$ and $\mathbf{P}_H^{\perp}\mathbf{x}$ are *independent Gaussian vectors* since

$$\begin{aligned} E\left[(\mathbf{P}_H\mathbf{w})(\mathbf{P}_H^{\perp}\mathbf{w})^T\right] &= \mathbf{P}_H\sigma^2\mathbf{I}\mathbf{P}_H^{\perp} \\ &= \sigma^2\mathbf{P}_H\mathbf{P}_H^{\perp} = \mathbf{0} \end{aligned}$$

and $\mathbf{P}_H$ has rank p, while $\mathbf{P}_H^{\perp}$ has rank $N-p$ [Kay-I 1993, Problem 8.12], yielding the $F_{p,N-p}$ distribution (see Chapter 2).

◇

The Rao test can also be useful for the problem of unknown signal parameters (not necessarily linear) since the MLE only under $\mathcal{H}_0$ need be found. An illustration

was previously given in Example 6.8. However, a cautionary note is illustrated by the following example.

Example 9.4 - Signal with Unknown Parameters in WGN with Unknown Variance

Assume that the signal is given by $As[n;\alpha]$, where $-\infty < A < \infty$ is unknown and α is an unknown signal parameter. An example might be the signal $A\cos 2\pi f_0 n$, where $\alpha = f_0$. The unknown parameter vector is $\boldsymbol{\theta} = [A\, f_0\, \sigma^2]^T$, so that in the context of the notation of Chapter 6, $\boldsymbol{\theta}_r = A$, $\boldsymbol{\theta}_{r_0} = 0$, and $\boldsymbol{\theta}_s = [f_0\, \sigma^2]^T$. Now the Rao test is from (6.35)

$$T_R(\mathbf{x}) = \left.\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}^T \left[\mathbf{I}^{-1}(\tilde{\boldsymbol{\theta}})\right]_{\theta_r\theta_r} \left.\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}$$

where $\tilde{\boldsymbol{\theta}} = [\boldsymbol{\theta}_{r_0}^T\, \tilde{\boldsymbol{\theta}}_{s_0}^T]^T = [0\, \hat{f}_0\, \hat{\sigma_0^2}]^T$ for $\hat{f}_0$ the MLE of f_0 under $\mathcal{H}_0$. But under $\mathcal{H}_0$ there is no signal since $A = 0$ and thus $\hat{f}_0$ cannot be expected to be estimable. This manifests itself as a singular Fisher information matrix for $A = 0$. For example, in a similar fashion to [Kay-I 1993, pp. 56–57], we can show that (see Problem 9.16)

$$\begin{aligned} \mathbf{I}(\boldsymbol{\theta}) &\approx \begin{bmatrix} \frac{N}{2\sigma^2} & 0 & 0 \\ 0 & \frac{2A^2\pi^2}{\sigma^2}\sum_{n=0}^{N-1} n^2 & 0 \\ 0 & 0 & \frac{N}{2\sigma^4} \end{bmatrix} \\ &\rightarrow \begin{bmatrix} \frac{N}{2\sigma^2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{N}{2\sigma^4} \end{bmatrix} \end{aligned}$$

as $A \to 0$. Clearly, $\mathbf{I}^{-1}(\tilde{\boldsymbol{\theta}})$ does not exist. Of course, the GLRT can always be used, but requires the MLE of $[A\, f_0\, \sigma^2]^T$ under $\mathcal{H}_1$. Hence, the Rao test is limited to a known signal except for its amplitude. But this problem is easily solved by employing Theorem 9.1. Thus, the Rao test becomes useful mainly for a linear signal model in *correlated Gaussian* noise with unknown noise parameters as shown in Section 9.5.

◇

9.4.4 Random Signal with Unknown PDF Parameters

In general, the GLRT for this problem is difficult to determine analytically. This was seen to be the case in Section 9.4.2 for a known signal PDF and an unknown noise variance. When unknown signal parameters are added, the difficulties only increase. Because of its limited utility, we will not pursue this case further.

9.5 Colored WSS Gaussian Noise

When the noise is correlated with unknown parameters, the MLE under $\mathcal{H}_1$ of the signal and noise parameters can be difficult to obtain. We therefore restrict our discussion to the case of a *deterministic signal* for which some solutions are possible. Before exploring this situation, deterministic signals with unknown parameters embedded in correlated Gaussian noise with unknown parameters, it should be mentioned that if the noise covariance is known, then the detectors described in Chapters 7 and 8 are directly applicable. This is because, for a known noise covariance matrix $\mathbf{C}_w$, we can prewhiten the data. Factoring $\mathbf{C}_w^{-1}$ as $\mathbf{D}^T\mathbf{D}$, where $\mathbf{D}$ is an $N \times N$ invertible matrix, we form $\mathbf{x}' = \mathbf{D}\mathbf{x}$. Then, $\mathbf{x}' = \mathbf{D}\mathbf{s} + \mathbf{D}\mathbf{w} = \mathbf{D}\mathbf{x} + \mathbf{w}'$, where $\mathbf{C}_{w'} = E(\mathbf{w}'\mathbf{w}'^T) = \mathbf{I}$ (see [Kay-I 1993, pp. 94–95]). The detection problem then reduces to one of detecting the signal $\mathbf{D}\mathbf{s}$, which may contain unknown parameters, in WGN with *known* variance $\sigma^2 = 1$.

In practice, we often assume that the noise is WSS. A simple model for WSS colored noise is the autoregressive (AR) model of order one [Kay 1988, Chapter 5]. It is defined as the output $w[n]$ of a one-pole recursive filter excited at the input by WGN $u[n]$ or as

$$w[n] = -a[1]w[n-1] + u[n] \tag{9.17}$$

where the AR filter parameter $a[1]$ satisfies $|a[1]| < 1$ (for stability of the filter) and $u[n]$ is WGN with variance σ_u^2. If $a[1] = 0$, then we have our usual WGN model with $\sigma^2 = \sigma_u^2$. The effect of the filter is to color the noise to produce a nonflat PSD whose ACF can be shown to be (see Problem 9.17)

$$r_{ww}[k] = \frac{\sigma_u^2}{1 - a^2[1]}\left(-a[1]\right)^{|k|}. \tag{9.18}$$

As $|a[1]| \to 1$, the noise becomes more heavily correlated. The PSD can also be shown to be (see Problem 9.17)

$$P_{ww}(f) = \frac{\sigma_u^2}{|1 + a[1]\exp(-j2\pi f)|^2} \qquad |f| \le \frac{1}{2} \tag{9.19}$$

some examples of which are shown in Figure 9.3. Hence, $a[1]$ and σ_u^2 control the bandwidth and total power of the process, respectively. Extensions to model bandpass processes are given in [Kay 1988, Chapter 5]. We will assume that $a[1]$ and σ_u^2 are unknown and consider deterministic signals.

9.5.1 Known Deterministic Signals

Using the AR model for the noise, we now determine the GLRT when $a[1], \sigma_u^2$ are unknown and the signal $s[n]$ is deterministic and known. We first require the PDF of the noise samples. An exact expression is difficult to work with, and hence, we use

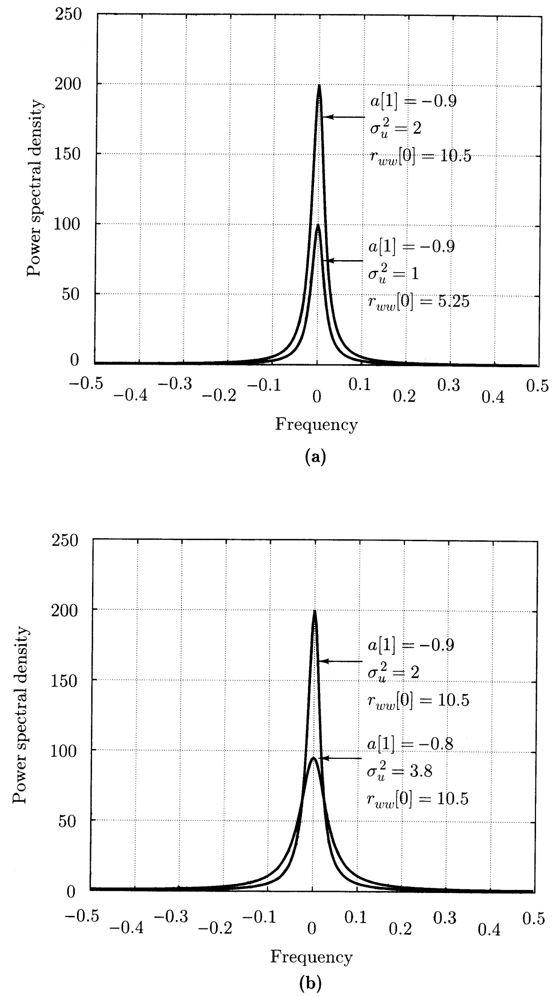


Figure 9.3. Power spectral density of AR process (a) Same bandwidth but different power (b) Same power but different bandwidths.

a large data record approximation. This has been derived in [Kay-I 1993, Example 7.18] as

$$\ln p(\mathbf{w}; a[1], \sigma_u^2) = -\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln\sigma_u^2 - \frac{N}{2\sigma_u^2}\int_{-\frac{1}{2}}^{\frac{1}{2}} |A(f)|^2 I_w(f)df$$

where $A(f) = 1 + a[1]\exp(-j2\pi f)$ and $I_w(f) = (1/N)|\sum_{n=0}^{N-1} w[n]\exp(-j2\pi fn)|^2$ is the periodogram. This may also be derived based on (2.36) if we note that $\int_{-\frac{1}{2}}^{\frac{1}{2}} \ln|A(f)|^2 df = 0$ [Kay 1988]. The MLE of $a[1], \sigma_u^2$ based on this approximation has been shown to be [Kay-I 1993, pp. 196–198]

$$\begin{aligned} \hat{a}[1] &= -\frac{\hat{r}_{ww}[1]}{\hat{r}_{ww}[0]} && (9.20)\\ \hat{\sigma_u^2} &= \hat{r}_{ww}[0] + \hat{a}[1]\hat{r}_{ww}[1] \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} |\hat{A}(f)|^2 I_w(f)df && (9.21)\end{aligned}$$

where

$$\hat{r}_{ww}[k] = \frac{1}{N}\sum_{n=0}^{N-1-k} w[n]w[n+k]$$

for $k \geq 0$ is an estimate of the ACF and

$$\hat{A}(f) = 1 + \hat{a}[1]\exp(-j2\pi f).$$

Thus, we have

$$\ln p(\mathbf{w}; \hat{a}[1], \hat{\sigma_u^2}) = -\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln\hat{\sigma_u^2} - \frac{N}{2}$$

which is equivalent to $p(\mathbf{x}; \hat{a}_0[1], \hat{\sigma^2_{u_0}}, \mathcal{H}_0)$ for $\hat{a}_0[1], \hat{\sigma^2_{u_0}}$ the MLE of $a[1], \sigma_u^2$ under $\mathcal{H}_0$. The only modification to the PDF when the signal is present is the replacement of $x[n]$ by $x[n] - s[n]$. The GLRT becomes

$$2\ln\frac{p(\mathbf{x}; \hat{a}_1[1], \hat{\sigma^2_{u_1}}, \mathcal{H}_1)}{p(\mathbf{x}; \hat{a}_0[1], \hat{\sigma^2_{u_0}}, \mathcal{H}_0)} = N\ln\frac{\hat{\sigma^2_{u_0}}}{\hat{\sigma^2_{u_1}}} \tag{9.22}$$

where

$$\begin{aligned} \hat{\sigma^2_{u_0}} &= \int_{-\frac{1}{2}}^{\frac{1}{2}} |\hat{A}_0(f)|^2 I_x(f)df \\ \hat{\sigma^2_{u_1}} &= \int_{-\frac{1}{2}}^{\frac{1}{2}} |\hat{A}_1(f)|^2 I_{x-s}(f)df && (9.23)\end{aligned}$$

and

$$\begin{aligned} \hat{a}_0[1] &= -\frac{\hat{r}_{xx}[1]}{\hat{r}_{xx}[0]} \\ \hat{a}_1[1] &= -\frac{\hat{r}_{x-s,x-s}[1]}{\hat{r}_{x-s,x-s}[0]}. && (9.24)\end{aligned}$$

Noting that

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} |A(f)|^2 I(f)df = \frac{1}{N}\int_{-\frac{1}{2}}^{\frac{1}{2}} |A(f)X(f)|^2 df$$

where $X(f) = \sum_{n=0}^{N-1} x[n]\exp(-j2\pi fn)$ is the Fourier transform of $x[n]$, we have that approximately for large N (see Problem 9.18)

$$\begin{aligned} \hat{\sigma^2_{u_0}} &= \frac{1}{N}\sum_{n=1}^{N-1}(x[n] + \hat{a}_0[1]x[n-1])^2 \\ \hat{\sigma^2_{u_1}} &= \frac{1}{N}\sum_{n=1}^{N-1}[(x[n]-s[n]) + \hat{a}_1[1](x[n-1]-s[n-1])]^2. && (9.25)\end{aligned}$$

The overall detector is shown in Figure 9.4. An intuitive explanation of the detector is as follows. Assume that a signal is present. Then, $\hat{a}_1[1]$ will be a better estimate than $\hat{a}_0[1]$, since the latter will be based on the data $s[n] + w[n]$. Futhermore, the output of the filter $\hat{A}_0(f)$ will be due to noise *as well as the signal*, leading to an inflated estimate of σ_u^2. The overall effect is to produce a large value of the test statistic or to decide that a signal is present. A similar argument holds for the noise-only case. We note that the filtering $x[n] + a[1]x[n-1]$ prewhitens the noise, since if $x[n] = s[n] + w[n]$, then $x[n] + a[1]x[n-1] = (w[n] + a[1]w[n-1]) + (s[n] + a[1]s[n-1]) = u[n] + (s[n] + a[1]s[n-1])$. The performance of this detector has been derived in [Kay 1983]. It does not conform to the usual asymptotic GLRT statistics.

9.5.2 Deterministic Signals with Unknown Parameters

When the signal has unknown parameters as well, the GLRT can be difficult to determine, even for the simple AR noise model. For signals such as DC levels, sinusoids, etc. with unknown amplitudes only, the approximate MLEs required are tractable and we refer the reader to [Kay and Nagesha 1994] for the details. However, for a general signal with unknown amplitude we cannot easily do so. An alternative approach is to use the Rao test. This is because the only MLE required is under $\mathcal{H}_0$, which amounts to the MLE of the noise or AR parameters, which have already been determined to be given by (9.20), (9.21) with $w[n]$ replaced by $x[n]$.

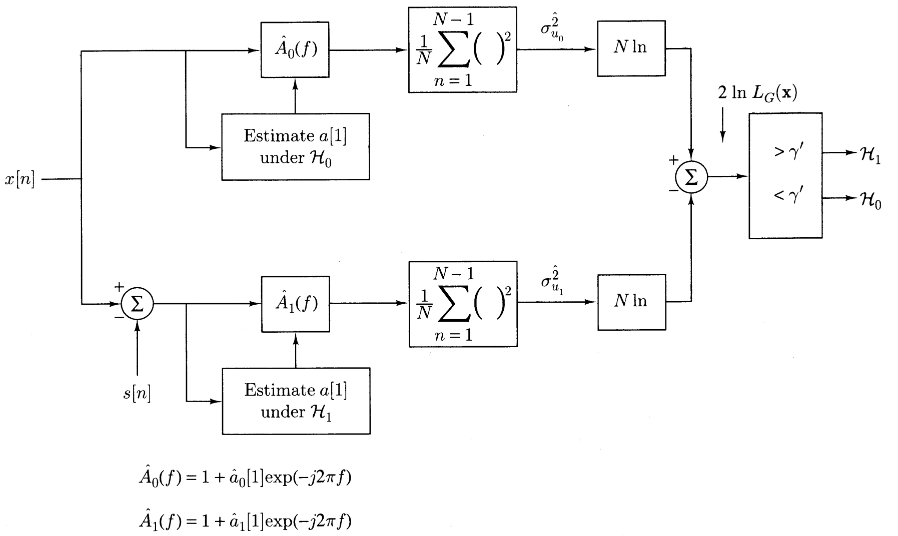


Figure 9.4. GLRT for known deterministic signal in colored autoregressive noise with unknown parameters.

Hence, we now apply the Rao test to the detection problem

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = As[n] + w[n] & n = 0, 1, \ldots, N-1 \end{array}$$

where A is an unknown amplitude, $s[n]$ is a known deterministic signal, and $w[n]$ is a Gaussian AR process of order one with unknown parameters $a[1]$ and σ_u^2. The equivalent parameter test is

$$\begin{array}{l} \mathcal{H}_0 : A = 0, |a[1]| < 1, \sigma_u^2 > 0 \\ \mathcal{H}_1 : A \neq 0, |a[1]| < 1, \sigma_u^2 > 0. \end{array}$$

This problem is a generalization of Example 6.10 since for $a[1] = 0$, the colored noise becomes WGN. To apply the Rao test of Section 6.6 we note that

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_r \\ \boldsymbol{\theta}_s \end{bmatrix}$$

where

$$\boldsymbol{\theta}_r = A$$

$$\boldsymbol{\theta}_s = \boldsymbol{\theta}_w = \begin{bmatrix} a[1] \\ \sigma_u^2 \end{bmatrix}$$

and $\boldsymbol{\theta}_{r_0} = 0$. In effect, $a[1], \sigma_u^2$ are nuisance parameters. The Rao test is

$$T_R(\mathbf{x}) = \left.\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r}\right|^T_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} [\mathbf{I}^{-1}(\tilde{\boldsymbol{\theta}})]_{\theta_r\theta_r} \left.\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}$$

where $\tilde{\boldsymbol{\theta}} = [\boldsymbol{\theta}_{r_0}^T \, \hat{\boldsymbol{\theta}}_{s_0}^T]^T = [0 \, \hat{\boldsymbol{\theta}}_{w_0}^T]^T$ is the MLE of $\boldsymbol{\theta}$ under $\mathcal{H}_0$ or the MLE of $a[1], \sigma_u^2$ when $A = 0$. Also,

$$[\mathbf{I}^{-1}(\boldsymbol{\theta})]_{\theta_r\theta_r} = \left(\mathbf{I}_{\theta_r\theta_r}(\boldsymbol{\theta}) - \mathbf{I}_{\theta_r\theta_s}(\boldsymbol{\theta})\mathbf{I}^{-1}_{\theta_s\theta_s}(\boldsymbol{\theta})\mathbf{I}_{\theta_s\theta_r}(\boldsymbol{\theta})\right)^{-1}.$$

Since $\boldsymbol{\theta}_r = A$ is a scalar, the Rao test reduces to

$$T_R(\mathbf{x}) = \frac{\left(\left.\dfrac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}\right)^2}{\mathbf{I}_{\theta_r\theta_r}(\tilde{\boldsymbol{\theta}}) - \mathbf{I}_{\theta_r\theta_s}(\tilde{\boldsymbol{\theta}})\mathbf{I}^{-1}_{\theta_s\theta_s}(\tilde{\boldsymbol{\theta}})\mathbf{I}_{\theta_s\theta_r}(\tilde{\boldsymbol{\theta}})}. \tag{9.26}$$

To evaluate the numerator we use (3C.5) in [Kay-I 1993, pg. 75], which is valid for $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \mathbf{C}(\boldsymbol{\theta}))$. For the problem at hand $\boldsymbol{\mu}$ depends only on $\boldsymbol{\theta}_r = \theta_r = A$ and $\mathbf{C}$ depends only on $\boldsymbol{\theta}_s = \boldsymbol{\theta}_w = [a[1] \, \sigma_u^2]^T$. It follows that

$$\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_r} = \frac{\partial \boldsymbol{\mu}(\theta_r)^T}{\partial \theta_r}\mathbf{C}^{-1}(\boldsymbol{\theta}_s)(\mathbf{x} - \boldsymbol{\mu}(\theta_r)). \tag{9.27}$$

Hence, we have

$$\begin{aligned} \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial A} &= \frac{\partial A\mathbf{s}^T}{\partial A}\mathbf{C}^{-1}(\boldsymbol{\theta}_w)(\mathbf{x} - A\mathbf{s}) \\ &= \mathbf{s}^T\mathbf{C}^{-1}(\boldsymbol{\theta}_w)(\mathbf{x} - A\mathbf{s}) \end{aligned}$$

where $\mathbf{s} = [s[0] \, s[1] \ldots s[N-1]]^T$ and therefore

$$\left.\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial A}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} = \mathbf{s}^T\mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0})\mathbf{x} \tag{9.28}$$

where $\mathbf{C}(\hat{\boldsymbol{\theta}}_{w_0})$ is the covariance matrix of $\mathbf{w}$ with $a[1], \sigma_u^2$ replaced by their MLEs under $\mathcal{H}_0$. Also, from (3.31) of [Kay-I 1993, pg. 47]

$$[\mathbf{I}(\boldsymbol{\theta})]_{ij} = \left[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_i}\right]^T \mathbf{C}^{-1}(\boldsymbol{\theta}) \left[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_j}\right] + \frac{1}{2}\mathrm{tr}\left[\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_i}\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_j}\right]. \tag{9.29}$$

Again, for our problem, $\boldsymbol{\theta} = [\boldsymbol{\theta}_r^T \, \boldsymbol{\theta}_s^T]^T$, where $\boldsymbol{\theta}_r = A$ is a signal parameter and $\boldsymbol{\theta}_s = [a[1] \, \sigma_u^2]^T$ are noise parameters (nuisance parameters). The signal parameter is contained only in $\boldsymbol{\mu}(\boldsymbol{\theta})$ while the noise parameters are contained only in $\mathbf{C}(\boldsymbol{\theta})$ or we have $\boldsymbol{\mu}(\theta_r)$ and $\mathbf{C}(\boldsymbol{\theta}_s)$. As a result, we see that $\mathbf{I}_{\theta_r\theta_s}(\boldsymbol{\theta}) = \mathbf{0}$ from (9.29) and thus the Fisher information matrix is block-diagonal (see Problem 9.20). From (9.29) we have

$$\begin{aligned} \mathbf{I}_{\theta_r\theta_r}(\boldsymbol{\theta}) = I_{AA}(\boldsymbol{\theta}) &= \frac{\partial \boldsymbol{\mu}(A)^T}{\partial A}\mathbf{C}^{-1}(\boldsymbol{\theta}_w)\frac{\partial \boldsymbol{\mu}(A)}{\partial A} \\ &= \mathbf{s}^T\mathbf{C}^{-1}(\boldsymbol{\theta}_w)\mathbf{s} \end{aligned}$$

since $\boldsymbol{\mu}(A) = A\mathbf{s}$ and

$$I_{AA}(\tilde{\boldsymbol{\theta}}) = \mathbf{s}^T\mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w0})\mathbf{s}. \tag{9.30}$$

Using (9.28) and (9.30) in (9.26) produces

$$T_R(\mathbf{x}) = \frac{(\mathbf{s}^T\mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w0})\mathbf{x})^2}{\mathbf{s}^T\mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w0})\mathbf{s}} \tag{9.31}$$

which is recognized as a incoherent generalized matched filter with normalization. The detector can be shown to be equivalent for large data records to (see Problem 9.21)

$$T_R(\mathbf{x}) = \frac{\left(\sum_{n=1}^{N-1}(s[n] + \hat{a}_0[1]s[n-1])(x[n] + \hat{a}_0[1]x[n-1])\right)^2}{\hat{\sigma^2_{u0}}\sum_{n=1}^{N-1}(s[n] + \hat{a}_0[1]s[n-1])^2} \tag{9.32}$$

where $\hat{a}_0[1]$ and $\hat{\sigma^2_{u0}}$ are given in (9.24) and (9.25), respectively, and is shown in Figure 9.5. The asymptotic performance is given by (6.23) and (6.24) with $r = 1$ as

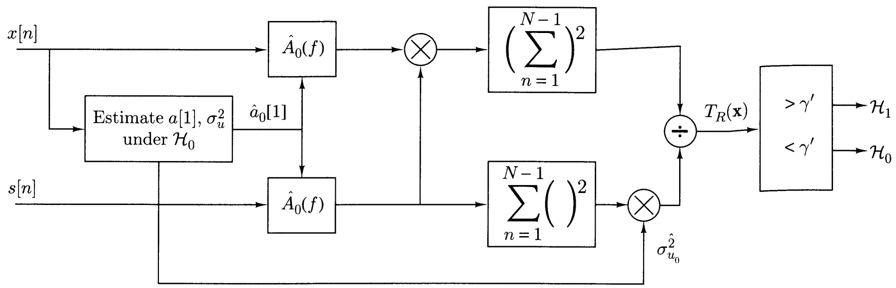
$$T_R(\mathbf{x}) \overset{a}{\sim} \begin{cases} \chi_1^2 & \text{under } \mathcal{H}_0 \\ \chi_1'^2(\lambda) & \text{under } \mathcal{H}_1 \end{cases} \tag{9.33}$$

where

$$\begin{aligned} \lambda &= A^2 I_{AA}(A=0, a[1], \sigma_u^2) \\ &= A^2\mathbf{s}^T\mathbf{C}^{-1}(\boldsymbol{\theta}_w)\mathbf{s} \end{aligned}$$

and A is the true value of the amplitude under $\mathcal{H}_1$, $\boldsymbol{\theta}_w$ is the true value of the noise parameter. It is interesting to note that if the noise parameters are known and only A is unknown, then it is easily shown (see Problem 9.22) that the Rao test statistic is

$$T_R(\mathbf{x}) = \frac{(\mathbf{s}^T\mathbf{C}^{-1}(\boldsymbol{\theta}_w)\mathbf{x})^2}{\mathbf{s}^T\mathbf{C}^{-1}(\boldsymbol{\theta}_w)\mathbf{s}} \tag{9.34}$$



$\hat{A}_0(f) = 1 + \hat{a}_0[1]\exp(-j2\pi f)$

Figure 9.5. Rao test for deterministic signal with unknown amplitude in colored autoregressive noise with unknown parameters.

and the asymptotic performance is again given by (9.33). For either problem $r = 1$ so that the PDFs are χ_1^2 and $\chi_1'^2(\lambda)$ under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. The only difference is in the noncentrality parameter. When the noise parameters are unknown, then λ is generally decreased since

$$\begin{aligned} \lambda &= (\boldsymbol{\theta}_{r_1} - \boldsymbol{\theta}_{r_0})^T\Big[\mathbf{I}_{\theta_r\theta_r}(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s) \\ &\quad - \mathbf{I}_{\theta_r\theta_s}(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s)\mathbf{I}^{-1}_{\theta_s\theta_s}(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s)\mathbf{I}_{\theta_s\theta_r}(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s)\Big](\boldsymbol{\theta}_{r_1} - \boldsymbol{\theta}_{r_0}). \end{aligned} \tag{9.35}$$

But for this example, $\mathbf{I}_{\theta_r\theta_s}(\boldsymbol{\theta}) = \mathbf{0}$, and so the noncentrality parameter does not decrease. Alternatively, the block-diagonal nature of the Fisher information matrix says that we can estimate the signal amplitude with the same accuracy when $\mathbf{C}$ is known as when it contains unknown parameters (of course only as $N \to \infty$ by using an MLE). In effect, $T_R(\mathbf{x})$ is a function of the signal amplitude estimate (see Problem 9.23) and its asymptotic performance is unchanged when noise parameters are present. Likewise, the GLRT, which has the same asymptotic performance, enjoys this property as well. We generalize these results in the next theorem. See Appendix 9B for a proof.

Theorem 9.2 (Rao Test for the General Linear Model with Unknown Noise Parameters)
Assume the data have the form $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$, *where* $\mathbf{H}$ *is a known* $N \times p$ $(N > p)$ *observation matrix of rank* p, $\boldsymbol{\theta}$ *is a* $p \times 1$ *parameter vector, and* w *is an* $N \times 1$ *noise vector with PDF* $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}(\boldsymbol{\theta}_w))$ *for* $\boldsymbol{\theta}_w$ *an unknown* $q \times 1$ *noise parameter*

vector. The Rao test for the hypothesis testing problem

$$\begin{array}{l}\mathcal{H}_0 : \boldsymbol{\theta} = \mathbf{0}, \boldsymbol{\theta}_w \\ \mathcal{H}_1 : \boldsymbol{\theta} \neq \mathbf{0}, \boldsymbol{\theta}_w\end{array}$$

decides $\mathcal{H}_1$ *if*

$$T_R(\mathbf{x}) = \mathbf{x}^T \mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0}) \mathbf{H} (\mathbf{H}^T \mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0}) \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0}) \mathbf{x} > \gamma' \tag{9.36}$$

where $\hat{\boldsymbol{\theta}}_{w_0}$ *is the MLE of* $\boldsymbol{\theta}_w$ *under* $\mathcal{H}_0$ *or the value obtained by maximizing*

$$p(\mathbf{x}; \boldsymbol{\theta}_w, \mathcal{H}_0) = \frac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}(\mathbf{C}(\boldsymbol{\theta}_w))} \exp\left[-\frac{1}{2}\mathbf{x}^T \mathbf{C}^{-1}(\boldsymbol{\theta}_w)\mathbf{x}\right].$$

The asymptotic (as $N \to \infty$*) performance is*

$$\begin{aligned} P_{FA} &= Q_{\chi_p^2}(\gamma') \\ P_D &= Q_{\chi_p'^2(\lambda)}(\gamma') \end{aligned}$$

where

$$\lambda = \boldsymbol{\theta}_1^T \mathbf{H}^T \mathbf{C}^{-1}(\boldsymbol{\theta}_w) \mathbf{H} \boldsymbol{\theta}_1$$

and $\boldsymbol{\theta}_1$ *is the true value of* $\boldsymbol{\theta}$ *under* $\mathcal{H}_1$*. The asymptotic performance is the same whether* $\boldsymbol{\theta}_w$ *is known or not. If* $\boldsymbol{\theta}_w$ *is known, then the Rao test becomes*

$$T_R(\mathbf{x}) = \mathbf{x}^T \mathbf{C}^{-1}(\boldsymbol{\theta}_w) \mathbf{H} (\mathbf{H}^T \mathbf{C}^{-1}(\boldsymbol{\theta}_w) \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1}(\boldsymbol{\theta}_w) \mathbf{x} > \gamma'.$$

9.6 Signal Processing Example

We now give an example of the application of Theorem 9.2 to the design of a detector for a wideband signal. When multiple targets are close together as in an active sonar or radar, the echos tend to be clumped together in time. To do so requires a wide bandwidth transmit signal that has good temporal resolution. Recall that the wider the bandwidth of a signal, the narrower its correlation width [Van Trees 1971]. Hence, for high resolution [Rihaczek 1985] we desire a wide bandwidth transmit signal such as a linear FM or

$$p(t) = \cos[2\pi(F_0 t + \tfrac{1}{2}kt^2)] \qquad 0 \le t \le T.$$

The instantaneous frequency is $F_i(t) = F_0 + kt$, where F_0 is the start frequency in Hz and k is the sweep rate in Hz/sec. The bandwidth is about kT Hz [Glisson, Black, and Sage 1970]. In this type of situation the noise coloration will be important in the detector design, since the noise background in frequency is not flat over the signal bandwidth (as it is in a narrowband system). Hence, a detector that accounts for the noise coloration will perform better. To design a detector we now assume the discrete-time detection problem

$$\begin{array}{ll}\mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = A\cos[2\pi(f_0 n + \frac{1}{2}mn^2) + \phi] + w[n] & n = 0, 1, \ldots, N-1\end{array}$$

where the amplitude A and phase ϕ are unknown, and f_0, the starting frequency, and m, the sweep rate, are known. The noise is assumed to be modeled by an AR process of order p_{AR} or [Kay 1988, Chapter 5]

$$x[n] = -\sum_{k=1}^{p_{\mathrm{AR}}} a[k]x[n-k] + u[n]$$

where $\{a[1], a[2], \ldots, a[p_{\mathrm{AR}}], \sigma_u^2\}$ are unknown AR parameters. This more general AR model allows us to model a wide range of PSDs. A GLRT will be difficult to implement due to the need for the MLE under $\mathcal{H}_1$. Hence, we resort to the Rao test as described in Theorem 9.2. We first show that the received signal has the form of the linear model. This is because

$$\begin{aligned} & A\cos[2\pi(f_0 n + \tfrac{1}{2}mn^2) + \phi] \\ = \;& A\cos\phi\cos[2\pi(f_0 n + \tfrac{1}{2}mn^2)] - A\sin\phi\sin[2\pi(f_0 n + \tfrac{1}{2}mn^2)] \\ = \;& \alpha_1\cos[2\pi(f_0 n + \tfrac{1}{2}mn^2)] + \alpha_2\sin[2\pi(f_0 n + \tfrac{1}{2}mn^2)] \end{aligned}$$

where $\alpha_1 = A\cos\phi, \alpha_2 = -A\sin\phi$. Thus, we have $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$ where $\boldsymbol{\theta} = [\alpha_1 \, \alpha_2]^T$ and

$$\mathbf{H} = \begin{bmatrix} 1 & 0 \\ \cos[2\pi(f_0 + \frac{1}{2}m)] & \sin[2\pi(f_0 + \frac{1}{2}m)] \\ \vdots & \vdots \\ \cos[2\pi(f_0(N-1) + \frac{1}{2}m(N-1)^2)] & \sin[2\pi(f_0(N-1) + \frac{1}{2}m(N-1)^2)] \end{bmatrix}. \tag{9.37}$$

Letting $\mathbf{a} = [a[1]\, a[2] \ldots a[p_{\mathrm{AR}}]]^T$, we have the parameter test

$$\begin{array}{l}\mathcal{H}_0 : \boldsymbol{\theta} = \mathbf{0}, \mathbf{a}, \sigma_u^2 > 0 \\ \mathcal{H}_1 : \boldsymbol{\theta} \neq \mathbf{0}, \mathbf{a}, \sigma_u^2 > 0\end{array}$$

which is exactly in the form assumed in Theorem 9.2 with $p = 2$ and $\boldsymbol{\theta}_w = [\mathbf{a}^T \, \sigma_u^2]^T$. The Rao test then follows from that theorem as

$$T_R(\mathbf{x}) = \mathbf{x}^T \mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0}) \mathbf{H} (\mathbf{H}^T \mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0}) \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0}) \mathbf{x} \tag{9.38}$$

where $\mathbf{C}(\hat{\boldsymbol{\theta}}_{w_0})$ is $\mathbf{C}$ (the covariance matrix of $\mathbf{w}$) with the AR parameters replaced by their MLE under $\mathcal{H}_0$ or $\hat{\boldsymbol{\theta}}_{w_0}$. The AR parameter estimator was given by the Yule-Walker equations in Example 7.18 [Kay-I 1993, pp. 196–198]. Also, from Theorem 9.2 the asymptotic PDF is

$$T_R(\mathbf{x}) \overset{a}{\sim} \begin{cases} \chi_2^2 & \text{under } \mathcal{H}_0 \\ \chi_2'^2(\lambda) & \text{under } \mathcal{H}_1 \end{cases} \tag{9.39}$$

where $\lambda = \boldsymbol{\theta}_1^T \mathbf{H}^T \mathbf{C}^{-1}(\boldsymbol{\theta}_w)\mathbf{H}\boldsymbol{\theta}_1$. In the case when the noise parameters are known, the Rao test produces the same statistic except that we use the known parameters in $\mathbf{C}$. Also, the asymptotic PDFs are the same. We can also write the test statistic as

$$T_R(\mathbf{x}) = \mathbf{x}^T \mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0})\mathbf{H}\hat{\boldsymbol{\theta}} = \mathbf{x}^T \mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0})\hat{\mathbf{s}}$$

where $\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0})\mathbf{H})^{-1}\mathbf{H}^T \mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0})\mathbf{x}$ is a "pseudo-MLE" of $\boldsymbol{\theta}$, which replaces $\mathbf{C}(\boldsymbol{\theta}_w)$ by an estimate. We see that $T_R(\mathbf{x})$ has the form of a prewhitener followed by an estimator-correlator.

In practice, $\mathbf{C}(\hat{\boldsymbol{\theta}}_{w_0})$ may be difficult to invert since it is $N \times N$. For N large an approximation derived in Appendix 9C shows that the statistic of (9.38) can be reduced to

$$T_R(\mathbf{x}) = \frac{\left| \sum_{n=p_{\mathrm{AR}}}^{N-1} \left(\sum_{k=0}^{p_{\mathrm{AR}}} \hat{a}_0[k]x[n-k] \right) \left(\sum_{l=0}^{p_{\mathrm{AR}}} \hat{a}_0[l]\tilde{s}[n-l] \right) \right|^2}{\hat{\sigma}_{u_0}^2 \sum_{n=p_{\mathrm{AR}}}^{N-1} \left(\sum_{l=0}^{p_{\mathrm{AR}}} \hat{a}_0[l]s[n-l] \right)^2} \tag{9.40}$$

where

$$\begin{aligned} \tilde{s}[n] &= \exp[j2\pi(f_0 n + \tfrac{1}{2}mn^2)] \\ s[n] &= \cos[2\pi(f_0 n + \tfrac{1}{2}mn^2)] \end{aligned}$$

and $\{\hat{a}_0[1], \hat{a}_0[2], \ldots, \hat{a}_0[p_{\mathrm{AR}}], \hat{\sigma}_{u_0}^2\}$ are obtained as the solution of the Yule-Walker equations (the approximate MLE under $\mathcal{H}_0$), and $\hat{a}_0[0] = 1$. The approximation takes advantage of the fact that the covariance matrix is Toeplitz and also that the columns of $\mathbf{H}$ are Hilbert transforms of each other and hence are orthogonal for large N. The same results can also be obtained by using the special form of the inverse of the covariance matrix for an AR process [Kay 1988]. The detector, as shown in Figure 9.6, is seen to consist of a prewhitener, using the estimated AR parameters, followed by a correlator, a squarer, and finally a normalizer. It is also shown in Appendix 9C that the noncentrality parameter reduces to

$$\lambda = \frac{A^2}{\sigma_u^2} \sum_{n=p_{\mathrm{AR}}}^{N-1} \left(\sum_{k=0}^{p_{\mathrm{AR}}} a[k]s[n-k] \right)^2. \tag{9.41}$$

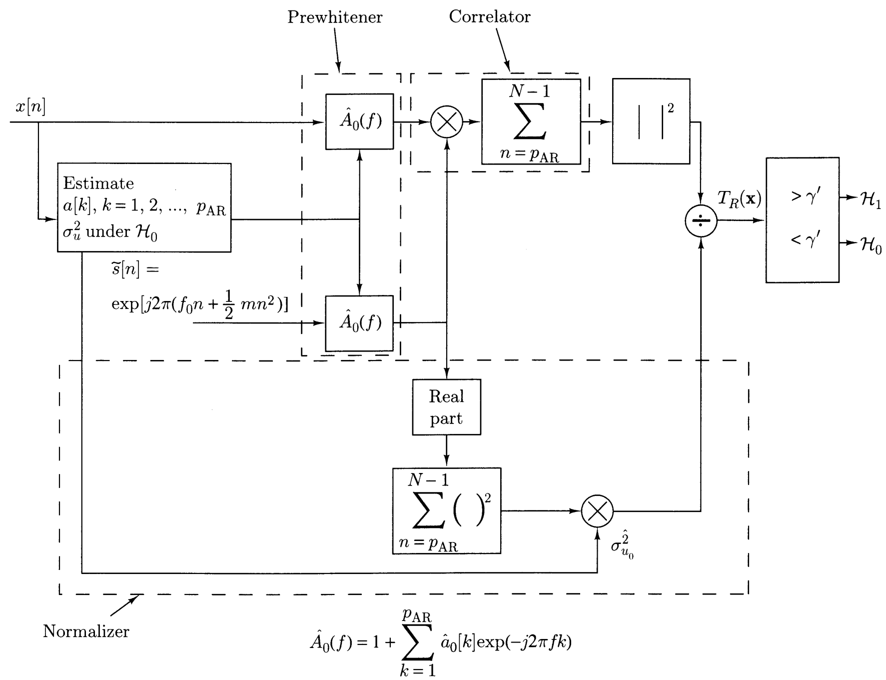
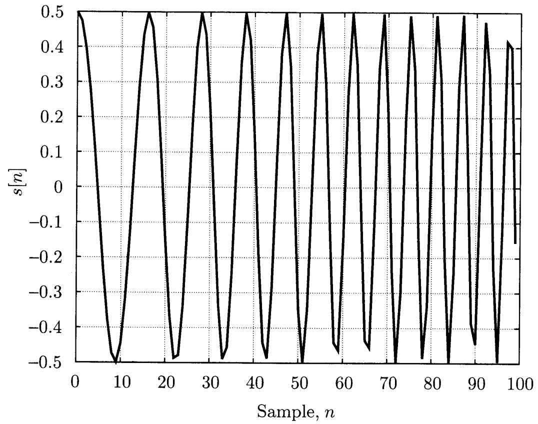


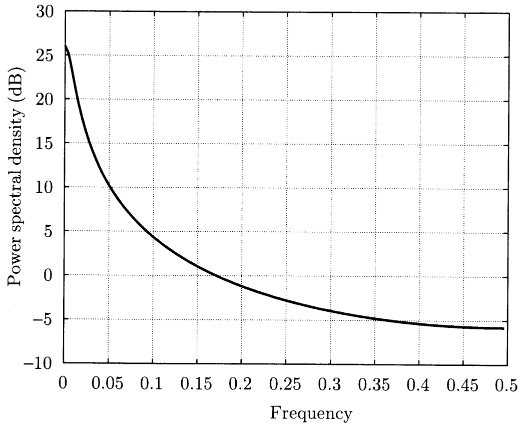
Figure 9.6. Rao test for signal processing example.

We now illustrate the detection performance with an example. We choose a signal with the parameters $A = 0.5$, $f_0 = 0.05$, $m = 0.0015$, $\phi = 0$, and $N = 100$. As shown in Figure 9.7a, the signal has a frequency that sweeps from $f_0 = 0.05$ to $f_0 + mN = 0.2$. The noise is modeled by an AR process of order $p_{\mathrm{AR}} = 1$ with parameters $a[1] = -0.95$ and $\sigma_u^2 = 1$ so that $w[n] = 0.95w[n-1] + u[n]$. Its PSD is shown in Figure 9.7b. Next, we plot the PSD of the noise (shown dashed) versus the power spectrum (magnitude-squared Fourier transform) of the signal in Figure 9.7c. It is seen that most of the frequency content of the signal is indeed over the swept frequency range [0.05, 0.2]. Also, over this band the noise power changes rapidly. Hence, it would be expected that a standard matched filter, which does not account for the noise coloration, would perform poorer than the Rao detector. To first verify the asymptotic performance of the Rao test we plot the results of a Monte Carlo simulation for the Rao test of (9.40) versus the asymptotic performance as obtained from (9.39) and (9.41). For this example, we have from (9.40)

$$T_R(\mathbf{x}) = \frac{\left|\sum_{n=1}^{N-1} \left(x[n] + \hat{a}_0[1]x[n-1]\right)\left(\tilde{s}[n] + \hat{a}_0[1]\tilde{s}[n-1]\right)\right|^2}{\hat{\sigma^2_{u_0}} \sum_{n=1}^{N-1} \left(s[n] + \hat{a}_0[1]s[n-1]\right)^2}$$

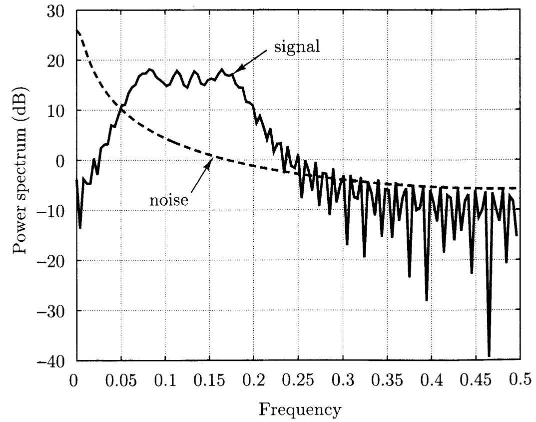


(a)



(b)

Figure 9.7. Signal and noise descriptions (a) Linear FM signal (b) Power spectral density of noise.



(c)

Figure 9.7. Continued (c) Power spectrum of signal and power spectral density of noise.

where

$$\begin{aligned} \hat{a}_0[1] &= -\frac{\hat{r}_{xx}[1]}{\hat{r}_{xx}[0]} \\ \hat{\sigma^2_{u_0}} &= \hat{r}_{xx}[0] + \hat{a}_0[1]\hat{r}_{xx}[1] \end{aligned}$$

and

$$\hat{r}_{xx}[k] = \frac{1}{N} \sum_{k=0}^{N-1-k} x[n]x[n+k]$$

for $k = 0, 1$. The asymptotic performance is given by (9.39) with

$$\lambda = \frac{A^2}{\sigma_u^2} \sum_{n=1}^{N-1} \left(s[n] + a[1]s[n-1]\right)^2$$

from (9.41). In Figure 9.8 we plot the theoretical asymptotic receiver operating characteristics (shown as a dashed curve) versus the Monte Carlo results for $T_R(\mathbf{x})$. Note that the agreement is quite close even for the short data record of $N = 100$ points. Finally, as a basis of comparison we examine the detection performance of

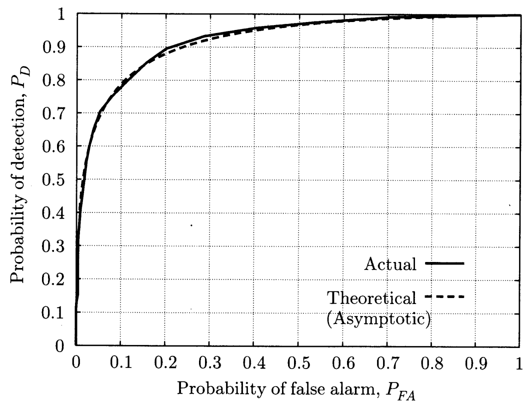


Figure 9.8. Receiver operating characteristics for Rao test.

an incoherent matched filter or

$$T(\mathbf{x}) = \left| \sum_{n=0}^{N-1} x[n]\tilde{s}[n] \right|^2$$

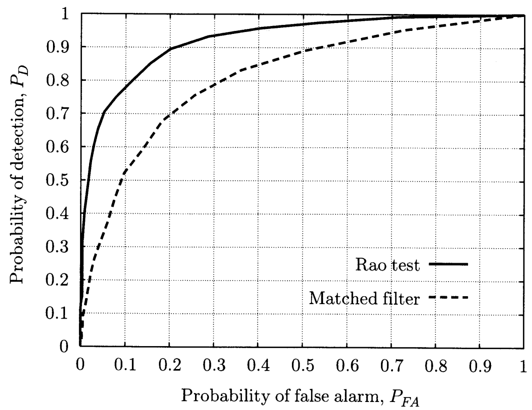


Figure 9.9. Receiver operating characteristics comparison.

which assumes that the noise is WGN (see Section 7.6.2 for the analogous result for a sinusoidal signal). The Monte Carlo results are shown in Figure 9.9. As expected, the matched filter (shown as the dashed curve) performs poorer. Actually, in practice the matched filter performance would be even poorer than that shown in Figure 9.9. This is because its lack of the CFAR property would require some ad-hoc normalization to control the false alarm rate. A degradation in P_D is therefore incurred. The Rao test does not suffer from this deficiency.

References

Bickel, P.J., K.A. Doksum, *Mathematical Statistics*, Holden-Day, San Francisco, 1977.

Glisson, T.H., C.I. Black, A.P. Sage, "On Sonar Signal Analysis," *IEEE Trans. Aero. Elect. Syst,*, Vol. AES-6, pp. 37–50, Jan. 1970.

Kay, S., "Asymptotically Optimal Detection in Unknown Colored Noise via Autoregressive Modeling," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-31, pp. 927–939, Aug. 1983.

Kay, S., V. Nagesha, "Maximum Likelihood Estimation of Signals in Autoregressive Noise," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-41, pp. 88–101, Jan. 1994.

Kay, S.M., *Modern Spectral Estimation: Theory and Application*, Prentice-Hall, Englewood Cliffs, N.J., 1988.

Kendall, Sir M., A. Stuart, *The Advanced Theory of Statistics*, Vol. 1, Macmillan, New York, 1976.

Rihaczek, A.W., *Principles of High-Resolution Radar*, Peninsula Pub., Los Altos, Calif., 1985.

Scharf, L.L., B. Friedlander, "Matched Subspace Detectors," *IEEE Trans. Signal Processing*, Vol. 42, pp. 2146–2157, Aug. 1994.

Van Trees, H.L., *Detection, Estimation, and Modulation*, Part III, J. Wiley, New York, 1971.

Problems

9.1 The Student's t PDF with ν degrees of freedom, denoted as t_ν, is defined as

$$p(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \frac{1}{\left(1+\frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}} \qquad -\infty < t < \infty.$$

It can be shown [Kendall and Stuart 1976] to be the PDF obtained for the transformed random variable $t = x/(\sqrt{y/\nu})$, where $x \sim \mathcal{N}(0,1)$, $y \sim \chi^2_\nu$, and x and y are independent. Show that

$$T(\mathbf{x}, \mathbf{w}_R) = \frac{\sqrt{N}\bar{x}}{\sqrt{\frac{1}{N}\sum_{n=0}^{N-1} w_R^2[n]}} \sim t_N$$

if $x[n]$ is WGN with variance σ^2 and $w_R[n]$ is WGN with variance σ^2 and independent of $w[n]$. Hence, the use of $T(\mathbf{x}, \mathbf{w}_R)$ produces a CFAR statistic.

9.2 For the detection of a known DC level A embedded in WGN with variance σ^2, we consider the statistics

$$T_1(\mathbf{x}) = \frac{\sqrt{N}\bar{x}}{\sqrt{\sigma^2}}$$

where σ^2 is known and

$$T_2(\mathbf{x}, \mathbf{w}_R) = \frac{\sqrt{N}\bar{x}}{\sqrt{\frac{1}{N}\sum_{n=0}^{N-1} w_R^2[n]}}$$

where σ^2 is unknown. Under $\mathcal{H}_0$ we have that $T_1(\mathbf{x}) \sim \mathcal{N}(0,1)$ and $T_2(\mathbf{x}, \mathbf{w}_R) \sim t_N$ (see Problem 9.1). Plot the PDFs of the two test statistics and compare for $N = 5$ and $N = 50$. Discuss your results.

9.3 A statistic $T(\mathbf{x})$ is said to be *scale invariant* if $T(a\mathbf{x}) = T(\mathbf{x})$ for any $a > 0$. If $\mathbf{x} = [x[0]\, x[1] \ldots x[N-1]]^T$, where $x[n]$ is WGN with variance σ^2, show that $P_{FA} = \Pr\{T(\mathbf{x}) > \gamma'; \mathcal{H}_0\}$ does not depend on σ^2 if $T(\mathbf{x})$ is scale invariant. It can also be shown that scale invariance is necessary for P_{FA} not to depend on σ^2.

9.4 We wish to detect an unknown DC level A embedded in WGN with unknown variance σ^2 based on N data samples. The DC level is known to take on one of the values ± 1. Derive the GLRT and determine if it yields a CFAR detector. Hint: See Problem 7.1 for the MLE of A.

9.5 For the statistic of (9.4) we determine the asymptotic PDF by using a first-order Taylor expansion, which assumes that A is small (weak signal assumption). To do so, first expand $T(\mathbf{x})$ (considered as a function of A) about $A = 0$. You should be able to show that

$$T(\mathbf{x}) \approx T'(\mathbf{x}) = \frac{\bar{x} - \frac{A}{2}}{\frac{1}{N}\sum_{n=0}^{N-1} x^2[n]}$$

where we neglect the term $A\bar{x}^2$ since $A\bar{x}^2 \approx A^3 << A$. Then use $T(\mathbf{x})$ under $\mathcal{H}_1$ and $T'(\mathbf{x})$ under $\mathcal{H}_0$ so that the denominator in each case converges to σ^2. Finally, verify the PDF of (9.5) for large N. Hint: By Slutsky's theorem [Bickel and Doksum 1977] the denominators of $T(\mathbf{x})$ and $T'(\mathbf{x})$ can be replaced by σ^2 before determining the PDF.

9.6 It was asserted that the asymptotic PDF of the GLRT holds for the hypothesis testing problem (see Section 6.5)

$$\begin{aligned} \mathcal{H}_0 &: A = 0, \sigma^2 > 0 \\ \mathcal{H}_1 &: A \neq 0, \sigma^2 > 0 \end{aligned}$$

but not for (see (9.5))

$$\begin{aligned} \mathcal{H}_0 &: A = 0, \sigma^2 > 0 \\ \mathcal{H}_1 &: A = A_0, \sigma^2 > 0. \end{aligned}$$

In each hypothesis test we are essentially testing whether the vector parameter $[A\, \sigma^2]^T$ lies in a particular subspace of the entire parameter space $\Omega = \{(A, \sigma^2) : -\infty < A < \infty, \sigma^2 > 0\}$. For each problem determine the subspaces corresponding to $\mathcal{H}_0$ and $\mathcal{H}_1$. Conjecture the conditions under which the asymptotic PDF of the GLRT is valid.

9.7 Determine the GLRT for the detection of a signal with alternating sign embedded in WGN or for the detection problem

$$\begin{aligned} \mathcal{H}_0 &: x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 &: x[n] = A(-1)^n + w[n] & n = 0, 1, \ldots, N-1 \end{aligned}$$

where A is unknown and $w[n]$ is WGN with unknown variance σ^2. Also, determine the asymptotic performance. What significance does the energy-to-noise ratio or NA^2/σ^2 have? Hint: The Fisher information matrix is

$$\mathbf{I}(A, \sigma^2) = \begin{bmatrix} N/\sigma^2 & 0 \\ 0 & 2N/\sigma^4 \end{bmatrix}.$$

9.8 Repeat Problem 9.7 but instead determine the Rao test and its asymptotic performance.

9.9 Show that the PDF of

$$T(\mathbf{x}) = \frac{\sum_{n=0}^{N-1} x[n]s[n]}{\sqrt{\frac{1}{N}\sum_{n=0}^{N-1} x^2[n]}}$$

where $x[n]$ is WGN with variance σ^2 does not depend on σ^2. Hint: See Problem 9.3.

9.10 Verify the result given in (9.9).

9.11 Find the GLRT and its exact performance for the detection of a deterministic signal with unknown amplitude embedded in WGN or for the detection problem

$$\begin{array}{ll}\mathcal{H}_0 : x[n] = w[n] & n = 0,1,\ldots,N-1 \\ \mathcal{H}_1 : x[n] = As[n] + w[n] & n = 0,1,\ldots,N-1\end{array}$$

where A is unknown, $s[n]$ is known with $\sum_{n=0}^{N-1} s^2[n] = 1$, and $w[n]$ is WGN with unknown variance σ^2.

9.12 Find the GLRT and its exact performance for the detection of a sinusoid in WGN or for the detection problem

$$\begin{array}{ll}\mathcal{H}_0 : x[n] = w[n] & n = 0,1,\ldots,N-1 \\ \mathcal{H}_1 : x[n] = A\cos(2\pi f_0 n + \phi) + w[n] & n = 0,1,\ldots,N-1\end{array}$$

where A, ϕ are unknown and can take on values $A > 0$ and $0 \le \phi \le 2\pi$, and $w[n]$ is WGN with unknown variance σ^2. Assume that $f_0 = k/N$ for k an integer taking on a value $1 \le k \le N/2 - 1$ (N even). Explain your results.

9.13 Find the GLRT and its exact performance for the detection problem

$$\begin{aligned}\mathcal{H}_0 : x[n] &= w[n] \qquad n = 0,1,\ldots,N-1 \\ \mathcal{H}_1 : x[n] &= \left\{\begin{array}{ll} A + w[n] & n = 0,1,\ldots,N/2-1 \\ B + w[n] & n = N/2, N/2+1,\ldots,N-1\end{array}\right.\end{aligned}$$

for N even, where A, B are unknown, and $w[n]$ is WGN with unknown variance σ^2. Compare your results to the case of an unknown DC level in WGN with unknown variance as in Example 9.2.

9.14 An $F_{r,N-p}$ random variable is equivalent to

$$\frac{\chi_r^2/r}{\chi_{N-p}^2/(N-p)}$$

where the χ_r^2 and χ_{N-p}^2 random variables are independent. Argue that as $N \to \infty$, $\chi_{N-p}^2/(N-p) \to 1$ and hence $F_{r,N-p} \to \chi_r^2/r$. Hint: Recall that a χ_ν^2 random variable is the sum of the squares of ν IID $\mathcal{N}(0,1)$ random variables.

9.15 In Example 9.3 interpret the test statistic

$$T(\mathbf{x}) = \frac{N-p}{p}\frac{||\mathbf{P}_H\mathbf{x}||^2}{||\mathbf{P}_H^\perp\mathbf{x}||^2}$$

if $\mathbf{H} = \mathbf{1}$ (i.e., the signal is a DC level).

9.16 Verify $\mathbf{I}(\boldsymbol{\theta})$ in Example 9.4. Assume that $0 < f_0 < 1/2$ and make any large data record length approximations such as

$$\begin{aligned}\frac{1}{N}\sum_{n=0}^{N-1}\sin 4\pi f_0 n &\approx 0 \\ \frac{1}{N}\sum_{n=0}^{N-1} n\sin 4\pi f_0 n &\approx 0.\end{aligned}$$

9.17 Verify the ACF and PSD of an AR process of order one given by (9.18) and (9.19), respectively. To do so first show that the ACF of the output $w[n]$ of a causal linear shift invariant system with impulse response $h[n]$ is

$$r_{ww}[k] = \sigma_u^2\sum_{n=0}^{\infty} h[n]h[n+|k|]$$

if the input to the system is white noise $u[n]$ with variance σ_u^2.

9.18 Show that for large N

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} |A(f)X(f)|^2 df \approx \sum_{n=1}^{N-1}(x[n] + a[1]x[n-1])^2$$

where $A(f) = 1 + a[1]\exp(-j2\pi f)$ and $X(f) = \sum_{n=0}^{N-1} x[n]\exp(-j2\pi f n)$.

9.19 Consider the detection of an unknown DC level A embedded in WGN with variance σ^2. Based on N data samples, determine the asymptotic performance of a GLRT when σ^2 is known and when σ^2 is unknown. Explain your results. Hint: See Problem 9.7 for the Fisher information matrix.

9.20 Let $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\alpha}_s), \mathbf{C}(\boldsymbol{\alpha}_w))$, where $\mathbf{x}$ is an $N \times 1$ random vector, so that the mean depends only on $\boldsymbol{\alpha}_s$ (the signal parameter) and the covariance matrix only on $\boldsymbol{\alpha}_w$ (the noise parameter). Show that the submatrices of the partitioned Fisher information matrix

$$\mathbf{I}(\boldsymbol{\theta}) = \mathbf{I}(\boldsymbol{\alpha}_s, \boldsymbol{\alpha}_w) = \begin{bmatrix}\mathbf{I}_{\alpha_s\alpha_s} & \mathbf{I}_{\alpha_s\alpha_w} \\ \mathbf{I}_{\alpha_w\alpha_s} & \mathbf{I}_{\alpha_w\alpha_w}\end{bmatrix}$$

are given by

$$\begin{aligned}[\mathbf{I}_{\alpha_s\alpha_s}]_{ij} &= \left[\frac{\partial\boldsymbol{\mu}(\boldsymbol{\alpha}_s)}{\partial\alpha_{s_i}}\right]^T \mathbf{C}^{-1}(\boldsymbol{\alpha}_w)\left[\frac{\partial\boldsymbol{\mu}(\boldsymbol{\alpha}_s)}{\partial\alpha_{s_j}}\right] \\ \mathbf{I}_{\alpha_s\alpha_w} &= \mathbf{I}_{\alpha_w\alpha_s} = \mathbf{0} \\ [\mathbf{I}_{\alpha_w\alpha_w}]_{ij} &= \frac{1}{2}\mathrm{tr}\left[\mathbf{C}^{-1}(\boldsymbol{\alpha}_w)\frac{\partial\mathbf{C}(\boldsymbol{\alpha}_w)}{\partial\alpha_{w_i}}\mathbf{C}^{-1}(\boldsymbol{\alpha}_w)\frac{\partial\mathbf{C}(\boldsymbol{\alpha}_w)}{\partial\alpha_{w_j}}\right].\end{aligned}$$

Use the result given in (3.31) of [Kay-I 1993, pg. 47], where $\boldsymbol{\theta} = [\boldsymbol{\alpha}_s^T \, \boldsymbol{\alpha}_w^T]^T$. An example of this result was given in Problem 9.19.

9.21 An AR process of order one has the $N \times N$ inverse covariance matrix [Kay 1988]

$$\mathbf{C}^{-1} = \frac{1}{\sigma_u^2} \begin{bmatrix} 1 & a[1] & 0 & 0 & \dots & 0 \\ a[1] & 1+a^2[1] & a[1] & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & a[1] & 1+a^2[1] & a[1] \\ 0 & 0 & \dots & 0 & a[1] & 1 \end{bmatrix}.$$

Use this to show that $T_R(\mathbf{x})$ as given by (9.31) is approximately equal to (9.32). Hint: First show that

$$\mathbf{C}^{-1} \approx \frac{1}{\sigma_u^2} \mathbf{A}^T \mathbf{A}$$

where $\mathbf{A}$ is the $N \times N$ matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ a[1] & 1 & 0 & 0 & \dots & 0 \\ 0 & a[1] & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & a[1] & 1 \end{bmatrix}.$$

9.22 Find the Rao test statistic for the detection problem

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \dots, N-1 \\ \mathcal{H}_1 : x[n] = As[n] + w[n] & n = 0, 1, \dots, N-1 \end{array}$$

where A is unknown and $w[n]$ is zero mean Gaussian noise with known covariance matrix $\mathbf{C}(\boldsymbol{\theta}_w)$.

9.23 Show that $T_R(\mathbf{x})$ as given by (9.34) (the noise covariance matrix is known) is equivalent to

$$T_R(\mathbf{x}) = \frac{\hat{A}^2}{\operatorname{var}(\hat{A})}$$

where $\hat{A}$ is the MVU estimator of A. Recall that for the classical general linear model the MVU estimator is $\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x}$ and its covariance matrix is $\mathbf{C}_{\hat{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}$.

9.24 Find the Rao test statistic for an unknown DC level A embedded in Gaussian noise with covariance matrix $\mathbf{C}(P_0) = P_0\mathbf{Q}$, where $P_0 > 0$ is unknown and $\mathbf{Q}$ is a known positive definite matrix. Assume that $\{x[0], x[1], \dots, x[N-1]\}$ are observed.

Appendix 9A

Derivation of GLRT for Classical Linear Model for σ^2 Unknown

The reader may wish to refer to Appendix 7B for the case when σ^2 is known. The GLRT decides $\mathcal{H}_1$ if

$$L_G(\mathbf{x}) = \frac{p(\mathbf{x}; \hat{\boldsymbol{\theta}}_1, \hat{\sigma_1^2})}{p(\mathbf{x}; \hat{\boldsymbol{\theta}}_0, \hat{\sigma_0^2})} > \gamma$$

where

$$p(\mathbf{x}; \boldsymbol{\theta}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})\right]$$

and $\hat{\boldsymbol{\theta}}_i, \hat{\sigma_i^2}$ is the MLE of $\boldsymbol{\theta}, \sigma^2$ under $\mathcal{H}_i$. Under $\mathcal{H}_1$ there are no constraints on $\boldsymbol{\theta}$ other than that we must exclude the set of $\boldsymbol{\theta}$ for which $\mathbf{A}\boldsymbol{\theta} = \mathbf{b}$. However, the unconstrained MLE of $\boldsymbol{\theta}$ or $\hat{\boldsymbol{\theta}}$ (where we allow $\boldsymbol{\theta}$ to take on any value in R^p) has zero probability of satisfying $\mathbf{A}\hat{\boldsymbol{\theta}} = \mathbf{b}$. The reader may wish to consider for the unknown signal amplitude case the probability that $\hat{A} = \sum_{n=0}^{N-1} x[n]s[n] / \sum_{n=0}^{N-1} s^2[n]$ satisfies $\hat{A} = 0$. Hence, $\hat{\boldsymbol{\theta}}_1$ is equivalent to the *unconstrained MLE* $\hat{\boldsymbol{\theta}}$, which for the linear model is

$$\hat{\boldsymbol{\theta}}_1 = \hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}.$$

The MLE of σ^2 under $\mathcal{H}_1$ is found by maximizing $p(\mathbf{x}; \hat{\boldsymbol{\theta}}_1, \sigma^2)$ over σ^2. This is easily shown to produce (see [Kay-I 1993, pp. 176–177])

$$\hat{\sigma_1^2} = \frac{1}{N}(\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}_1)^T(\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}_1).$$

Hence

$$p(\mathbf{x};\hat{\boldsymbol\theta}_1,\hat{\sigma_1^2}) = \frac{1}{(2\pi\hat{\sigma_1^2})^{\frac{N}{2}}}\exp\left(-\frac{N}{2}\right).$$

To find $\hat{\boldsymbol\theta}_0$ we must find the *constrained* MLE or the MLE of $\boldsymbol\theta$ for which $\mathbf{A}\boldsymbol\theta = \mathbf{b}$. This is equivalent to the constrained least squares estimator (since $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$), which has been shown to be [Kay-I 1993, pp. 251–252]

$$\hat{\boldsymbol\theta}_0 = \hat{\boldsymbol\theta}_1 - (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\left[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\right]^{-1}(\mathbf{A}\hat{\boldsymbol\theta}_1 - \mathbf{b}).$$

Also, it can easily be shown that

$$\hat{\sigma_0^2} = \frac{1}{N}(\mathbf{x}-\mathbf{H}\hat{\boldsymbol\theta}_0)^T(\mathbf{x}-\mathbf{H}\hat{\boldsymbol\theta}_0)$$

so that

$$p(\mathbf{x};\hat{\boldsymbol\theta}_0,\hat{\sigma_0^2}) = \frac{1}{(2\pi\hat{\sigma_0^2})^{\frac{N}{2}}}\exp\left(-\frac{N}{2}\right).$$

Thus, the GLRT is

$$L_G(\mathbf{x}) = \left(\frac{\hat{\sigma_0^2}}{\hat{\sigma_1^2}}\right)^{\frac{N}{2}}.$$

If we let $T'(\mathbf{x}) = L_G(\mathbf{x})^{2/N} - 1$, which is a monotonically increasing function of $L_G(\mathbf{x})$, then

$$\begin{aligned}T'(\mathbf{x}) &= \frac{\hat{\sigma_0^2}-\hat{\sigma_1^2}}{\hat{\sigma_1^2}}\\ &= \frac{(\mathbf{x}-\mathbf{H}\hat{\boldsymbol\theta}_0)^T(\mathbf{x}-\mathbf{H}\hat{\boldsymbol\theta}_0) - (\mathbf{x}-\mathbf{H}\hat{\boldsymbol\theta}_1)^T(\mathbf{x}-\mathbf{H}\hat{\boldsymbol\theta}_1)}{(\mathbf{x}-\mathbf{H}\hat{\boldsymbol\theta}_1)^T(\mathbf{x}-\mathbf{H}\hat{\boldsymbol\theta}_1)}.\end{aligned}$$

But as shown in Appendix 7B, the numerator reduces to

$$(\mathbf{A}\hat{\boldsymbol\theta}_1 - \mathbf{b})^T[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}(\mathbf{A}\hat{\boldsymbol\theta}_1 - \mathbf{b})$$

and the denominator can be reduced as follows

$$\begin{aligned}(\mathbf{x}-\mathbf{H}\hat{\boldsymbol\theta}_1)^T(\mathbf{x}-\mathbf{H}\hat{\boldsymbol\theta}_1) &= (\mathbf{x}-\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x})^T(\mathbf{x}-\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x})\\ &= \mathbf{x}^T(\mathbf{I}-\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T)^T(\mathbf{I}-\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T)\mathbf{x}\\ &= \mathbf{x}^T(\mathbf{I}-\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T)\mathbf{x}\end{aligned}$$

so that

$$T'(\mathbf{x}) = \frac{(\mathbf{A}\hat{\boldsymbol\theta}_1 - \mathbf{b})^T[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}(\mathbf{A}\hat{\boldsymbol\theta}_1 - \mathbf{b})}{\mathbf{x}^T(\mathbf{I}-\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T)\mathbf{x}}.$$

To determine the detector performance we consider the equivalent statistic

$$T'(\mathbf{x}) = \frac{(\mathbf{A}\hat{\boldsymbol\theta}_1 - \mathbf{b})^T[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}(\mathbf{A}\hat{\boldsymbol\theta}_1 - \mathbf{b})/\sigma^2}{\mathbf{x}^T(\mathbf{I}-\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T)\mathbf{x}/\sigma^2} = \frac{N(\mathbf{x})}{D(\mathbf{x})}$$

and show that the numerator $N(\mathbf{x})$ has the PDF

$$N(\mathbf{x}) \sim \begin{cases}\chi_r^2 & \text{under } \mathcal{H}_0\\ \chi_r'^2(\lambda) & \text{under } \mathcal{H}_1\end{cases} \tag{9A.1}$$

where

$$\lambda = \frac{(\mathbf{A}\boldsymbol\theta_1 - \mathbf{b})^T[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}(\mathbf{A}\boldsymbol\theta_1 - \mathbf{b})}{\sigma^2}$$

and the denominator $D(\mathbf{x})$ has the PDF

$$D(\mathbf{x}) \sim \begin{cases}\chi_{N-p}^2 & \text{under } \mathcal{H}_0\\ \chi_{N-p}^2 & \text{under } \mathcal{H}_1\end{cases}$$

and also that $N(\mathbf{x})$ and $D(\mathbf{x})$ are independent. Hence, normalizing the numerator and denominator by the corresponding degrees of freedom produces

$$T(\mathbf{x}) = \frac{N(\mathbf{x})/r}{D(\mathbf{x})/(N-p)} \sim \begin{cases}F_{r,N-p} & \text{under } \mathcal{H}_0\\ F'_{r,N-p}(\lambda) & \text{under } \mathcal{H}_1.\end{cases}$$

Now (9A.1) follows from Theorem 7.1. To find the PDF of $D(\mathbf{x})$ we note that it is a quadratic form in $\mathbf{x}$ or $D(\mathbf{x}) = \mathbf{x}^T\mathbf{B}\mathbf{x}$, where the matrix $\mathbf{B}$ is $\mathbf{B} = (\mathbf{I} - \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T)/\sigma^2 = \mathbf{P}_H^\perp/\sigma^2$. The matrix $\mathbf{P}_H^\perp$ is the orthogonal projection matrix (see [Kay-I 1993, pg. 231]). It can be shown (see Problem 8.12 in [Kay-I 1993, pg. 277]) that the projection matrix is idempotent and has rank $N-p$. Furthermore, since $\mathbf{P}_H^\perp\mathbf{H} = \mathbf{0}$, we have under $\mathcal{H}_i$ that

$$\begin{aligned}D(\mathbf{x}) &= \frac{1}{\sigma^2}(\mathbf{H}\boldsymbol\theta_i + \mathbf{w})^T\mathbf{P}_H^\perp(\mathbf{H}\boldsymbol\theta_i + \mathbf{w})\\ &= \frac{1}{\sigma^2}\mathbf{w}^T\mathbf{P}_H^\perp\mathbf{w}\\ &= \left(\frac{\mathbf{w}}{\sigma}\right)^T\mathbf{P}_H^\perp\left(\frac{\mathbf{w}}{\sigma}\right)\end{aligned}$$

where $\mathbf{w}/\sigma \sim \mathcal{N}(\mathbf{0},\mathbf{I})$. From (2.29) it follows that

$$D(\mathbf{x}) \sim \begin{cases}\chi_{N-p}^2 & \text{under } \mathcal{H}_0\\ \chi_{N-p}^2 & \text{under } \mathcal{H}_1.\end{cases}$$

Finally, to show that $N(\mathbf{x})$ and $D(\mathbf{x})$ are independent under either hypothesis, we first note that $N(\mathbf{x})$ is a function of $\hat{\boldsymbol{\theta}}_1$ only. It thus remains to be shown that $\hat{\boldsymbol{\theta}}_1$ is independent of $D(\mathbf{x})$. To do so we appeal to the theorem that asserts the independence of a quadratic form $\mathbf{x}^T\mathbf{B}\mathbf{x}$ and a linear form $\mathbf{d}^T\mathbf{x}$, where $\mathbf{d}$ is a column vector, if and only if $\mathbf{B}\mathbf{d} = \mathbf{0}$ for $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2\mathbf{I})$ [Kendall and Stuart 1976]. To apply the theorem we let $z = \boldsymbol{\alpha}^T\hat{\boldsymbol{\theta}}_1$, where $\boldsymbol{\alpha}$ is an arbitrary $p\times 1$ vector, so that z is a linear form in $\mathbf{x}$. Then $z = \boldsymbol{\alpha}^T\hat{\boldsymbol{\theta}}_1 = \boldsymbol{\alpha}^T(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$, where $\mathbf{d} = \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\boldsymbol{\alpha}$. But $\mathbf{B}\mathbf{d} = \mathbf{P}_H^{\perp}\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\boldsymbol{\alpha} = \mathbf{0}$ since $\mathbf{P}_H^{\perp}\mathbf{H} = \mathbf{0}$. This shows that $\boldsymbol{\alpha}^T\hat{\boldsymbol{\theta}}_1$ is independent of $D(\mathbf{x})$ for all $\boldsymbol{\alpha}$. Finally, by appealing to characteristic functions it can be shown that $\hat{\boldsymbol{\theta}}_1$ is independent of $D(\mathbf{x})$ and hence $N(\mathbf{x})$ is independent of $D(\mathbf{x})$.

Appendix 9B

Rao Test for General Linear Model with Unknown Noise Parameters

Assume the general linear model or $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$, where $\mathbf{H}$ is a known $N \times p$ matrix, $\boldsymbol{\theta}$ is an unknown $p \times 1$ parameter vector, and $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}(\boldsymbol{\theta}_w))$. The covariance matrix $\mathbf{C}$ is assumed to depend on an unknown $q \times 1$ parameter vector $\boldsymbol{\theta}_w$. The hypothesis test is

$$\begin{aligned}\mathcal{H}_0 &: \boldsymbol{\theta} = \mathbf{0}, \boldsymbol{\theta}_w \\ \mathcal{H}_1 &: \boldsymbol{\theta} \neq \mathbf{0}, \boldsymbol{\theta}_w.\end{aligned}$$

The Rao test and its asymptotic performance follow from Chapter 6. To avoid confusion in the definition of $\boldsymbol{\theta}$, we replace the $\boldsymbol{\theta}$ used in Chapter 6 to represent the total vector of unknown parameters by the vector $\boldsymbol{\xi}$, where $\boldsymbol{\xi} = [\boldsymbol{\theta}_r^T\ \boldsymbol{\theta}_s^T]^T$. Then, we let $\boldsymbol{\theta}_r = \boldsymbol{\theta}$ and $\boldsymbol{\theta}_s = \boldsymbol{\theta}_w$ so that

$$T_R(\mathbf{x}) = \left.\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta}, \boldsymbol{\theta}_w)}{\partial \boldsymbol{\theta}}\right|^T_{\boldsymbol{\theta}=\mathbf{0}, \boldsymbol{\theta}_w=\hat{\boldsymbol{\theta}}_{w_0}} [\mathbf{I}^{-1}(\tilde{\boldsymbol{\xi}})]_{\theta\theta} \left.\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta}, \boldsymbol{\theta}_w)}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\mathbf{0}, \boldsymbol{\theta}_w=\hat{\boldsymbol{\theta}}_{w_0}}$$

where $\hat{\boldsymbol{\theta}}_{w_0}$ is the MLE of $\boldsymbol{\theta}_w$ under $\mathcal{H}_0$ and

$$T_R(\mathbf{x}) \overset{a}{\sim} \begin{cases} \chi_p^2 & \text{under } \mathcal{H}_0 \\ \chi_p'^2(\lambda) & \text{under } \mathcal{H}_1 \end{cases}$$

where $\lambda = \boldsymbol{\theta}_1^T\mathbf{I}_{\theta\theta}(\mathbf{0}, \boldsymbol{\theta}_w)\boldsymbol{\theta}_1$ (see (6.24)). The noncentrality parameter is a result of the block-diagonal nature of the Fisher information matrix. Now for the general linear model it can be shown that (in a similar fashion to [Kay-I 1993, pp. 84–85])

$$\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta}, \boldsymbol{\theta}_w)}{\partial \boldsymbol{\theta}} = \mathbf{H}^T\mathbf{C}^{-1}(\boldsymbol{\theta}_w)\mathbf{H}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$$

where $\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}(\boldsymbol{\theta}_w)\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}(\boldsymbol{\theta}_w)\mathbf{x}$ so that

$$\left.\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta},\boldsymbol{\theta}_w)}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\mathbf{0},\boldsymbol{\theta}_w=\hat{\boldsymbol{\theta}}_{w_0}} = \mathbf{H}^T\mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0})\mathbf{H}\hat{\boldsymbol{\theta}}.$$

Also, because of the block-diagonal Fisher information matrix we have

$$[\mathbf{I}^{-1}(\tilde{\boldsymbol{\xi}})]_{\theta\theta} = \mathbf{I}^{-1}_{\theta\theta}(\tilde{\boldsymbol{\xi}}) = (\mathbf{H}^T\mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0})\mathbf{H})^{-1}.$$

This yields

$$\begin{aligned}
T_R(\mathbf{x}) &= \hat{\boldsymbol{\theta}}^T\mathbf{H}^T\mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0})\mathbf{H}(\mathbf{H}^T\mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0})\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0})\mathbf{H}\hat{\boldsymbol{\theta}} \\
&= \hat{\boldsymbol{\theta}}^T\mathbf{H}^T\mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0})\mathbf{H}\hat{\boldsymbol{\theta}} \\
&= \mathbf{x}^T\mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0})\mathbf{H}(\mathbf{H}^T\mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0})\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0})\mathbf{H} \\
&\quad \cdot (\mathbf{H}^T\mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0})\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0})\mathbf{x} \\
&= \mathbf{x}^T\mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0})\mathbf{H}(\mathbf{H}^T\mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0})\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0})\mathbf{x}
\end{aligned}$$

and

$$\lambda = \boldsymbol{\theta}_1^T\mathbf{H}^T\mathbf{C}^{-1}(\boldsymbol{\theta}_w)\mathbf{H}\boldsymbol{\theta}_1.$$

When $\boldsymbol{\theta}_w$ is known, then using the Rao test for no nuisance parameters, we arrive at the same test statistic except that the true value of $\boldsymbol{\theta}_w$ replaces its MLE. Also, the asymptotic performance is the same.

Appendix 9C

Asymptotically Equivalent Rao Test for Signal Processing Example

Since $\mathbf{C}$ is a covariance matrix for a WSS random process, we have from Chapter 2 that as $N \to \infty$

$$\mathbf{C}^{-1} = \sum_{k=0}^{N-1} \frac{1}{P_{ww}(f_k)}\mathbf{v}_k\mathbf{v}_k^H$$

where $P_{ww}(f_k)$ is the PSD of $w[n]$ evaluated at $f = f_k = k/N$ and also $\mathbf{v}_k = (1/\sqrt{N})[1\ \exp(j2\pi f_k)\ldots\exp[j2\pi f_k(N-1)]]^T$. Hence, letting $\mathbf{H} = [\mathbf{h}_1\, \mathbf{h}_2]$, where $\mathbf{h}_i$ is the ith column of $\mathbf{H}$

$$\begin{aligned}
\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H} &= [\ \mathbf{h}_1\ \ \mathbf{h}_2\ ]^T \sum_{k=0}^{N-1}\frac{1}{P_{ww}(f_k)}\mathbf{v}_k\mathbf{v}_k^H[\ \mathbf{h}_1\ \ \mathbf{h}_2\ ] \\
&= \begin{bmatrix} \mathbf{h}_1^T \\ \mathbf{h}_2^T \end{bmatrix} \sum_{k=0}^{N-1}\frac{1}{P_{ww}(f_k)}\mathbf{v}_k[\ \mathbf{v}_k^H\mathbf{h}_1\ \ \mathbf{v}_k^H\mathbf{h}_2\ ] \\
&= \sum_{k=0}^{N-1}\frac{1}{P_{ww}(f_k)}\begin{bmatrix} \mathbf{h}_1^T\mathbf{v}_k \\ \mathbf{h}_2^T\mathbf{v}_k \end{bmatrix}[\ \mathbf{v}_k^H\mathbf{h}_1\ \ \mathbf{v}_k^H\mathbf{h}_2\ ] \\
&= \sum_{k=0}^{N-1}\frac{1}{P_{ww}(f_k)}\begin{bmatrix} \mathbf{h}_1^T\mathbf{v}_k\mathbf{v}_k^H\mathbf{h}_1 & \mathbf{h}_1^T\mathbf{v}_k\mathbf{v}_k^H\mathbf{h}_2 \\ \mathbf{h}_2^T\mathbf{v}_k\mathbf{v}_k^H\mathbf{h}_1 & \mathbf{h}_2^T\mathbf{v}_k\mathbf{v}_k^H\mathbf{h}_2 \end{bmatrix}.
\end{aligned}$$

But as $N \to \infty$

$$\begin{aligned}
\sum_{k=0}^{N-1} \frac{1}{P_{ww}(f_k)} \mathbf{h}_i^T \mathbf{v}_k \mathbf{v}_k^H \mathbf{h}_j &= \frac{1}{N} \sum_{k=0}^{N-1} \frac{H_i^*(f_k) H_j(f_k)}{P_{ww}(f_k)} \\
&\to \int_0^1 \frac{H_i^*(f) H_j(f)}{P_{ww}(f)} df \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{H_i^*(f) H_j(f)}{P_{ww}(f)} df \\
&= \begin{cases} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{|H_i(f)|^2}{P_{ww}(f)} df & i = j \\ 0 & i \neq j \end{cases}
\end{aligned}$$

where $H_i(f)$ is the Fourier transform of the elements in $\mathbf{h}_i$ or more explicitly $H_i(f) = \sum_{n=0}^{N-1} [\mathbf{h}_i]_n \exp(-j2\pi f n)$. The last result follows from the fact that $\mathbf{h}_1$ and $\mathbf{h}_2$ are Hilbert transforms of each other for large N so that $H_2(f) = -j\,\mathrm{sgn}(f) H_1(f)$. Hence

$$\begin{aligned}
\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{H_1^*(f) H_2(f)}{P_{ww}(f)} df &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{-j\,\mathrm{sgn}(f) |H_1(f)|^2}{P_{ww}(f)} df \\
&= 0
\end{aligned}$$

since $|H_1(f)|^2 / P_{ww}(f)$ is even in f and $\mathrm{sgn}(f)$ is odd in f. Also, note that $|H_1(f)|^2 = |H_2(f)|^2 = |S(f)|^2$, where $S(f) = \mathcal{F}\{\cos[2\pi(f_0 n + \frac{1}{2} m^2)]\}$. We have then as $N \to \infty$

$$(\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} = \frac{1}{\beta} \mathbf{I} \tag{9C.1}$$

where

$$\beta = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{|S(f)|^2}{P_{ww}(f)} df.$$

Now

$$\begin{aligned}
T_R(\mathbf{x}) &= \mathbf{x}^T \mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0}) \mathbf{H} (\mathbf{H}^T \mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0}) \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0}) \mathbf{x} \\
&= \frac{1}{\hat{\beta}} \mathbf{x}^T \mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0}) \mathbf{H} \mathbf{H}^T \mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0}) \mathbf{x} \\
&= \frac{1}{\hat{\beta}} ||\mathbf{H}^T \mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0}) \mathbf{x}||^2 = \frac{1}{\hat{\beta}} [(\mathbf{h}_1^T \mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0}) \mathbf{x})^2 + (\mathbf{h}_2^T \mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0}) \mathbf{x})^2] \\
&= \frac{1}{\hat{\beta}} |(\mathbf{h}_1 + j\mathbf{h}_2)^T \mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0}) \mathbf{x}|^2 \\
&= \frac{1}{\hat{\beta}} |\tilde{\mathbf{h}}^T \mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0}) \mathbf{x}|^2
\end{aligned}$$

where $\tilde{\mathbf{h}} = \mathbf{h}_1 + j\mathbf{h}_2$ and $\hat{\beta}$ is obtained from β by using $\hat{\boldsymbol{\theta}}_{w_0}$ in $P_{ww}(f)$. But

$$\begin{aligned}
\tilde{\mathbf{h}}^T \mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0}) \mathbf{x} &= \tilde{\mathbf{h}}^T \sum_{k=0}^{N-1} \frac{1}{\hat{P}_{ww}(f_k)} \mathbf{v}_k \mathbf{v}_k^H \mathbf{x} \\
&= \sum_{k=0}^{N-1} \frac{1}{\hat{P}_{ww}(f_k)} \tilde{\mathbf{h}}^T \mathbf{v}_k \mathbf{v}_k^H \mathbf{x} \\
&\to \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{X(f) \tilde{S}^*(f)}{\hat{P}_{ww}(f)} df
\end{aligned}$$

where $\tilde{S}(f) = \mathcal{F}\{\exp[j2\pi(f_0 n + \frac{1}{2} m n^2)]\}$. Finally, for the detection of a linear FM signal in WSS random noise, we have the approximate Rao test statistic

$$T_R(\mathbf{x}) = \frac{\left| \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{X(f) \tilde{S}^*(f)}{\hat{P}_{ww}(f)} df \right|^2}{\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{|S(f)|^2}{\hat{P}_{ww}(f)} df}.$$

Specifically, for the AR noise model we have

$$P_{ww}(f) = \frac{\sigma_u^2}{|A(f)|^2}$$

so that after substitution

$$T_R(\mathbf{x}) = \frac{\left| \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\hat{A}_0(f) X(f) (\hat{A}_0(f) \tilde{S}(f))^*}{\hat{\sigma}_{u_0}^2} df \right|^2}{\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{|\hat{A}_0(f) S(f)|^2}{\hat{\sigma}_{u_0}^2} df}$$

and by Parseval's theorem this is approximately

$$T_R(\mathbf{x}) = \frac{\left| \sum_{n=p_{\mathrm{AR}}}^{N-1} \left(\sum_{k=0}^{p_{\mathrm{AR}}} \hat{a}_0[k] x[n-k] \right) \left(\sum_{l=0}^{p_{\mathrm{AR}}} \hat{a}_0[l] \tilde{s}[n-l] \right) \right|^2}{\hat{\sigma}_{u_0}^2 \sum_{n=p_{\mathrm{AR}}}^{N-1} \left(\sum_{l=0}^{p_{\mathrm{AR}}} \hat{a}_0[l] s[n-l] \right)^2}.$$

Also, the noncentrality parameter or $\lambda = (\mathbf{H}\boldsymbol{\theta}_1)^T \mathbf{C}^{-1}(\boldsymbol{\theta}_w)(\mathbf{H}\boldsymbol{\theta}_1)$ follows from (9C.1) as

$$\begin{aligned}
\lambda &= \beta \boldsymbol{\theta}_1^T \boldsymbol{\theta}_1 \\
&= A^2 \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{|S(f)|^2 |A(f)|^2}{\sigma_u^2} df
\end{aligned}$$

since $\boldsymbol{\theta}_1^T\boldsymbol{\theta}_1 = \alpha_1^2 + \alpha_2^2 = A^2$. By Parseval's theorem we have for large N

$$\lambda = \frac{A^2}{\sigma_u^2} \sum_{n=p_{\mathrm{AR}}}^{N-1} \left(\sum_{k=0}^{p_{\mathrm{AR}}} a[k]s[n-k] \right)^2 .$$

Chapter 10

NonGaussian Noise

10.1 Introduction

The Gaussian noise assumption is usually made for mathematical simplicity and justified by the central limit theorem. For many practical problems of interest it is entirely adequate and results in easily implemented detectors. However, some noise cannot be characterized as being Gaussian due to infrequent but high level events. These events, sometimes referred to as "noise spikes," are encountered in extremely low frequency (ELF) electromagnetic noise due to thunderstorms or in under ice acoustic noise due to iceberg breakup, as examples. In these cases it is important to model these noise spikes more accurately than the description afforded by the Gaussian model. A failure to do so leads to poor detection performance. In this chapter we study the detection of deterministic signals in nonGaussian noise. As this is a current area of research, we limit our discussion to the key ideas. The interested reader is referred to [Middleton 1984, Kassam 1988, Poor 1988, Kay and Sengupta 1991].

10.2 Summary

A general class of nonGaussian PDFs is given by (10.4), which includes the Gaussian PDF as a special case. For a known deterministic signal in IID nonGaussian noise, the NP detector is given by (10.7) or by (10.8) for a symmetrized version. In either case the asymptotic performance can be determined from (10.11). A weak signal equivalent or the locally optimum detector is given by (10.10). For a deterministic signal known except for amplitude, the Rao test produces the statistic of (10.19) with asymptotic performance given by (10.17). The Rao test for a deterministic signal with unknown parameters as described by the linear model form is given in Theorem 10.1 (see (10.25)–(10.27)). It is applied in Section 10.6 to the detection of a sinusoid of unknown amplitude and phase in IID nonGaussian noise. This produces the Rao test of (10.28) and (10.29) for a generalized Gaussian PDF. Its asymptotic performance is found from (10.31) and (10.32).

10.3 NonGaussian Noise Characteristics

The simplest type of nonGaussian noise is one for which the samples $w[n]$ are IID and for which each sample has a first-order PDF that does not obey the Gaussian form. We have already encountered several of these PDFs. As an example, the Laplacian PDF is

$$p(w[n]) = \frac{1}{\sqrt{2\sigma^2}} \exp\left(-\sqrt{\frac{2}{\sigma^2}}|w[n]|\right) \qquad -\infty < w[n] < \infty \tag{10.1}$$

where σ^2 is the variance or noise power. A Gaussian PDF with the same variance is

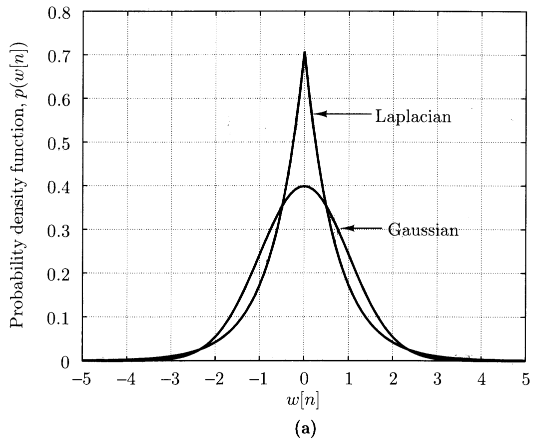
$$p(w[n]) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}w^2[n]\right). \tag{10.2}$$

Since each PDF is zero mean, the first two moments agree. In Figure 10.1 we plot the two PDFs, with Figure 10.1a using a linear vertical axis and Figure 10.1b using a logarithmic scale. It is seen that the main difference is a heavier-tailed characteristic for the Laplacian PDF. This has the effect of producing noise samples that are larger in magnitude than those of the Gaussian PDF. This follows because $\Pr\{|w[n]| > \gamma\}$ is larger for the Laplacian PDF (see Problem 10.1). To illustrate this we plot a time series realization of IID noise samples in Figure 10.2 for each of the PDFs. The nonGaussian time series exhibits *spikes* or *outliers* due to the heavy PDF tails. It would be expected that any detector would need to account for these outliers so that the false alarms would not be excessive. We will see shortly that good detectors for nonGaussian noise typically include nonlinearities or *clippers* to reduce the noise spikes.

The degree of nonGaussianity of a zero mean PDF is typically measured by its *kurtosis* relative to a Gaussian PDF. This is defined as

$$\gamma_2 = \frac{E(w^4[n])}{E^2(w^2[n])} - 3. \tag{10.3}$$

A heavier-tailed PDF will have a larger value of $E(w^4[n])$ for a given noise power and hence will have a larger kurtosis. For a Gaussian PDF, $\gamma_2 = 0$ since $E(w^4[n]) = 3\sigma^4$, while for a nonGaussian PDF, γ_2 will depart from zero. For the Laplacian PDF example it can be shown that $E(w^4[n]) = 6\sigma^4$ and hence $\gamma_2 = 3$ (see Problem 10.2). It is also possible for the kurtosis to be negative, for which $E(w^4[n]) < 3E^2(w^2[n])$ or whose fourth-order moment is smaller than that of the Gaussian. Such PDFs



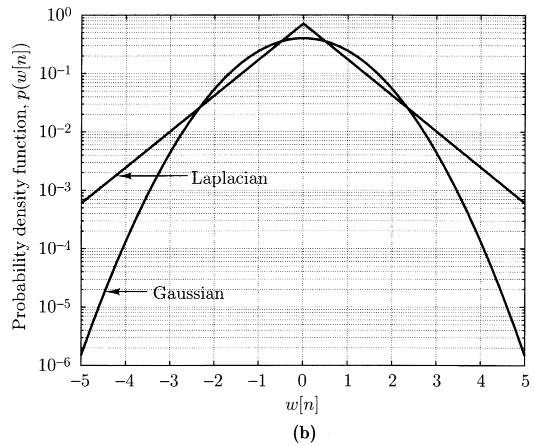


Figure 10.1. Gaussian versus nonGaussian PDF ($\sigma^2 = 1$) (a) Linear scale (b) Logarithmic scale.

have tails that fall off more quickly than the Gaussian. As an example, if $w[n] \sim \mathcal{U}[-\sqrt{3\sigma^2}, \sqrt{3\sigma^2}]$, which has a variance of σ^2, the fourth-order moment is

$$E(w^4[n]) = \int_{-\sqrt{3\sigma^2}}^{\sqrt{3\sigma^2}} w^4[n] \frac{1}{2\sqrt{3\sigma^2}} dw[n]$$

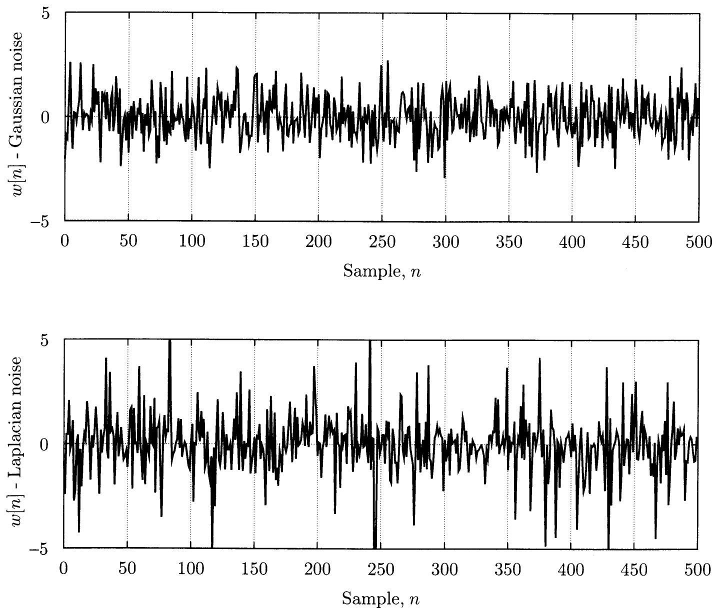


Figure 10.2. Realization of Gaussian and nonGaussian noise ($\sigma^2 = 1$).

$$= \frac{9\sigma^4}{5}$$

so that $\gamma_2 = -1.2$.

A general family of PDFs that encompass the Gaussian, Laplacian, and uniform PDFs is the *generalized Gaussian* or exponential power distribution [Box and Tiao 1973]. In Example 6.9 we studied the detection of an unknown DC level in this type of noise. The generalized Gaussian PDF is

$$p(w) = \frac{c_1(\beta)}{\sqrt{\sigma^2}} \exp\left(-c_2(\beta)\left|\frac{w}{\sqrt{\sigma^2}}\right|^{\frac{2}{1+\beta}}\right) \tag{10.4}$$

where

$$c_1(\beta) = \frac{\Gamma^{\frac{1}{2}}(\frac{3}{2}(1+\beta))}{(1+\beta)\Gamma^{\frac{3}{2}}(\frac{1}{2}(1+\beta))}$$

$$c_2(\beta) = \left[\frac{\Gamma(\frac{3}{2}(1+\beta)}{\Gamma(\frac{1}{2}(1+\beta))}\right]^{\frac{1}{1+\beta}}$$

for $\beta > -1$ and $\Gamma(x)$ is the Gamma function or

$$\Gamma(x) = \int_0^\infty u^{x-1}\exp(-u)du.$$

For $\beta = 0$ the PDF is Gaussian, while for $\beta = 1$ it is Laplacian. As $\beta \to -1$, it tends to the uniform PDF (see Problem 10.3). In Example 6.9 we had chosen $\beta = -1/2$ for which the tails fall off more quickly than the Gaussian or as $\exp[-c_2(\beta)|w/\sqrt{\sigma^2}|^4]$ from (10.4). The Rao detector obtained was found to be based on third-order moments.

Another useful family of nonGaussian PDFs is the mixture family. An example of a Gaussian mixture PDF has already been explored in [Kay-I 1993, pp. 150, 289–291] and further illustrated in Problem 10.7. Other families of PDFs are described in [Johnson and Kotz 1994] and [Ord 1972].

10.4 Known Deterministic Signals

We now consider the detection of a known signal in IID nonGaussian noise with known PDF. This discussion extends that of Chapter 4 for which the noise was WGN (IID with first-order Gaussian PDF). Our usual example of a DC level is examined first.

Example 10.1 - DC Level in IID NonGaussian Noise

The detection problem is the following

$$\begin{aligned}\mathcal{H}_0 &: x[n] = w[n] \qquad n = 0,1,\ldots,N-1\\ \mathcal{H}_1 &: x[n] = A + w[n] \quad n = 0,1,\ldots,N-1\end{aligned}$$

where A is known with $A > 0$ and the $w[n]$ are IID noise samples with known PDF $p(w[n])$. The NP detector decides $\mathcal{H}_1$ if the likelihood ratio exceeds a threshold or if

$$L(\mathbf{x}) = \frac{p(\mathbf{x};\mathcal{H}_1)}{p(\mathbf{x};\mathcal{H}_0)} > \gamma.$$

Due to the IID assumption we have

$$L(\mathbf{x}) = \frac{\prod_{n=0}^{N-1} p(x[n];\mathcal{H}_1)}{\prod_{n=0}^{N-1} p(x[n];\mathcal{H}_0)}$$

$$= \frac{\prod_{n=0}^{N-1} p(x[n]-A)}{\prod_{n=0}^{N-1} p(x[n])}$$

and we decide $\mathcal{H}_1$ if

$$\ln L(\mathbf{x}) = \sum_{n=0}^{N-1} \ln \frac{p(x[n]-A)}{p(x[n])} > \ln \gamma = \gamma'.$$

If we let $g(x) = \ln[p(x-A)/p(x)]$, we decide $\mathcal{H}_1$ if

$$\sum_{n=0}^{N-1} g(x[n]) > \gamma'. \tag{10.5}$$

For the WGN problem, where $p(w[n])$ is given by (10.2),

$$\begin{aligned} g(x) &= \ln \left(\frac{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-A)^2\right]}{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}x^2\right]} \right) \\ &= \frac{Ax}{\sigma^2} - \frac{A^2}{2\sigma^2}. \end{aligned} \tag{10.6}$$

It is seen that $g(x)$ is linear with x, leading to the sample mean statistic $\bar{x}$ (substitute (10.6) in (10.5)). Otherwise, $g(x)$ will be a nonlinear function. As an example, for the Laplacian PDF we have from (10.1)

$$\begin{aligned} g(x) &= \ln \left(\frac{\frac{1}{\sqrt{2\sigma^2}} \exp\left(-\sqrt{\frac{2}{\sigma^2}}|x-A|\right)}{\frac{1}{\sqrt{2\sigma^2}} \exp\left(-\sqrt{\frac{2}{\sigma^2}}|x|\right)} \right) \\ &= \sqrt{\frac{2}{\sigma^2}}(|x| - |x-A|). \end{aligned}$$

This is shown in Figure 10.3a. A somewhat more intuitive form can be obtained by letting $y[n] = x[n] - A/2$ so that under $\mathcal{H}_0$, $E(y[n]) = -A/2$ and under $\mathcal{H}_1$, $E(y[n]) = A/2$, effectively symmetrizing the data. Then from (10.5) we decide $\mathcal{H}_1$ if

$$\sum_{n=0}^{N-1} g(y[n] + A/2) > \gamma'$$

or

$$\sum_{n=0}^{N-1} h(y[n]) > \gamma'$$

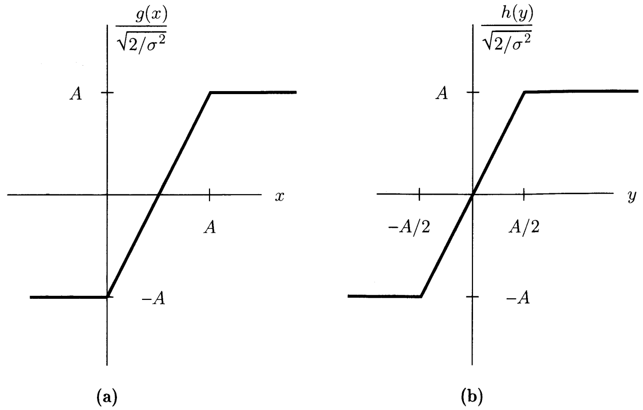


Figure 10.3. Laplacian PDF nonlinearity (a) Version applied to original data (b) Symmetric version applied to $y[n] = x[n] - A/2$.

where

$$\begin{aligned} h(y) &= g(y + A/2) \\ &= \ln \frac{p(y - \frac{A}{2})}{p(y + \frac{A}{2})}. \end{aligned}$$

Then the nonlinearity appears as shown in Figure 10.3b for the Laplacian PDF. The overall detector is shown in Figure 10.4. The nonlinearity $h(y)$ acts to limit the samples that are large in magnitude. This is an attempt by the NP detector to

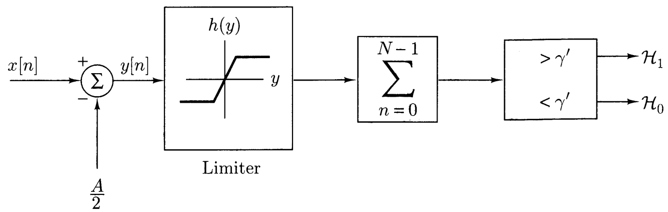


Figure 10.4. NP Detector for DC level in IID Laplacian noise.

reduce the effect of the noise outliers. Without the limiter the P_D could be reduced substantially.

$\diamond$

More generally, for the detection of a known deterministic signal $s[n]$ in IID nonGaussian noise with PDF $p(w[n])$, we decide $\mathcal{H}_1$ if

$$\sum_{n=0}^{N-1} g_n(x[n]) > \gamma' \tag{10.7}$$

where

$$g_n(x) = \ln \frac{p(x - As[n])}{p(x)}.$$

The nonlinearity depends on the sample to which it is applied. By defining $y[n] = x[n] - As[n]/2$, we decide $\mathcal{H}_1$ if

$$\sum_{n=0}^{N-1} h_n(y[n]) > \gamma' \tag{10.8}$$

where

$$h_n(y) = \ln \frac{p\left(y - \frac{As[n]}{2}\right)}{p\left(y + \frac{As[n]}{2}\right)} \tag{10.9}$$

is a symmetric limiter. This is shown in Figure 10.5. Note that if $p(w)$ is an even function (symmetric about $w = 0$), then $h_n(y)$ will be an odd function (antisymmetric about $y = 0$). This follows from (10.9) by showing that $h_n(-y) = -h_n(y)$.

The determination of P_{FA} and P_D for the detector of (10.8) is difficult due to the nonlinearity. Hence, we resort to an asymptotic analysis. In the process we also obtain an equivalent asymptotic detector. We consider the known signal

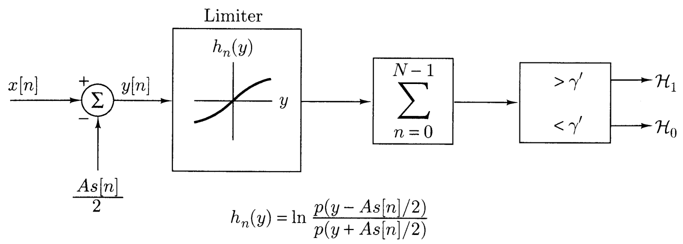


Figure 10.5. NP Detector for known deterministic signal in IID nonGaussian noise.

$As[n]$, where $A > 0$, and examine the NP detector as $A \to 0$ or the signal is weak. Using (10.7) we view $g_n(x)$ as a function of A and expand it in a first-order Taylor expansion about $A = 0$. Doing so produces

$$\begin{aligned} g_n(x) &= \ln \frac{p(x - As[n])}{p(x)} \\ &\approx 0 + \left.\frac{\frac{dp(w)}{dw}}{p(w)}\right|_{w=x-As[n],A=0} (-s[n])A \\ &= -\frac{\frac{dp(x)}{dx}}{p(x)} s[n]A. \end{aligned}$$

Now from (10.7) we decide $\mathcal{H}_1$ if

$$\sum_{n=0}^{N-1} g_n(x[n]) \approx \sum_{n=0}^{N-1} -\frac{\frac{dp(x[n])}{dx[n]}}{p(x[n])} s[n]A > \gamma'$$

or

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} -\frac{\frac{dp(x[n])}{dx[n]}}{p(x[n])} s[n] > \gamma'' \tag{10.10}$$

since $A > 0$. This is also the LMP detector for A unknown and $A > 0$ (to within a scale factor) since $A > 0$ is a one-sided hypothesis test (see Problem 10.5). As a result, it can be said to be asymptotically optimum [Kassam 1988] and is therefore referred to as the *locally optimum* (LO) detector. The weak signal NP detector or LO detector as shown in Figure 10.6 consists of the nonlinearity

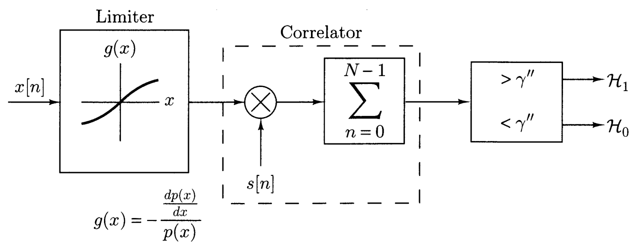


Figure 10.6. NP Detector for known weak deterministic signal and $A > 0$.

$g(x) = -(dp(x)/dx)/p(x)$ followed by a replica-correlator. Assuming $p(x)$ is even, $g(x)$ will be odd (see Problem 10.6). Note that for Gaussian noise (actually WGN since the samples are IID)

$$\begin{aligned} g(x) &= -\frac{\frac{dp(x)}{dx}}{p(x)} \\ &= -\frac{d\ln p(x)}{dx} \\ &= -\frac{d}{dx}\ln\left[\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}x^2\right)\right] = \frac{1}{\sigma^2}x \end{aligned}$$

so that

$$T(\mathbf{x}) = \frac{1}{\sigma^2}\sum_{n=0}^{N-1} x[n]s[n]$$

is a linear function of the data and is a replica-correlator, as expected. The asymptotic PDF of $T(\mathbf{x})$ is shown in Appendix 10A to be

$$T(\mathbf{x}) \overset{a}{\sim} \begin{cases} \mathcal{N}\left(0, i(A)\sum_{n=0}^{N-1} s^2[n]\right) & \text{under } \mathcal{H}_0 \\ \mathcal{N}\left(Ai(A)\sum_{n=0}^{N-1} s^2[n], i(A)\sum_{n=0}^{N-1} s^2[n]\right) & \text{under } \mathcal{H}_1 \end{cases} \tag{10.11}$$

where

$$i(A) = \int_{-\infty}^{\infty} \frac{\left(\frac{dp(w)}{dw}\right)^2}{p(w)}dw \tag{10.12}$$

is the Fisher information for the estimator of a DC level A in nonGaussian noise with PDF $p(w)$ based on a *single sample.* To verify this let $x[0] = A + w[0]$, where $w[0]$ has the PDF $p(w[0])$. Then, the PDF of $x[0]$ is $p(x[0] - A; A)$. The Fisher information for A is

$$\begin{aligned} i(A) &= E\left[\left(\frac{\partial \ln p(x[0]-A;A)}{\partial A}\right)^2\right] \\ &= E\left[\left(\left.\frac{\frac{dp(w)}{dw}}{p(w)}\right|_{w=x[0]-A}(-1)\right)^2\right] \\ &= \int_{-\infty}^{\infty} \left.\frac{\left(\frac{dp(w)}{dw}\right)^2}{p^2(w)}\right|_{w=x[0]-A} p(x[0]-A;A)dx[0] \end{aligned}$$

$$\begin{aligned} &= \int_{-\infty}^{\infty} \frac{\left(\frac{dp(w)}{dw}\right)^2}{p^2(w)}p(w)dw \\ &= \int_{-\infty}^{\infty} \frac{\left(\frac{dp(w)}{dw}\right)^2}{p(w)}dw. \end{aligned}$$

The quantity $i(A)$ is also referred to as the *intrinsic accuracy* of a PDF. From (10.11) we see that the test statistic asymptotically has the performance of the NP detector for the mean-shifted Gauss-Gauss problem (see Section 3.3). Consequently, from (3.10) we have that

$$P_D = Q\left(Q^{-1}(P_{FA}) - \sqrt{d^2}\right) \tag{10.13}$$

where d^2 is the deflection coefficient or from (10.11)

$$\begin{aligned} d^2 &= \frac{(E(T;\mathcal{H}_1) - E(T;\mathcal{H}_0))^2}{\mathrm{var}(T;\mathcal{H}_0)} \\ &= A^2 i(A)\sum_{n=0}^{N-1} s^2[n]. \end{aligned} \tag{10.14}$$

The deflection coefficient for this detector is also related to the *efficacy* of the test statistic as defined in Problem 10.10. Also, in comparing two detectors for large data records we can use the *ratio* of the deflection coefficients. This measure is called the *asymptotic relative efficiency* and is defined in Problem 10.11. It is seen that the effect of the noise PDF on the asymptotic detection performance is only via $i(A)$. It is interesting to note that the PDF that yields the *smallest* $i(A)$ and hence the poorest detection performance is the *Gaussian* PDF (see Problem 6.20). An example follows.

Example 10.2 - Weak Signal Detection in Laplacian Noise

We now assume that the noise is IID and Laplacian as in Example 10.1. The signal is a known DC level with $A > 0$. The weak signal NP detector from (10.10) with $s[n] = 1$ decides $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} -\frac{\frac{dp(x[n])}{dx[n]}}{p(x[n])} > \gamma''.$$

But from (10.1)

$$g(x) = -\frac{\frac{dp(x)}{dx}}{p(x)} = -\frac{d\ln p(x)}{dx} = \sqrt{\frac{2}{\sigma^2}}\frac{d|x|}{dx}.$$

Since $d|x|/dx = \text{sgn}(x)$, we have

$$T(\mathbf{x}) = \sqrt{\frac{2}{\sigma^2}} \sum_{n=0}^{N-1} \text{sgn}(x[n]).$$

The weak signal detector simply adds the signs of the data samples together. The nonlinearity is now an infinite clipper since

$$\frac{g(x)}{\sqrt{\frac{2}{\sigma^2}}} = \begin{cases} 1 & x > 0 \\ -1 & x < 0. \end{cases}$$

The asymptotic detection performance is obtained from (10.13) and (10.14). It can be shown that $i(A) = 2/\sigma^2$ [Kay-I 1993, pg. 63] so that the deflection coefficient is

$$d^2 = \frac{2NA^2}{\sigma^2}.$$

If we compare the performance to the Gaussian case, which was derived in Example 3.2 (there it is the exact finite data record performance), we see that the deflection coefficient in the Laplacian case is doubled. Hence, the asymptotic performance in Laplacian noise is better than in Gaussian noise. As asserted earlier, the poorest performance occurs for Gaussian noise.

◇

10.5 Deterministic Signals with Unknown Parameters

We confine our discussion to the detection of deterministic signals with unknown signal parameters. For random signals the approaches are the same except that the implementation becomes more difficult. When noise parameters are also unknown, the implementation difficulties increase as well [Kassam 1988]. We refer the interested reader to [Kay and Sengupta 1991] for the problem of detection of deterministic signals in *correlated* nonGaussian noise with *unknown noise parameters.*

The problem that we will address is the detection of a deterministic signal known except for amplitude in IID nonGaussian noise or

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = As[n] + w[n] & n = 0, 1, \ldots, N-1 \end{array}$$

where A is unknown, $s[n]$ is known, and $w[n]$ is IID nonGaussian noise with known PDF $p(w[n])$. Different cases arise depending upon any partial knowledge of A. If it is known that $A > 0$, then the test is a one-sided hypothesis test for which as $A \to 0$ the optimal NP detector of (10.10) can be implemented. Equivalently, we can apply the LMP test of (6.36), which can be said to be asymptotically optimal (for large data records and hence, weak signals). If, on the other hand, A can take on any value, then the GLRT or the asymptotically equivalent Rao test can be used. The Rao test has the advantage of not requiring the MLE of A in its implementation. This is of considerable importance in the nonGaussian problem for which the MLE can be difficult to obtain. We now illustrate the GLRT and Rao tests when A is unknown and can take on any value $-\infty < A < \infty$. The reader may also wish to refer to Example 6.9 for the specific case when $s[n] = 1$ (signal is a DC level) and $p(w[n])$ is a member of the generalized Gaussian family. The GLRT decides $\mathcal{H}_1$ if

$$L_G(\mathbf{x}) = \frac{p(\mathbf{x}; \hat{A}, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > \gamma$$

where $\hat{A}$ is the MLE of A under $\mathcal{H}_1$. But

$$\begin{aligned} p(\mathbf{x}; A, \mathcal{H}_1) &= \prod_{n=0}^{N-1} p(x[n] - As[n]) & (10.15) \\ p(\mathbf{x}; \mathcal{H}_0) &= \prod_{n=0}^{N-1} p(x[n]) \end{aligned}$$

and thus

$$\begin{aligned} 2\ln L_G(\mathbf{x}) &= 2\sum_{n=0}^{N-1} \ln \frac{p(x[n] - \hat{A}s[n])}{p(x[n])} \\ &= 2\max_A \sum_{n=0}^{N-1} \ln \frac{p(x[n] - As[n])}{p(x[n])} & (10.16) \end{aligned}$$

since the PDF under $\mathcal{H}_0$ does not depend on A. This is as far as we can carry the derivation. The asymptotic detection peformance will be given by (6.23) and (6.27) which for $\boldsymbol{\theta} = A$, $r = 1$ is

$$2\ln L_G(\mathbf{x}) \overset{a}{\sim} \begin{cases} \chi_1^2 & \text{under } \mathcal{H}_0 \\ \chi_1'^2(\lambda) & \text{under } \mathcal{H}_1. \end{cases} \qquad (10.17)$$

where $\lambda = A^2 I(A = 0)$ and $I(A)$ is the Fisher information. The Rao test (for no nuisance parameters) can be more easily implemented since we do not require $\hat{A}$. As per (6.31) we decide $\mathcal{H}_1$ if

$$T_R(\mathbf{x}) = \frac{\left(\left. \frac{\partial \ln p(\mathbf{x}; A, \mathcal{H}_1)}{\partial A} \right|_{A=0} \right)^2}{I(A = 0)}.$$

But from (10.15)

$$\begin{aligned}
\frac{\partial \ln p(\mathbf{x}; A, \mathcal{H}_1)}{\partial A} &= \frac{\partial}{\partial A} \sum_{n=0}^{N-1} \ln p(x[n] - As[n]) \\
&= \sum_{n=0}^{N-1} \left. \frac{\frac{dp(w)}{dw}}{p(w)} \right|_{w=x[n]-As[n]} (-s[n]) \qquad (10.18)
\end{aligned}$$

which, when evaluated, at $A = 0$ yields the Rao test statistic

$$T_R(\mathbf{x}) = \frac{\left(\sum_{n=0}^{N-1} -\frac{\frac{dp(x[n])}{dx[n]}}{p(x[n])} s[n] \right)^2}{I(A=0)}. \qquad (10.19)$$

The Rao test statistic is similar to the LO detector of (10.10) except for the squaring operation (since A can be positive or negative) and the denominator normalization (required to ensure a CFAR detector). The normalization factor $I(A = 0)$ is found as follows. Using (10.18) we have

$$\begin{aligned}
I(A) &= -E\left[\frac{\partial^2 \ln p(\mathbf{x}; A, \mathcal{H}_1)}{\partial A^2}\right] \\
&= -E\left[\frac{\partial}{\partial A} \sum_{n=0}^{N-1} \left. \frac{\frac{dp(w)}{dw}}{p(w)} \right|_{w=x[n]-As[n]} (-s[n])\right] \\
&= \sum_{n=0}^{N-1} E\left[\frac{\partial}{\partial A} \left. \frac{\frac{dp(w)}{dw}}{p(w)} \right|_{w=x[n]-As[n]}\right] s[n]. \qquad (10.20)
\end{aligned}$$

But

$$\frac{\partial}{\partial A} \left. \frac{\frac{dp(w)}{dw}}{p(w)} \right|_{w=x[n]-As[n]} = \left. \frac{\frac{d^2p(w)}{dw^2}(-s[n])}{p(w)} - \frac{\left(\frac{dp(w)}{dw}\right)^2 (-s[n])}{p^2(w)} \right|_{w=x[n]-As[n]}$$

The first term has zero expectation since

$$\begin{aligned}
E\left[\left. \frac{\frac{d^2p(w)}{dw^2}}{p(w)} \right|_{w=x[n]-As[n]} \right] &= \int_{-\infty}^{\infty} \left. \frac{\frac{d^2p(w)}{dw^2}}{p(w)} \right|_{w=x[n]-As[n]} p(x[n] - As[n]) dx[n] \\
&= \int_{-\infty}^{\infty} \frac{\frac{d^2p(w)}{dw^2}}{p(w)} p(w) dw \\
&= \int_{-\infty}^{\infty} \frac{d^2p(w)}{dw^2} dw \\
&= \left. \frac{dp(w)}{dw} \right|_{-\infty}^{\infty}
\end{aligned}$$

which we assume to be zero. Typical PDFs will satisfy this property. Thus,

$$\begin{aligned}
E\left[\frac{\partial}{\partial A} \left. \frac{\frac{dp(w)}{dw}}{p(w)} \right|_{w=x[n]-As[n]} \right] & \\
&= s[n] \int_{-\infty}^{\infty} \left. \frac{\left(\frac{dp(w)}{dw}\right)^2}{p^2(w)} \right|_{w=x[n]-As[n]} p(x[n] - As[n]) dx[n] \\
&= s[n] \int_{-\infty}^{\infty} \frac{\left(\frac{dp(w)}{dw}\right)^2}{p(w)} dw \\
&= s[n] i(A).
\end{aligned}$$

Finally, we have from (10.20)

$$I(A = 0) = I(A) = i(A) \sum_{n=0}^{N-1} s^2[n]. \qquad (10.21)$$

Therefore, we decide $\mathcal{H}_1$ if

$$T_R(\mathbf{x}) = \frac{\left(\displaystyle\sum_{n=0}^{N-1} -\frac{\frac{dp(x[n])}{dx[n]}}{p(x[n])}s[n]\right)^2}{i(A)\displaystyle\sum_{n=0}^{N-1} s^2[n]} > \gamma'. \qquad (10.22)$$

The asymptotic detection performance, which is the same as that of the GLRT, is given by (10.17). The noncentrality parameter is from (10.21) $\lambda = A^2 i(A) \sum_{n=0}^{N-1} s^2[n]$. Some examples follow.

Example 10.3 - GLRT for DC Level in IID Laplacian Noise - A Unknown

We assume that $s[n] = 1$ (the signal is a DC level) and that the N IID noise samples have a first-order PDF given by (10.1). Then, to implement the GLRT we require the MLE of A or we need to maximize over A

$$p(\mathbf{x}; A, \mathcal{H}_1) = \left(\frac{1}{2\sigma^2}\right)^{\frac{N}{2}} \exp\left(-\sqrt{\frac{2}{\sigma^2}} \sum_{n=0}^{N-1} |x[n] - A|\right)$$

or we must minimize

$$J(A) = \sum_{n=0}^{N-1} |x[n] - A|.$$

To do so we assume N is even to simplify the derivation. We first observe that $J(A)$ is differentiable except at the points $\{x[0], x[1], \ldots, x[N-1]\}$. Excluding these points for the present we have

$$\frac{dJ(A)}{dA} = -\sum_{n=0}^{N-1} \operatorname{sgn}(x[n] - A).$$

Recalling that $\operatorname{sgn}(x[n]-A) = 1$ if $A < x[n]$ and $\operatorname{sgn}(x[n]-A) = -1$ if $A > x[n]$, we can achieve $dJ/dA = 0$ if A is chosen as the *median* of the data samples. We denote the median by x_{med}. Then, half of the samples will yield $\operatorname{sgn}(x[n] - A) = -1$ and the other half will yield $\operatorname{sgn}(x[n] - A) = 1$. Note that the median is not unique. For example, if $x[n] = 1, 2, 4, 10$, then the median is any value between 2 and 4. The function to be minimized $J(A)$ is shown in Figure 10.7. The reader should verify that if $2 < A < 4$, then $J(A)$ is minimized and the minimum value is $J(A) = 11$. The exclusion of the data points from the minimization can be justified due to the

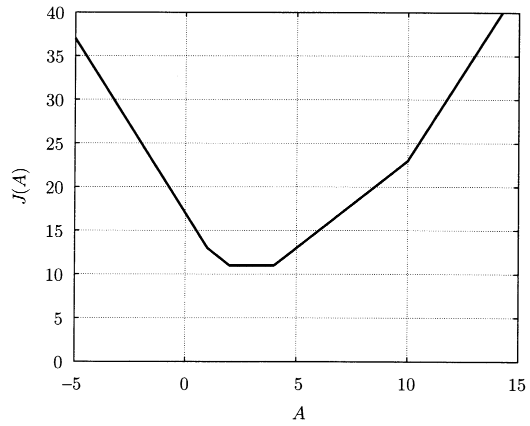


Figure 10.7. Example of function to be minimized for MLE of A in Laplacian noise.

convexity of $J(A)$. As a result, the possibly local minimum found is actually the global minimum. The MLE is therefore $\hat{A} = x_{\text{med}}$ and the GLRT decides $\mathcal{H}_1$ if

$$\begin{aligned} 2\ln L_G(\mathbf{x}) &= 2\ln \frac{p(\mathbf{x}; x_{\text{med}}, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} \\ &= -2\sqrt{\frac{2}{\sigma^2}} \sum_{n=0}^{N-1} (|x[n] - x_{\text{med}}| - |x[n]|) > 2\ln\gamma. \end{aligned}$$

We can simplify the test statistic somewhat if we "rank order" the data samples as $\{x_0, x_1, \ldots, x_{N-1}\}$, where x_0 is the smallest value of $x[n]$ and x_{N-1} is the largest value of $x[n]$. Then, for N even we choose the median as the midpoint between the $N/2 - 1$ and $N/2$ samples or

$$x_{\text{med}} = \frac{x_{\frac{N}{2}-1} + x_{\frac{N}{2}}}{2}.$$

Noting that the sum of a set of numbers does not depend on the ordering, the GLRT statistic then reduces to

$$2\ln L_G(\mathbf{x}) = \sqrt{\frac{8}{\sigma^2}} \sum_{n=0}^{N-1} \left[|x_n| - |x_n - \tfrac{1}{2}(x_{\frac{N}{2}-1} + x_{\frac{N}{2}})|\right]$$

$$
\begin{aligned}
&= \sqrt{\frac{8}{\sigma^2}} \sum_{n=0}^{\frac{N}{2}-1} [|x_n| + (x_n - \tfrac{1}{2}(x_{\frac{N}{2}-1} + x_{\frac{N}{2}}))] \\
&\quad + \sqrt{\frac{8}{\sigma^2}} \sum_{n=\frac{N}{2}}^{N-1} [|x_n| - (x_n - \tfrac{1}{2}(x_{\frac{N}{2}-1} + x_{\frac{N}{2}}))] \\
&= \sqrt{\frac{8}{\sigma^2}} \sum_{n=0}^{\frac{N}{2}-1} (|x_n| + x_n) + \sqrt{\frac{8}{\sigma^2}} \sum_{n=\frac{N}{2}}^{N-1} (|x_n| - x_n).
\end{aligned}
$$

If $x_{\text{med}} > 0$, the samples $\{x_{N/2}, x_{N/2+1}, \ldots, x_{N-1}\}$ are all positive, and thus, the second sum is zero. Only the positive samples from the first sum contribute. Similarly, for $x_{\text{med}} < 0$ only the negative samples from the second sum contribute. Finally, we have

$$
2\ln L_G(\mathbf{x}) = \begin{cases} \sqrt{\frac{32}{\sigma^2}} \displaystyle\sum_{\{n:0<x_n<x_{\text{med}}\}} x_n & \text{if } x_{\text{med}} > 0 \\ -\sqrt{\frac{32}{\sigma^2}} \displaystyle\sum_{\{n:x_{\text{med}}<x_n<0\}} x_n & \text{if } x_{\text{med}} < 0. \end{cases}
$$

The detector (to within a scale factor) sums the magnitudes of the samples that lie in the interval between zero and the median. Hence, if the median is positive, we sum the magnitudes from 0 to x_{med} and discard the samples exceeding the median. Any outliers will then not be important. In fact, any sample with magnitude greater than the median could be arbitrarily increased in value and still not affect the sum. For Gaussian noise, on the other hand, we would sum *all* the samples.

The asymptotic detection performance is given by (10.17) as

$$
2\ln L_G(\mathbf{x}) \overset{a}{\sim} \begin{cases} \chi_1^2 & \text{under } \mathcal{H}_0 \\ \chi_1'^2(\lambda) & \text{under } \mathcal{H}_1 \end{cases} \tag{10.23}
$$

where by using (10.21) the noncentrality parameter is $\lambda = A^2 I(A=0) = A^2 i(A) \sum_{n=0}^{N-1} s^2[n] = 2NA^2/\sigma^2$ since $i(A) = 2/\sigma^2$.

◇

Example 10.4 - Rao Test for DC Level in IID Laplacian Noise - *A* Unknown

Continuing the problem of the previous example we now determine the Rao test. From (10.22) we decide $\mathcal{H}_1$ if

$$
T_R(\mathbf{x}) = \frac{\left(\displaystyle\sum_{n=0}^{N-1} - \frac{\frac{dp(x[n])}{dx[n]}}{p(x[n])} \right)^2}{2N/\sigma^2} > \gamma'
$$

since $s[n] = 1$ and $i(A) = 2/\sigma^2$. But from (10.1) we have

$$
\begin{aligned}
-\frac{\frac{dp(x[n])}{dx[n]}}{p(x[n])} &= -\frac{d\ln p(x[n])}{dx[n]} \\
&= \sqrt{\frac{2}{\sigma^2}} \frac{d|x[n]|}{dx[n]} \\
&= \sqrt{\frac{2}{\sigma^2}} \operatorname{sgn}(x[n])
\end{aligned}
$$

except for $x[n] = 0$ for which the derivative is not defined. Since the probability of actually obtaining $x[n] = 0$ is zero, we can ignore this possibility. We have

$$
\begin{aligned}
T_R(\mathbf{x}) &= \frac{\left(\displaystyle\sum_{n=0}^{N-1} \sqrt{\frac{2}{\sigma^2}} \operatorname{sgn}(x[n]) \right)^2}{\frac{2N}{\sigma^2}} \\
&= N \left(\frac{1}{N} \sum_{n=0}^{N-1} \operatorname{sgn}(x[n]) \right)^2.
\end{aligned} \tag{10.24}
$$

To within a scale factor the Rao test averages the *signs of the samples* and squares the result (since A can be positive or negative). The asymptotic detection performance is the same as the GLRT (see (10.23)). In contrast to the GLRT, however, the Rao detector does not sum the *samples* but only their signs. This approach also avoids the large contributions due to noise outliers.

◇

Before concluding this section, we state a general theorem for the detection of a signal of the linear model form. This will be useful for a wide variety of practical problems. The reader should note that this theorem is a generalization of Theorem 7.1 for WGN.

Theorem 10.1 Rao Test for Linear Model Signal in IID NonGaussian Noise

Assume the data have the form $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$, *where* $\mathbf{H}$ *is a known* $N \times p$ $(N > p)$ *observation matrix of rank* p, $\boldsymbol{\theta}$ *is a* $p \times 1$ *vector of parameters, and* $\mathbf{w}$ *is an* $N \times 1$ *noise vector whose elements are IID random variables with known PDF* $p(w[n])$. *The Rao test for the hypothesis testing problem*

$$\begin{aligned}\mathcal{H}_0 &: \boldsymbol{\theta} = \mathbf{0}\\ \mathcal{H}_1 &: \boldsymbol{\theta} \neq \mathbf{0}\end{aligned}$$

is to decide $\mathcal{H}_1$ *if*

$$T_R(\mathbf{x}) = \frac{\mathbf{y}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{y}}{i(A)} > \gamma' \tag{10.25}$$

where $\mathbf{y} = [y[0]\, y[1] \ldots y[N-1]]^T$ *with* $y[n] = g(x[n])$ *and*

$$g(w) = -\frac{\dfrac{dp(w)}{dw}}{p(w)} \tag{10.26}$$

and

$$i(A) = \int_{-\infty}^{\infty} \frac{\left(\dfrac{dp(w)}{dw}\right)^2}{p(w)} dw.$$

The asymptotic (as $N \to \infty$*) detection performance is given by*

$$\begin{aligned}P_{FA} &= Q_{\chi^2_p}(\gamma')\\ P_D &= Q_{\chi'^2_p(\lambda)}(\gamma')\end{aligned} \tag{10.27}$$

where

$$\lambda = i(A)\boldsymbol{\theta}_1^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}_1$$

for $\boldsymbol{\theta}_1$ *the true value of* $\boldsymbol{\theta}$ *under* $\mathcal{H}_1$.

The proof is given in Appendix 10B. The Rao test for a signal known except for amplitude is a special case of the linear model and has already been given by (10.22). An example of the application of this theorem is given in the next section.

10.6 Signal Processing Example

We now apply Theorem 10.1 to the detection of a sinusoid of unknown amplitude and phase in IID nonGaussian noise. Specifically, the detection problem is

$$\begin{aligned}\mathcal{H}_0 &: x[n] = w[n] & n = 0, 1, \ldots, N-1\\ \mathcal{H}_1 &: x[n] = A\cos(2\pi f_0 n + \phi) + w[n] & n = 0, 1, \ldots, N-1\end{aligned}$$

where A, ϕ are unknown ($A > 0$, $0 \leq \phi \leq 2\pi$ for identifiability as discussed in Section 7.6.2), f_0 is known and $0 < f_0 < 1/2$, and $w[n]$ are IID noise samples with PDF given by the generalized Gaussian PDF of (10.4). Note that for $\beta = 0$ we have WGN with variance σ^2 (the case discussed in Section 7.6.2). We assume that the noise PDF is known. As in Example 7.2, we first rewrite the data in the linear model form as $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$, where

$$\begin{aligned}\mathbf{H} &= \begin{bmatrix} 1 & 0 \\ \cos 2\pi f_0 & \sin 2\pi f_0 \\ \vdots & \vdots \\ \cos[2\pi f_0(N-1)] & \sin[2\pi f_0(N-1)] \end{bmatrix}\\ \boldsymbol{\theta} &= \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}\end{aligned}$$

for $\alpha_1 = A\cos\phi$ and $\alpha_2 = -A\sin\phi$. Since $0 < f_0 < 1/2$, we have $\mathbf{H}^T\mathbf{H} \approx (N/2)\mathbf{I}$ (see [Kay-I 1993, pg. 194]) so that from (10.25)

$$\begin{aligned}T_R(\mathbf{x}) &= \frac{\frac{2}{N}\mathbf{y}^T\mathbf{H}\mathbf{H}^T\mathbf{y}}{i(A)}\\ &= \frac{2}{Ni(A)}\left[\left(\sum_{n=0}^{N-1} y[n]\cos 2\pi f_0 n\right)^2 + \left(\sum_{n=0}^{N-1} y[n]\sin 2\pi f_0 n\right)^2\right]\\ &= \frac{2}{Ni(A)}\left|\sum_{n=0}^{N-1} y[n]\exp(-j2\pi f_0 n)\right|^2\\ &= \frac{2}{i(A)} I_y(f_0)\end{aligned} \tag{10.28}$$

where $I_y(f_0)$ is the periodogram of $y[n]$ evaluated at $f = f_0$. To find $y[n]$ we use (10.26) and (10.4)

$$\begin{aligned}g(w) &= -\frac{\dfrac{dp(w)}{dw}}{p(w)} = -\frac{d\ln p(w)}{dw}\\ &= c_2(\beta)\frac{d}{dw}\left|\frac{w}{\sqrt{\sigma^2}}\right|^{\frac{2}{1+\beta}}.\end{aligned}$$

But

$$\frac{d|w|^{\frac{2}{1+\beta}}}{dw} = \begin{cases} \dfrac{d}{dw} w^{\frac{2}{1+\beta}} & w > 0 \\ \dfrac{d}{dw}(-w)^{\frac{2}{1+\beta}} & w < 0 \end{cases}$$

$$
\begin{aligned}
&= \begin{cases} \frac{2}{1+\beta} w^{\frac{1-\beta}{1+\beta}} & w > 0 \\ -\frac{2}{1+\beta}(-w)^{\frac{1-\beta}{1+\beta}} & w < 0 \end{cases} \\
&= \frac{2}{1+\beta} |w|^{\frac{1-\beta}{1+\beta}} \operatorname{sgn}(w)
\end{aligned}
$$

so that

$$g(w) = \frac{2c_2(\beta)}{(1+\beta)(\sigma^2)^{\frac{1}{1+\beta}}} |w|^{\frac{1-\beta}{1+\beta}} \operatorname{sgn}(w)$$

and finally

$$y[n] = \frac{2\left[\frac{\Gamma(\frac{3}{2}(1+\beta))}{\sigma^2\Gamma(\frac{1}{2}(1+\beta))}\right]^{\frac{1}{1+\beta}}}{1+\beta} |x[n]|^{\frac{1-\beta}{1+\beta}} \operatorname{sgn}(x[n]). \tag{10.29}$$

The normalized nonlinearity

$$h(x) = |x|^{\frac{1-\beta}{1+\beta}} \operatorname{sgn}(x)$$

is plotted in Figure 10.8 for various values of β. Lastly, we require $i(A)$. After some manipulation (see Problem 10.14) this can be shown to be

$$i(A) = \frac{4/\sigma^2}{(1+\beta)^2} \frac{\Gamma(\frac{3}{2}(1+\beta))\Gamma(\frac{3}{2} - \frac{1}{2}\beta)}{\Gamma^2(\frac{1}{2}(1+\beta))}. \tag{10.30}$$

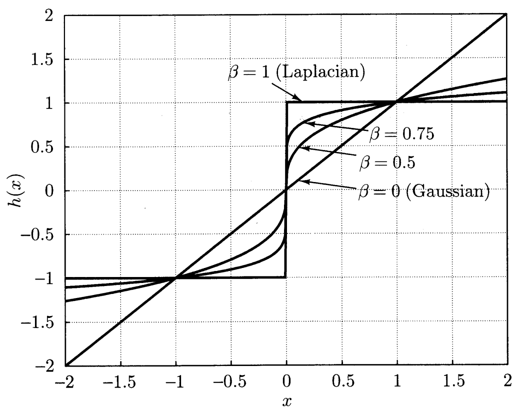


Figure 10.8. Limiter for generalized Gaussian noise.

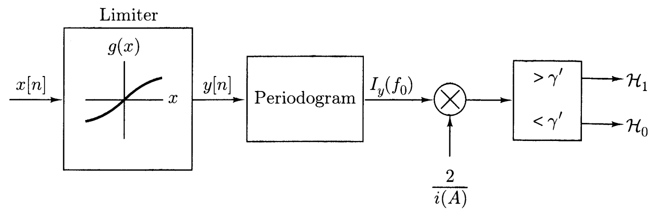


Figure 10.9. Rao detector for sinusoid (unknown A, ϕ) in IID generalized Gaussian noise.

The detector is shown in Figure 10.9. The performance is given by (see (10.27))

$$
\begin{aligned}
P_{FA} &= Q_{\chi_2^2}(\gamma') = \exp(-\gamma'/2) \\
P_D &= Q_{\chi_2'^2(\lambda)}(\gamma')
\end{aligned} \tag{10.31}
$$

where the noncentrality parameter is

$$
\begin{aligned}
\lambda &= i(A)||\mathbf{H}\boldsymbol{\theta}_1||^2 \\
&\approx i(A)\left(\sum_{n=0}^{N-1} \alpha_1^2 \cos^2 2\pi f_0 n + \sum_{n=0}^{N-1} \alpha_2^2 \sin^2 2\pi f_0 n\right) \\
&\approx i(A)\frac{N}{2}(\alpha_1^2 + \alpha_2^2) \\
&= \frac{NA^2 i(A)}{2}.
\end{aligned} \tag{10.32}
$$

It is of interest to determine the effect of the nonGaussian noise. We see that the detection performance is influenced by the PDF of the noise only via $i(A)$. Also, recall that P_D is monotonically increasing with λ and hence with $i(A)$. For Gaussian noise $\beta = 0$ and from (10.30), $i(A) = 1/\sigma^2$. Thus, for Gaussian noise $\lambda = NA^2/(2\sigma^2)$. For nonGaussian noise it is $\lambda = NA^2 i(A)/2$, where $i(A)$ is given by (10.30). The gain in performance in dB due to the nonGaussian noise is

$$10\log_{10} \sigma^2 i(A) = 10\log_{10}\left[\frac{4}{(1+\beta)^2} \frac{\Gamma(\frac{3}{2}(1+\beta))\Gamma(\frac{3}{2} - \frac{1}{2}\beta)}{\Gamma^2(\frac{1}{2}(1+\beta))}\right] \quad \text{dB}.$$

This is shown in Figure 10.10 for $-1 < \beta < 3$. As expected, the minimum is for $\beta = 0$ or for Gaussian noise. Also, it is observed that the improvement in performance can be substantial, especially for the typical heavy-tailed PDFs encountered in

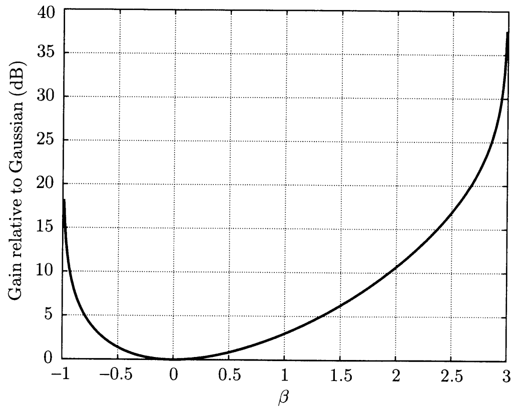


Figure 10.10. Asymptotic gain of nonGaussian PDF designed detector to linear detector.

practice ($\beta > 0$). This result can be interpreted in two ways. First, it is easier to detect a signal in nonGaussian noise than in Gaussian noise *of the same variance.* In essence because of the heavy tails, the nonGaussian noise will have a PDF that is narrower at $w = 0$, as seen in Figure 10.1 for the Laplacian PDF. As such, it is easier to detect a slight shift in the mean due to a signal. But a more important interpretation says that in the nonGaussian noise environment the Rao test will yield a substantial performance improvement over a detector designed for Gaussian noise. The performance improvement, as quantified in Figure 10.10, is the gain of the Rao detector over a incoherent linear detector that decides $\mathcal{H}_1$ if

$$\frac{I_x(f_0)}{\sigma^2/2} > \gamma'.$$

This result is discussed further in Problem 10.15.

Finally, if the frequency is unknown, a common situation in practice, a reasonable approach is to decide $\mathcal{H}_1$ if

$$\max_{f_0} T_R(\mathbf{x}) > \gamma''$$

or

$$\max_{f_0} \frac{2}{i(A)} I_y(f_0) > \gamma''.$$

This is because the GLRT for an unknown frequency decides $\mathcal{H}_1$ if

$$2\ln\frac{p(\mathbf{x};\hat{A},\hat{\phi},\hat{f}_0,\mathcal{H}_1)}{p(\mathbf{x};\mathcal{H}_0)} > 2\ln\gamma$$

or equivalently if

$$2\ln\max_{f_0}\frac{p(\mathbf{x};\hat{A},\hat{\phi},f_0,\mathcal{H}_1)}{p(\mathbf{x};\mathcal{H}_0)} > 2\ln\gamma$$

or since the logarithm is a monotonic function, we decide $\mathcal{H}_1$ if (see Problem 7.21)

$$\max_{f_0}\underbrace{2\ln\frac{p(\mathbf{x};\hat{A},\hat{\phi},f_0,\mathcal{H}_1)}{p(\mathbf{x};\mathcal{H}_0)}}_{T(\mathbf{x})} > 2\ln\gamma.$$

But $T(\mathbf{x})$ is the GLRT statistic for the *known* frequency case, and is therefore, asymptotically equivalent to the Rao statistic. Hence, we decide $\mathcal{H}_1$ if

$$\max_{f_0} T_R(\mathbf{x}) > \gamma''.$$

Note, however, that the asymptotic performance as given by (10.31) *does not apply* in this case. A cautionary note as illustrated by Example 9.4 indicates why the Rao test cannot be directly applied when the frequency is unknown.

References

Box, G.E.P., G.C. Tiao, *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading, MA, 1973.

Huber, P.J., *Robust Statistics*, J. Wiley, New York, 1981.

Johnson, N.L., S. Kotz, N. Balakrishnan, *Continuous Univariate Distributions*, Vols. 1,2,, J. Wiley, New York, 1994.

Kassam, S.A, H.V. Poor, "Robust Techniques for Signal Processing: A Survey," *Proc. IEEE*, Vol. 73, pp. 433-481, March 1985.

Kassam, S.A., *Signal Detection in Non-Gaussian Noise*, Springer-Verlag, New York, 1988.

Kay, S., D. Sengupta, "Recent Advances in Non-Gaussian Autoregressive Processes," in *Advances in Spectrum Analysis and Array Processing*, Vol. I, Prentice-Hall, Englewood Cliffs, N.J., 1991.

Middleton D., "Threshold Detection in non-Gaussian Interference Environments," *IEEE Trans. Electromagnet. Compat.*, Vol. EC-26, pp. 19–28, 1984.

Ord, J.K., *Families of Frequency Distributions*, Griffin's Stat. Monographs and Courses, London, 1972.

Poor, H.V., *An Introduction to Signal Detection and Estimation*, Springer-Verlag, New York, 1988.

Problems

10.1 Determine the probability that the random variable $w[n]$ exceeds $3\sqrt{\sigma^2}$ if the PDF is Gaussian (see (10.2)) and if the PDF is a Laplacian (see (10.1)). How do the PDF "tails" describe the high level events?

10.2 Verify that the fourth moment of a Laplacian PDF is $E(w^4[n]) = 6\sigma^4$. Hint: $\int_0^\infty x^k \exp(-ax)dx = k!/a^{k+1}$ for k an integer and $a > 0$.

10.3 Plot the generalized Gaussian PDF for $\beta = -0.5, -0.75, -0.99$ and $\sigma^2 = 1$. What does the PDF converge to as $\beta \to -1$?

10.4 Consider the detection problem

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = A + w[n] & n = 0, 1, \ldots, N-1 \end{array}$$

where A is known and $w[n]$ are IID samples with Cauchy PDF

$$p(w) = \frac{1}{\pi(1+w^2)} \qquad -\infty < w < \infty.$$

Determine the NP test statistic and plot the nonlinearity $g(x) = \ln(p(x-A)/p(x))$ for $A = 1$.

10.5 Show that for the detection problem

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = As[n] + w[n] & n = 0, 1, \ldots, N-1 \end{array}$$

where A is unknown with $A > 0$, $s[n]$ is known, and $w[n]$ is IID nonGaussian noise with PDF $p(w[n])$, the LMP detector is equivalent to (10.10) (to within a scale factor).

10.6 Show that if the noise PDF $p(w)$ is even or $p(-w) = p(w)$, then $g(w) = -(dp(w)/dw)/p(w)$ is odd or $g(-w) = -g(w)$.

10.7 An example of a Gaussian mixture PDF is

$$p(w) = \frac{1}{2}\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}w^2\right) + \frac{1}{2}\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}w^2\right)$$

for which the random variable is distributed according to a $\mathcal{N}(0,1)$ half the time and according to a $\mathcal{N}(0,\sigma^2)$ the other half. Plot the PDFs and the nonlinearity $g(w) = -(dp(w)/dw)/p(w)$ for $\sigma^2 = 1$ and $\sigma^2 = 100$ and explain your results.

10.8 For the detection problem of Problem 10.4 determine the weak signal NP test statistic assuming that $A > 0$. Describe the operation of the detector.

10.9 Consider the detection problem

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = A + w[n] & n = 0, 1, \ldots, N-1 \end{array}$$

where A is known and $A > 0$, $w[n]$ are IID samples with PDF $p(w[n])$. The mean of $w[n]$ is zero and the variance of $w[n]$ is σ^2. The sample mean or linear detector is proposed or we decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \frac{1}{N}\sum_{n=0}^{N-1} x[n] > \gamma'.$$

Find P_{FA} and P_D for this detector by invoking the central limit theorem. Discuss the *robustness* of this detector or how its performance changes if the form of $p(w[n])$ changes but the mean and variance remain the same. If $w[n]$ is Laplacian with PDF given by (10.1), how much loss does this detector incur over the weak signal NP detector of Example 10.2? A further discussion of the robustness issue can be found in [Huber 1981, Kassam and Poor 1985].

10.10 In this problem we describe the *efficacy* of a test statistic $T(\mathbf{x})$ (i.e., detector). For the one-sided parameter testing problem

$$\begin{array}{l} \mathcal{H}_0 : \theta = 0 \\ \mathcal{H}_1 : \theta > 0 \end{array}$$

we decide $\mathcal{H}_1$ if $T(\mathbf{x}) > \gamma'$. The efficacy of a test statistic T, which measures its "weak signal" discriminating ability for large data records between the hypotheses $\theta = 0$ and $\theta > 0$, is

$$\xi(T) = \lim_{N\to\infty}\lim_{\theta\to 0} \frac{\left(\dfrac{dE(T;\theta)}{d\theta}\right)^2}{N\mathrm{var}(T;\theta)}$$

where $E(T;\theta)$ and $\mathrm{var}(T;\theta)$ are the mean and variance of T when the expectations are with respect to $p(\mathbf{x};\theta)$. For the test statistic given by (10.10), let $\theta = A$ and $s[n] = 1$ to find the efficacy and then by using (10.11), relate it to the deflection coefficient.

10.11 In this problem we discuss the *asymptotic relative efficiency* (ARE), which compares the discriminating ability of two *different* test statistics for the parameter testing problem described in Problem 10.10. It is defined as the ratio of data samples required to attain a given P_{FA} and P_D as the data record length goes to infinity. Specifically, the ARE of the test statistic T_2 with respect to the test statistic T_1 is

$$\text{ARE}_{2,1} = \lim_{N\to\infty} \frac{N_1}{N_2}$$

where $N_1 < N_2$ so that $0 < \text{ARE}_{2,1} < 1$. Consider the detection problem

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = A + w[n] & n = 0, 1, \ldots, N-1 \end{array}$$

where A is known and $A > 0$, and $w[n]$ are IID samples with PDF $p(w[n])$. The mean of $w[n]$ is zero and the variance of $w[n]$ is σ^2. The LO detector was given by (10.10). We let this statistic be $T_1(\mathbf{x})$. Compare its performance as given by (10.11) to that of a sample mean test statistic (i.e., a linear detector) or one that decides $\mathcal{H}_1$ if

$$T_2(\mathbf{x}) = \frac{1}{N}\sum_{n=0}^{N-1} x[n] > \gamma'$$

by computing the ARE. Assume that for large N the sample mean statistic $T_2(\mathbf{x})$ is Gaussian. Also, show that the ARE is related to the efficacies as $\text{ARE}_{2,1} = \xi(T_2)/\xi(T_1)$.

10.12 Verify that the test statistic of (10.24) has the asymptotic PDF of (10.17) under $\mathcal{H}_0$. Use the central limit theorem and evaluate just the mean and variance of the test statistic assuming $w[n]$ is Laplacian noise.

10.13 Evaluate the Rao test statistic given by (10.25) if $w[n]$ is WGN with variance σ^2. You should obtain the same results as in Theorem 10.1 for the GLRT by letting $\mathbf{A} = \mathbf{I}$ and $\mathbf{b} = \mathbf{0}$. Explain why.

10.14 Verify (10.30) for the intrinsic accuracy of a generalized Gaussian PDF. Hint:

$$\int_0^\infty x^\nu \exp(-ax)dx = \frac{\Gamma(\nu+1)}{a^{\nu+1}}$$

for $\nu > 0$ and $a > 0$.

10.15 For the signal processing example of Section 10.6 consider the detector that decides $\mathcal{H}_1$ if

$$\frac{I_x(f_0)}{\sigma^2/2} > \gamma'$$

where

$$I_x(f_0) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x[n]\exp(-j2\pi f_0 n)\right|^2$$

is the periodogram of the data $x[n]$. This detector was found in Section 7.6.2 to be the GLRT for the WGN case. In our problem, however, the noise samples are IID with nonGaussian PDF $p(w)$. The noise mean is zero and its variance is σ^2. Argue that as $N \to \infty$

$$\frac{I_x(f_0)}{\sigma^2/2} \overset{a}{\sim} \begin{cases} \chi_2^2 & \text{under } \mathcal{H}_0 \\ \chi_2'^2(\lambda) & \text{under } \mathcal{H}_1 \end{cases}$$

where $\lambda = NA^2/2\sigma^2$, by using the central limit theorem. Then, show that for large data records the loss in performance relative to the detector of (10.28) is $10\log_{10}\sigma^2 i(A)$ dB. Hint: See Section 7.6.2.

Appendix 10A

Asymptotic Performance of NP Detector for Weak Signals

From (10.10) we have

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} -\frac{d\ln p(x[n])}{dx[n]} s[n].$$

Since the $x[n]$'s are IID, so are the random variables $-d\ln p(x[n])/dx[n]$ for $n = 0, 1, \ldots, N-1$. Applying the central limit theorem, we have that $T(\mathbf{x})$ is an asymptotically Gaussian random variable. We need only find its mean and variance. Under $\mathcal{H}_0$

$$\begin{aligned} E\left(-\frac{d\ln p(x[n])}{dx[n]}\right) &= -\int_{-\infty}^{\infty} \frac{d\ln p(x)}{dx} p(x)dx \\ &= -\int_{-\infty}^{\infty} \frac{dp(x)}{dx} dx = -\, p(x)|_{-\infty}^{\infty} = 0 \end{aligned}$$

since the PDF $p(x)$ must approach zero for large x. Hence, $E(T(\mathbf{x}); \mathcal{H}_0) = 0$. The variance under $\mathcal{H}_0$ is

$$\operatorname{var}(T(\mathbf{x}); \mathcal{H}_0) = \sum_{n=0}^{N-1} s^2[n] \operatorname{var}\left(\frac{d\ln p(x[n])}{dx[n]}\right).$$

Since the mean is zero, we have

$$\operatorname{var}\left(\frac{d\ln p(x[n])}{dx[n]}\right) = \int_{-\infty}^{\infty} \left(\frac{d\ln p(x)}{dx}\right)^2 p(x)dx$$

$$\begin{aligned} &= \int_{-\infty}^{\infty} \frac{\left(\frac{dp(x)}{dx}\right)^2}{p(x)} dx \\ &= i(A) \end{aligned}$$

and therefore

$$\operatorname{var}(T(\mathbf{x}); \mathcal{H}_0) = i(A) \sum_{n=0}^{N-1} s^2[n].$$

Under $\mathcal{H}_1$

$$E\left(-\frac{d\ln p(x[n])}{dx[n]}\right) = -\int_{-\infty}^{\infty} \frac{d\ln p(x)}{dx} p(x - As[n])dx.$$

For A small we have upon using a first-order Taylor expansion about $A = 0$

$$p(x - As[n]) \approx p(x) - \frac{dp(x)}{dx} s[n]A.$$

Using this we have

$$\begin{aligned} E\left(-\frac{d\ln p(x[n])}{dx[n]}\right) &= -\int_{-\infty}^{\infty} \frac{d\ln p(x)}{dx} p(x)dx + \int_{-\infty}^{\infty} \frac{d\ln p(x)}{dx} \frac{dp(x)}{dx} s[n]Adx \\ &= \int_{-\infty}^{\infty} \frac{dp(x)}{dx} dx + As[n] \int_{-\infty}^{\infty} \frac{\left(\frac{dp(x)}{dx}\right)^2}{p(x)} dx \\ &= As[n]i(A) \end{aligned}$$

so that

$$E(T(\mathbf{x}); \mathcal{H}_1) = Ai(A) \sum_{n=0}^{N-1} s^2[n].$$

The variance under $\mathcal{H}_1$ is found as follows. Noting the independence of the $d\ln p(x[n])/dx[n]$ random variables

$$\operatorname{var}(T(\mathbf{x}); \mathcal{H}_1) = \sum_{n=0}^{N-1} s^2[n] \operatorname{var}\left(\frac{d\ln p(x)}{dx}\right).$$

Now

$$\begin{aligned} E\left[\left(\frac{d\ln p(x)}{dx}\right)^2\right] &= \int_{-\infty}^{\infty} \left(\frac{d\ln p(x)}{dx}\right)^2 p(x - As[n])dx \\ &\approx \int_{-\infty}^{\infty} \left(\frac{d\ln p(x)}{dx}\right)^2 \left(p(x) - \frac{dp(x)}{dx} s[n]A\right) dx \\ &\approx \int_{-\infty}^{\infty} \left(\frac{d\ln p(x)}{dx}\right)^2 p(x)dx = i(A) \end{aligned}$$

for A small. Using this result we have

$$\begin{aligned}\operatorname{var}\left[\left(\frac{d\ln p(x[n])}{dx[n]}\right)^2\right] &= i(A) - A^2 s^2[n] i^2(A) \\ &\approx i(A)\end{aligned}$$

for A small. Thus,

$$\operatorname{var}(T(\mathbf{x});\mathcal{H}_1) \approx i(A)\sum_{n=0}^{N-1} s^2[n] = \operatorname{var}(T(\mathbf{x});\mathcal{H}_0)$$

and

$$T(\mathbf{x}) \stackrel{a}{\sim} \begin{cases} \mathcal{N}\left(0, i(A)\sum_{n=0}^{N-1} s^2[n]\right) & \text{under } \mathcal{H}_0 \\ \mathcal{N}\left(Ai(A)\sum_{n=0}^{N-1} s^2[n], i(A)\sum_{n=0}^{N-1} s^2[n]\right) & \text{under } \mathcal{H}_1. \end{cases}$$

Appendix 10B

Rao Test for Linear Model Signal with IID NonGaussian Noise

The results of Theorem 10.1 are derived in this appendix. The Rao test is from (6.31) (no nuisance parameters) with $\boldsymbol{\theta}_0 = \mathbf{0}$

$$T_R(\mathbf{x}) = \left.\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\mathbf{0}}^T \mathbf{I}^{-1}(\boldsymbol{\theta}=\mathbf{0}) \left.\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\mathbf{0}}.$$

The PDF is

$$p(\mathbf{x};\boldsymbol{\theta}) = \prod_{n=0}^{N-1} p(x[n]-s[n])$$

where $s[n] = [\mathbf{s}]_n$ and noting that the $w[n]$'s are IID with PDF $p(w[n])$. But for the linear model $\mathbf{s} = \mathbf{H}\boldsymbol{\theta}$ and thus, $s[n] = \sum_{i=1}^{p} h_{ni}\theta_i$, where $h_{ni} = [\mathbf{H}]_{ni}$. Now

$$\begin{aligned}\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_k} &= \frac{\partial}{\partial \theta_k}\sum_{n=0}^{N-1} \ln p(x[n]-s[n]) \\ &= \sum_{n=0}^{N-1} \left.\frac{\frac{dp(w)}{dw}}{p(w)}\right|_{w=x[n]-s[n]} (-h_{nk}). \end{aligned} \tag{10B.1}$$

Evaluating this at $\boldsymbol{\theta} = \mathbf{0}$ and hence $s[n] = 0$ produces

$$\left.\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_k}\right|_{\boldsymbol{\theta}=\mathbf{0}} = \sum_{n=0}^{N-1} g(x[n])h_{nk}$$

where

$$g(w) = -\frac{\frac{dp(w)}{dw}}{p(w)}.$$

Let $\mathbf{y} = [g(x[0])\, g(x[1]) \ldots g(x[N-1])]^T$. Then

$$\left.\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\mathbf{0}} = \begin{bmatrix} \mathbf{h}_1^T\mathbf{y} \\ \mathbf{h}_2^T\mathbf{y} \\ \vdots \\ \mathbf{h}_p^T\mathbf{y} \end{bmatrix}$$

where $\mathbf{h}_i$ is the ith column of $\mathbf{H}$, and thus

$$\left.\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\mathbf{0}} = \mathbf{H}^T\mathbf{y}. \tag{10B.2}$$

Now from (10B.1)

$$\begin{aligned}
[\mathbf{I}(\boldsymbol{\theta})]_{kl} &= E\left[\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_k}\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_l}\right] \\
&= E\left[\sum_{m=0}^{N-1} g(x[m]-s[m])h_{mk} \sum_{n=0}^{N-1} g(x[n]-s[n])h_{nl}\right] \\
&= \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} E\left[g(x[m]-s[m])g(x[n]-s[n])\right] h_{mk}h_{nl}.
\end{aligned}$$

But the $x[n]$'s are independent, so that for $m \neq n$

$$E\left[g(x[m]-s[m])g(x[n]-s[n])\right] = E\left[g(x[m]-s[m])\right] E\left[g(x[n]-s[n])\right]$$

and

$$\begin{aligned}
E\left[g(x[n]-s[n])\right] &= -\int_{-\infty}^{\infty} \left.\frac{\frac{dp(w)}{dw}}{p(w)}\right|_{w=x[n]-s[n]} p(x[n]-s[n])dx[n] \\
&= \int_{-\infty}^{\infty} \frac{dp(x[n]-s[n])}{d(x[n]-s[n])}dx[n] \\
&= \int_{-\infty}^{\infty} \frac{dp(w)}{dw}dw \\
&= p(w)|_{-\infty}^{\infty} = 0
\end{aligned}$$

since the PDF $p(w)$ must approach zero for large w. The only contribution to the sum is for $m = n$ so that

$$[\mathbf{I}(\boldsymbol{\theta})]_{kl} = \sum_{n=0}^{N-1} h_{nk}h_{nl}E\left[(g(x[n]-s[n]))^2\right].$$

But

$$\begin{aligned}
E\left[(g(x[n]-s[n]))^2\right] &= \int_{-\infty}^{\infty} \left.\left(\frac{\frac{dp(w)}{dw}}{p(w)}\right)^2\right|_{w=x[n]-s[n]} p(x[n]-s[n])dx[n] \\
&= \int_{-\infty}^{\infty} \left.\frac{\left(\frac{dp(w)}{dw}\right)^2}{p(w)}\right|_{w=x[n]-s[n]} dx[n] \\
&= \int_{-\infty}^{\infty} \frac{\left(\frac{dp(w)}{dw}\right)^2}{p(w)}dw \\
&= i(A)
\end{aligned}$$

so that $[\mathbf{I}(\boldsymbol{\theta})]_{kl} = i(A)\sum_{n=0}^{N-1} h_{nk}h_{nl}$ or

$$\mathbf{I}(\boldsymbol{\theta}=\mathbf{0}) = \mathbf{I}(\boldsymbol{\theta}) = i(A)\mathbf{H}^T\mathbf{H}.$$

Finally, using (10B.2) we have

$$\begin{aligned}
T_R(\mathbf{x}) &= (\mathbf{H}^T\mathbf{y})^T(i(A)\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{y} \\
&= \frac{\mathbf{y}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{y}}{i(A)}.
\end{aligned}$$

The asymptotic detection performance is the same as for the GLRT for no nuisance parameters, and hence, follows from the more general results of Section 6.5. Specifically,

$$T_R(\mathbf{x}) \overset{a}{\sim} \begin{cases} \chi_p^2 & \text{under } \mathcal{H}_0 \\ \chi_p'^2(\lambda) & \text{under } \mathcal{H}_1. \end{cases}$$

From (6.27) the noncentrality parameter is given as

$$\lambda = \boldsymbol{\theta}_1^T\mathbf{I}(\boldsymbol{\theta}=\mathbf{0})\boldsymbol{\theta}_1 = i(A)\boldsymbol{\theta}_1^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}_1.$$

Chapter 11

Summary of Detectors

11.1 Introduction

The choice of a detector that will perform well for a particular application depends upon many considerations. Of primary concern is the selection of an optimality criterion and a good model for describing the data statistically. The optimality criterion is usually dictated by the problem goals but may be modified by practical considerations. Similarly, the data model should be complex enough to describe the principal features of the data, but at the same time simple enough to allow the use of a detector that is optimal and easily implemented. For a particular problem we are neither assured of finding an optimal detector, or if we are fortunate enough to do so, of being able to implement it. Therefore, it becomes critical to have at one's disposal a knowledge of the various detection approaches and the conditions under which their use may be justified. To this end we now summarize the approaches, assumptions, and for the linear data model, the explicit detectors obtained. Then, we will illustrate the decision-making process that one must go through in order to choose a good detector. Finally, some other detection approaches that are not discussed in this text are briefly described.

11.2 Detection Approaches

Simple Binary Hypothesis Test
(No unknown parameters)

We wish to decide between the hypotheses $\mathcal{H}_0$ or $\mathcal{H}_1$ based on the observations $\mathbf{x} = [x[0]\, x[1] \ldots x[N-1]]^T$.

1. Neyman-Pearson (NP)

a. *Data Model/Assumptions*
PDFs $p(\mathbf{x};\mathcal{H}_0)$, $p(\mathbf{x};\mathcal{H}_1)$ are assumed known.

b. *Detector*
Decide $\mathcal{H}_1$ if

$$L(\mathbf{x}) = \frac{p(\mathbf{x};\mathcal{H}_1)}{p(\mathbf{x};\mathcal{H}_0)} > \gamma$$

where the threshold γ is found from the constraint that the probability of false alarm P_{FA} satisfies

$$P_{FA} = \Pr\{L(\mathbf{x}) > \gamma; \mathcal{H}_0\} = \alpha.$$

c. *Optimality Criterion*
Maximizes probability of detection $P_D = \Pr\{L(\mathbf{x}) > \gamma; \mathcal{H}_1\}$ for a given $P_{FA} = \alpha$.

d. *Performance*
No general results.

e. *Comments*
The test statistic $L(\mathbf{x})$ is known as the likelihood ratio and the detector is termed the likelihood ratio test (LRT).

f. *Reference*
Chapter 3

2. Minimum Probability of Error

a. *Data Model/Assumptions*
Hypotheses are modeled as random events with known prior probabilities $P(\mathcal{H}_0)$, $P(\mathcal{H}_1)$. Also, the conditional PDFs $p(\mathbf{x}|\mathcal{H}_0)$, $p(\mathbf{x}|\mathcal{H}_1)$ are assumed known.

b. *Detector*
Decide $\mathcal{H}_1$ if

$$L(\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{H}_1)}{p(\mathbf{x}|\mathcal{H}_0)} > \frac{P(\mathcal{H}_0)}{P(\mathcal{H}_1)} = \gamma$$

or equivalently if

$$P(\mathcal{H}_1|\mathbf{x}) > P(\mathcal{H}_0|\mathbf{x}). \tag{11.1}$$

c. *Optimality Criterion*
Minimizes the probability of error or

$$P_e = \Pr\{L(\mathbf{x}) > \gamma|\mathcal{H}_0\}P(\mathcal{H}_0) + \Pr\{L(\mathbf{x}) < \gamma|\mathcal{H}_1\}P(\mathcal{H}_1).$$

d. *Performance*
No general results.

e. *Comments*
The decision rule of (11.1) is termed the maximum a posteriori (MAP) rule. If $P(\mathcal{H}_0) = P(\mathcal{H}_1) = 1/2$, then it reduces to deciding $\mathcal{H}_1$ if $p(\mathbf{x}|\mathcal{H}_1) > p(\mathbf{x}|\mathcal{H}_0)$ and is called the conditional maximum likelihood (ML) rule.

f. *Reference*
Chapter 3

3. Bayes Risk

a. *Data Model/Assumptions*
Hypotheses are modeled as random events with known prior probabilities $P(\mathcal{H}_0)$, $P(\mathcal{H}_1)$. The conditional PDFs $p(\mathbf{x}|\mathcal{H}_0)$, $p(\mathbf{x}|\mathcal{H}_1)$ are assumed known. Finally, costs are assigned to various errors, where C_{ij} is the cost of deciding $\mathcal{H}_i$ when $\mathcal{H}_j$ is true.

b. *Detector*
Decide $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x}|\mathcal{H}_1)}{p(\mathbf{x}|\mathcal{H}_0)} > \frac{(C_{10}-C_{00})}{(C_{01}-C_{11})}\frac{P(\mathcal{H}_0)}{P(\mathcal{H}_1)} = \gamma$$

where $C_{10} > C_{00}, C_{01} > C_{11}$.

c. *Optimality Criterion*
Minimizes the Bayes risk or expected cost

$$\mathcal{R} = E(C) = \sum_{i=0}^{1}\sum_{j=0}^{1} C_{ij}P(\mathcal{H}_i|\mathcal{H}_j)P(\mathcal{H}_j)$$

where $P(\mathcal{H}_i|\mathcal{H}_j)$ is the probability of deciding $\mathcal{H}_i$ when $\mathcal{H}_j$ is true.

d. *Performance*
No general results.

e. *Comments*
If $C_{00} = C_{11} = 0$ and $C_{10} = C_{01} = 1$, then $\mathcal{R} = P_e$, and decision rule reduces to the previous MAP rule of (11.1).

f. *Reference*
Chapter 3

Simple Multiple Hypothesis Test
(No unknown parameters)

We wish to decide among the hypotheses $\mathcal{H}_0, \mathcal{H}_1, \ldots, \mathcal{H}_{M-1}$.

4. Minimum Probability of Error

a. *Data Model/Assumptions*
Hypotheses are modeled as random events with known prior probabilities $P(\mathcal{H}_0), P(\mathcal{H}_1), \ldots, P(\mathcal{H}_{M-1})$. Also, the conditional PDFs $p(\mathbf{x}|\mathcal{H}_0)$, $p(\mathbf{x}|\mathcal{H}_1), \ldots, p(\mathbf{x}|\mathcal{H}_{M-1})$ are assumed known.

b. *Detector*
Decide the hypothesis that maximizes $P(\mathcal{H}_i|\mathbf{x})$ or decide $\mathcal{H}_k$ if

$$P(\mathcal{H}_k|\mathbf{x}) > P(\mathcal{H}_i|\mathbf{x}) \qquad i \neq k \tag{11.2}$$

or equivalently if

$$\ln p(\mathbf{x}|\mathcal{H}_k) + \ln P(\mathcal{H}_k)$$

is maximum.

c. *Optimality Criterion*
Minimizes the probability of error or

$$P_e = \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} P(\mathcal{H}_i|\mathcal{H}_j)P(\mathcal{H}_j).$$

d. *Performance*
No general results.

e. *Comments*
The decision rule of (11.2) is termed the MAP rule. If the prior probabilities are equal or $P(\mathcal{H}_i) = 1/M$, then the rule reduces to deciding $\mathcal{H}_k$ if

$$p(\mathbf{x}|\mathcal{H}_k) > p(\mathbf{x}|\mathcal{H}_i) \qquad i \neq k \tag{11.3}$$

and is termed the conditional ML decision rule.

f. *Reference*
Chapter 3

5. Bayes Risk

a. *Data Model/Assumptions*
Hypotheses are modeled as random events with known prior probabilities $P(\mathcal{H}_0), P(\mathcal{H}_1), \ldots, P(\mathcal{H}_{M-1})$. The conditional PDFs $p(\mathbf{x}|\mathcal{H}_0), p(\mathbf{x}|\mathcal{H}_1), \ldots, p(\mathbf{x}|\mathcal{H}_{M-1})$ are assumed known. Finally, costs are assigned to various errors, where C_{ij} is the cost of deciding $\mathcal{H}_i$ when $\mathcal{H}_j$ is true.

b. *Detector*
Decide $\mathcal{H}_k$ if

$$C_k(\mathbf{x}) > C_i(\mathbf{x}) \qquad i \neq k$$

where

$$C_i(\mathbf{x}) = \sum_{j=0}^{M-1} C_{ij}P(\mathcal{H}_j|\mathbf{x}).$$

c. *Optimality Criterion*
Minimizes the Bayes risk or expected cost

$$\mathcal{R} = E(C) = \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} C_{ij} P(\mathcal{H}_i|\mathcal{H}_j)P(\mathcal{H}_j)$$

d. *Performance*
No general results.

e. *Comments*
If $C_{ii} = 0$ for $i = 0, 1, \ldots, M-1$ and $C_{ij} = 1$ for $i \neq j$, then $\mathcal{R} = P_e$, and the decision rule reduces to the MAP rule of (11.2).

f. *Reference*
Chapter 3

Composite Binary Hypothesis Test

(Unknown parameters present)

6. Generalized Likelihood Ratio Test (GLRT)

a. *Data Model/Assumptions*
PDFs under $\mathcal{H}_0$ and $\mathcal{H}_1$ contain unknown parameters, denoted by $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$, respectively. The PDFs are given by $p(\mathbf{x}; \boldsymbol{\theta}_0, \mathcal{H}_0)$ and $p(\mathbf{x}; \boldsymbol{\theta}_1, \mathcal{H}_1)$ and are assumed known except for $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$.

b. *Detector*
Decide $\mathcal{H}_1$ if

$$L_G(\mathbf{x}) = \frac{p(\mathbf{x}; \hat{\boldsymbol{\theta}}_1, \mathcal{H}_1)}{p(\mathbf{x}; \hat{\boldsymbol{\theta}}_0, \mathcal{H}_0)} > \gamma$$

where $\hat{\boldsymbol{\theta}}_i$ is the maximum likelihood estimator (MLE) or $\hat{\boldsymbol{\theta}}_i$ maximizes $p(\mathbf{x}; \boldsymbol{\theta}_i, \mathcal{H}_i)$.

c. *Optimality Criterion*
None.

d. *Performance*
No general results.

e. *Comments*
An equivalent form is to decide $\mathcal{H}_1$ if

$$L_G(\mathbf{x}) = \frac{\max_{\boldsymbol{\theta}_1} p(\mathbf{x}; \boldsymbol{\theta}_1, \mathcal{H}_1)}{\max_{\boldsymbol{\theta}_0} p(\mathbf{x}; \boldsymbol{\theta}_0, \mathcal{H}_0)} > \gamma.$$

The quantity $L_G(\mathbf{x})$ is termed the generalized likelihood ratio.

f. *Reference*
Chapter 6

7. Bayesian

a. *Data Model/Assumptions*
The unknown parameter vectors $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ are modeled as random vectors with known prior PDFs $p(\boldsymbol{\theta}_0)$ and $p(\boldsymbol{\theta}_1)$. The conditional PDFs $p(\mathbf{x}|\boldsymbol{\theta}_0; \mathcal{H}_0)$ and $p(\mathbf{x}|\boldsymbol{\theta}_1; \mathcal{H}_1)$ are assumed known.

b. *Detector*
Decide $\mathcal{H}_1$ if

$$L(\mathbf{x}) = \frac{p(\mathbf{x}; \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} = \frac{\int p(\mathbf{x}|\boldsymbol{\theta}_1; \mathcal{H}_1)p(\boldsymbol{\theta}_1)d\boldsymbol{\theta}_1}{\int p(\mathbf{x}|\boldsymbol{\theta}_0; \mathcal{H}_1)p(\boldsymbol{\theta}_0)d\boldsymbol{\theta}_0} > \gamma.$$

c. *Optimality Criterion*
Same as Neyman-Pearson (item 1) since the unknown parameters are "integrated out."

d. *Performance*
No general results.

e. *Comments*
See Neyman-Pearson for determination of γ. Integration may be difficult to implement, depending on choice of priors.

f. *Reference*
Chapter 6

Composite Binary Parameter Test

(Unknown parameters present but no nuisance parameters)

The PDFs under $\mathcal{H}_0$ and $\mathcal{H}_1$ are the *same* except that the value of the unknown parameter vector $\boldsymbol{\theta}$ is different. The PDF is denoted by $p(\mathbf{x}; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is $r \times 1$. The hypothesis (or parameter) test is

$$\begin{aligned} \mathcal{H}_0 &: \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ \mathcal{H}_1 &: \boldsymbol{\theta} \neq \boldsymbol{\theta}_0. \end{aligned}$$

8. Generalized Likelihood Ratio Test (GLRT)

a. *Data Model/Assumptions*
PDF $p(\mathbf{x}; \boldsymbol{\theta})$ is assumed known except for $\boldsymbol{\theta}$ under $\mathcal{H}_1$.

b. *Detector*
Decide $\mathcal{H}_1$ if

$$L_G(\mathbf{x}) = \frac{p(\mathbf{x}; \hat{\boldsymbol{\theta}}_1)}{p(\mathbf{x}; \boldsymbol{\theta}_0)} > \gamma$$

where $\hat{\boldsymbol{\theta}}_1$ is the MLE of $\boldsymbol{\theta}$ under $\mathcal{H}_1$ (maximizes $p(\mathbf{x}; \boldsymbol{\theta})$).

c. *Optimality Criterion*
None.

d. *Performance*
Asymptotic statistics (as $N \to \infty$) given by

$$2 \ln L_G(\mathbf{x}) \overset{a}{\sim} \begin{cases} \chi_r^2 & \text{under } \mathcal{H}_0 \\ \chi_r'^2(\lambda) & \text{under } \mathcal{H}_1 \end{cases}$$

where

$$\lambda = (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^T \mathbf{I}(\boldsymbol{\theta}_0)(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0).$$

$\mathbf{I}(\boldsymbol{\theta})$ denotes the $r \times r$ Fisher information matrix and $\boldsymbol{\theta}_1$ is the true value under $\mathcal{H}_1$.

e. *Comments*
Requires MLE under $\mathcal{H}_1$.

f. *Reference*
Chapter 6

9. Wald test

a. *Data Model/Assumptions*
Same as GLRT (see item 8a).

b. *Detector*
Decide $\mathcal{H}_1$ if

$$T_{\mathrm{W}}(\mathbf{x}) = (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0)^T \mathbf{I}(\hat{\boldsymbol{\theta}}_1)(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0) > \gamma.$$

c. *Optimality Criterion*
None.

d. *Performance*
Same as GLRT (see item 8d).

e. *Comments*
Requires MLE under $\mathcal{H}_1$.

f. *Reference*
Chapter 6

10. Rao test

a. *Data Model/Assumptions*
Same as GLRT (see item 8a).

b. *Detector*
Decide $\mathcal{H}_1$ if

$$T_{\mathrm{R}}(\mathbf{x}) = \left.\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}^T \mathbf{I}^{-1}(\boldsymbol{\theta}_0) \left.\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} > \gamma.$$

c. *Optimality Criterion*
None.

d. *Performance*
Same as GLRT (see item 8d).

e. *Comments*
Requires no MLEs.

f. *Reference*
Chapter 6

Composite Binary Parameter Test
(Unknown parameters present and nuisance parameters)

The PDFs under $\mathcal{H}_0$ and $\mathcal{H}_1$ are the same except that the value of the unknown parameter vector $\boldsymbol{\theta}$ is different. The parameter vector is $\boldsymbol{\theta} = [\boldsymbol{\theta}_r^T \; \boldsymbol{\theta}_s^T]^T$, where $\boldsymbol{\theta}_r$ is $r \times 1$ and $\boldsymbol{\theta}_s$ (the nuisance parameter vector) is $s \times 1$. The PDF is denoted by $p(\mathbf{x}; \boldsymbol{\theta}_r, \boldsymbol{\theta}_s)$. The hypothesis test is

$$\begin{aligned} \mathcal{H}_0 &: \boldsymbol{\theta} = \boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s \\ \mathcal{H}_1 &: \boldsymbol{\theta} \neq \boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s . \end{aligned}$$

The nuisance parameters $\boldsymbol{\theta}_s$ are unknown under $\mathcal{H}_0$ and $\mathcal{H}_1$.

11. Generalized likelihood ratio test (GLRT)

a. *Data Model/Assumptions*
The PDF $p(\mathbf{x}; \boldsymbol{\theta}_r, \boldsymbol{\theta}_s)$ is assumed known except for $\boldsymbol{\theta}_s$ under $\mathcal{H}_0$ and $\boldsymbol{\theta}_r, \boldsymbol{\theta}_s$ under $\mathcal{H}_1$.

b. *Detector*
Decide $\mathcal{H}_1$ if

$$L_G(\mathbf{x}) = \frac{p(\mathbf{x}; \hat{\boldsymbol{\theta}}_{r_1}, \hat{\boldsymbol{\theta}}_{s_1})}{p(\mathbf{x}; \boldsymbol{\theta}_{r_0}, \hat{\boldsymbol{\theta}}_{s_0})} > \gamma$$

where $\hat{\boldsymbol{\theta}}_{r_1}, \hat{\boldsymbol{\theta}}_{s_1}$ is the MLE under $\mathcal{H}_1$ (or the unrestricted MLE, which is found by maximizing $p(\mathbf{x}; \boldsymbol{\theta}_r, \boldsymbol{\theta}_s)$ over $\boldsymbol{\theta}_r, \boldsymbol{\theta}_s$) and $\boldsymbol{\theta}_{s_0}$ is the MLE under $\mathcal{H}_0$ (or the restricted MLE, which is found by maximizing $p(\mathbf{x}; \boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s)$ over $\boldsymbol{\theta}_s$).

c. *Optimality Criterion*
None.

d. *Performance*
Asymptotic statistics (as $N \to \infty$) given by

$$2 \ln L_G(\mathbf{x}) \overset{a}{\sim} \begin{cases} \chi_r^2 & \text{under } \mathcal{H}_0 \\ \chi_r'^2(\lambda) & \text{under } \mathcal{H}_1 \end{cases}$$

where

$$\begin{aligned}\lambda &= (\boldsymbol{\theta}_{r_1} - \boldsymbol{\theta}_{r_0})^T \left[\mathbf{I}_{\theta_r\theta_r}(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s)\right. \\ &\quad \left. - \mathbf{I}_{\theta_r\theta_s}(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s)\mathbf{I}^{-1}_{\theta_s\theta_s}(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s)\mathbf{I}_{\theta_s\theta_r}(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s)\right] (\boldsymbol{\theta}_{r_1} - \boldsymbol{\theta}_{r_0}).\end{aligned}$$

$\boldsymbol{\theta}_{r_1}$ is the true value of $\boldsymbol{\theta}_r$ under $\mathcal{H}_1$ and $\boldsymbol{\theta}_s$ is the true value, which is the same under either hypothesis. The Fisher information matrix has been partitioned as

$$\begin{aligned}\mathbf{I}(\boldsymbol{\theta}) = \mathbf{I}(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s) &= \begin{bmatrix} \mathbf{I}_{\theta_r\theta_r}(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s) & \mathbf{I}_{\theta_r\theta_s}(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s) \\ \mathbf{I}_{\theta_s\theta_r}(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s) & \mathbf{I}_{\theta_s\theta_s}(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s) \end{bmatrix} \\ &= \begin{bmatrix} r \times r & r \times s \\ s \times r & s \times s \end{bmatrix}.\end{aligned}$$

e. *Comments*
Requires MLEs under $\mathcal{H}_0$ and $\mathcal{H}_1$.

f. *Reference*
Chapter 6

12. Wald Test

a. *Data Model/Assumptions*
Same as GLRT (see item 11a).

b. *Detector*
Decide $\mathcal{H}_1$ if

$$T_{\mathrm{W}}(\mathbf{x}) = (\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0})^T \left(\left[\mathbf{I}^{-1}(\hat{\boldsymbol{\theta}}_1)\right]_{\theta_r\theta_r}\right)^{-1} (\hat{\boldsymbol{\theta}}_{r_1} - \boldsymbol{\theta}_{r_0}) > \gamma$$

where $\hat{\boldsymbol{\theta}}_1 = [\hat{\boldsymbol{\theta}}_{r_1}^T \, \hat{\boldsymbol{\theta}}_{s_1}^T]^T$ is the MLE under $\mathcal{H}_1$ and

$$\left[\mathbf{I}^{-1}(\boldsymbol{\theta})\right]_{\theta_r\theta_r} = \left(\mathbf{I}_{\theta_r\theta_r}(\boldsymbol{\theta}) - \mathbf{I}_{\theta_r\theta_s}(\boldsymbol{\theta})\mathbf{I}^{-1}_{\theta_s\theta_s}(\boldsymbol{\theta})\mathbf{I}_{\theta_s\theta_r}(\boldsymbol{\theta})\right)^{-1}.$$

c. *Optimality Criterion*
None.

d. *Performance*
Same as GLRT (see item 11d).

e. *Comments*
Requires MLE under $\mathcal{H}_1$.

f. *Reference*
Chapter 6

13. Rao Test

a. *Data Model/Assumptions*
Same as GLRT (see item 11a).

b. *Detector*
Decide $\mathcal{H}_1$ if

$$T_{\mathrm{R}}(\mathbf{x}) = \left.\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r}\right|^T_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} \left[\mathbf{I}^{-1}(\tilde{\boldsymbol{\theta}})\right]_{\theta_r\theta_r} \left.\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} > \gamma$$

where $\tilde{\boldsymbol{\theta}} = [\boldsymbol{\theta}_{r_0}^T \, \hat{\boldsymbol{\theta}}_{s_0}^T]^T$ is the MLE of $\boldsymbol{\theta}$ under $\mathcal{H}_0$ (restricted MLE found by maximizing $p(\mathbf{x}; \boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s)$ over $\boldsymbol{\theta}_s$).

c. *Optimality Criterion*
None.

d. *Performance*
Same as GLRT (see item 11d).

e. *Comments*
Requires MLE under $\mathcal{H}_0$ only.

f. *Reference*
Chapter 6

Composite Binary One-Sided Parameter Test

(Scalar unknown parameter present and no nuisance parameters)

This test is used for a one-sided hypothesis test whose PDF is the same under $\mathcal{H}_0$ and $\mathcal{H}_1$ but differs in the scalar parameter value. The PDF is denoted by $p(\mathbf{x}; \theta)$. The hypothesis test is

$$\begin{aligned}\mathcal{H}_0 &: \theta = \theta_0 \\ \mathcal{H}_1 &: \theta > \theta_0.\end{aligned}$$

14. Locally Most Powerful (LMP) Test

a. *Data Model/Assumptions*
The PDF $p(\mathbf{x}; \theta)$ is assumed known except for θ under $\mathcal{H}_1$.

b. *Detector*
Decide $\mathcal{H}_1$ if

$$T_{\mathrm{LMP}}(\mathbf{x}) = \frac{\left.\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta}\right|_{\theta=\theta_0}}{\sqrt{I(\theta_0)}} > \gamma$$

where $I(\theta)$ is the Fisher information and the threshold γ is given by (as $N \to \infty$) $\gamma = Q^{-1}(P_{FA})$.

c. *Optimality Criterion*
Maximizes P_D for a given P_{FA} if $\theta_1 - \theta_0$ is positive and small, where θ_1 is the true value of θ under $\mathcal{H}_1$.

d. *Performance*
Asymptotic statistics (as $N \to \infty$) given by

$$T_{\mathrm{LMP}}(\mathbf{x}) \overset{a}{\sim} \begin{cases} \mathcal{N}(0,1) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(\sqrt{I(\theta_0)}(\theta_1 - \theta_0), 1) & \text{under } \mathcal{H}_1 \end{cases}$$

where θ_1 is the value under $\mathcal{H}_1$.

e. *Comments*
Can be thought of as the one-sided equivalent of the Rao test (see item 10b) for a scalar $\boldsymbol{\theta}$.

f. *Reference*
Chapter 6

Composite Multiple Hypothesis Test

We wish to decide among the hypotheses $\mathcal{H}_0, \mathcal{H}_1, \ldots, \mathcal{H}_{M-1}$.

15. Generalized Maximum Likelihood Rule

a. *Data Model/Assumptions*
The data PDFs $p(\mathbf{x}; \boldsymbol{\theta}_i | \mathcal{H}_i)$ are known except for the parameter vector $\boldsymbol{\theta}_i$, whose dimension may change with the hypothesis.

b. *Detector*
The generalized ML rule decides $\mathcal{H}_k$ if

$$\xi_i = \ln p(\mathbf{x}; \hat{\boldsymbol{\theta}}_i | \mathcal{H}_i) - \frac{1}{2} \ln \det \left(\mathbf{I}(\hat{\boldsymbol{\theta}}_i) \right)$$

is maximized for $i = k$, where $\hat{\boldsymbol{\theta}}_i$ is the MLE of $\boldsymbol{\theta}$ assuming $\mathcal{H}_i$ is true (maximizes $p(\mathbf{x}; \boldsymbol{\theta} | \mathcal{H}_i)$). Also, $\mathbf{I}(\boldsymbol{\theta})$ is the Fisher information matrix assuming $\mathcal{H}_i$ is true.

c. *Optimality Criterion*
None.

d. *Performance*
No general results.

e. *Comments*
The approach is a hybrid between Bayesian and classical. The first term of ξ_i is an estimated likelihood of the data while the second term is a penalty factor for having to estimate unknown parameters. It extends the ML rule of (11.3).

f. *Reference*
Chapter 6

11.3 Linear Model

When the linear model can be used to describe the data under $\mathcal{H}_1$, explicit detectors and their performances can be obtained. The linear model can be either classical or Bayesian. We describe each one separately. Then, we summarize an extension, termed the nonGaussian linear model.

Classical Linear Model

For the classical linear model the data under $\mathcal{H}_1$ are given by

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

where $\mathbf{x}$ is an $N \times 1$ vector of observations, $\mathbf{H}$ is a known $N \times p$ observation matrix with $N > p$ and of rank p, $\boldsymbol{\theta}$ is a $p \times 1$ vector or parameters (which may be known or not), and $\mathbf{w}$ is an $N \times 1$ noise vector with PDF $\mathcal{N}(\mathbf{0}, \mathbf{C})$. The PDF of $\mathbf{x}$ is

$$p(\mathbf{x}; \boldsymbol{\theta}) = \frac{1}{(2\pi)^{N/2} \det^{\frac{1}{2}}(\mathbf{C})} \exp\left[-\frac{1}{2} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) \right].$$

16. Known Deterministic Signal

a. *Hypothesis Test*

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{x} = \mathbf{w} \\ \mathcal{H}_1 &: \mathbf{x} = \mathbf{H}\boldsymbol{\theta}_1 + \mathbf{w} \end{aligned}$$

or

$$\begin{aligned} \mathcal{H}_0 &: \boldsymbol{\theta} = \mathbf{0} \\ \mathcal{H}_1 &: \boldsymbol{\theta} = \boldsymbol{\theta}_1 \end{aligned}$$

where $\boldsymbol{\theta}_1$ is the known value of $\boldsymbol{\theta}$ under $\mathcal{H}_1$.

b. *Detector*
Decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \mathbf{x}^T \mathbf{C}^{-1} \mathbf{s} > \gamma'$$

where $\mathbf{s} = \mathbf{H}\boldsymbol{\theta}_1$ and $\gamma' = \sqrt{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}}\, Q^{-1}(P_{FA})$.

c. *Optimality Criterion*
Maximizes P_D for a a given P_{FA} (Neyman-Pearson).

d. *Performance*

$$P_D = Q(Q^{-1}(P_{FA}) - \sqrt{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}}).$$

e. *Comments*
This is a special case of the generalized matched filter or generalized replica-correlator. See Example 4.9 for an illustration.

f. *Reference*
Section 4.6

17. Deterministic Signal with Unknown Parameters

a. *Hypothesis Test*

$$\begin{aligned}\mathcal{H}_0 &: \mathbf{x} = \mathbf{w} \\ \mathcal{H}_1 &: \mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}\end{aligned}$$

or

$$\begin{aligned}\mathcal{H}_0 &: \boldsymbol{\theta} = \mathbf{0} \\ \mathcal{H}_1 &: \boldsymbol{\theta} \neq \mathbf{0}\end{aligned}$$

where $\boldsymbol{\theta}$ is unknown under $\mathcal{H}_1$ and $\mathbf{C} = \sigma^2\mathbf{I}$.

b. *Detector*

Decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \frac{\mathbf{x}^T\mathbf{H}\hat{\boldsymbol{\theta}}_1}{\sigma^2} = \frac{\mathbf{x}^T\hat{\mathbf{s}}}{\sigma^2} > \gamma'$$

where $\hat{\boldsymbol{\theta}}_1 = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$ is the MLE of $\boldsymbol{\theta}$ under $\mathcal{H}_1$, $\hat{\mathbf{s}} = \mathbf{H}\hat{\boldsymbol{\theta}}_1$ is the signal MLE, and $\gamma' = Q^{-1}_{\chi^2_p}(P_{FA})$.

c. *Optimality Criterion*

None since this is a GLRT detector.

d. *Performance*

$$\begin{aligned}P_{FA} &= Q_{\chi^2_p}(\gamma') \\ P_D &= Q_{\chi'^2_p(\lambda)}(\gamma')\end{aligned}$$

where

$$\lambda = \frac{\boldsymbol{\theta}_1^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}_1}{\sigma^2} = \frac{\mathbf{s}^T\mathbf{s}}{\sigma^2}$$

and $\boldsymbol{\theta}_1$ is the true value of $\boldsymbol{\theta}$ under $\mathcal{H}_1$.

e. *Comments*

This is a special case of Theorem 7.1 with $\mathbf{A} = \mathbf{I}$, $\mathbf{b} = \mathbf{0}$, and $r = p$. It can be extended to the case of any known $\mathbf{C}$. The test statistic has the form of an estimator-correlator. See Example 7.3 for an illustration.

f. *Reference*

Section 7.7

18. Deterministic Signal with Unknown Parameters and Unknown Noise Variance

a. *Hypothesis Test*

$$\begin{aligned}\mathcal{H}_0 &: \mathbf{x} = \mathbf{w} \\ \mathcal{H}_1 &: \mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}\end{aligned}$$

or

$$\begin{aligned}\mathcal{H}_0 &: \boldsymbol{\theta} = \mathbf{0}, \sigma^2 > 0 \\ \mathcal{H}_1 &: \boldsymbol{\theta} \neq \mathbf{0}, \sigma^2 > 0\end{aligned}$$

where $\boldsymbol{\theta}$ is unknown under $\mathcal{H}_1$ and $\mathbf{C} = \sigma^2\mathbf{I}$ with σ^2 unknown under $\mathcal{H}_0$ and $\mathcal{H}_1$.

b. *Detector*

Decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \frac{1}{p}\frac{\mathbf{x}^T\mathbf{H}\hat{\boldsymbol{\theta}}_1}{\hat{\sigma_1^2}} = \frac{1}{p}\frac{\mathbf{x}^T\hat{\mathbf{s}}}{\hat{\sigma_1^2}} > \gamma'$$

where $\hat{\boldsymbol{\theta}}_1 = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$ is the MLE of $\boldsymbol{\theta}$ under $\mathcal{H}_1$, $\hat{\mathbf{s}} = \mathbf{H}\hat{\boldsymbol{\theta}}_1$ is the signal MLE, and $\hat{\sigma_1^2} = (\mathbf{x}^T\mathbf{x} - \mathbf{x}^T\mathbf{H}\hat{\boldsymbol{\theta}}_1)/(N-p)$ is the minimum variance unbiased (MVU) estimator of σ^2 under $\mathcal{H}_1$. The threshold is given by $\gamma' = Q^{-1}_{F_{p,N-p}}(P_{FA})$.

c. *Optimality Criterion*

None since this is the GLRT.

d. *Performance*

$$\begin{aligned}P_{FA} &= Q_{F_{p,N-p}}(\gamma') \\ P_D &= Q_{F'_{p,N-p}(\lambda)}(\gamma')\end{aligned}$$

where

$$\lambda = \frac{\boldsymbol{\theta}_1^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}_1}{\sigma^2} = \frac{\mathbf{s}^T\mathbf{s}}{\sigma^2}$$

and $\boldsymbol{\theta}_1$ is the true value of $\boldsymbol{\theta}$ under $\mathcal{H}_1$.

e. *Comments*

This is a special case of Theorem 9.1 with $\mathbf{A} = \mathbf{I}$, $\mathbf{b} = \mathbf{0}$, and $r = p$. It can be extended to $\mathbf{C} = \sigma^2\mathbf{V}$, where $\mathbf{V}$ is known and σ^2 is unknown. The test statistic has the form of a normalized estimator-correlator. See Example 9.2 for an illustration.

f. *Reference*

Section 9.4.3

General Linear Model - Unknown Noise Parameters

For the general linear model the data under $\mathcal{H}_1$ are given by

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

where $\mathbf{x}$ is an $N \times 1$ vector of observations, $\mathbf{H}$ is a known $N \times p$ observation matrix with $N > p$ and of rank p, $\boldsymbol{\theta}$ is a $p \times 1$ vector of unknown parameters, and $\mathbf{w}$ is an $N \times 1$ noise vector with PDF $\mathcal{N}(\mathbf{0}, \mathbf{C}(\boldsymbol{\theta}_w))$, for $\boldsymbol{\theta}_w$ an unknown $q \times 1$ noise parameter vector. The PDF of $\mathbf{x}$ is

$$p(\mathbf{x}; \boldsymbol{\theta}, \boldsymbol{\theta}_w) = \frac{1}{(2\pi)^{N/2}\det^{\frac{1}{2}}(\mathbf{C}(\boldsymbol{\theta}_w))}\exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T\mathbf{C}^{-1}(\boldsymbol{\theta}_w)(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})\right].$$

19. Rao Test

a. *Hypothesis Test*

$$\begin{aligned}\mathcal{H}_0 &: \mathbf{x} = \mathbf{w} \\ \mathcal{H}_1 &: \mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}\end{aligned}$$

or

$$\begin{aligned}\mathcal{H}_0 &: \boldsymbol{\theta} = \mathbf{0}, \boldsymbol{\theta}_w \\ \mathcal{H}_1 &: \boldsymbol{\theta} \neq \mathbf{0}, \boldsymbol{\theta}_w.\end{aligned}$$

b. *Detector*
The Rao detector decides $\mathcal{H}_1$ if

$$T_R(\mathbf{x}) = \mathbf{x}^T\mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0})\mathbf{H}\hat{\boldsymbol{\theta}}_1 = \mathbf{x}^T\mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0})\hat{\mathbf{s}} > \gamma'$$

where

$$\hat{\boldsymbol{\theta}}_1 = (\mathbf{H}^T\mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0})\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}(\hat{\boldsymbol{\theta}}_{w_0})\mathbf{x}$$

is an estimator of $\boldsymbol{\theta}$ under $\mathcal{H}_1$, $\hat{\boldsymbol{\theta}}_{w_0}$ is the MLE of $\boldsymbol{\theta}_w$ under $\mathcal{H}_0$, and $\hat{\mathbf{s}} = \mathbf{H}\hat{\boldsymbol{\theta}}_1$ is an estimator of the signal. The threshold is given by (as $N \to \infty$) $\gamma' = Q^{-1}_{\chi^2_p}(P_{FA})$.

c. *Optimality Criterion*
None.

d. *Performance*
The asymptotic performance (as $N \to \infty$) is

$$\begin{aligned}P_{FA} &= Q_{\chi^2_p}(\gamma') \\ P_D &= Q_{\chi'^2_p(\lambda)}(\gamma')\end{aligned}$$

where

$$\lambda = \boldsymbol{\theta}_1^T\mathbf{H}^T\mathbf{C}^{-1}(\boldsymbol{\theta}_w)\mathbf{H}\boldsymbol{\theta}_1$$

and $\boldsymbol{\theta}_1$ is the true value of $\boldsymbol{\theta}$ under $\mathcal{H}_1$, and also $\boldsymbol{\theta}_w$ is the true value, which is assumed to be the same under either hypothesis.

e. *Comments*
The test statistic has the form of a generalized estimator-correlator. See Section 9.6 for an illustration.

f. *Reference*
Section 9.5

Bayesian Linear Model

For the Bayesian linear model the data under $\mathcal{H}_1$ are given by

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

where $\mathbf{x}$ is an $N \times 1$ vector of observations, $\mathbf{H}$ is a known $N \times p$ observation matrix with $N > p$, $\boldsymbol{\theta}$ is a $p \times 1$ random vector with $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_\theta)$, and $\mathbf{w}$ is an $N \times 1$ noise vector with PDF $\mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ and is independent of $\boldsymbol{\theta}$. The PDF of $\mathbf{x}$ is

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{N/2}\det^{\frac{1}{2}}(\mathbf{C}_\theta + \sigma^2\mathbf{I})}\exp\left[-\frac{1}{2}\mathbf{x}^T(\mathbf{C}_\theta + \sigma^2\mathbf{I})^{-1}\mathbf{x}\right].$$

20. Neyman-Pearson (NP)

a. *Hypothesis Test*

$$\begin{aligned}\mathcal{H}_0 &: \mathbf{x} = \mathbf{w} \\ \mathcal{H}_1 &: \mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}.\end{aligned}$$

b. *Detector*
Decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \mathbf{x}^T\hat{\mathbf{s}} > \gamma'$$

where

$$\hat{\mathbf{s}} = \mathbf{H}\mathbf{C}_\theta\mathbf{H}^T(\mathbf{H}\mathbf{C}_\theta\mathbf{H}^T + \sigma^2\mathbf{I})^{-1}\mathbf{x}$$

is the minimum mean square error (MMSE) estimator of $\mathbf{s} = \mathbf{H}\boldsymbol{\theta}$.

c. *Optimality Criterion*
Maximizes P_D for a given P_{FA} (Neyman-Pearson).

d. *Performance*
See Section 5.3.

e. *Comments*
Can be extended to the case of a general known noise covariance matrix $\mathbf{C}$. See Example 5.5 for an illustration.

f. *Reference*
Section 5.4

NonGaussian Linear Model

For the nonGaussian linear model the data under $\mathcal{H}_1$ are given by

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

where $\mathbf{x}$ is an $N \times 1$ vector of observations, $\mathbf{H}$ is a known $N \times p$ observation matrix with $N > p$ and of rank p, $\boldsymbol{\theta}$ is a $p \times 1$ vector of unknown parameters, and $\mathbf{w}$ is an $N \times 1$ noise vector with IID components. The PDF of $w[n]$ is $p(w)$ and is assumed known.

21. Rao Test

a. *Hypothesis Test*

$$\mathcal{H}_0 : \mathbf{x} = \mathbf{w}$$
$$\mathcal{H}_1 : \mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

or

$$\mathcal{H}_0 : \boldsymbol{\theta} = \mathbf{0}$$
$$\mathcal{H}_1 : \boldsymbol{\theta} \neq \mathbf{0}$$

where $\boldsymbol{\theta}$ is unknown under $\mathcal{H}_1$.

b. *Detector*

The Rao test decides $\mathcal{H}_1$ if

$$T_R(\mathbf{x}) = \frac{\mathbf{y}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{y}}{i(A)} > \gamma'$$

where $\mathbf{y} = [y[0]\, y[1] \dots y[N-1]]^T$ with $y[n] = g(x[n])$ and

$$\begin{aligned} g(w) &= -\frac{d\ln p(w)}{dw} \\ i(A) &= \int_{-\infty}^{\infty} \frac{\left(\frac{dp(w)}{dw}\right)^2}{p(w)} dw. \end{aligned}$$

The threshold is determined by (as $N \to \infty$) $\gamma' = Q^{-1}_{\chi^2_p}(P_{FA})$.

c. *Optimality Criterion*

None.

d. *Performance*

The asymptotic performance (as $N \to \infty$) is

$$\begin{aligned} P_{FA} &= Q_{\chi^2_p}(\gamma') \\ P_D &= Q_{\chi'^2_p(\lambda)}(\gamma') \end{aligned}$$

where

$$\lambda = i(A)\boldsymbol{\theta}_1^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}_1$$

for $\boldsymbol{\theta}_1$ the true value of $\boldsymbol{\theta}$ under $\mathcal{H}_1$.

e. *Comments*

See Section 10.6 for an illustration.

f. *Reference*

Section 10.5

11.4 Choosing a Detector

We now discuss some of the considerations involved in choosing a detector. Although not meant to be all inclusive, the general decision making process is illustrated. First consider the problem of binary hypothesis testing or a decision between a signal present ($\mathcal{H}_1$ true) or noise only ($\mathcal{H}_0$ true). A flowchart is shown in Figure 11.1. If prior probabilities for the hypotheses $P(\mathcal{H}_0)$, $P(\mathcal{H}_1)$ are available, then we can pursue a Bayesian approach. Otherwise, a Neyman-Pearson (NP) approach in which we attempt to maximize the probability of detection P_D for a given probability of false alarm P_{FA} is appropriate as shown in Figure 11.2. Next we inquire if the costs of a decision C_{ij} for $i = 0, 1; j = 0, 1$ are known. If they are unknown, then we cannot proceed. If the costs are known (and are arbitrary given values), then the Bayes risk criterion as described in item 3 is implemented. To do so assumes knowledge of the conditional data PDFs $p(\mathbf{x}|\mathcal{H}_0)$ and $p(\mathbf{x}|\mathcal{H}_1)$. If these are not

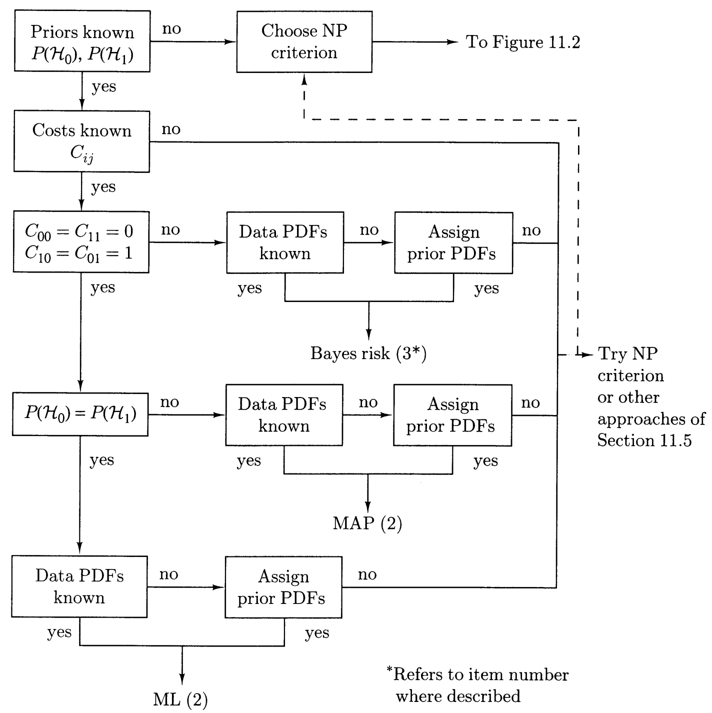


Figure 11.1. Optimal Bayesian approaches for binary hypothesis testing.

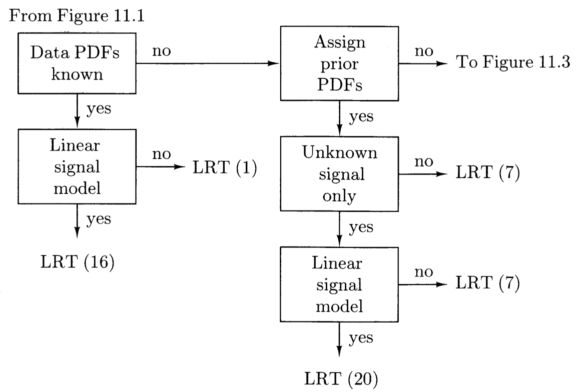


Figure 11.2. Optimal Neyman-Pearson approaches for binary hypothesis testing.

known, they we may still be able to implement the Bayes risk test by assigning prior PDFs to the unknown parameters. Then, the unknown parameters are "integrated out." If not, then the Bayesian approach cannot be implemented. Special cases occur when there is no cost for a correct decision ($C_{00} = C_{11} = 0$) and an equal cost of an incorrect decision ($C_{10} = C_{01} = 1$). Then, the maximum a posteriori (MAP) rule results as described in item 2. If, furthermore, the prior probabilities for the hypotheses are equal, the conditional maximum likelihood (ML) rule results as described in item 2. In all cases, the conditional data PDFs must be known or priors assigned to any unknown parameters. Otherwise, the Bayesian approach cannot be implemented and one must resort to the NP criterion. Note that the Bayes risk test, and MAP/ML rules are optimal in that they minimize the Bayes risk and probability of error, respectively.

Next, we describe the NP approach as shown in Figure 11.2. When it is possible to implement, it maximizes P_D for a given P_{FA}. We first inquire as to whether the data PDFs $p(\mathbf{x}; \mathcal{H}_0)$ and $p(\mathbf{x}; \mathcal{H}_1)$ are known. If they are not, then we can attempt to assign prior PDFs to the unknown parameters. If this is not possible, then we proceed to the suboptimal approaches of Figure 11.3. Otherwise, if the data PDFs are known, and also the signal obeys the linear model, we are led to a likelihood ratio test described in item 16. If, however, the linear signal model is not appropriate, then the more general LRT of item 1 must be implemented. When the data PDFs are known except for signal parameters, and the signal is linear with parameters that can be assigned a prior PDF, we have the Bayesian linear model. Consequently, the LRT of item 20 is appropriate. If there are unknown noise parameters or if the linear signal model does not apply, then the more general LRT of item 7 is used. For all of these cases, the resultant LRT is optimal.

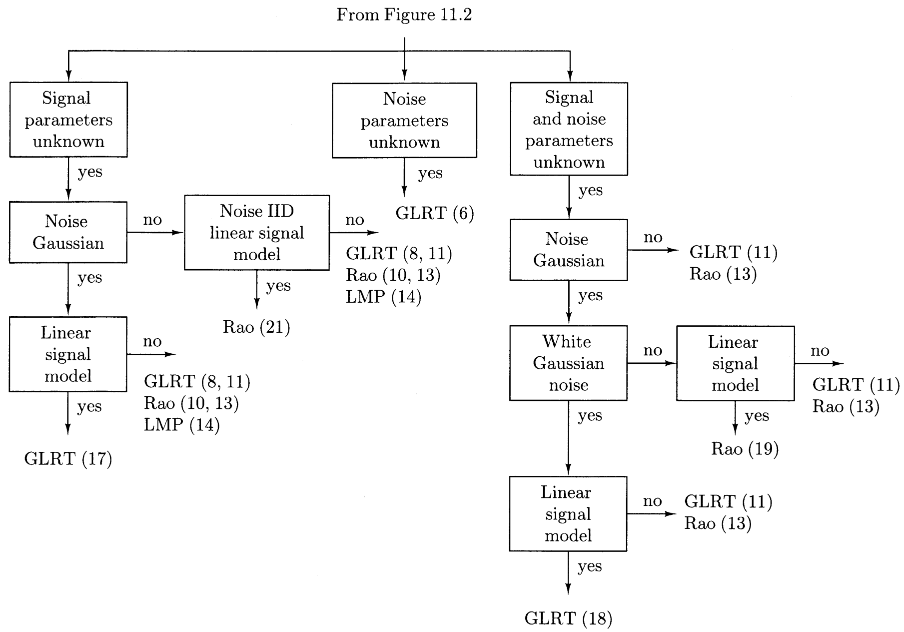


Figure 11.3. Suboptimal approaches for composite binary hypothesis testing.

Now assume an NP criterion but with data PDFs that contain unknown parameters. The decision-making process is shown in Figure 11.3. Prior PDFs cannot be assigned to the unknown parameters. The detectors to be described are suboptimal in that there may exist others that have a higher P_D for a given P_{FA}. We begin by assuming that only signal parameters are unknown. If the noise is Gaussian and the linear signal model applies, then the preferred approach is to use the GLRT as described in item 17. If, however, the noise is Gaussian but the signal parameters do not follow the linear model, then the GLRT (items 8 or 11), the Rao test (items 10 or 13), or the locally most powerful (LMP) test (item 14) should be tried. For nonGaussian noise with IID samples and the linear signal model, the Rao test described in item 21 can be applied. Otherwise, for more general nonGaussian noise and/or more general unknown signal parameters, the GLRT (items 8 or 11), the Rao test (items 10 or 13), or the locally most powerful (LMP) test (item 14) should be tried. When only noise parameters are unknown, the GLRT described in item 6 is appropriate. Finally, for unknown signal and noise parameters we first determine if the noise is Gaussian. If it is, and furthermore it is white Gaussian noise (WGN) with unknown variance σ^2, then for a linear signal model the GLRT of item 18 applies. For an arbitrary signal model the GLRT (item 11) or Rao test (item 13) should be used. If the noise is Gaussian but nonwhite so that there are multiple unknown noise parameters $\boldsymbol{\theta}_w$ and if the signal model is linear, then the Rao test described in item 19 is preferred. Lastly, for either nonGaussian noise or correlated/WGN with an arbitrary signal model, the GLRT (item 11) or Rao test (item 13) are to be tried.

It should be noted that although the Wald test was described in items 9 and 12 for completeness, it is seldom used for the detection problem. This is because it requires the MLE under $\mathcal{H}_1$. If that MLE is attainable, then usually the MLE under $\mathcal{H}_0$ is known as well, being a simpler problem. Hence, the GLRT is then the preferred approach.

Next we refer to Figure 11.4 for the Bayesian approach to multiple hypothesis testing. The flowchart is nearly identical to the binary case of Figure 11.1. The final

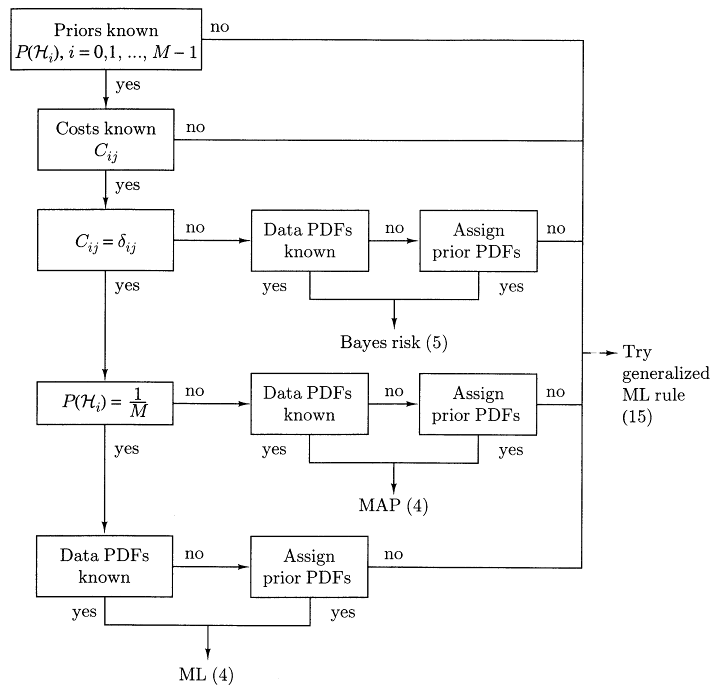


Figure 11.4. Optimal Bayesian approaches for multiple hypothesis testing.

decision rules are simple generalizations of the binary ones. Note that the Bayes risk and MAP/ML rules are optimal according to the Bayes risk and minimum probability of error criteria, respectively. If these approaches cannot be applied (usually due to unknown parameters in the data PDFs), then the generalized ML rule described in item 15 can be tried. There is, however, no optimality associated with it.

11.5 Other Approaches and Other Texts

In addition to the detection methods we have already described there are several other well-known approaches. We have omitted a discussion of these as they are beyond the scope of this text. We now briefly allude to these alternative approaches. In the Neyman-Pearson formulation with unknown parameters it is sometimes possible to restrict the class of detectors to ones that are invariant to the unknown parameters. Such an approach, although at times difficult to implement, leads to a test that is optimal within a restricted class. This test, termed the *uniformly most powerful invariant test*, is described in [Lehmann 1959, Scharf 1991]. A test that can be implemented sequentially, and which sometimes allows better detectability, is the *Wald sequential probability ratio test*. It is described in [Kendall and Stuart 1976–1979, Zacks 1981, Helstrom 1995]. In the Bayesian risk approach to detection, the mandatory choice of the priors for the hypotheses can be alleviated by employing the *minimax* approach. In essence, the worst possible risk is minimized. This technique is described in [Van Trees 1968, McDonough and Whalen 1995]. Finally, a class of detectors that do not rely on explicit knowledge of the PDF of the data are termed *nonparametric detectors*. These detectors will perform much poorer than the optimal Neyman-Pearson and Bayesian approaches since the assumed knowledge is less. However, the performance will be much more robust, in that similar performance will be obtained for a wide class of PDFs. Nonparametric detectors are described in [Cox and Hinkley 1974, Kendall and Stuart 1976–1979, Huber 1981].

For a broad description of detection theory from an engineering viewpoint recommended texts are [Van Trees 1968, Scharf 1991, McDonough and Whalen 1995, Helstrom 1995]. A more theoretical discussion of hypothesis testing from a statistical viewpoint can be found in [Lehmann 1959, Rao 1973, Cox and Hinkley 1974, Graybill 1976, Kendall and Stuart 1976–1979, Zacks 1981].

References

Cox, D.R., D.V. Hinkley, *Theoretical Statistics*, Chapman and Hall, New York, 1974.

Graybill, F.A., *Theory and Application of the Linear Model*, Duxbury Press, North Scituate, Mass., 1976.

Helstrom, C.W., *Elements of Signal Detection and Estimation*, Prentice-Hall, Englewood Cliffs, N.J., 1995.

Huber, P.J., *Robust Statistics*, J. Wiley, New York, 1981.

Kendall, Sir M., A. Stuart, *The Advanced Theory of Statistics*, Vols. 1–3, Macmillan, New York, 1976–1979.

Lehmann, E.L, *Testing Statistical Hypotheses*, J. Wiley, New York, 1959.

McDonough, R.N., A.D. Whalen, *Detection of Signals in Noise*, Academic Press, New York, 1995.

Rao, C.R., *Linear Statistical Inference and its Applications*, J. Wiley, New York, 1973.

Scharf, L.L, *Statistical Signal Processing*, Addison-Wesley, Reading, Mass., 1991.

Van Trees, H.L., *Detection, Estimation, and Modulation Theory*, Vol. I, J. Wiley, New York, 1968.

Zacks, S., *Parametric Statistical Inference*, Pergamon, New York, 1981.

Chapter 12

Model Change Detection

12.1 Introduction

Detection of temporal or spatial changes in the characteristics of physical systems is an important practical problem. In speech the vocal tract changes either smoothly or abruptly in response to the sound being uttered. Machinery that is operating satisfactorily may all of a sudden experience difficulty. This may be due, for example, to a loss of oil pressure, precipitating a machine failure. In both of these cases, system parameters have changed in time. Changes in system parameters can occur spatially as well. Examples are the change in the speed of sound as a sound wave traverses a boundary such as an air-water interface and the variation of temperature with altitude. These changes may either be abrupt (sound speed at an interface) or smooth (temperature with altitude). We will concentrate our discussion on the detection of *abrupt changes*. Our description will of necessity be brief. For additional approaches and applications the reader is referred to [Basseville and Nikiforov 1993].

12.2 Summary

Some basic problems of detecting the change time of a DC level are explored. If the change time is known as well as the DC levels before and after the jump, then the NP statistic is given by (12.5) and its performance by (12.6) and (12.7). For a known change time but unknown levels, the GLRT leads to (12.8) as the test and (12.9), (12.10) as its performance. When additionally the change time is unknown, the GLRT produces the test statistic of (12.11). For unknown multiple change times a dynamic programming algorithm is used to reduce the computation of a GLRT. Specifically, for unknown DC levels and unknown change times, the dynamic programming algorithm is given by (12.14) and (12.13). An example for three change times is given. A more general dynamic programming solution to the change time problem is described in Appendix 12A. Finally, the theory is applied in Section 12.6 to the detection of a maneuvering object as well as the detection of a change in the PSD of a random process.

12.3 Description of Problem

To illustrate the basic problem assume that we wish to detect a parameter jump at a *known time*. If the *parameters before and after the jump are also known*, then the solution is relatively straightforward as is now illustrated.

Example 12.1 - Known DC Level Jump at Known Time

Assume that we observe a known DC level of amplitude $A = A_0$ embedded in WGN of known variance σ^2. At some known time $n = n_0$, the DC level may jump from $A = A_0$ to $A = A_0 + \Delta A$. An example is shown in Figure 12.1 in which $A_0 = 1$, $\Delta A = 3$, $n_0 = 50$, and $\sigma^2 = 1$. The size of the jump ΔA is also assumed known and $\Delta A > 0$. We wish to design a detector so that with high probability we will detect the jump but at the same time control the amount of false alarms. This problem is exactly in the framework of the Neyman-Pearson (NP) approach and is therefore easily solved. Our hypothesis test is

$$\begin{aligned}\mathcal{H}_0 : x[n] &= \quad A_0 + w[n] \qquad\qquad n = 0,1,\ldots,N-1\\ \mathcal{H}_1 : x[n] &= \begin{cases} A_0 + w[n] & n = 0,1,\ldots,n_0-1\\ A_0 + \Delta A + w[n] & n = n_0, n_0+1,\ldots,N-1\end{cases}\end{aligned} \tag{12.1}$$

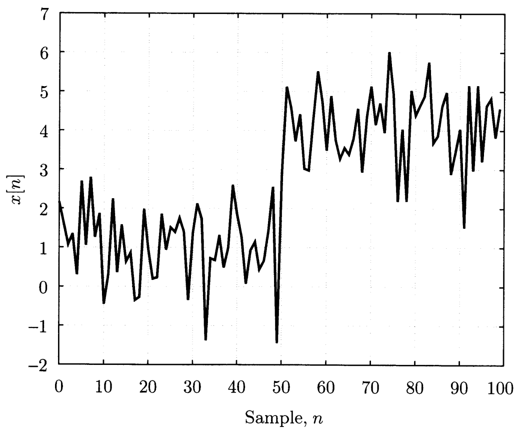


Figure 12.1. DC level in WGN with change in DC level from $A = 1$ to $A = 4$.

where $w[n]$ is WGN with known variance σ^2, A_0 and $\Delta A > 0$ are known, and the jump time n_0 is known. Equivalently, we have the parameter test

$$\begin{aligned}\mathcal{H}_0 : A &= \quad A_0 \qquad\qquad n = 0,1,\ldots,N-1\\ \mathcal{H}_1 : A &= \begin{cases} A_0 & n = 0,1,\ldots,n_0-1\\ A_0 + \Delta A & n = n_0, n_0+1,\ldots,N-1.\end{cases}\end{aligned} \tag{12.2}$$

If we let A_1 be the value of the DC level before the jump and A_2 be its value after the jump, the PDF that describes this data is

$$p(\mathbf{x}; A_1, A_2) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{n=0}^{n_0-1}(x[n]-A_1)^2 + \sum_{n=n_0}^{N-1}(x[n]-A_2)^2\right)\right]. \tag{12.3}$$

The detector that maximizes P_D subject to a constraint on P_{FA} is the NP detector that decides $\mathcal{H}_1$ if

$$L(\mathbf{x}) = \frac{p(\mathbf{x}; \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > \gamma$$

which in the notation of (12.3) becomes

$$L(\mathbf{x}) = \frac{p(\mathbf{x}; A_1 = A_0, A_2 = A_0 + \Delta A)}{p(\mathbf{x}; A_1 = A_0, A_2 = A_0)}.$$

We then have

$$\begin{aligned}\ln L(\mathbf{x}) &= -\frac{1}{2\sigma^2}\sum_{n=n_0}^{N-1}\left[(x[n]-A_0-\Delta A)^2 - (x[n]-A_0)^2\right]\\ &= -\frac{1}{2\sigma^2}\sum_{n=n_0}^{N-1}[-2\Delta A(x[n]-A_0) + \Delta A^2]\\ &= \frac{\Delta A}{\sigma^2}\sum_{n=n_0}^{N-1}(x[n]-A_0) - \frac{(N-n_0)\Delta A^2}{2\sigma^2}\end{aligned} \tag{12.4}$$

or we decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \frac{1}{N-n_0}\sum_{n=n_0}^{N-1}(x[n]-A_0) > \gamma'. \tag{12.5}$$

It is not surprising that the test statistic computes the average deviation of the data from A_0 only over the assumed jump interval. This is because the noise samples are independent so that knowledge of the data before the jump is irrelevant (see Section 3.5). To determine the performance of the detector we note that

$$T(\mathbf{x}) \sim \begin{cases} \mathcal{N}(0, \sigma^2/(N-n_0)) & \text{under } \mathcal{H}_0\\ \mathcal{N}(\Delta A, \sigma^2/(N-n_0)) & \text{under } \mathcal{H}_1\end{cases}$$

which is in the form of the mean shifted Gauss-Gauss problem. Hence, the performance is from Chapter 3

$$P_D = Q(Q^{-1}(P_{FA}) - \sqrt{d^2}) \tag{12.6}$$

where

$$\begin{aligned} d^2 &= \frac{\Delta A^2}{\sigma^2/(N-n_0)} \\ &= \frac{(N-n_0)\Delta A^2}{\sigma^2} \end{aligned} \tag{12.7}$$

is the deflection coefficient. We see that as expected the performance improves as the length of the data interval *after the jump* $N - n_0$ increases and as the jump size ΔA increases. As we observe more data, P_D increases. We can view the required $N - n_0$ as the *delay time* in detecting a jump. For example, if the jump to noise power is $\Delta A^2/\sigma^2 = 1$ and we desire $P_{FA} = 0.001$ and $P_D = 0.99$, then from (12.6) and (12.7) the delay must be at least $N - n_0 = 30$ samples. Note that the threshold is determined from

$$P_{FA} = Q\left(\frac{\gamma'}{\sqrt{\sigma^2/(N-n_0)}}\right)$$

as

$$\gamma' = \sqrt{\frac{\sigma^2}{N-n_0}}Q^{-1}(P_{FA})$$

and decreases with the data record length. As an example, in Figure 12.2a we have plotted a realization of $x[n]$ in which $A_0 = 1$, $\Delta A = 1$, $n_0 = 50$, and $\sigma^2 = 1$. In Figure 12.2b $T(\mathbf{x})$ (see (12.5)) as a function of $N - 1$ (the last sample in the data record) has been plotted for the realization of Figure 12.2a. Also shown in the figure are the thresholds γ' for $P_{FA} = 10^{-3}$ and $P_{FA} = 10^{-6}$. As N increases, the threshold γ' decreases since we can expect a more accurate estimate of the DC level. If we wish to detect the change earlier, our only option is to allow more false alarms. In Figure 12.2b it is seen that a delay of about 18 samples is incurred for $P_{FA} = 10^{-6}$ but only about 4 samples for $P_{FA} = 10^{-3}$.

This example illustrates the typical tradeoffs in parameter change detection. To detect quickly we require either a large jump and/or a small amount of noise. For better performance we can choose to allow more delay (assuming, of course, that the signal remains constant after the jump (see Problem 12.2)). These characteristics are summarized by the deflection coefficient of (12.7). Also, in practice, we will need to compute $T(\mathbf{x})$ sequentially in time. Thus, computational procedures such as sequential least squares [Kay-I 1993, Chapter 8] (see Problem 12.1) and the Kalman filter are useful [Kay-I 1993, Chapter 13]. More generally, if a sufficient statistic exists for the parameter of interest, then the computation may be simplified [Birdsall and Gobien 1973].

◇

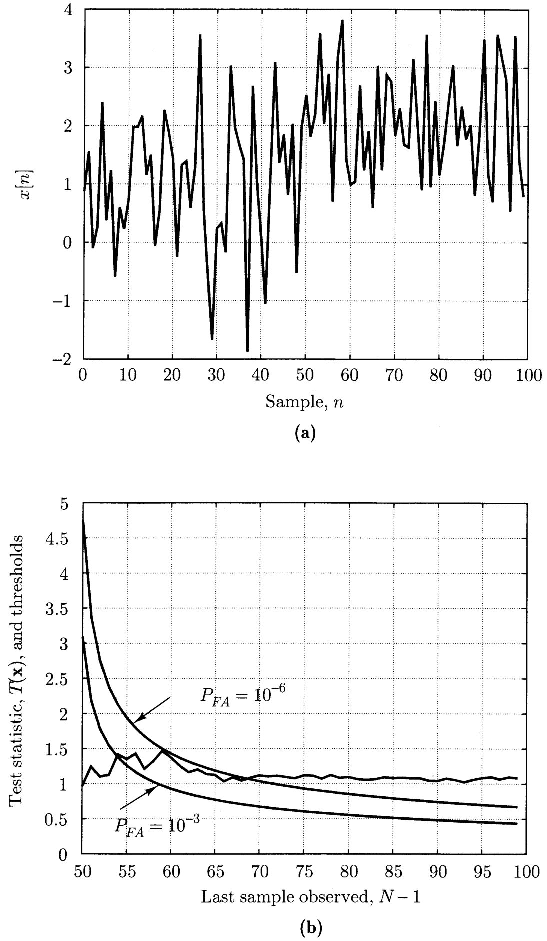


Figure 12.2. Detection of a DC level jump in WGN (a) Realization (b) Test statistic and threshold.

The next example addresses the problem of detecting a change in the variance of a WGN process.

Example 12.2 - Known Variance Jump at Known Time

We now wish to detect a jump in the variance or power of WGN at a known time n_0. An example is shown in Figure 12.3 in which the variance jumps from $\sigma^2 = 1$ to $\sigma^2 = 4$ at $n_0 = 50$. Similar to the previous example, we consider the hypothesis test

$$\begin{aligned}
\mathcal{H}_0 : x[n] &= \quad w[n] \qquad n = 0, 1, \ldots, N-1 \\
\mathcal{H}_1 : x[n] &= \begin{cases} w_1[n] & n = 0, 1, \ldots, n_0 - 1 \\ w_2[n] & n = n_0, n_0 + 1, \ldots, N-1. \end{cases}
\end{aligned}$$

where $w[n]$ is WGN with known variance σ_0^2, $w_1[n]$ is WGN with known variance σ_0^2, and $w_2[n]$ is WGN with known variance $\sigma_0^2 + \Delta\sigma^2$ for $\Delta\sigma^2 > 0$. We further assume that the noise processes $w_1[n]$ and $w_2[n]$ are independent of each other. In terms of a parameter test we have

$$\begin{aligned}
\mathcal{H}_0 : \sigma^2 &= \quad \sigma_0^2 \qquad\qquad n = 0, 1, \ldots, N-1 \\
\mathcal{H}_1 : \sigma^2 &= \begin{cases} \sigma_0^2 & n = 0, 1, \ldots, n_0 - 1 \\ \sigma_0^2 + \Delta\sigma^2 & n = n_0, n_0 + 1, \ldots, N-1 \end{cases}
\end{aligned}$$

where σ_0^2 and $\Delta\sigma^2$ are known and $\Delta\sigma^2 > 0$. An NP detector decides $\mathcal{H}_1$ if

$$L(\mathbf{x}) = \frac{p(\mathbf{x}; \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > \gamma$$

where

$$L(\mathbf{x}) = \frac{\frac{1}{(2\pi\sigma_0^2)^{\frac{n_0}{2}}} \exp\left(-\frac{1}{2\sigma_0^2} \sum_{n=0}^{n_0-1} x^2[n]\right) \frac{1}{(2\pi(\sigma_0^2+\Delta\sigma^2))^{\frac{N-n_0}{2}}} \exp\left(-\frac{1}{2(\sigma_0^2+\Delta\sigma^2)} \sum_{n=n_0}^{N-1} x^2[n]\right)}{\frac{1}{(2\pi\sigma_0^2)^{\frac{N}{2}}} \exp\left(-\frac{1}{2\sigma_0^2} \sum_{n=0}^{N-1} x^2[n]\right)}.$$

Taking the logarithm and retaining only the data-dependent terms we have

$$\begin{aligned}
T'(\mathbf{x}) &= -\frac{1}{2}\left(\sum_{n=0}^{n_0-1} \frac{x^2[n]}{\sigma_0^2} + \sum_{n=n_0}^{N-1} \frac{x^2[n]}{\sigma_0^2 + \Delta\sigma^2} - \sum_{n=0}^{N-1} \frac{x^2[n]}{\sigma_0^2}\right) \\
&= -\frac{1}{2}\left(-\sum_{n=n_0}^{N-1} \frac{x^2[n]}{\sigma_0^2} + \sum_{n=n_0}^{N-1} \frac{x^2[n]}{\sigma_0^2 + \Delta\sigma^2}\right) \\
&= \frac{1}{2}\left(\sum_{n=n_0}^{N-1} x^2[n] \left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_0^2 + \Delta\sigma^2}\right)\right)
\end{aligned}$$

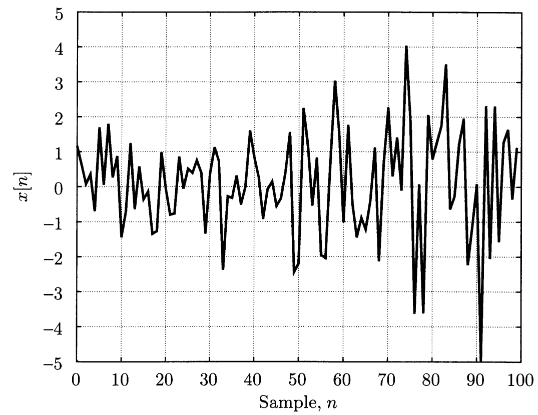


Figure 12.3. Realization of WGN whose variance jumps from $\sigma^2 = 1$ to $\sigma^2 = 4$.

or we decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \sum_{n=n_0}^{N-1} x^2[n] > \gamma'.$$

This is just the energy detector that was the NP detector for detecting a white Gaussian random signal in WGN (see Example 5.1). In effect, the jump in variance can be attributed to a white Gaussian signal that arrives at $n = n_0$. Its performance is given in Chapter 5 and is shown to improve as $\Delta\sigma^2/\sigma_0^2$ increases.

$\diamond$

In both Examples 12.1 and 12.2 we do not actually need to know ΔA and $\Delta\sigma^2$, respectively, to implement the NP detector. It is enough to know that they are both positive. In both these cases the NP test is UMP (see Chapter 3 and Problem 12.4).

12.4 Extensions to the Basic Problem

We now begin to relax the strict assumptions of the previous two examples to allow for an unknown change time and also unknown parameters before and after the jump. These assumptions are more representative of the actual situation in practice. We begin our discussion by examining the case of unknown parameters.

Example 12.3 - Unknown DC Levels and Known Jump Time

If we assume that n_0 is known but that the DC levels before the jump A_1 and after the jump A_2 are unknown, then the hypothesis testing problem is

$$\begin{array}{l}\mathcal{H}_0 : A_1 = A_2 \\ \mathcal{H}_1 : A_1 \neq A_2.\end{array}$$

Hence, we do not assume knowledge of the DC level before the jump A_1 or the jump size $\Delta A = A_2 - A_1$. Since this is a composite hypothesis test, we will apply the GLRT (see Chapter 6). We need to retain all of the samples in this case since the MLE of A_1 under $\mathcal{H}_0$ and under $\mathcal{H}_1$ will be different. A GLRT decides $\mathcal{H}_1$ if

$$L_G(\mathbf{x}) = \frac{p(\mathbf{x}; A_1 = \hat{A}_1, A_2 = \hat{A}_2)}{p(\mathbf{x}; A_1 = \hat{A}, A_2 = \hat{A})} > \gamma$$

where $\hat{A}$ is the MLE of the DC level under $\mathcal{H}_0$ (no jump) and $\hat{A}_1, \hat{A}_2$ are the MLEs of the DC level before the jump and after the jump, respectively. We have from (12.3) that

$$p(\mathbf{x}; A_1, A_2) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{n=0}^{n_0-1}(x[n]-A_1)^2 + \sum_{n=n_0}^{N-1}(x[n]-A_2)^2\right)\right].$$

Finding the MLEs is an easy task with the results being

$$\begin{aligned}\hat{A} &= \frac{1}{N}\sum_{n=0}^{N-1} x[n] = \bar{x} \\ \hat{A}_1 &= \frac{1}{n_0}\sum_{n=0}^{n_0-1} x[n] \\ \hat{A}_2 &= \frac{1}{N-n_0}\sum_{n=n_0}^{N-1} x[n].\end{aligned}$$

The GLRT becomes

$$\begin{aligned}&2\ln L_G(\mathbf{x}) \\ &= \frac{1}{\sigma^2}\left[\sum_{n=0}^{N-1}(x[n]-\bar{x})^2 - \sum_{n=0}^{n_0-1}(x[n]-\hat{A}_1)^2 - \sum_{n=n_0}^{N-1}(x[n]-\hat{A}_2)^2\right] \\ &= \frac{1}{\sigma^2}\left[\sum_{n=0}^{N-1} x^2[n] - N\bar{x}^2 - \left(\sum_{n=0}^{n_0-1} x^2[n] - n_0\hat{A}_1^2\right)\right.\end{aligned}$$

$$\begin{aligned}&\qquad \left.- \left(\sum_{n=n_0}^{N-1} x^2[n] - (N-n_0)\hat{A}_2^2\right)\right] \\ &= \frac{1}{\sigma^2}\left[(N-n_0)\hat{A}_2^2 + n_0\hat{A}_1^2 - N\bar{x}^2\right] \\ &= \frac{N}{\sigma^2}\left[\frac{n_0}{N}(\hat{A}_1^2 - \hat{A}_2^2) + \hat{A}_2^2 - \bar{x}^2\right].\end{aligned}$$

Noting that

$$\begin{aligned}\bar{x} &= \frac{n_0}{N}\hat{A}_1 + \frac{N-n_0}{N}\hat{A}_2 \\ &= \frac{n_0}{N}(\hat{A}_1 - \hat{A}_2) + \hat{A}_2\end{aligned}$$

we have after some simplification that we decide $\mathcal{H}_1$ if

$$2\ln L_G(\mathbf{x}) = \frac{(\hat{A}_1 - \hat{A}_2)^2}{\sigma^2\left(\frac{1}{n_0} + \frac{1}{N-n_0}\right)} > \gamma'. \tag{12.8}$$

The GLRT is seen to compute the difference of the estimated levels before and after the jump (normalized by the standard deviation) and compares the square of this value to a threshold. The exact PDF is easily found since $\hat{A}_1, \hat{A}_2$ are Gaussian and independent of each other. As usual

$$\hat{A}_1 - \hat{A}_2 \sim \begin{cases} \mathcal{N}\left(0, \sigma^2\left(\frac{1}{n_0} + \frac{1}{N-n_0}\right)\right) & \text{under } \mathcal{H}_0 \\ \mathcal{N}\left(A_1 - A_2, \sigma^2\left(\frac{1}{n_0} + \frac{1}{N-n_0}\right)\right) & \text{under } \mathcal{H}_1\end{cases}$$

so that

$$2\ln L_G(\mathbf{x}) \sim \begin{cases} \chi_1^2 & \text{under } \mathcal{H}_0 \\ \chi_1'^2(\lambda) & \text{under } \mathcal{H}_1\end{cases} \tag{12.9}$$

where

$$\lambda = \frac{(A_1 - A_2)^2}{\sigma^2\left(\frac{1}{n_0} + \frac{1}{N-n_0}\right)}. \tag{12.10}$$

As expected, the performance increases monotonically with λ, which in turn, increases with the jump power-to-noise ratio and the data record lengths of the two segments. It is also interesting to observe that the best performance occurs when the jump is at the midpoint of the data record (see Problem 12.9).

$\diamond$

Next we extend Example 12.1 for known DC levels to accommodate an unknown jump time.

Example 12.4 - Known DC Levels and Unknown Jump Time

Our hypothesis is now given by (12.2) with A_0, $\Delta A > 0$ known but with n_0 *unknown*. We assume $n_{0_{\min}} \le n_0 \le n_{0_{\max}}$, where presumably $n_{0_{\min}} \gg 1$ and $n_{0_{\max}} \ll N-1$ so that the transition, if it occurs, is not too close to the endpoints of the observation interval. Using a GLRT we decide $\mathcal{H}_1$ if

$$L_G(\mathbf{x}) = \frac{p(\mathbf{x}; \hat{n}_0, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > \gamma$$

where $\hat{n}_0$ is the MLE under $\mathcal{H}_1$. Equivalently, we have (see Chapter 6)

$$\begin{aligned} L_G(\mathbf{x}) &= \max_{n_0} \frac{p(\mathbf{x}; n_0, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} \\ &= \max_{n_0} L(\mathbf{x}; n_0) \end{aligned}$$

or

$$\ln L_G(\mathbf{x}) = \ln \max_{n_0} L(\mathbf{x}; n_0) = \max_{n_0} \ln L(\mathbf{x}; n_0)$$

where $L(\mathbf{x}; n_0)$ is just the likelihood ratio from Example 12.1. Thus, from (12.4)

$$\begin{aligned} \ln L(\mathbf{x}; n_0) &= \frac{\Delta A}{\sigma^2} \sum_{n=n_0}^{N-1} (x[n] - A_0) - \frac{(N-n_0)\Delta A^2}{2\sigma^2} \\ &= \frac{\Delta A}{\sigma^2} \sum_{n=n_0}^{N-1} \left(x[n] - A_0 - \frac{\Delta A}{2} \right) \end{aligned}$$

so that

$$\ln L_G(\mathbf{x}) = \frac{\Delta A}{\sigma^2} \max_{n_0} \sum_{n=n_0}^{N-1} \left(x[n] - A_0 - \frac{\Delta A}{2} \right)$$

or we decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \max_{n_0} \sum_{n=n_0}^{N-1} \left(x[n] - A_0 - \frac{\Delta A}{2} \right) > \gamma'.$$

The usual asymptotic GLRT statistics do not hold here. Comparing the results to Example 12.1 we see that the basic difference is that the test statistic is now maximized over all possible values of n_0. This is the same result as for the detection of a signal with an unknown arrival time (see Chapter 7). Here the signal that we are trying to detect is actually the jump or $s[n] = \Delta A$, which is the difference of the means between the hypotheses.

$\diamond$

The final case occurs when the DC levels as well as the jump time are unknown. Then it can be shown (see Problem 12.10) that the GLRT decides $\mathcal{H}_1$ if

$$\max_{n_0} \frac{(\hat{A}_1 - \hat{A}_2)^2}{\sigma^2 \left(\frac{1}{n_0} + \frac{1}{N-n_0} \right)} > \gamma' \qquad (12.11)$$

where

$$\begin{aligned} \hat{A}_1 &= \frac{1}{n_0} \sum_{n=0}^{n_0-1} x[n] \\ \hat{A}_2 &= \frac{1}{N-n_0} \sum_{n=n_0}^{N-1} x[n]. \end{aligned}$$

The analogous results for a jump in variance are explored in Problems 12.11 and 12.12.

12.5 Multiple Change Times

A problem of some practical importance occurs when a parameter's value changes more than once in a data record. Then, we wish to determine the multiple change times. An example is given in Figure 12.4, for the DC level in WGN. The change

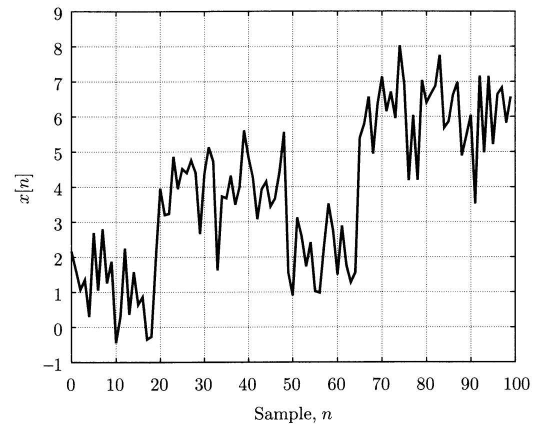


Figure 12.4. DC level in WGN with multiple jumps in DC level.

times are at $n_0 = 20$, $n_1 = 50$, and $n_2 = 65$, the DC levels are $A = 1, 4, 2, 6$, respectively, and $\sigma^2 = 1$. The problem can be quite complex and is further complicated when there are unknown parameters. The complexity stems from the large number of likelihood ratios that must be computed due to the combinatorial explosion with the number of change times. For example, in Figure 12.4 if we were to enumerate all possible change times, it would be about $N^3/6$ or in general $O(N^M)$ for M changes. To reduce the computational load to a more manageable level we can employ the technique of *dynamic programming* (DP). The computation for DP increases *linearly* with M, and the algorithm is easily implemented on a digital computer. We illustrate the general approach with an example.

Example 12.5 - Unknown DC Levels and Unknown Jump Times

Assume that the DC level changes three times as in Figure 12.4. Then, the problem is to detect or equivalently to estimate the change times. If we can determine the MLE of the change times, then any hypothesis test we wish to propose can be carried out using a GLRT. For example, we might wish to test whether there are two changes or three changes. To do so would require the evaluation of the likelihood ratio over all these possibilities. We leave the hypothesis testing problem to the reader (see Problem 12.13 for an example) and concentrate on the estimation of the levels and change times.

Assume we have the signal

$$s[n] = \begin{cases} A_0 & n = 0, 1, \ldots, n_0 - 1 \\ A_1 & n = n_0, n_0 + 1, \ldots, n_1 - 1 \\ A_2 & n = n_1, n_1 + 1, \ldots, n_2 - 1 \\ A_3 & n = n_2, n_2 + 1, \ldots, N - 1 \end{cases}$$

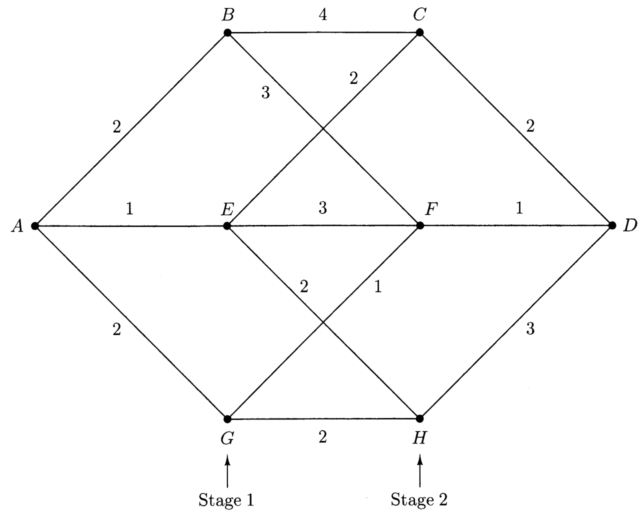
embedded in WGN of variance σ^2. We do not assume knowledge of the levels so that they must be jointly estimated along with the change times. Clearly, if the change times n_i's were known, the MLE of each level A_i would be given by the sample mean of the data over each segment.

The joint MLE of $\mathbf{A} = [A_0\, A_1\, A_2\, A_3]^T$ and $\mathbf{n} = [n_0\, n_1\, n_2]^T$ is found by minimizing

$$\begin{aligned} J(\mathbf{A}, \mathbf{n}) &= \sum_{n=0}^{n_0-1} (x[n] - A_0)^2 + \sum_{n=n_0}^{n_1-1} (x[n] - A_1)^2 \\ &\quad + \sum_{n=n_1}^{n_2-1} (x[n] - A_2)^2 + \sum_{n=n_2}^{N-1} (x[n] - A_3)^2. \end{aligned} \tag{12.12}$$

Dynamic programming, which is a minimization approach, exploits the observation that not all combinations of n_0, n_1, n_2 need be evaluated to find the minimum of $J(\mathbf{A}, \mathbf{n})$. The basic principle can be explained by referring to Figure 12.5. The objective is for a person to travel from point A to point D by the shortest path.

The distances for each segment are indicated. By evaluating the distances for all the possible paths, we see that the path AGFD has the minimum distance. However, it is possible to determine this optimal path without an exhaustive search by the following strategy. Assume first that we have traveled via the optimal path to stage 2. Then we must be at either C, F, or H. If we pass through C, then we must have traveled via path ABC or path AEC. The shorter of these two paths is AEC. Consequently, if C is on the optimal path, the solution must be AECD. This eliminates the computation of the path ABCD. Similarly, if F is on the optimal path, then we must have traversed AGF since it is the minimum distance of ABF, AEF, and AGF. We have now eliminated the computation for the additional paths ABFD and AEFD. Continuing in this manner, we ultimately determine AGFD as the optimal path. A decision tree diagram of this process is depicted in Figure 12.6. Note that at stage 2 we effectively eliminate many of the paths from consideration.



Paths	Distance
ABCD	8
AECD	5
AEFD	5
ABFD	6
* *AGFD*	4
AGHD	7
AEHD	6

* Optimal path

Figure 12.5. Enumeration of possible paths and their distances.

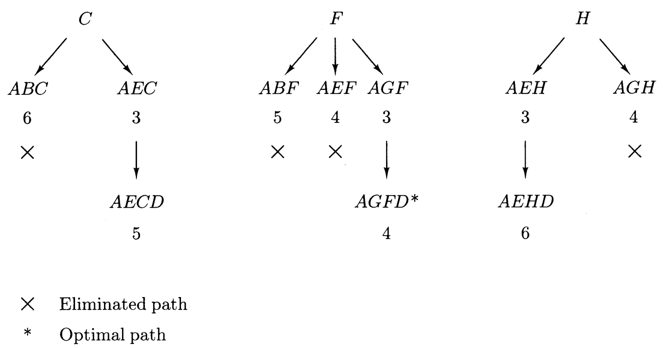


Figure 12.6. Dynamic programming search for optimal path.

In fact, at stage 2 we retain only the three best paths, since the optimal paths must be via C or F or H. The same process can be repeated for any number of stages. Since we retain only a fixed number of paths per stage, the computation grows linearly with the number of stages.

We now apply DP to the multiple change time problem. In doing so, we must point out that not all problems can be solved this way. Inherent in the approach is the assumption that the error to be minimized has a certain form that is amenable to a recursive evaluation and, furthermore, that it exhibits the *Markov* property. In terms of the previous example, the Markov property requires that the distance from C to D does not depend upon how we arrived at C, for example. This allows us to add the distances without regard to the specifics of the previous paths. Also, the mathematical formulation of an optimization problem in terms of DP may not be apparent to the novice and appears to be quite "tricky." The reader is referred to [Bellman and Dreyfus 1962, Larson and Casti 1982] for additional discussions and examples of DP.

To minimize (12.12) over $\mathbf{A}$ and $\mathbf{n}$ we first note that for given change times

$$\hat{A}_i = \frac{1}{n_i - n_{i-1}} \sum_{n=n_{i-1}}^{n_i - 1} x[n]$$

for $i = 0, 1, 2, 3$, where $n_{-1} = 0$ and $n_3 = N$. This is just the sample mean for the data interval $[n_{i-1}, n_i - 1]$. Then, we define

$$\Delta_i[n_{i-1}, n_i - 1] = \sum_{n=n_{i-1}}^{n_i - 1} (x[n] - \hat{A}_i)^2. \qquad (12.13)$$

Note that the minimization of $J(\mathbf{A}, \mathbf{n})$ over $\mathbf{A}$ and $\mathbf{n}$ is equivalent to the minimization of

$$J(\hat{\mathbf{A}}, \mathbf{n}) = \sum_{i=0}^{3} \Delta_i[n_{i-1}, n_i - 1]$$

over $\mathbf{n}$. To minimize $J(\hat{\mathbf{A}}, \mathbf{n})$ over n_0, n_1, n_2 using DP we define

$$I_k[L] = \min_{\substack{n_0, n_1, \ldots, n_{k-1} \\ n_{-1}=0, n_k=L+1}} \sum_{i=0}^{k} \Delta_i[n_{i-1}, n_i - 1]$$

where $0 < n_0 < n_1 < \ldots < n_{k-1} < L+1$. Note that $I_3[N-1]$ is the minimum of $J(\hat{\mathbf{A}}, \mathbf{n})$. We next develop a recursion for the minimum, much the same as the minimum from A to D in Figure 12.5 was the minimum distance from A to stage 2 plus the distance from stage 2 to D.

$$\begin{aligned}
I_k[L] &= \min_{\substack{n_{k-1} \\ n_k=L+1}} \min_{\substack{n_0, n_1, \ldots, n_{k-2} \\ n_{-1}=0}} \sum_{i=0}^{k} \Delta_i[n_{i-1}, n_i - 1] \\
&= \min_{\substack{n_{k-1} \\ n_k=L+1}} \min_{\substack{n_0, n_1, \ldots, n_{k-2} \\ n_{-1}=0}} \sum_{i=0}^{k-1} \Delta_i[n_{i-1}, n_i - 1] + \Delta_k[n_{k-1}, n_k - 1] \\
&= \min_{\substack{n_{k-1} \\ n_k=L+1}} \left[\left(\min_{\substack{n_0, n_1, \ldots, n_{k-2} \\ n_{-1}=0}} \sum_{i=0}^{k-1} \Delta_i[n_{i-1}, n_i - 1] \right) + \Delta_k[n_{k-1}, n_k - 1] \right]
\end{aligned}$$

or finally we have that

$$I_k[L] = \min_{n_{k-1}} \left(I_{k-1}[n_{k-1} - 1] + \Delta_k[n_{k-1}, L] \right). \qquad (12.14)$$

This says that for a data record $[0, L]$, the minimum error for $k+1$ segments (k change times) is the minimum error for the first k segments that end at $n = n_{k-1} - 1$ and the error contributed by the last segment from $n = n_{k-1}$ to $n = L$. The solution to our problem is for $k = 3$ and $L = N - 1$.

The procedure to compute the solution, assuming a minimum length segment of one sample, is as follows:

1. Let $k = 0$ and compute $I_0[L]$ for $L = 0, 1, \ldots, N-4$ as

$$I_0[L] = \Delta_0[n_{-1} = 0, n_0 - 1 = L] = \sum_{n=0}^{L} \left(x[n] - \hat{A}_0 \right)^2$$

where $\hat{A}_0 = (1/(L+1)) \sum_{n=0}^{L} x[n]$.

2. Let $k = 1$ and compute $I_1[L]$ for $L = 1, 2, \ldots, N-3$ as

$$I_1[L] = \min_{1 \le n_0 \le L} (I_0[n_0 - 1] + \Delta_1[n_0, L])$$

where from step 1 we have already determined $I_0[n_0 - 1]$ for $n_0 = 1, 2, \ldots, N-3$. The term $\Delta_1[n_0, L]$ is found from (12.13). Also, for each L determine the value of n_0 that minimizes $I_1[L]$ and call it $n_0(L)$.

3. Let $k = 2$ and compute $I_2[L]$ for $L = 2, 3, \ldots, N-2$ as

$$I_2[L] = \min_{2 \le n_1 \le L} (I_1[n_1 - 1] + \Delta_2[n_1, L]).$$

To find $I_1[n_1 - 1]$ use the results of step 2. Also, $\Delta_2[n_1, L]$ is found from (12.13). Again for each L determine the value of n_1 that minimizes $I_2[L]$ and call it $n_1(L)$.

4. Finally let $k = 3$ and compute $I_3[L]$ for $L = N - 1$ as

$$I_3[L] = \min_{3 \le n_2 \le L} (I_2[n_2 - 1] + \Delta_3[n_2, L]).$$

We denote the minimizing value of n_2 as $n_2(N-1)$. Now the minimum of the function $J(\hat{\mathbf{A}}, \mathbf{n})$ is $I_3[N-1]$.

5. The change times are found by using the backward recursion

$$\begin{aligned}
\hat{n}_2 &= n_2(N-1) \\
\hat{n}_1 &= n_1(\hat{n}_2 - 1) \\
\hat{n}_0 &= n_0(\hat{n}_1 - 1).
\end{aligned}$$

To appreciate the recursions involved in the implementation of the DP algorithm the reader should carry out a simple numerical example. For this example with $A = 1, 4, 2, 6$ and $n_0 = 20, n_1 = 50, n_2 = 65$, the results are $\hat{n}_0 = 20, \hat{n}_1 = 49, \hat{n}_2 = 65$. The MATLAB program dp.m implements this algorithm and can be found in Appendix 12B. Note that from (12.14) the bulk of the computation is in finding

$$\Delta_k[n_{k-1}, L] = \sum_{n=n_{k-1}}^{L} (x[n] - \hat{A}_k)^2. \tag{12.15}$$

This is the minimum least squares error when the data record $[n_{k-1}, L]$ is used to estimate the mean. As such, the computation may be done recursively using sequential least squares formulas [Kay-I 1993, pp. 242–244]. Denoting $\hat{A}[m, n]$ as the sample mean and $J_{\min}[m, n]$ as the minimum least squares error for the data record $[m, n]$, we have the recursions

$$\begin{aligned}
\hat{A}[m, n] &= \hat{A}[m, n-1] + \frac{1}{n - m + 1}(x[n] - \hat{A}[m, n-1]) \\
J_{\min}[m, n] &= J_{\min}[m, n-1] + \frac{n - m}{n - m + 1}(x[n] - \hat{A}[m, n-1])^2.
\end{aligned}$$

The recursion is initialized with $\hat{A}[0, 0] = x[0]$ and $J_{\min}[0, 0] = 0$ and proceeds for each $n \ge m$, where $m = 0, 1, \ldots, N-1$. Note that this approach is easily extendible to any number of segments if the MLE is easily found once the data record beginning and end points are known. A more general formulation of the DP approach to detection of change times is given in Appendix 12A. Speech segmentation via AR modeling is one application [Svendsen and Soong 1987]. Another is a linear model whose parameters change from segment to segment. If we have an unknown number of segments, then techniques such as the MDL (see Chapter 6) are applicable.

12.6 Signal Processing Examples

We now apply our results to some signal processing problems of interest. The first example addresses the detection of an object's deviation from a nominal straight line trajectory. The second example concerns the detection of a change in the PSD of a random process, from that of a white process to a colored one. The first problem has application to motion control and also tracking of maneuvering targets while the second is important for accurate statistical modeling in detection/estimation algorithms.

12.6.1 Maneuver Detection

For an object moving in a straight line we wish to detect if and when it departs from its nominal straight line trajectory. The object is assumed to travel in a plane so that its nominal position at time $t = n\Delta$ is given by

$$\begin{aligned}
r_x[n] &= r_x[0] + v_x n\Delta \\
r_y[n] &= r_y[0] + v_y n\Delta
\end{aligned}$$

where $n \ge 0$ and Δ is the time interval between samples. Its initial position $(r_x[0], r_y[0])$ is known, and the velocity (v_x, v_y) is assumed to be constant and also known. (This model is slightly different from that in [Kay-I 1993, pp. 456–466] in which a Kalman tracking filter was applied. Here there is no plant noise. However, this example may be extended to the former case, although it is tedious.) If at some time $n = n_0$, the object departs from its straight line trajectory, then it will experience an acceleration. Its position then becomes for $n \geq n_0$

$$\begin{aligned} r_x[n] &= r_x[0] + v_x n\Delta + \frac{1}{2}a_x(n-n_0)^2\Delta^2 \\ r_y[n] &= r_y[0] + v_y n\Delta + \frac{1}{2}a_y(n-n_0)^2\Delta^2 \end{aligned}$$

where (a_x, a_y) denotes the acceleration. To detect this acceleration we will need to test if $a_x \neq 0$ and/or $a_y \neq 0$ for $n \geq n_0$. Since our knowledge of $r_x[n], r_y[n]$ will not be perfect due to sensor errors, we will include these errors in our model as WGN. A model for our observations then becomes

$$\begin{aligned} r_x[n] &= r_x[0] + v_x n\Delta + w_x[n] \\ r_y[n] &= r_y[0] + v_y n\Delta + w_y[n] \end{aligned}$$

for $n = 0, 1, \ldots, n_0 - 1$ and

$$\begin{aligned} r_x[n] &= r_x[0] + v_x n\Delta + \frac{1}{2}a_x(n-n_0)^2\Delta^2 + w_x[n] \\ r_y[n] &= r_y[0] + v_y n\Delta + \frac{1}{2}a_y(n-n_0)^2\Delta^2 + w_y[n] \end{aligned}$$

for $n = n_0, n_0+1, \ldots, N-1$. The noise processes $w_x[n]$ and $w_y[n]$ are assumed to be WGN with the same known variance σ^2 and independent of each other. With these assumptions we have a linear model for the data, and consequently, our detection problem reduces to a parameter test of the model. The GLRT for the classical linear model (Theorem 7.1) applies directly if we make a decision at $n = N-1 > n_0$ based on the observed data set $\{n_0, n_0+1, \ldots, N-1\}$. The data prior to $n = n_0$ is irrelevant to the problem (since the observation noise is WGN and hence independent) and can be discarded. Hence, by defining

$$\begin{aligned} \epsilon_x[n] &= r_x[n] - r_x[0] - v_x n\Delta \\ \epsilon_y[n] &= r_y[n] - r_y[0] - v_y n\Delta \end{aligned}$$

for $n_0 \leq n \leq N-1$ the linear model becomes

$$\underbrace{\begin{bmatrix} \boldsymbol{\epsilon}_x \\ \boldsymbol{\epsilon}_y \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} \mathbf{h} & \mathbf{0} \\ \mathbf{0} & \mathbf{h} \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} a_x \\ a_y \end{bmatrix}}_{\boldsymbol{\theta}} + \underbrace{\begin{bmatrix} \mathbf{w}_x \\ \mathbf{w}_y \end{bmatrix}}_{\mathbf{w}}$$

where

$$\begin{aligned} \boldsymbol{\epsilon}_x &= [\epsilon_x[n_0]\, \epsilon_x[n_0+1] \ldots \epsilon_x[N-1]]^T \\ \boldsymbol{\epsilon}_y &= [\epsilon_y[n_0]\, \epsilon_y[n_0+1] \ldots \epsilon_y[N-1]]^T \\ \mathbf{h} &= [0\ \Delta^2/2 \ldots (N-1-n_0)^2\Delta^2/2]^T \\ \mathbf{w}_x &= [w_x[n_0]\, w_x[n_0+1] \ldots w_x[N-1]]^T \\ \mathbf{w}_y &= [w_y[n_0]\, w_y[n_0+1] \ldots w_y[N-1]]^T \end{aligned}$$

and $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. The dimensionalities of $\mathbf{x}$ and $\mathbf{H}$ are $2(N-n_0) \times 1$ and $2(N-n_0) \times 2$, respectively. Applying Theorem 7.1 (see Chapter 7) we wish to test

$$\begin{aligned} &\mathcal{H}_0 : \boldsymbol{\theta} = \mathbf{0} \\ &\mathcal{H}_1 : \boldsymbol{\theta} \neq \mathbf{0}. \end{aligned}$$

Letting $\mathbf{A} = \mathbf{I}$, $\mathbf{b} = \mathbf{0}$ in Theorem 7.1, the GLRT decides $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \frac{\hat{\boldsymbol{\theta}}_1^T \mathbf{H}^T \mathbf{H} \hat{\boldsymbol{\theta}}_1}{\sigma^2} > \gamma' \tag{12.16}$$

where $\hat{\boldsymbol{\theta}}_1 = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$ so that

$$T(\mathbf{x}) = \frac{\mathbf{x}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}}{\sigma^2}.$$

But $\mathbf{H}^T\mathbf{H} = \operatorname{diag}(\mathbf{h}^T\mathbf{h}, \mathbf{h}^T\mathbf{h}) = \mathbf{h}^T\mathbf{h}\mathbf{I}$ so that

$$\begin{aligned} T(\mathbf{x}) &= \frac{||\mathbf{H}^T\mathbf{x}||^2}{\sigma^2\mathbf{h}^T\mathbf{h}} \\ &= \frac{(\mathbf{h}^T\boldsymbol{\epsilon}_x)^2 + (\mathbf{h}^T\boldsymbol{\epsilon}_y)^2}{\sigma^2\mathbf{h}^T\mathbf{h}} \end{aligned}$$

or finally we decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \frac{\left(\sum_{n=n_0}^{N-1} \frac{(n-n_0)^2\Delta^2}{2}\epsilon_x[n]\right)^2 + \left(\sum_{n=n_0}^{N-1} \frac{(n-n_0)^2\Delta^2}{2}\epsilon_y[n]\right)^2}{\sigma^2 \sum_{n=n_0}^{N-1} \left(\frac{(n-n_0)^2\Delta^2}{2}\right)^2} > \gamma'.$$

From (12.16) this can be shown to be equivalent to

$$T(\mathbf{x}) = \left[\frac{1}{\sigma^2}\sum_{n=n_0}^{N-1}\left(\frac{(n-n_0)^2\Delta^2}{2}\right)^2\right](\hat{a}_x^2 + \hat{a}_y^2)$$

where

$$\begin{bmatrix} \hat{a}_x \\ \hat{a}_y \end{bmatrix} = \begin{bmatrix} (\mathbf{h}^T\mathbf{h})^{-1}\mathbf{h}^T\boldsymbol{\epsilon}_x \\ (\mathbf{h}^T\mathbf{h})^{-1}\mathbf{h}^T\boldsymbol{\epsilon}_y \end{bmatrix}$$

is the minimum variance unbiased estimator of the acceleration. Note that under $\mathcal{H}_0$ we have from Theorem 7.1 that $T(\mathbf{x}) \sim \chi_2^2$. Therefore, it easily follows that

$$P_{FA} = \exp(-\gamma'/2).$$

In practice, n_0 is usually not known so that $T(\mathbf{x})$ must be computed for all possible change times. Furthermore, to convert the statistic into one that will quickly detect the change, it is customary to use a fixed length M-point data window $[n_0, n_0+M-1]$ (where $M \ll N$) to compute the statistic. The test statistic is then determined for all possible n_0. The tradeoff, as usual, will be to quickly detect the change, for which M should be small, while at the same time to prevent excessive false alarms, for which M should be large. Thus, the statistic that examines the data in the interval $[n_0, n_0 + M - 1]$ is

$$T_{n_0}(\mathbf{x}) = \frac{\left(\sum_{n=n_0}^{n_0+M-1} \frac{(n-n_0)^2\Delta^2}{2}\epsilon_x[n]\right)^2 + \left(\sum_{n=n_0}^{n_0+M-1} \frac{(n-n_0)^2\Delta^2}{2}\epsilon_y[n]\right)^2}{\sigma^2 \sum_{n=n_0}^{n_0+M-1} \left(\frac{(n-n_0)^2\Delta^2}{2}\right)^2} > \gamma'$$

and is computed for each possible manuever time $n_0 \geq 0$. An example is now described. The initial position of an object is $(r_x[0], r_y[0]) = (0,0)$, its velocity is $(v_x, v_y) = (1,1)$, and the sample interval is $\Delta = 1$. This nominal straight line trajectory is maintained for $n = 0, 1, \ldots, n_0 - 1$, where $n_0 = 50$. For $n \geq n_0$ the object experiences an acceleration of $(a_x, a_y) = (0.03, 0.05)$. If WGN of variance $\sigma^2 = 10$ is added to the position to represent the measurement error, then a typical realization is shown in Figure 12.7. Note that the nominal straight line trajectory is shown as a dashed line and that the position begins to deviate from it at $n_0 = 50$ due to the acceleration. In Figure 12.8 we plot $T_{n_0}(\mathbf{x})$ versus n_0 for a window width of $M = 20$. Depending on the threshold chosen, the acceleration will be detected after a delay of about 30 samples. For example, if the threshold were set at 100, then the change would be first detected for $\hat{n}_0 = 65$. Since the detection window $[n_0, n_0 + M - 1]$ is $[n_0, n_0 + 19]$, a detection occurs for the data record $[65, 84]$ for a delay of 35 samples.

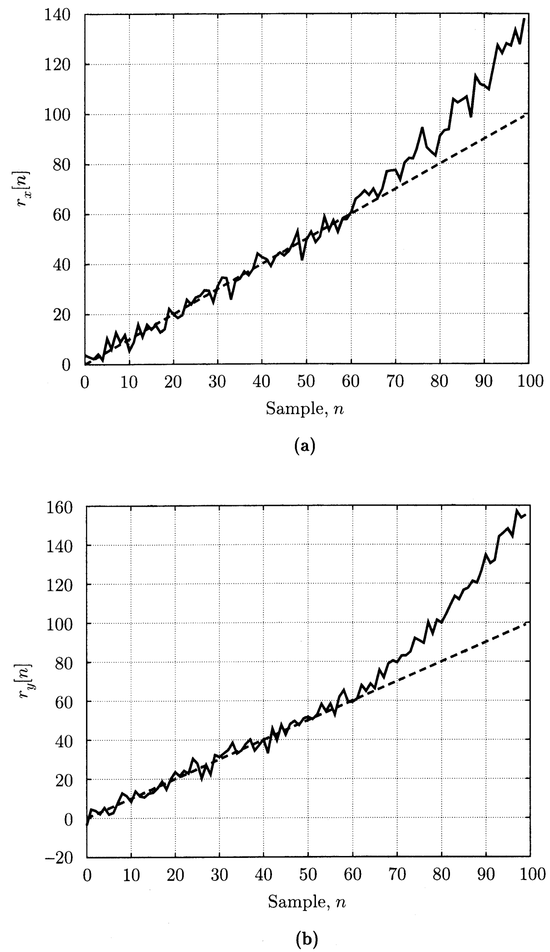


Figure 12.7. Position of accelerating object embedded in WGN (a) $r_x[n]$ (b) $r_y[n]$.

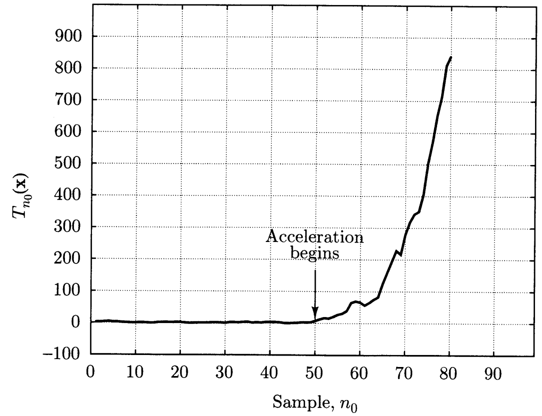


Figure 12.8. Test statistic used to detect acceleration.

12.6.2 Time Varying PSD Detection

We next examine the detection of a change in the PSD of a Gaussian random process from a white PSD to a colored one. The process may represent either a random signal or noise. To model the colored process we assume that the PSD has the AR(1) form (see Section 9.5 and Appendix 1)

$$P_{xx}(f) = \frac{\sigma_u^2}{|1 + a[1]\exp(-j2\pi f)|^2} \tag{12.17}$$

where $|a[1]| < 1$ and $\sigma_u^2 > 0$. The results to be derived are easily extended to an AR PSD of any order. Before the change, the PSD is that of a white process that corresponds to setting $a[1] = 0$ in (12.17), while afterward the colored process is characterized by $a[1] \neq 0$. We also allow different σ_u^2 before and after the change and denote these by σ_1^2 and σ_2^2. Thus, power level changes can also be detected. No knowledge is assumed of $\sigma_u^2 = \sigma_0^2$, the power of the process under $\mathcal{H}_0$, or of σ_1^2, the power of the process before the change under $\mathcal{H}_1$, and $a[1], \sigma_2^2$, the AR parameters of the process after the change under $\mathcal{H}_1$. Therefore, the data before the change must be retained to obtain variance estimates. The hypothesis test becomes

$$\begin{array}{lll} \mathcal{H}_0 : a[1] = 0, \sigma_u^2 = \sigma_0^2 & & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : a[1] = 0, \sigma_u^2 = \sigma_1^2 & & n = 0, 1, \ldots, n_0 - 1 \\ \phantom{\mathcal{H}_1 :} a[1] \neq 0, \sigma_u^2 = \sigma_2^2 \neq \sigma_1^2 & & n = n_0, n_0+1, \ldots N-1 \end{array}$$

where under $\mathcal{H}_1$ the processes before and after the change are assumed independent. A GLRT would decide $\mathcal{H}_1$ if

$$L_G(\mathbf{x}) = \frac{p(\mathbf{x}_1; a[1] = 0, \hat{\sigma_1^2})p(\mathbf{x}_2; \hat{a}[1], \hat{\sigma_2^2})}{p(\mathbf{x}; \hat{\sigma_0^2})} > \gamma$$

where $\mathbf{x}_1 = [x[0]\, x[1] \ldots x[n_0 - 1]]^T$, $\mathbf{x}_2 = [x[n_0]\, x[n_0+1] \ldots x[N-1]]^T$. Under $\mathcal{H}_1$, $\hat{\sigma_1^2}$ is the MLE of σ_u^2 based on $\mathbf{x}_1$ and $\hat{a}[1], \hat{\sigma_2^2}$ is the MLE of $a[1], \sigma_u^2$ based on $\mathbf{x}_2$. Under $\mathcal{H}_0$, $\hat{\sigma_0^2}$ is the MLE of σ_u^2 based on $\mathbf{x} = [x[0]\, x[1] \ldots x[N-1]]^T$. The MLE of σ_u^2 under $\mathcal{H}_0$ is easily shown to be

$$\hat{\sigma_0^2} = \frac{1}{N}\sum_{n=0}^{N-1} x^2[n]$$

so that

$$p(\mathbf{x}; \hat{\sigma_0^2}) = \frac{1}{(2\pi\hat{\sigma_0^2})^{\frac{N}{2}}}\exp(-N/2).$$

Under $\mathcal{H}_1$ for $0 \le n \le n_0 - 1$ we have just a white Gaussian process so that

$$p(\mathbf{x}_1; a[1] = 0, \sigma_1^2) = \frac{1}{(2\pi\sigma_1^2)^{\frac{n_0}{2}}}\exp\left(-\frac{1}{2\sigma_1^2}\sum_{n=0}^{n_0-1} x^2[n]\right)$$

and it follows that

$$\hat{\sigma_1^2} = \frac{1}{n_0}\sum_{n=0}^{n_0-1} x^2[n].$$

To find the MLE under $\mathcal{H}_1$ for $n \ge n_0$ we need the PDF of an AR(1) process. For the observed data set $\mathbf{x}' = [x[k]\, x[k+1] \ldots x[k+K-1]]^T$ this can be shown to be approximately (for a large data record or large K) [Kay 1988]

$$p(\mathbf{x}'; a[1], \sigma_u^2) = \frac{1}{(2\pi\sigma_u^2)^{\frac{K}{2}}}\exp\left[-\frac{1}{2\sigma_u^2}\sum_{n=k+1}^{k+K-1}(x[n] + a[1]x[n-1])^2\right]$$

so that

$$p(\mathbf{x}_2; a[1], \sigma_2^2) = \frac{1}{(2\pi\sigma_2^2)^{\frac{N-n_0}{2}}}\exp\left(-\frac{1}{2\sigma_2^2}\sum_{n=n_0+1}^{N-1}(x[n] + a[1]x[n-1])^2\right). \tag{12.18}$$

It can be shown that (see Problem 12.15)

$$\hat{a}[1] = -\frac{\sum_{n=n_0+1}^{N-1} x[n]x[n-1]}{\sum_{n=n_0+1}^{N-1} x^2[n-1]} \tag{12.19}$$

$$\hat{\sigma_2^2} = \frac{1}{N-n_0}\sum_{n=n_0+1}^{N-1}(x[n]+\hat{a}[1]x[n-1])^2. \tag{12.20}$$

Using (12.19) we have

$$\begin{aligned}\hat{\sigma_2^2} &\approx \left[\frac{1}{N-n_0}\sum_{n=n_0+1}^{N-1} x^2[n]\right]\left[\left(1-\hat{a}^2[1]\right)\frac{\sum_{n=n_0+1}^{N-1} x^2[n-1]}{\sum_{n=n_0+1}^{N-1} x^2[n]}\right] \\ &\approx \left[\frac{1}{N-n_0}\sum_{n=n_0}^{N-1} x^2[n]\right](1-\hat{a}^2[1]).\end{aligned} \tag{12.21}$$

Substituting the MLEs into the numerator of the GLRT test statistic produces

$$p(\mathbf{x}_1; a[1]=0, \hat{\sigma_1^2})p(\mathbf{x}_2; \hat{a}[1], \hat{\sigma_2^2}) =$$

$$\frac{1}{(2\pi\hat{\sigma_1^2})^{\frac{n_0}{2}}}\exp(-n_0/2)\frac{1}{(2\pi\hat{\sigma_2^2})^{\frac{N-n_0}{2}}}\exp(-(N-n_0)/2)$$

and thus

$$L_G(\mathbf{x}) = \frac{(\hat{\sigma_0^2})^{\frac{N}{2}}}{(\hat{\sigma_1^2})^{\frac{n_0}{2}}(\hat{\sigma_2^2})^{\frac{N-n_0}{2}}}$$

or using (12.21) this becomes

$$L_G(\mathbf{x}) = \frac{(\hat{\sigma_0^2})^{\frac{N}{2}}}{(\hat{\sigma_1^2})^{\frac{n_0}{2}}\left(\frac{1}{N-n_0}\sum_{n=n_0}^{N-1} x^2[n]\right)^{\frac{N-n_0}{2}}}\left(\frac{1}{1-\hat{a}^2[1]}\right)^{\frac{N-n_0}{2}}.$$

Finally, the GLRT decides $\mathcal{H}_1$ if

$$L_G(\mathbf{x}) = \left(\frac{(\hat{r}_0[0])^{\frac{N}{2}}}{(\hat{r}_1[0])^{\frac{n_0}{2}}(\hat{r}_2[0])^{\frac{N-n_0}{2}}}\right)\left(\frac{1}{1-\hat{a}^2[1]}\right)^{\frac{N-n_0}{2}} > \gamma. \tag{12.22}$$

To interpret the results more intuitively, we have redefined the power estimates as autocorrelation estimates for lag zero. Thus, $\hat{r}_0[0] = \hat{\sigma_0^2}$ is the estimate of the process power assuming no change, while $\hat{r}_1[0] = (1/n_0)\sum_{n=0}^{n_0-1} x^2[n]$ and $\hat{r}_2[0] = (1/(N-n_0))\sum_{n=n_0}^{N-1} x^2[n]$ are the process power estimates before the change and after the change, respectively. It is seen then that the first term in $L_G(\mathbf{x})$ measures the change in process *power* while the second term measures the change in *shape of the PSD*. In either case a change will cause the factor in parentheses to be much larger than one (see also Problem 12.16).

As an example, consider a random process that changes from a white Gaussian process with variance $\sigma_u^2 = 5$ to an AR(1) process with parameters $a[1] = -0.9$, $\sigma_u^2 = 1$ at $n_0 = 50$. A realization of the process is shown in Figure 12.9. Note that after the change the process exhibits more correlation between samples and hence does not fluctuate as rapidly. Also, the power after the change increases slightly to $\sigma_u^2/(1-a^2[1]) = 5.3$. If we do not assume knowledge of the change time, then we must compute the GLRT for each possible n_0. In Figure 12.10 we have plotted $2\ln L_G(\mathbf{x})$ using (12.22) for each data record $[n_0, N-1]$ for $10 \le n_0 \le 90$ and $N = 100$. The realization of Figure 12.9 has been used. Note that the peak is at $\hat{n}_0 = 51$ and also that the estimated value of $a[1]$ is near the true value once n_0 exceeds the true change time as shown in Figure 12.11. This is because in computing $2\ln L_G(\mathbf{x})$ for $n_0 = 50$, we are using the data record $[n_0, N-1] = [50, 99]$ to estimate $a[1]$. From Figure 12.11 the value is about $\hat{a}[1] = -0.8$. Hence, not only can we detect the change but we also have access to the new parameters after the change.

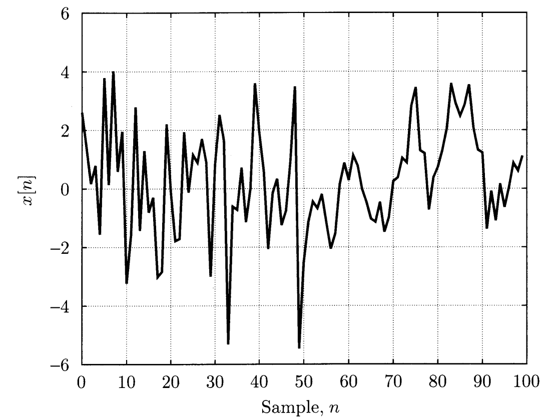


Figure 12.9. Random process realization whose PSD changes from white to colored.

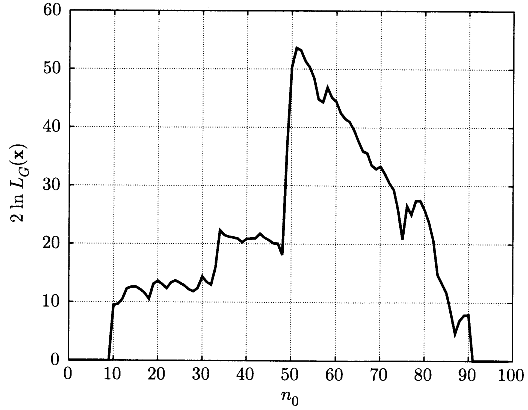


Figure 12.10. GLRT detection of PSD change time.

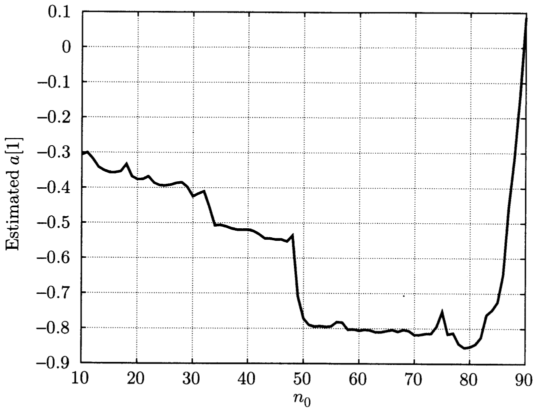


Figure 12.11. Estimated AR(1) filter parameter.

References

Basseville, M., I.V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*, Prentice-Hall, Englewood Cliffs, N.J. 1993.

Bellman, R., S. Dreyfus, *Applied Dynamic Programming*, Princeton Univ. Press, Princeton, N.J., 1962.

Birdsall, T.G., J.O. Gobien, "Sufficient Statistics and Reproducing Densities in Simultaneous Detection and Estimation," *IEEE Trans. on Info. Theory*, Vol. IT-19, pp. 760–768, 1973.

Kay, S.M., *Modern Spectral Estimation: Theory and Application*, Prentice-Hall, Englewood Cliffs, N.J., 1988.

Larson, R.E., J.L. Casti, *Principles of Dynamic Programming*, Vol. I,II, Marcel Dekker, New York, 1982.

Svendsen, F.K., F.K. Soong, "On the Automatic Segmentation of Speech Signals," 1987 Int. Conf. Acoust., Speech, and Signal Processing, Dallas, TX, pp. 77–80.

Problems

12.1 Let the test statistic of (12.5) be denoted by $T[N-1]$, indicating that it is based on the data record $[0, N-1]$. Then show that it can be evaluated sequentially in time as the data record length N increases by finding g so that $T[N] = T[N-1] + g(x[N], T[N-1])$. Note that

$$T[N] = \frac{1}{N - n_0 + 1} \sum_{n=n_0}^{N} (x[n] - A_0). \tag{12.23}$$

12.2 Evaluate the test statistic $T[N]$ of (12.23) for $5 \leq N \leq 14$ for the signal

$$s[n] = \begin{cases} 2 & n = 0, 1, \ldots, 4 \\ 3 & n = 5, 6, \ldots, 9 \\ 0 & n = 10, 11, \ldots, 14. \end{cases}$$

Note that the signal jumps in level at $n_0 = 5$ and *also* at $n_1 = 10$. Assume that $A_0 = 2$ and that there is no noise. What happens for $N > 9$?

12.3 For the problem described in Example 12.1, plot the delay time as a function of $|\Delta A|$. Assume that the noise variance is $\sigma^2 = 1$ and the desired operating conditions are $P_{FA} = 10^{-3}$ and $P_D = 0.99$.

12.4 For Example 12.1 show that the test is UMP with respect to ΔA as long as $\Delta A > 0$. In other words, show that the NP detector can be implemented (test statistic and threshold can be found) without knowledge of ΔA.

12.5 If in Example 12.3 the DC level before the jump is *known*, find the GLRT to detect a change in level.

12.6 A sinusoid $s[n] = \cos 2\pi f_0 n$ is observed in WGN with known variance σ^2 for $n = 0, 1, \ldots, N-1$. Find a test statistic to determine if the frequency at a known time n_0 changes from the known frequency f_0 to some unknown frequency. You may assume that N is large and that the frequencies are not near 0 or 1/2.

12.7 A DC level of known amplitude A is observed in IID noise with PDF $p(w[n])$ for $n = 0, 1, \ldots, N-1$. Find $L_G(\mathbf{x})$ for a jump in level at a known time n_0 and having an unknown size ΔA.

12.8 To verify the operation of a communications link, a known pilot signal $s[n]$ is inserted into the transmitted information. At the receiver the pilot signal is extracted and verified to be present so as to ensure that the link has "not gone down." If the link is operating correctly, the pilot signal is received undistorted. Propose a detector to determine when the pilot signal is absent, assuming that $\mathcal{H}_1$ is the hypothesis that the link has "gone down." Use any reasonable assumptions about the noise.

12.9 Find n_0 so that the detection performance is maximized in Example 12.3 and explain your results.

12.10 Verify that for unknown DC levels and an unknown jump time the GLRT produces (12.11). Hint: Use results of Example 12.3 as well as those of Problem 7.21.

12.11 A WGN process may exhibit a jump in its variance at a known time n_0. If the variance σ_0^2 before the jump is known but the jump size $\Delta\sigma^2$ is unknown, show that the GLRT for this problem is to decide $\mathcal{H}_1$ if

$$2\ln L_G(\mathbf{x}) = (N - n_0)\left[\frac{\sigma_0^2 + \widehat{\Delta\sigma^2}}{\sigma_0^2} - \ln\left(\frac{\sigma_0^2 + \widehat{\Delta\sigma^2}}{\sigma_0^2}\right) - 1\right] > \gamma'$$

where

$$\widehat{\Delta\sigma^2} = \frac{1}{N - n_0}\sum_{n=n_0}^{N-1} x^2[n] - \sigma_0^2.$$

Next, since $x - \ln x - 1$ is monotonic with x for $x > 1$, simplify the test statistic. Assume that $\widehat{\Delta\sigma^2} > 0$.

12.12 Repeat Problem 12.11, but now assume that under $\mathcal{H}_1$ the variance before the jump or σ_1^2 and the variance after the jump or σ_2^2 are both unknown. Also, assume that under $\mathcal{H}_0$ the variance σ^2 is unknown as well. You should be able to show that the GLRT decides $\mathcal{H}_1$ if

$$2\ln L_G(\mathbf{x}) = N\ln\left(\frac{\hat{\sigma^2}}{(\hat{\sigma_1^2})^{\frac{n_0}{N}}(\hat{\sigma_2^2})^{\frac{N-n_0}{N}}}\right) > \gamma'$$

where $\hat{\sigma^2}$ is the variance estimator based on the entire data record, $\hat{\sigma_1^2}$ is that based on the data over the interval $[0, n_0 - 1]$, and $\hat{\sigma_2^2}$ is that based on the data over the interval $[n_0, N-1]$. Simplify and explain your results if $n_0 = N/2$ for N even.

12.13 We consider the following multiple change problem

$$\begin{aligned}
\mathcal{H}_0 : x[n] &= \quad w[n] \qquad\qquad n = 0, 1, \ldots, N-1 \\
\mathcal{H}_1 : x[n] &= \begin{cases} A_0 + w[n] & n = 0, 1, \ldots, n_0 - 1 \\ A_1 + w[n] & n = n_0, n_0 + 1, \ldots, n_1 - 1 \\ A_2 + w[n] & n = n_1, n_1 + 1 \ldots, N-1 \end{cases}
\end{aligned}$$

where A_0, A_1, A_2 and n_0, n_1, n_2 are unknown and $w[n]$ is WGN with *unknown* variance σ^2. Show that the GLRT may be written as

$$L_G(\mathbf{x}) = \max_{n_0, n_1, n_2}\left(\frac{\hat{\sigma_0^2}}{\hat{\sigma_1^2}}\right)^{\frac{N}{2}}$$

where

$$\begin{aligned}
\hat{\sigma_0^2} &= \frac{1}{N}\sum_{n=0}^{N-1} x^2[n] \\
\hat{\sigma_1^2} &= \frac{1}{N}\left(\sum_{n=0}^{n_0-1}(x[n] - \hat{A}_0)^2 + \sum_{n=n_0}^{n_1-1}(x[n] - \hat{A}_1)^2 + \sum_{n=n_1}^{N-1}(x[n] - \hat{A}_2)^2\right)
\end{aligned}$$

and

$$\begin{aligned}
\hat{A}_0 &= \frac{1}{n_0}\sum_{n=0}^{n_0-1} x[n] \\
\hat{A}_1 &= \frac{1}{n_1 - n_0}\sum_{n=n_0}^{n_1-1} x[n] \\
\hat{A}_2 &= \frac{1}{N - n_1}\sum_{n=n_1}^{N-1} x[n].
\end{aligned}$$

12.14 Explain how the DP algorithm in Example 12.5 would have to be modified if the DC levels for each segment were replaced by straight lines or $A + Bn$, where A and B were unknown.

12.15 Verify (12.19) and (12.20) by maximizing $p(\mathbf{x}_2; a[1], \sigma_2^2)$ given in (12.18).

12.16 Show that the first factor is (12.22) is greater than one for $\hat{r}_1[0] \neq \hat{r}_2[0]$. The second factor is also greater than one since for large N, $|\hat{a}[1]| < 1$. Hint: First show that $\hat{r}_0[0] = (n_0/N)\hat{r}_1[0] + ((N - n_0)/N)\hat{r}_2[0]$. Then use the inequality

$$\frac{\alpha a + (1-\alpha)b}{a^{\alpha}b^{1-\alpha}} \geq 1$$

for $a \geq 0$, $b \geq 0$ and $0 < \alpha < 1$. Equality holds if and only if $a = b$. For $\alpha = 1/2$ this is the arithmetic mean divided by the geometric mean.

Appendix 12A

General Dynamic Programming Approach to Segmentation

We wish to segment a time series into N_s segments by choosing the set of change times $\{n_0, n_1, \ldots, n_{N_s-2}\}$, an example of which is shown in Figure 12.12. To do so we assume that the ith segment for $i = 0, 1, \ldots, N_s - 1$ is characterized by the PDF $p_i(x[n_{i-1}], x[n_{i-1}+1], \ldots, x[n_i - 1]; \boldsymbol{\theta}_i)$, where $\boldsymbol{\theta}_i$ is a vector of unknown parameters. Furthermore, each segment is assumed to be statistically independent of all the other segments. With these assumptions the PDF of the data set can be written as

$$\prod_{i=0}^{N_s-1} p_i(x[n_{i-1}], x[n_{i-1}+1], \ldots, x[n_i - 1]; \boldsymbol{\theta}_i) \tag{12A.1}$$

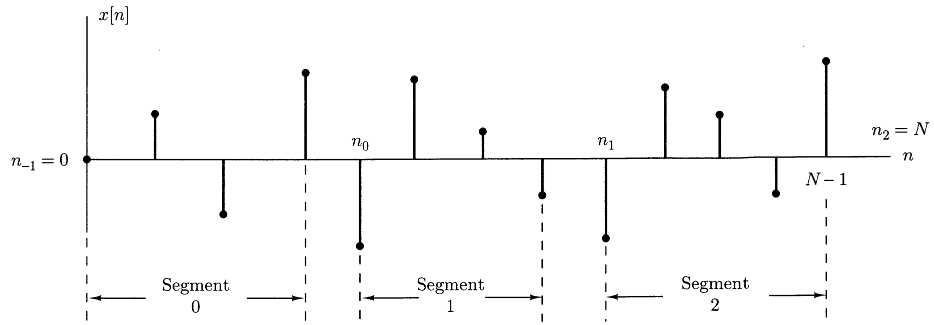


Figure 12.12. Definitions of segments and change times ($N_s = 3$).

where we have defined $n_{-1} = 0$ and $n_{N_s-1} = N$. The MLE segmenter chooses $\{n_0, n_1, \ldots, n_{N_s-2}, \boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{N_s-1}\}$ as those values that maximize (12A.1) or by defining

$$\mathbf{x}[i,j] = [x[i]\, x[i+1] \ldots x[j]]^T$$

the MLE segmenter must maximize

$$\sum_{i=0}^{N_s-1} \ln p_i(\mathbf{x}[n_{i-1}, n_i - 1]; \boldsymbol{\theta}_i).$$

To apply DP let $\hat{\boldsymbol{\theta}}_i$ be the MLE of $\boldsymbol{\theta}_i$ based on the ith data segment or $\mathbf{x}[n_{i-1}, n_i - 1] = [x[n_{i-1}]\, x[n_{i-1}+1] \ldots x[n_i - 1]]^T$. Defining

$$\Delta_i[n_{i-1}, n_i - 1] = -\ln p_i(\mathbf{x}[n_{i-1}, n_i - 1]; \hat{\boldsymbol{\theta}}_i)$$

we wish to minimize

$$\sum_{i=0}^{N_s-1} \Delta_i[n_{i-1}, n_i - 1].$$

We can now apply the results of Section 12.5 to generate the recursion

$$I_k[L] = \min_{n_{k-1}} \left(I_{k-1}[n_{k-1} - 1] + \Delta_k[n_{k-1}, L]\right).$$

If we include the constraint that each segment must be at least one sample in length, then we have that

$$I_k[L] = \min_{k \le n_{k-1} \le L} \left(I_{k-1}[n_{k-1} - 1] + \Delta_k[n_{k-1}, L]\right).$$

The solution to the original problem is obtained for $k = N_s - 1$ and $L = N - 1$. To begin the recursion we need to compute

$$\begin{aligned} I_0[L] &= \Delta_0[n_{-1}, L] \\ &= -\ln p_0(\mathbf{x}[n_{-1}, L]; \hat{\boldsymbol{\theta}}_0) \\ &= -\ln p_0(\mathbf{x}[0, L]; \hat{\boldsymbol{\theta}}_0) \end{aligned}$$

where $\hat{\boldsymbol{\theta}}_0$ is the MLE based on $\mathbf{x}[0, L]$ and $L = 0, 1, \ldots, N - N_s$.

Appendix 12B

MATLAB Program for Dynamic Programming

dp.m

```
% dp.m
%
% This program implements the dynamic programming algorithm
% for determining the three change times of a signal that
% consists of four unknown DC levels in WGN (see Figure 12.4).
% The DC levels and change times are unknown.
%
  clear all
% Generate data
  randn('seed',0)
  A=[1;4;2;6];varw=1;sig=sqrt(varw);
  x=[sig*randn(20,1)+A(1); sig*randn(30,1)+A(2);...
     sig*randn(15,1)+A(3);sig*randn(35,1)+A(4)];
  N=length(x);
% Begin DP algorithm
% Since MATLAB cannot accommodate matrix/vector indices of zero,
% augment L,k,n by one when necessary.
% Initialize DP algorithm
  for L=0:N-4
  LL=L+1;
  I(1,LL)=(x(1:LL)-mean(x(1:LL)))'*(x(1:LL)-mean(x(1:LL)));
  end
% Begin DP recursions
  for k=1:3
    kk=k+1;
    if k<3
```

```
      for L=k:N-4+k
      LL=L+1;
%  Load in large number to prevent minimizing value of J
%  to occur for a value of J(1:k), which is not computed
      J(1:k)=10000*ones(k,1);
%  Compute least squares error for all possible change times
        for n=k:L
        nn=n+1;
        Del=(x(nn:LL)-mean(x(nn:LL)))'*(x(nn:LL)-mean(x(nn:LL)));
        J(nn)=I(kk-1,nn-1)+Del;
        end
%  Determine minimum of least squares error and change time that
%  yields the minimum
      [I(kk,LL),ntrans(L,k)]=min(J(1:LL));
      end
    else
%  Final stage computation
    L=N-1;LL=L+1;J(1:k)=10000*ones(k,1);
      for n=k:N-1
      nn=n+1;
      Del=(x(nn:LL)-mean(x(nn:LL)))'*(x(nn:LL)-mean(x(nn:LL)));
      J(nn)=I(kk-1,nn-1)+Del;
      end
    [Imin,ntrans(N-1,k)]=min(J(1:N));
    end
  end
%  Determine change times that minimize least squares error
  n2est=ntrans(N-1,3);
  n1est=ntrans(n2est-1,2);
  n0est=ntrans(n1est-1,1);
%  Reference change times to [0,N-1] interval instead of
%  MATLAB's [1,N]
  n0est=n0est-1
  n1est=n1est-1
  n2est=n2est-1
```

Chapter 13

Complex/Vector Extensions, and Array Processing

13.1 Introduction

In many practical problems of interest the received data samples are actually vectors. Examples are in applications employing multiple sensors, such as radar, sonar, communications, infrared imaging, and biomedical signal processing. The detector structures derived in previous chapters can be extended to process these data sample vectors. The resultant approaches are commonly referred to as *array processing* and are described in this chapter. Furthermore, because many of these applications involve bandpass signals, it becomes convenient to consider the complex envelope of the signal. As such, we will generalize some of our previous results to the complex data case. Much of the algebraic manipulations involve the complex extension of the multivariate Gaussian PDF, which has already been described in [Kay-I 1993, Chapter 15]. The reader may wish to review that chapter before continuing with the material of this chapter.

13.2 Summary

The Neyman-Pearson (NP) detectors for complex data are derived in Section 13.3. The matched filter or replica-correlator is given by (13.3) and its detection performance by (13.6) and (13.7). The generalized matched filter is given by (13.10) and its detection performance by (13.11) and (13.12). Lastly, the estimator-correlator detector is summarized by (13.14) and (13.15). Its detection performance is described in Problem 13.8. When the signal is deterministic with unknown parameters and can be modeled by the complex classical linear model, the GLRT detector is given by (13.19) and its exact detection performance by (13.20). On the other hand, if the signal obeys the Bayesian linear model (the unknown parameters are random), then Section 13.4.2 describes the Neyman-Pearson detector. The extension of the usual detectors to complex vector observations is discussed in Section 13.5 with the detector structures summarized in Section 13.6. In particular, for a known deterministic signal, the detectors for various covariance assumptions for the noise are given by (13.26), (13.27), (13.29), and (13.30). If the signal is random, then Section 13.6.5 describes various detectors. In the case of a deterministic signal with unknown parameters and a classical linear model assumption, the GLRT is given by either (13.34) or (13.35). The corresponding exact detection performance is summarized in (13.36) and (13.37). If the data record is large and the signal is a WSS multichannel Gaussian random process, then an approximate estimator-correlator is given by (13.43). Finally, the application of the results of this chapter is made to active sonar/radar and broadband passive sonar in Section 13.8. The detectors are given by (13.50) for active sonar/radar and by (13.57) and (13.56) for broadband passive sonar.

13.3 Known PDFs

We now consider the complex data extensions of the matched filter (or equivalently the replica-correlator), generalized matched filter, and estimator-correlator detectors. In each case we assume that our goal is to maximize P_D, subject to a constraint on P_{FA}, and hence, we employ the NP criterion. As a result, the LRT will be the optimal detector.

13.3.1 Matched Filter

We begin by considering the detection of a known complex deterministic signal in complex white Gaussian noise (CWGN). The hypothesis testing problem is

$$\begin{array}{ll} \mathcal{H}_0 : \tilde{x}[n] = \tilde{w}[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : \tilde{x}[n] = \tilde{s}[n] + \tilde{w}[n] & n = 0, 1, \ldots, N-1 \end{array} \qquad (13.1)$$

where $\tilde{s}[n]$ is a known complex signal and $\tilde{w}[n]$ is CWGN with variance σ^2 or $\tilde{w}[n] \sim \mathcal{CN}(0, \sigma^2)$, and all samples are uncorrelated and hence independent. We will use a "tilde" (for example, $\tilde{x}[n]$) to emphasize that a quantity is complex when it is necessary to avoid confusion with its real counterpart (for example, $x[n]$). The LRT decides $\mathcal{H}_1$ if

$$L(\tilde{\mathbf{x}}) = \frac{p(\tilde{\mathbf{x}}; \mathcal{H}_1)}{p(\tilde{\mathbf{x}}; \mathcal{H}_0)} > \gamma$$

where $\tilde{\mathbf{x}} = [\tilde{x}[0]\, \tilde{x}[1] \ldots \tilde{x}[N-1]]^T$. But

$$\begin{aligned} p(\tilde{\mathbf{x}}; \mathcal{H}_1) &= \frac{1}{\pi^N \sigma^{2N}} \exp\left[-\frac{1}{\sigma^2}(\tilde{\mathbf{x}} - \tilde{\mathbf{s}})^H (\tilde{\mathbf{x}} - \tilde{\mathbf{s}})\right] \\ p(\tilde{\mathbf{x}}; \mathcal{H}_0) &= \frac{1}{\pi^N \sigma^{2N}} \exp\left[-\frac{1}{\sigma^2}\tilde{\mathbf{x}}^H \tilde{\mathbf{x}}\right] \end{aligned}$$

where H denotes the complex conjugate transpose and $\tilde{\mathbf{s}} = [\tilde{s}[0]\, \tilde{s}[1] \ldots \tilde{s}[N-1]]^T$. Thus,

$$\begin{aligned} \ln L(\tilde{\mathbf{x}}) &= -\frac{1}{\sigma^2}\left[(\tilde{\mathbf{x}} - \tilde{\mathbf{s}})^H (\tilde{\mathbf{x}} - \tilde{\mathbf{s}}) - \tilde{\mathbf{x}}^H \tilde{\mathbf{x}}\right] \\ &= -\frac{1}{\sigma^2}\left[-\tilde{\mathbf{x}}^H \tilde{\mathbf{s}} - \tilde{\mathbf{s}}^H \tilde{\mathbf{x}} + \tilde{\mathbf{s}}^H \tilde{\mathbf{s}}\right] \\ &= \frac{2}{\sigma^2}\mathrm{Re}(\tilde{\mathbf{s}}^H \tilde{\mathbf{x}}) - \frac{1}{\sigma^2}\tilde{\mathbf{s}}^H \tilde{\mathbf{s}}. \end{aligned}$$

Since $\tilde{\mathbf{s}}$ is known, we decide $\mathcal{H}_1$ if

$$T(\tilde{\mathbf{x}}) = \mathrm{Re}(\tilde{\mathbf{s}}^H \tilde{\mathbf{x}}) > \gamma' \qquad (13.2)$$

or equivalently

$$T(\tilde{\mathbf{x}}) = \mathrm{Re}\left(\sum_{n=0}^{N-1} \tilde{x}[n]\tilde{s}^*[n]\right) > \gamma' \qquad (13.3)$$

as shown in Figure 13.1. This is the complex version of a replica-correlator. It can also be expressed in matched filter form (see Problem 13.2). The detection performance is easily found, since the sum of independent complex Gaussian random variables is a complex Gaussian random variable whose real part is a real Gaussian random variable. Let $\tilde{z} = \sum_{n=0}^{N-1} \tilde{x}[n]\tilde{s}^*[n]$ and note that it is a complex Gaussian random variable with moments

$$\begin{aligned} E(\tilde{z}; \mathcal{H}_0) &= \sum_{n=0}^{N-1} E(\tilde{x}[n])\tilde{s}^*[n] = 0 \\ E(\tilde{z}; \mathcal{H}_1) &= \sum_{n=0}^{N-1} E(\tilde{x}[n])\tilde{s}^*[n] = \sum_{n=0}^{N-1} |\tilde{s}[n]|^2 \\ \mathrm{var}(\tilde{z}; \mathcal{H}_0) &= \mathrm{var}\left(\sum_{n=0}^{N-1} \tilde{x}[n]\tilde{s}^*[n]\right) \end{aligned}$$

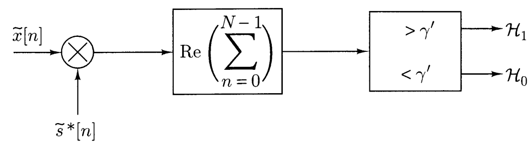


Figure 13.1. Replica-correlator for complex data.

$$
\begin{aligned}
&= \sum_{n=0}^{N-1} \operatorname{var}(\tilde{x}[n])|\tilde{s}[n]|^2 \\
&= \sigma^2 \sum_{n=0}^{N-1} |\tilde{s}[n]|^2
\end{aligned}
$$

since $\operatorname{var}(\tilde{a}\tilde{z}) = |\tilde{a}|^2 \operatorname{var}(\tilde{z})$ and the $\tilde{x}[n]$'s are uncorrelated. Similarly, under $\mathcal{H}_1$ we have the same variance. Thus,

$$
\tilde{z} \sim \begin{cases} \mathcal{CN}(0, \sigma^2\mathcal{E}) & \text{under } \mathcal{H}_0 \\ \mathcal{CN}(\mathcal{E}, \sigma^2\mathcal{E}) & \text{under } \mathcal{H}_1 \end{cases} \tag{13.4}
$$

where $\mathcal{E} = \sum_{n=0}^{N-1} |\tilde{s}[n]|^2$ is the signal energy. Since the real and imaginary parts of a complex Gaussian random variable are real Gaussian random variables, independent of one other, and having the same variance (one-half the total variance of the complex Gaussian random variable), we have

$$
T(\tilde{\mathbf{x}}) = \operatorname{Re}(\tilde{z}) \sim \begin{cases} \mathcal{N}(0, \sigma^2\mathcal{E}/2) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(\mathcal{E}, \sigma^2\mathcal{E}/2) & \text{under } \mathcal{H}_1 \end{cases} \tag{13.5}
$$

where we have noted that $\mathcal{E}$ is real. This is just the mean-shifted Gauss-Gauss problem so that it immediately follows from Chapter 4 that

$$
P_{FA} = Q\left(\frac{\gamma'}{\sqrt{\sigma^2\mathcal{E}/2}}\right) \tag{13.6}
$$

$$
P_D = Q\left(\frac{\gamma' - \mathcal{E}}{\sqrt{\sigma^2\mathcal{E}/2}}\right) \tag{13.7}
$$

or by eliminating the threshold γ' in P_D via $\gamma' = \sqrt{\sigma^2\mathcal{E}/2}\, Q^{-1}(P_{FA})$, we have finally that

$$
P_D = Q\left(Q^{-1}(P_{FA}) - \sqrt{d^2}\right) \tag{13.8}
$$

where the deflection coefficient is

$$
d^2 = \frac{2\mathcal{E}}{\sigma^2}. \tag{13.9}
$$

Note that d^2 is twice the value obtained for the real case. This may be attributed to the fact that the mean of $\tilde{z}$ under $\mathcal{H}_1$ is real (see (13.4)). Hence, the test statistic retains only the real part of $\tilde{z}$, causing the variance to be *halved*. An example follows.

Example 13.1 - Complex DC Level in CWGN

If $\tilde{s}[n] = \tilde{A}$, where $\tilde{A}$ is complex and known, then from (13.3) the replica-correlator decides $\mathcal{H}_1$ if

$$
T(\tilde{\mathbf{x}}) = \operatorname{Re}\left(\sum_{n=0}^{N-1} \tilde{x}[n]\tilde{A}^*\right) > \gamma'.
$$

If we let $\tilde{x}[n] = u[n] + jv[n]$ and $\tilde{A} = A_R + jA_I$, then this becomes

$$
\begin{aligned}
T(\tilde{\mathbf{x}}) &= \operatorname{Re}\left(\sum_{n=0}^{N-1}(u[n] + jv[n])(A_R - jA_I)\right) \\
&= A_R \sum_{n=0}^{N-1} u[n] + A_I \sum_{n=0}^{N-1} v[n]
\end{aligned}
$$

or

$$
T'(\tilde{\mathbf{x}}) = \frac{1}{N} T(\tilde{\mathbf{x}}) = A_R\bar{u} + A_I\bar{v}
$$

where $\bar{u}, \bar{v}$ are the sample means of $u[n]$ and $v[n]$, respectively. If we now let $\bar{\tilde{x}} = \bar{u} + j\bar{v}$, we have

$$
T'(\tilde{\mathbf{x}}) = \operatorname{Re}(\tilde{A}^*\bar{\tilde{x}}).
$$

No further simplification is possible unless $\tilde{A}$ is known to be real, in which case, $T'(\tilde{\mathbf{x}}) = A_R\bar{u}$. This example serves to illustrate that the NP detector for complex data and parameters does not always mimic the real case. The complex case is more closely aligned with the 2×1 vector case, consistent with the isomorphism that exists between complex numbers and 2×1 vectors [Kay-I 1993, Problem 15.5].

The detection performance is given by (13.8) and (13.9) with

$$
\begin{aligned}
d^2 &= \frac{2\sum_{n=0}^{N-1} |\tilde{s}[n]|^2}{\sigma^2} \\
&= \frac{2N|\tilde{A}|^2}{\sigma^2}.
\end{aligned}
$$

Note that the performance depends only on the magnitude of $\tilde{A}$ and not its phase. This is due to the circularly symmetric PDF for a $\mathcal{CN}(0, \sigma^2)$ random variable, i.e., no discrimination is possible in angle. See also Problem 13.4.

◇

13.3.2 Generalized Matched Filter

Next consider the hypothesis testing problem of (13.1) but with $\tilde{\mathbf{w}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$, where $\mathbf{C}$ is not necessarily $\sigma^2\mathbf{I}$. This is the problem of detection of a known deterministic signal in correlated Gaussian noise. Then the LRT decides $\mathcal{H}_1$ if

$$L(\tilde{\mathbf{x}}) = \frac{p(\tilde{\mathbf{x}}; \mathcal{H}_1)}{p(\tilde{\mathbf{x}}; \mathcal{H}_0)} > \gamma$$

where

$$\begin{aligned} p(\tilde{\mathbf{x}}; \mathcal{H}_1) &= \frac{1}{\pi^N \det(\mathbf{C})} \exp\left[-(\tilde{\mathbf{x}} - \tilde{\mathbf{s}})^H \mathbf{C}^{-1}(\tilde{\mathbf{x}} - \tilde{\mathbf{s}})\right] \\ p(\tilde{\mathbf{x}}; \mathcal{H}_0) &= \frac{1}{\pi^N \det(\mathbf{C})} \exp\left[-\tilde{\mathbf{x}}^H \mathbf{C}^{-1}\tilde{\mathbf{x}}\right]. \end{aligned}$$

One can easily show that this reduces to deciding $\mathcal{H}_1$ if

$$T(\tilde{\mathbf{x}}) = \mathrm{Re}(\tilde{\mathbf{s}}^H \mathbf{C}^{-1}\tilde{\mathbf{x}}) > \gamma' \tag{13.10}$$

as shown in Figure 13.2. To determine the detection performance we first note that $\tilde{z} = \tilde{\mathbf{s}}^H \mathbf{C}^{-1}\tilde{\mathbf{x}}$ is a complex Gaussian random variable since it is a linear tranformation of the complex Gaussian random vector $\tilde{\mathbf{x}}$. Then, it can be shown that

$$\begin{aligned} E(\tilde{z}; \mathcal{H}_0) &= 0 \\ E(\tilde{z}; \mathcal{H}_1) &= \tilde{\mathbf{s}}^H \mathbf{C}^{-1}\tilde{\mathbf{s}} \\ \mathrm{var}(\tilde{z}; \mathcal{H}_0) &= \mathrm{var}(\tilde{z}; \mathcal{H}_1) = \tilde{\mathbf{s}}^H \mathbf{C}^{-1}\tilde{\mathbf{s}} \end{aligned}$$

so that

$$T(\tilde{\mathbf{x}}) \sim \begin{cases} N(0, \tilde{\mathbf{s}}^H \mathbf{C}^{-1}\tilde{\mathbf{s}}/2) & \text{under } \mathcal{H}_0 \\ N(\tilde{\mathbf{s}}^H \mathbf{C}^{-1}\tilde{\mathbf{s}}, \tilde{\mathbf{s}}^H \mathbf{C}^{-1}\tilde{\mathbf{s}}/2) & \text{under } \mathcal{H}_1 \end{cases}$$

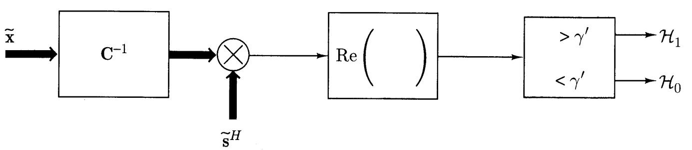


Figure 13.2. Generalized matched filter for complex data.

and therefore

$$P_{FA} = Q\left(\frac{\gamma'}{\sqrt{\tilde{\mathbf{s}}^H \mathbf{C}^{-1}\tilde{\mathbf{s}}/2}}\right) \tag{13.11}$$

$$P_D = Q\left(\frac{\gamma' - \tilde{\mathbf{s}}^H \mathbf{C}^{-1}\tilde{\mathbf{s}}}{\sqrt{\tilde{\mathbf{s}}^H \mathbf{C}^{-1}\tilde{\mathbf{s}}/2}}\right) \tag{13.12}$$

or using $\gamma' = \sqrt{\tilde{\mathbf{s}}^H \mathbf{C}^{-1}\tilde{\mathbf{s}}/2}\, Q^{-1}(P_{FA})$ we have that

$$P_D = Q\left(Q^{-1}(P_{FA}) - \sqrt{d^2}\right)$$

where the deflection coefficient is

$$d^2 = 2\tilde{\mathbf{s}}^H \mathbf{C}^{-1}\tilde{\mathbf{s}}.$$

Note that $\tilde{\mathbf{s}}^H \mathbf{C}^{-1}\tilde{\mathbf{s}}$ is real. The reader is asked to verify these results in Problem 13.5. If $\mathbf{C} = \sigma^2\mathbf{I}$, then the detector and its performance reduces to the previous case. As in Chapter 4, an equivalent form for the detector follows by factoring $\mathbf{C}^{-1}$ as $\mathbf{D}^H\mathbf{D}$. Then, from (13.10)

$$\begin{aligned} T(\tilde{\mathbf{x}}) &= \mathrm{Re}(\tilde{\mathbf{s}}^H \mathbf{D}^H \mathbf{D}\tilde{\mathbf{x}}) \\ &= \mathrm{Re}(\tilde{\mathbf{s}}'^H \tilde{\mathbf{x}}') \end{aligned}$$

where $\tilde{\mathbf{s}}' = \mathbf{D}\tilde{\mathbf{s}}$ and $\tilde{\mathbf{x}}' = \mathbf{D}\tilde{\mathbf{x}}$ is a whitened version of $\tilde{\mathbf{x}}$, since $\mathbf{C}_{\tilde{x}'} = \mathbf{I}$.

13.3.3 Estimator-Correlator

Consider the detection problem

$$\begin{aligned} \mathcal{H}_0 &: \tilde{x}[n] = \tilde{w}[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 &: \tilde{x}[n] = \tilde{s}[n] + \tilde{w}[n] & n = 0, 1, \ldots, N-1 \end{aligned}$$

where $\tilde{s}[n]$ is a complex Gaussian random process with zero mean and covariance matrix $\mathbf{C}_{\tilde{s}}$ and $\tilde{w}[n]$ is CWGN with variance σ^2. This is the problem of detecting a random signal in CWGN. The LRT decides $\mathcal{H}_1$ if

$$L(\tilde{\mathbf{x}}) = \frac{p(\tilde{\mathbf{x}}; \mathcal{H}_1)}{p(\tilde{\mathbf{x}}; \mathcal{H}_0)} > \gamma$$

where the PDFs are given by

$$\begin{aligned} p(\tilde{\mathbf{x}}; \mathcal{H}_1) &= \frac{1}{\pi^N \det(\mathbf{C}_{\tilde{s}} + \sigma^2\mathbf{I})} \exp\left[-\tilde{\mathbf{x}}^H (\mathbf{C}_{\tilde{s}} + \sigma^2\mathbf{I})^{-1}\tilde{\mathbf{x}}\right] \\ p(\tilde{\mathbf{x}}; \mathcal{H}_0) &= \frac{1}{\pi^N \sigma^{2N}} \exp\left[-\frac{1}{\sigma^2}\tilde{\mathbf{x}}^H \tilde{\mathbf{x}}\right]. \end{aligned}$$

The log-likelihood ratio is

$$\ln L(\tilde{\mathbf{x}}) = -\tilde{\mathbf{x}}^H\left[(\mathbf{C}_{\tilde{s}}+\sigma^2\mathbf{I})^{-1}-\frac{1}{\sigma^2}\mathbf{I}\right]\tilde{\mathbf{x}} - \ln\det(\mathbf{C}_{\tilde{s}}+\sigma^2\mathbf{I}) + \ln\det\sigma^{2N}. \quad (13.13)$$

Using the matrix inversion lemma (see Section 5.3)

$$(\mathbf{C}_{\tilde{s}}+\sigma^2\mathbf{I})^{-1} = \frac{1}{\sigma^2}\mathbf{I} - \frac{1}{\sigma^4}\left(\frac{1}{\sigma^2}\mathbf{I}+\mathbf{C}_{\tilde{s}}^{-1}\right)^{-1}$$

we have after discarding the non-data-dependent terms in (13.13) and scaling by σ^2

$$T(\tilde{\mathbf{x}}) = \tilde{\mathbf{x}}^H\frac{1}{\sigma^2}\left(\frac{1}{\sigma^2}\mathbf{I}+\mathbf{C}_{\tilde{s}}^{-1}\right)^{-1}\tilde{\mathbf{x}} = \tilde{\mathbf{x}}^H\hat{\tilde{\mathbf{s}}}$$

where

$$\begin{aligned}\hat{\tilde{\mathbf{s}}} &= \frac{1}{\sigma^2}\left(\frac{1}{\sigma^2}\mathbf{I}+\mathbf{C}_{\tilde{s}}^{-1}\right)^{-1}\tilde{\mathbf{x}}\\ &= \frac{1}{\sigma^2}\left[\frac{1}{\sigma^2}(\mathbf{C}_{\tilde{s}}+\sigma^2\mathbf{I})\mathbf{C}_{\tilde{s}}^{-1}\right]^{-1}\tilde{\mathbf{x}}\\ &= \mathbf{C}_{\tilde{s}}(\mathbf{C}_{\tilde{s}}+\sigma^2\mathbf{I})^{-1}\tilde{\mathbf{x}}.\end{aligned}$$

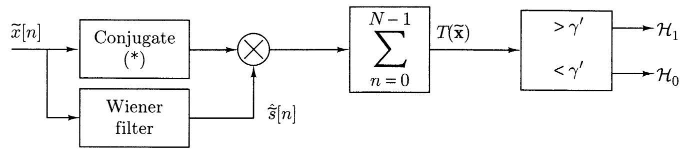
Thus, we decide $\mathcal{H}_1$ if

$$T(\tilde{\mathbf{x}}) = \tilde{\mathbf{x}}^H\hat{\tilde{\mathbf{s}}} > \gamma' \quad (13.14)$$

where

$$\hat{\tilde{\mathbf{s}}} = \mathbf{C}_{\tilde{s}}(\mathbf{C}_{\tilde{s}}+\sigma^2\mathbf{I})^{-1}\tilde{\mathbf{x}} \quad (13.15)$$

is the complex MMSE estimator of $\tilde{\mathbf{s}}$ (see [Kay-I 1993, pp. 532–535]). Note that the test statistic $T(\tilde{\mathbf{x}})$ is real since it is a Hermitian form $\tilde{\mathbf{x}}^H\mathbf{A}\tilde{\mathbf{x}}$ where $\mathbf{A}^H=\mathbf{A}$. This ensures that $\tilde{\mathbf{x}}^H\mathbf{A}\tilde{\mathbf{x}}$ is real since $(\tilde{\mathbf{x}}^H\mathbf{A}\tilde{\mathbf{x}})^H = \tilde{\mathbf{x}}^H\mathbf{A}^H\tilde{\mathbf{x}} = \tilde{\mathbf{x}}^H\mathbf{A}\tilde{\mathbf{x}}$. To verify that $\mathbf{A}=\mathbf{C}_{\tilde{s}}(\mathbf{C}_{\tilde{s}}+\sigma^2\mathbf{I})^{-1}$ is Hermitian we use the results that $(\mathbf{AB})^H=\mathbf{B}^H\mathbf{A}^H$, the sum of two Hermitian matrices is a Hermitian matrix, and the inverse of a Hermitian matrix is a Hermitian matrix (see Problem 13.7). The detector is shown in Figure 13.3. The detection performance of the estimator-correlator can be found analytically by first transforming to a canonical detector as in Section 5.3. Then, in the complex case the test statistic is a weighted sum of independent χ_2^2 random variables, where the weights depend on the eigenvalues of $\mathbf{C}_{\tilde{s}}$ and σ^2. The general case is explored in Problem 13.8, while the simple example of a rank one signal covariance matrix is given next.



$\hat{\tilde{\mathbf{s}}} = \mathbf{C}_{\tilde{s}}(\mathbf{C}_{\tilde{s}}+\sigma^2\mathbf{I})^{-1}\tilde{\mathbf{x}}$

Figure 13.3. Estimator-correlator for complex data.

Example 13.2 - Rank One Signal Covariance Matrix

A standard model for the complex envelope of an active sonar or radar return from a *nonfluctuating point target* is [Van Trees 1971]

$$\tilde{s}[n] = \tilde{A}\tilde{h}[n]$$

where $\tilde{h}[n]$ is a known complex deterministic signal and $\tilde{A}$ is a complex Gaussian random variable with $\tilde{A}\sim\mathcal{CN}(0,\sigma_A^2)$, which is assumed independent of the observation noise $\tilde{w}[n]$. The transmitted signal is $\tilde{h}[n]$, which after reflection from a target, results in a received signal that is modified in gain and phase. The unknown gain and phase are modeled via $\tilde{A}$. Note that this model is a special case of the complex Bayesian linear model (see [Kay-I 1993, Example 15.11]). With these assumptions the received signal is a complex Gaussian random process with zero mean and covariance matrix

$$\begin{aligned}[\mathbf{C}_{\tilde{s}}]_{mn} &= E(\tilde{s}[m]\tilde{s}^*[n]) = E(\tilde{A}\tilde{h}[m]\tilde{A}^*\tilde{h}^*[n])\\ &= E(|\tilde{A}|^2)\tilde{h}[m]\tilde{h}^*[n] = \sigma_A^2\tilde{h}[m]\tilde{h}^*[n]\end{aligned}$$

or

$$\mathbf{C}_{\tilde{s}} = \sigma_A^2\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H$$

where $\tilde{\mathbf{h}} = [\tilde{h}[0]\,\tilde{h}[1]\ldots\tilde{h}[N-1]]^T$. The signal covariance matrix is seen to be rank one, which allows the use of Woodbury's identity in determining $\hat{\tilde{\mathbf{s}}}$. Thus, from (13.15)

$$\begin{aligned}\hat{\tilde{\mathbf{s}}} &= \mathbf{C}_{\tilde{s}}(\mathbf{C}_{\tilde{s}}+\sigma^2\mathbf{I})^{-1}\tilde{\mathbf{x}}\\ &= \sigma_A^2\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H(\sigma_A^2\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H+\sigma^2\mathbf{I})^{-1}\tilde{\mathbf{x}}.\end{aligned}$$

Using Woodbury's identity, which for complex matrices is

$$(\mathbf{A}+\mathbf{u}\mathbf{u}^H)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{u}^H\mathbf{A}^{-1}}{1+\mathbf{u}^H\mathbf{A}^{-1}\mathbf{u}} \quad (13.16)$$

where $\mathbf{u}$ is a column vector, we have

$$\begin{aligned}\hat{\tilde{\mathbf{s}}} &= \sigma_A^2 \tilde{\mathbf{h}}\tilde{\mathbf{h}}^H \left(\frac{1}{\sigma^2}\mathbf{I} - \frac{1}{\sigma^4} \frac{\sigma_A^2 \tilde{\mathbf{h}}\tilde{\mathbf{h}}^H}{1+\frac{\sigma_A^2}{\sigma^2}\tilde{\mathbf{h}}^H\tilde{\mathbf{h}}} \right) \tilde{\mathbf{x}} \\ &= \left(\frac{\sigma_A^2}{\sigma^2} - \frac{\sigma_A^4}{\sigma^4} \frac{\tilde{\mathbf{h}}^H\tilde{\mathbf{h}}}{1+\frac{\sigma_A^2}{\sigma^2}\tilde{\mathbf{h}}^H\tilde{\mathbf{h}}} \right) \tilde{\mathbf{h}}\tilde{\mathbf{h}}^H \tilde{\mathbf{x}}.\end{aligned}$$

Letting $\bar{\mathcal{E}} = E(\sum_{n=0}^{N-1} |\tilde{s}[n]|^2) = \sigma_A^2 \tilde{\mathbf{h}}^H\tilde{\mathbf{h}}$ be the *expected* signal energy of the received signal $\tilde{s}[n]$, this reduces to

$$\hat{\tilde{\mathbf{s}}} = \frac{\sigma_A^2}{\bar{\mathcal{E}}+\sigma^2} \tilde{\mathbf{h}}\tilde{\mathbf{h}}^H \tilde{\mathbf{x}}$$

so that from (13.14)

$$T(\tilde{\mathbf{x}}) = \frac{\sigma_A^2}{\bar{\mathcal{E}}+\sigma^2} |\tilde{\mathbf{h}}^H \tilde{\mathbf{x}}|^2$$

or finally we decide $\mathcal{H}_1$ if

$$T'(\tilde{\mathbf{x}}) = \left| \sum_{n=0}^{N-1} \tilde{x}[n]\tilde{h}^*[n] \right|^2 > \gamma''. \tag{13.17}$$

The test statistic is seen to be a quadrature or incoherent matched filter and is the complex extension of the real rank one detector discussed in Problem 5.14. If the transmitted signal is $\tilde{h}[n] = \exp(j2\pi f_0 n)$, then this detector is the complex extension of the NP detector for a Rayleigh fading sinusoid. Note that in this case the test statistic $T'(\tilde{\mathbf{x}})/N$ is exactly the periodogram, i.e., no approximations have been made as were required in Example 5.5.

The detection performance is easily found since $\tilde{z} = \sum_{n=0}^{N-1} \tilde{x}[n]\tilde{h}^*[n]$ is a zero mean complex Gaussian random variable, and hence, $T'(\tilde{\mathbf{x}})$ is a scaled χ_2^2 random variable under either hypothesis. Computing the moments we have

$$E(\tilde{z};\mathcal{H}_0) = E(\tilde{z};\mathcal{H}_1) = 0$$

since $E(\tilde{s}[n]) = E(\tilde{A})\tilde{h}[n] = 0$ and

$$\begin{aligned}\operatorname{var}(\tilde{z};\mathcal{H}_0) &= \operatorname{var}\left(\sum_{n=0}^{N-1} \tilde{w}[n]\tilde{h}^*[n] \right) = \sigma^2 \sum_{n=0}^{N-1} |\tilde{h}[n]|^2 = \frac{\sigma^2\bar{\mathcal{E}}}{\sigma_A^2} \\ \operatorname{var}(\tilde{z};\mathcal{H}_1) &= \operatorname{var}\left(\sum_{n=0}^{N-1} (\tilde{A}\tilde{h}[n] + \tilde{w}[n])\tilde{h}^*[n] \right)\end{aligned}$$

$$\begin{aligned} &= \operatorname{var}\left(\sum_{n=0}^{N-1} \tilde{A}|\tilde{h}[n]|^2 \right) + \operatorname{var}\left(\sum_{n=0}^{N-1} \tilde{w}[n]\tilde{h}^*[n] \right) \\ &= \sigma_A^2 \left(\sum_{n=0}^{N-1} |\tilde{h}[n]|^2 \right)^2 + \frac{\sigma^2\bar{\mathcal{E}}}{\sigma_A^2} \\ &= \frac{\bar{\mathcal{E}}^2}{\sigma_A^2} + \frac{\sigma^2\bar{\mathcal{E}}}{\sigma_A^2}.\end{aligned}$$

Thus,

$$\tilde{z} \sim \begin{cases} \mathcal{CN}(0, \sigma^2\bar{\mathcal{E}}/\sigma_A^2) & \text{under } \mathcal{H}_0 \\ \mathcal{CN}(0, \sigma^2\bar{\mathcal{E}}/\sigma_A^2 + \bar{\mathcal{E}}^2/\sigma_A^2) & \text{under } \mathcal{H}_1 \end{cases}$$

and $T'(\tilde{\mathbf{x}}) = |\tilde{z}|^2 = \operatorname{Re}^2(\tilde{z}) + \operatorname{Im}^2(\tilde{z})$ is a scaled χ_2^2 random variable under $\mathcal{H}_0$ and $\mathcal{H}_1$. Recall that the real and imaginary parts of a zero mean complex Gaussian random variable are each zero mean with the same variance (half of the total variance) and independent of each other. Then,

$$\begin{aligned}\frac{|\tilde{z}|^2}{\sigma_0^2/2} &\sim \chi_2^2 \quad \text{under } \mathcal{H}_0 \\ \frac{|\tilde{z}|^2}{\sigma_1^2/2} &\sim \chi_2^2 \quad \text{under } \mathcal{H}_1\end{aligned}$$

where $\sigma_0^2 = \sigma^2\bar{\mathcal{E}}/\sigma_A^2$ and $\sigma_1^2 = \sigma_0^2 + \bar{\mathcal{E}}^2/\sigma_A^2$. It follows that

$$\begin{aligned}P_{FA} &= \exp\left(-\frac{\gamma''}{\sigma_0^2} \right) \\ P_D &= \exp\left(-\frac{\gamma''}{\sigma_1^2} \right)\end{aligned}$$

or by eliminating γ''

$$\begin{aligned}P_D &= P_{FA}^{\frac{1}{1+\bar{\mathcal{E}}/\sigma^2}} \\ &= P_{FA}^{\frac{1}{1+\bar{\eta}}} \end{aligned} \tag{13.18}$$

where $\bar{\eta} = \bar{\mathcal{E}}/\sigma^2$ is the average ENR. Finally, this same problem is examined for the case when $\tilde{A}$ is an unknown deterministic constant and $\tilde{h}[n] = \exp(j2\pi f_0 n)$ in the signal processing example of Section 13.8.1.

◇

13.4 PDFs with Unknown Parameters

We will limit our discussion to the detection of deterministic and random signals with unknown parameters in CWGN with known variance σ^2. Furthermore, because of its broad generality we will assume a linear model for the signal. In the first case we model the unknown signal parameters as deterministic so that the signal is deterministic. The complex classical linear model then applies (see [Kay-I 1993, Example 15.9]). In the second case we will assign a prior PDF to the unknown signal parameters. This transforms the problem into the detection of a random signal with *known* PDF, for which the complex Bayesian linear model can be applied (see [Kay-I 1993, pp. 533–534]). Example 13.2 is a special case of this latter approach.

13.4.1 Deterministic Signal

The detection problem to be considered is

$$\begin{array}{ll} \tilde{x}[n] = \tilde{w}[n] & n = 0, 1, \ldots, N-1 \\ \tilde{x}[n] = \tilde{s}[n] + \tilde{w}[n] & n = 0, 1, \ldots, N-1 \end{array}$$

where $\tilde{s}[n]$ obeys the linear model or $\tilde{\mathbf{s}} = \mathbf{H}\boldsymbol{\theta}$ for $\mathbf{H}$ a known complex $N \times p$ observation matrix with $N > p$ and rank p, $\boldsymbol{\theta}$ is an unknown $p \times 1$ complex parameter vector, and $\tilde{w}[n]$ is CWGN with known variance σ^2. Making use of the complex classical linear model framework we have under $\mathcal{H}_0$ that $\tilde{\mathbf{x}} = \tilde{\mathbf{w}}$ and under $\mathcal{H}_1$ that $\tilde{\mathbf{x}} = \mathbf{H}\boldsymbol{\theta} + \tilde{\mathbf{w}}$. The equivalent parameter test is

$$\begin{array}{l} \mathcal{H}_0 : \boldsymbol{\theta} = \mathbf{0} \\ \mathcal{H}_1 : \boldsymbol{\theta} \neq \mathbf{0}. \end{array}$$

The GLRT decides $\mathcal{H}_1$ if

$$L_G(\tilde{\mathbf{x}}) = \frac{p(\tilde{\mathbf{x}}; \hat{\boldsymbol{\theta}}_1)}{p(\tilde{\mathbf{x}}; \boldsymbol{\theta} = \mathbf{0})} > \gamma$$

where $\hat{\boldsymbol{\theta}}_1$ is the MLE of $\boldsymbol{\theta}$ under $\mathcal{H}_1$ or the value that maximizes

$$p(\tilde{\mathbf{x}}; \boldsymbol{\theta}) = \frac{1}{\pi^N \sigma^{2N}} \exp\left[-\frac{1}{\sigma^2}(\tilde{\mathbf{x}} - \mathbf{H}\boldsymbol{\theta})^H(\tilde{\mathbf{x}} - \mathbf{H}\boldsymbol{\theta})\right].$$

This can be shown to be given by $\hat{\boldsymbol{\theta}}_1 = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{x}$ [Kay-I 1993, pp. 521–522] so that the GLRT becomes

$$\begin{aligned} 2\ln L_G(\mathbf{x}) \\ &= -\frac{2}{\sigma^2}\left[(\tilde{\mathbf{x}} - \mathbf{H}\hat{\boldsymbol{\theta}}_1)^H(\tilde{\mathbf{x}} - \mathbf{H}\hat{\boldsymbol{\theta}}_1) - \tilde{\mathbf{x}}^H\tilde{\mathbf{x}}\right] \\ &= -\frac{2}{\sigma^2}\left[-\tilde{\mathbf{x}}^H\mathbf{H}\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_1^H\mathbf{H}^H\tilde{\mathbf{x}} + \hat{\boldsymbol{\theta}}_1^H\mathbf{H}^H\mathbf{H}\hat{\boldsymbol{\theta}}_1\right] \\ &= -\frac{2}{\sigma^2}\left[-\tilde{\mathbf{x}}^H\mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{H}\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_1^H\mathbf{H}^H\mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\tilde{\mathbf{x}} + \hat{\boldsymbol{\theta}}_1^H\mathbf{H}^H\mathbf{H}\hat{\boldsymbol{\theta}}_1\right] \\ &= \frac{\hat{\boldsymbol{\theta}}_1^H\mathbf{H}^H\mathbf{H}\hat{\boldsymbol{\theta}}_1}{\sigma^2/2}. \end{aligned}$$

Except for the factor of 2, the GLRT is identical to the real case (let $\mathbf{A} = \mathbf{I}, \mathbf{b} = \mathbf{0}$ in Theorem 7.1). Hence, we decide $\mathcal{H}_1$ if

$$T(\tilde{\mathbf{x}}) = \frac{\hat{\boldsymbol{\theta}}_1^H\mathbf{H}^H\mathbf{H}\hat{\boldsymbol{\theta}}_1}{\sigma^2/2} > \gamma' \tag{13.19}$$

where $T(\tilde{\mathbf{x}})$ is real. It is shown in Appendix 13A that

$$T(\tilde{\mathbf{x}}) \sim \begin{cases} \chi^2_{2p} & \text{under } \mathcal{H}_0 \\ \chi'^2_{2p}(\lambda) & \text{under } \mathcal{H}_1 \end{cases}$$

where

$$\lambda = \frac{\boldsymbol{\theta}_1^H\mathbf{H}^H\mathbf{H}\boldsymbol{\theta}_1}{\sigma^2/2}$$

and $\boldsymbol{\theta}_1$ is the true value of $\boldsymbol{\theta}$ under $\mathcal{H}_1$. Thus the detection performance becomes

$$\begin{aligned} P_{FA} &= Q_{\chi^2_{2p}}(\gamma') \\ P_D &= Q_{\chi'^2_{2p}(\lambda)}(\gamma'). \end{aligned} \tag{13.20}$$

These results are nearly identical to the real case (see Theorem 7.1) except that there is a factor of 2 in λ, and the degrees of freedom are doubled since we are now testing $2p$ *real* parameters.

13.4.2 Random Signal

We next consider the same detection problem except that we assign a prior PDF to the unknown parameter vector $\boldsymbol{\theta}$. In doing so, we will transform the problem into one whose PDF is known. In particular, along with the linear model assumption for the signal or $\tilde{\mathbf{s}} = \mathbf{H}\boldsymbol{\theta}$, we assume that $\boldsymbol{\theta} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_\theta)$, resulting in the Bayesian linear model. To summarize this model, we assume that under $\mathcal{H}_1$ we observe $\tilde{\mathbf{x}} = \mathbf{H}\boldsymbol{\theta} + \tilde{\mathbf{w}}$, where $\tilde{\mathbf{x}}$ is an $N \times 1$ complex data vector, $\mathbf{H}$ is a known $N \times p$ complex observation matrix with $N > p$, and $\boldsymbol{\theta}$ is a $p \times 1$ complex random vector with $\boldsymbol{\theta} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_\theta)$ and independent of $\tilde{\mathbf{w}}$, which is a $N \times 1$ complex noise vector with $\tilde{\mathbf{w}} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$. In [Kay-I 1993, Section 15.8] we studied a slightly more general form of the Bayesian linear model for which $E(\boldsymbol{\theta}) = \boldsymbol{\mu}_\theta \neq \mathbf{0}$. Note that the signal $\tilde{\mathbf{s}} = \mathbf{H}\boldsymbol{\theta}$ is now *random* and has the *known* PDF $\tilde{\mathbf{s}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{H}\mathbf{C}_\theta\mathbf{H}^H)$. As such, the detection problem reduces to the one already discussed in Section 13.3.3. The optimal NP detector is an estimator-correlator or we decide $\mathcal{H}_1$ if

$$T(\tilde{\mathbf{x}}) = \tilde{\mathbf{x}}^H \hat{\tilde{\mathbf{s}}} > \gamma'$$

where

$$\begin{aligned}\hat{\tilde{\mathbf{s}}} &= \mathbf{C}_{\tilde{s}}(\mathbf{C}_{\tilde{s}} + \sigma^2\mathbf{I})^{-1}\tilde{\mathbf{x}} \\ &= \mathbf{H}\mathbf{C}_\theta\mathbf{H}^H(\mathbf{H}\mathbf{C}_\theta\mathbf{H}^H + \sigma^2\mathbf{I})^{-1}\tilde{\mathbf{x}}.\end{aligned}$$

The detection performance is derived in Problem 13.8 with a simple example already given in Example 13.2.

13.5 Vector Observations and PDFs

In many signal processing systems the observations or data samples are vectors. A typical example and the one that we will focus on is in array processing in which multiple sensors each output a voltage at each sample time. If there are M sensors, then at the nth sample time we would observe the data vector

$$\tilde{\mathbf{x}}[n] = \begin{bmatrix} \tilde{x}_0[n] \\ \tilde{x}_1[n] \\ \vdots \\ \tilde{x}_{M-1}[n] \end{bmatrix}$$

where $\tilde{x}_m[n]$ is the nth sample at the mth sensor. The data vector $\tilde{\mathbf{x}}[n]$ for a given value of n is sometimes referred to as a *snapshot* in that it corresponds to the outputs of an array of sensors *at a given time.* If we observe the outputs for the time interval $[0, N-1]$, then the accumulated data set will be $\{\tilde{\mathbf{x}}[0], \tilde{\mathbf{x}}[1], \ldots, \tilde{\mathbf{x}}[N-1]\}$.

We assume for sake of generality that the data samples are complex. Now it is convenient to arrange all the data samples in a large vector as

$$\tilde{\mathbf{x}} = \begin{bmatrix} \tilde{\mathbf{x}}[0] \\ \tilde{\mathbf{x}}[1] \\ \vdots \\ \tilde{\mathbf{x}}[N-1] \end{bmatrix} = \begin{bmatrix} \tilde{x}_0[0] \\ \tilde{x}_1[0] \\ \vdots \\ \tilde{x}_{M-1}[0] \\ \cdots\cdots \\ \tilde{x}_0[1] \\ \tilde{x}_1[1] \\ \vdots \\ \tilde{x}_{M-1}[1] \\ \cdots\cdots \\ \vdots \\ \cdots\cdots \\ \tilde{x}_0[N-1] \\ \tilde{x}_1[N-1] \\ \vdots \\ \tilde{x}_{M-1}[N-1] \end{bmatrix} \tag{13.21}$$

which has dimension $MN \times 1$. The large data vector $\tilde{\mathbf{x}}$ is composed of the snapshots arranged in column order. We will term this arrangement a *temporal ordering.* It is the assumed ordering for *vector* or equivalently *multichannel* time series processing [Robinson 1967, Hannan 1970, Kay 1988]. See Figure 13.4a,b for an example, where this ordering is also referred to as a *column rollout* [Graybill 1969]. Once we have done so, our large data vector can be used exactly as before in determining the different detectors. The only difference is in the structure of the covariance matrix. It is this structure that allows us at times to simplify and interpret the resulting detector for array processing. To appreciate this structure we consider noise only. Then we have that $\tilde{\mathbf{x}} = \tilde{\mathbf{w}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$, where $\mathbf{C}$ is $MN \times MN$. By definition the covariance matrix is $\mathbf{C} = E(\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H)$ so that from (13.21)

$$\begin{aligned}\mathbf{C} &= E\left(\begin{bmatrix} \tilde{\mathbf{x}}[0] \\ \tilde{\mathbf{x}}[1] \\ \vdots \\ \tilde{\mathbf{x}}[N-1] \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}^H[0] & \tilde{\mathbf{x}}^H[1] & \ldots & \tilde{\mathbf{x}}^H[N-1] \end{bmatrix}\right) \\ &= \begin{bmatrix} E(\tilde{\mathbf{x}}[0]\tilde{\mathbf{x}}^H[0]) & E(\tilde{\mathbf{x}}[0]\tilde{\mathbf{x}}^H[1]) & \ldots & E(\tilde{\mathbf{x}}[0]\tilde{\mathbf{x}}^H[N-1]) \\ E(\tilde{\mathbf{x}}[1]\tilde{\mathbf{x}}^H[0]) & E(\tilde{\mathbf{x}}[1]\tilde{\mathbf{x}}^H[1]) & \ldots & E(\tilde{\mathbf{x}}[1]\tilde{\mathbf{x}}^H[N-1]) \\ \vdots & \vdots & \ddots & \vdots \\ E(\tilde{\mathbf{x}}[N-1]\tilde{\mathbf{x}}^H[0]) & E(\tilde{\mathbf{x}}[N-1]\tilde{\mathbf{x}}^H[1]) & \ldots & E(\tilde{\mathbf{x}}[N-1]\tilde{\mathbf{x}}^H[N-1]) \end{bmatrix}\end{aligned}$$

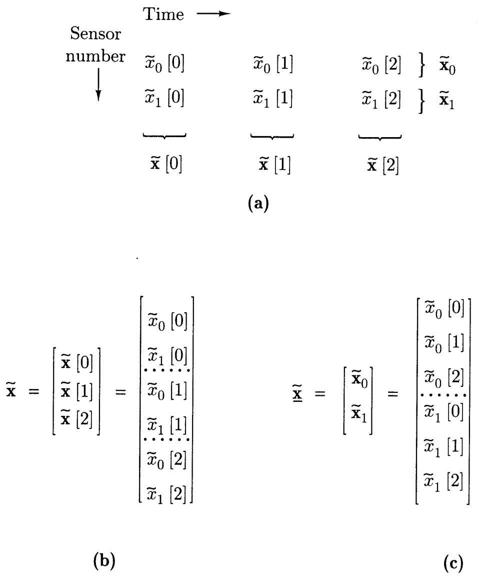


Figure 13.4. Various arrangements for array data (a) Array data (b) Temporal ordering (column rollout) (c) Spatial ordering (row rollout).

$$= \begin{bmatrix} \mathbf{C}[0,0] & \mathbf{C}[0,1] & \cdots & \mathbf{C}[0,N-1] \\ \mathbf{C}[1,0] & \mathbf{C}[1,1] & \cdots & \mathbf{C}[1,N-1] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}[N-1,0] & \mathbf{C}[N-1,1] & \cdots & \mathbf{C}[N-1,N-1] \end{bmatrix} \tag{13.22}$$

where

$$\mathbf{C}[i,j] = E(\tilde{\mathbf{x}}[i]\tilde{\mathbf{x}}^H[j]) \tag{13.23}$$

is the $M \times M$ "sub"-covariance matrix for the ith and jth *vector* samples. Note that $\mathbf{C}^H[i,j] = \mathbf{C}[j,i]$, and thus $\mathbf{C}$ is Hermitian as expected. For the case of scalar CWGN we have already seen that $\mathbf{C} = \sigma^2\mathbf{I}$. The latter covariance matrix is obtained if $\mathbf{C}[i,j] = \mathbf{0}$ for $i \neq j$ and also that $\mathbf{C}[i,i] = \sigma^2\mathbf{I}_M$, where $\mathbf{I}_M$ is the $M \times M$ identity matrix. In the context of array processing $\mathbf{C} = \sigma^2\mathbf{I}$ means that the noise processes at each sensor are CWGN with variance σ^2 and the processes are uncorrelated between sensors. It is also possible to have $\mathbf{C}[i,j] = \mathbf{0}$ for $i \neq j$ and $\mathbf{C}[i,i]$ not be a scaled identity matrix. Then the noise processes are uncorrelated at each sensor and between sensors for nonzero time lags, but correlated between

sensors at the same time instant. An example of this for $M = 2$ and $N = 3$ is the covariance matrix

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}[0,0] & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}[1,1] & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C}[2,2] \end{bmatrix}$$

which is block-diagonal. This reduces to *multichannel white noise* if $\mathbf{C}[i,i]$ is the same for all i or if $\mathbf{C}[i,j] = \mathbf{\Sigma}\delta_{ij}$ [Kay 1988]. Another possibility is that the noise process at each sensor is correlated but the noise processes are uncorrelated from sensor to sensor. As an example, for $M = 2$ and $N = 3$ we have that

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}[0,0] & \mathbf{C}[0,1] & \mathbf{C}[0,2] \\ \mathbf{C}[1,0] & \mathbf{C}[1,1] & \mathbf{C}[1,2] \\ \mathbf{C}[2,0] & \mathbf{C}[2,1] & \mathbf{C}[2,2] \end{bmatrix}$$

where the $[i,j]$th block is

$$\begin{aligned} \mathbf{C}[i,j] &= E(\tilde{\mathbf{x}}[i]\tilde{\mathbf{x}}^H[j]) = E\left(\begin{bmatrix} \tilde{x}_0[i] \\ \tilde{x}_1[i] \end{bmatrix} [\ \tilde{x}_0^*[j] \quad \tilde{x}_1^*[j]\ ]\right) \\ &= \begin{bmatrix} E(\tilde{x}_0[i]\tilde{x}_0^*[j]) & E(\tilde{x}_0[i]\tilde{x}_1^*[j]) \\ E(\tilde{x}_1[i]\tilde{x}_0^*[j]) & E(\tilde{x}_1[i]\tilde{x}_1^*[j]) \end{bmatrix} \\ &= \begin{bmatrix} E(\tilde{x}_0[i]\tilde{x}_0^*[j]) & 0 \\ 0 & E(\tilde{x}_1[i]\tilde{x}_1^*[j]) \end{bmatrix}. \end{aligned}$$

Hence, each $M \times M$ sub-covariance matrix $\mathbf{C}[i,j]$ is diagonal. To take advantage of this structure we will need to define a data vector as

$$\tilde{\mathbf{x}}_m = \begin{bmatrix} \tilde{x}_m[0] \\ \tilde{x}_m[1] \\ \vdots \\ \tilde{x}_m[N-1] \end{bmatrix}$$

where $\tilde{\mathbf{x}}_m$ is a vector containing the observed data set at the mth sensor. Then, due to the uncorrelated assumption between sensors we will have $E(\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_j^H) = \mathbf{0}$ for $i \neq j$. Now the preferred data sample ordering is a *spatial ordering* in which we arrange the samples from each sensor into the $MN \times 1$ large data vector

$$\underline{\tilde{\mathbf{x}}} = \begin{bmatrix} \tilde{\mathbf{x}}_0 \\ \tilde{\mathbf{x}}_1 \\ \vdots \\ \tilde{\mathbf{x}}_{M-1} \end{bmatrix}$$

an example of which is shown in Figure 13.4a,c. This arrangement is also termed a *row rollout.* The underbar notation will distinguish this ordering from our previous one. The covariance matrix then becomes for the $M = 2$ and $N = 3$ example

$$\begin{aligned}\underline{\mathbf{C}} &= E(\underline{\tilde{\mathbf{x}}}\,\underline{\tilde{\mathbf{x}}}^H) \\ &= \begin{bmatrix} E(\tilde{\mathbf{x}}_0\tilde{\mathbf{x}}_0^H) & \mathbf{0} \\ \mathbf{0} & E(\tilde{\mathbf{x}}_1\tilde{\mathbf{x}}_1^H) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{C}_{00} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{11} \end{bmatrix}\end{aligned}$$

where $\mathbf{C}_{ii} = E(\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^H)$ is the $N \times N$ ($N = 3$ for this example) covariance matrix for the data samples from the ith sensor. In this case, since the sensor outputs are uncorrelated between sensors, the covariance matrix $\underline{\mathbf{C}}$ is block-diagonal with $N \times N$ blocks.

The reason for the two possible orderings stems from the two-dimensional (2-D) nature of the data set as shown in Figure 13.4a. Since the data sample from the mth sensor at time n is $\tilde{x}_m[n]$ and we have samples for $m = 0, 1, \ldots, M-1$; $n = 0, 1, \ldots, N-1$, the data form a 2-D $M \times N$ array. The large data vector $\tilde{\mathbf{x}}$ corresponds to a "rolling out" of the columns while $\underline{\tilde{\mathbf{x}}}$ is a rolling out of the rows as illustrated in Figure 13.4. We will henceforth refer to $\tilde{\mathbf{x}}$ as a temporal ordering and $\underline{\tilde{\mathbf{x}}}$ as a spatial ordering to distinguish between the two possibilities. The former is advantageous when the samples are temporally uncorrelated while the latter is useful for uncorrelated samples between sensors or for spatially uncorrelated samples.

In summary, there are four cases of interest. They are:

1. General covariance matrix
2. Scaled identity covariance matrix, as in CWGN
3. Block-diagonal covariance matrix obtained for uncorrelated samples in time based on $\tilde{\mathbf{x}}$
4. Block-diagonal covariance matrix obtained for uncorrelated samples in space based on $\underline{\tilde{\mathbf{x}}}$.

We now determine the PDF for each of these cases for use later on. The data set will be assumed to be complex Gaussian with a nonzero mean and a given covariance matrix.

13.5.1 General Covariance Matrix

The PDF is

$$p(\tilde{\mathbf{x}}) = \frac{1}{\pi^{MN}\det(\mathbf{C})}\exp\left[-(\tilde{\mathbf{x}}-\tilde{\boldsymbol{\mu}})^H\mathbf{C}^{-1}(\tilde{\mathbf{x}}-\tilde{\boldsymbol{\mu}})\right] \tag{13.24}$$

where $\tilde{\boldsymbol{\mu}}$ is the $MN \times 1$ complex mean vector and $\mathbf{C}$ is the $MN \times MN$ covariance matrix.

13.5.2 Scaled Identity Matrix

If $\mathbf{C} = \sigma^2\mathbf{I}$, then the PDF of (13.24) reduces to

$$p(\tilde{\mathbf{x}}) = \frac{1}{\pi^{MN}\sigma^{2MN}}\exp\left[-\frac{1}{\sigma^2}(\tilde{\mathbf{x}}-\tilde{\boldsymbol{\mu}})^H(\tilde{\mathbf{x}}-\tilde{\boldsymbol{\mu}})\right]$$

and can also be written in temporally ordered form as

$$p(\tilde{\mathbf{x}}) = \frac{1}{\pi^{MN}\sigma^{2MN}}\exp\left(-\frac{1}{\sigma^2}\sum_{n=0}^{N-1}(\tilde{\mathbf{x}}[n]-\tilde{\boldsymbol{\mu}}[n])^H(\tilde{\mathbf{x}}[n]-\tilde{\boldsymbol{\mu}}[n])\right) \tag{13.25}$$

where $\tilde{\boldsymbol{\mu}}[n] = E(\tilde{\mathbf{x}}[n])$ or in spatially ordered form as

$$p(\underline{\tilde{\mathbf{x}}}) = \frac{1}{\pi^{MN}\sigma^{2MN}}\exp\left(-\frac{1}{\sigma^2}\sum_{m=0}^{M-1}(\tilde{\mathbf{x}}_m-\tilde{\boldsymbol{\mu}}_m)^H(\tilde{\mathbf{x}}_m-\tilde{\boldsymbol{\mu}}_m)\right)$$

where $\tilde{\boldsymbol{\mu}}_m = E(\tilde{\mathbf{x}}_m)$.

13.5.3 Uncorrelated from Temporal Sample to Sample

In this case we have that $\mathbf{C}[i,j] = E(\tilde{\mathbf{x}}[i]\tilde{\mathbf{x}}^H[j]) = \mathbf{0}$ for $i \neq j$, so that from (13.22)

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}[0,0] & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{C}[1,1] & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{C}[N-1,N-1] \end{bmatrix}.$$

Thus, we have

$$\begin{aligned}(\tilde{\mathbf{x}}-\tilde{\boldsymbol{\mu}})^H\mathbf{C}^{-1}(\tilde{\mathbf{x}}-\tilde{\boldsymbol{\mu}}) &= \sum_{n=0}^{N-1}(\tilde{\mathbf{x}}[n]-\tilde{\boldsymbol{\mu}}[n])^H\mathbf{C}^{-1}[n,n](\tilde{\mathbf{x}}[n]-\tilde{\boldsymbol{\mu}}[n]) \\ \det(\mathbf{C}) &= \prod_{n=0}^{N-1}\det(\mathbf{C}[n,n])\end{aligned}$$

and from (13.24)

$$\begin{aligned}&p(\tilde{\mathbf{x}}) \\ &= \frac{1}{\pi^{MN}\prod_{n=0}^{N-1}\det(\mathbf{C}[n,n])}\exp\left[-\sum_{n=0}^{N-1}(\tilde{\mathbf{x}}[n]-\tilde{\boldsymbol{\mu}}[n])^H\mathbf{C}^{-1}[n,n](\tilde{\mathbf{x}}[n]-\tilde{\boldsymbol{\mu}}[n])\right].\end{aligned}$$

13.5.4 Uncorrelated from Spatial Sample to Sample

We now assume that $\mathbf{C}_{ij} = E(\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_j^H) = \mathbf{0}$ for $i \neq j$, where $\mathbf{C}_{ij}$ is $N \times N$. Then, using a spatial ordering for the data samples the covariance matrix for $\underline{\tilde{\mathbf{x}}}$ becomes

$$\underline{\mathbf{C}} = \begin{bmatrix} \mathbf{C}_{00} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{11} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}_{M-1,M-1} \end{bmatrix}$$

and in like fashion to the previous case we have

$$p(\underline{\tilde{\mathbf{x}}}) = \frac{1}{\pi^{MN} \prod_{m=0}^{M-1} \det(\mathbf{C}_{mm})} \exp\left[-\sum_{m=0}^{M-1} (\tilde{\mathbf{x}}_m - \tilde{\boldsymbol{\mu}}_m)^H \mathbf{C}_{mm}^{-1} (\tilde{\mathbf{x}}_m - \tilde{\boldsymbol{\mu}}_m)\right].$$

13.6 Detectors for Vector Observations

We now determine the detectors for various cases of interest.

13.6.1 Known Deterministic Signal in CWGN

The hypothesis test is

$$\begin{array}{ll} \mathcal{H}_0 : \tilde{\mathbf{x}}[n] = \tilde{\mathbf{w}}[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : \tilde{\mathbf{x}}[n] = \tilde{\mathbf{s}}[n] + \tilde{\mathbf{w}}[n] & n = 0, 1, \ldots, N-1 \end{array}$$

where $\tilde{\mathbf{s}}[n]$ is a known complex deterministic signal and $\tilde{\mathbf{w}}[n]$ is *vector* CWGN with variance σ^2. The latter is defined as $\tilde{\mathbf{w}}[n] \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_M)$, where $\tilde{\mathbf{w}}[i]$ is uncorrelated with $\tilde{\mathbf{w}}[j]$ for $i \neq j$ or $E(\tilde{\mathbf{w}}[i]\tilde{\mathbf{w}}^H[j]) = \mathbf{0}$ for $i \neq j$. Hence, the covariance matrix becomes $\mathbf{C} = E(\tilde{\mathbf{w}}\tilde{\mathbf{w}}^H) = \sigma^2\mathbf{I}$. The NP detector decides $\mathcal{H}_1$ if

$$L(\tilde{\mathbf{x}}) = \frac{p(\tilde{\mathbf{x}}; \mathcal{H}_1)}{p(\tilde{\mathbf{x}}; \mathcal{H}_0)} > \gamma.$$

Using (13.25) with $\tilde{\boldsymbol{\mu}}[n] = \tilde{\mathbf{s}}[n]$, we have

$$\begin{aligned} \ln L(\tilde{\mathbf{x}}) &= -\frac{1}{\sigma^2}\left[\sum_{n=0}^{N-1} (\tilde{\mathbf{x}}[n] - \tilde{\mathbf{s}}[n])^H (\tilde{\mathbf{x}}[n] - \tilde{\mathbf{s}}[n]) - \sum_{n=0}^{N-1} \tilde{\mathbf{x}}^H[n]\tilde{\mathbf{x}}[n]\right] \\ &= -\frac{1}{\sigma^2}\left[\sum_{n=0}^{N-1} \left((\tilde{\mathbf{x}}[n] - \tilde{\mathbf{s}}[n])^H (\tilde{\mathbf{x}}[n] - \tilde{\mathbf{s}}[n]) - \tilde{\mathbf{x}}^H[n]\tilde{\mathbf{x}}[n]\right)\right] \\ &= -\frac{1}{\sigma^2}\left[\sum_{n=0}^{N-1} \left(-\tilde{\mathbf{x}}^H[n]\tilde{\mathbf{s}}[n] - \tilde{\mathbf{s}}^H[n]\tilde{\mathbf{x}}[n] + \tilde{\mathbf{s}}^H[n]\tilde{\mathbf{s}}[n]\right)\right] \end{aligned}$$

or by eliminating the non-data-dependent terms we decide $\mathcal{H}_1$ if

$$T(\tilde{\mathbf{x}}) = \mathrm{Re}\left(\sum_{n=0}^{N-1} \tilde{\mathbf{s}}^H[n]\tilde{\mathbf{x}}[n]\right) > \gamma'. \tag{13.26}$$

This is the replica-correlator for complex array data and is shown in Figure 13.5a. Clearly, this reduces to (13.3) in the scalar case. Note that it is possible to view

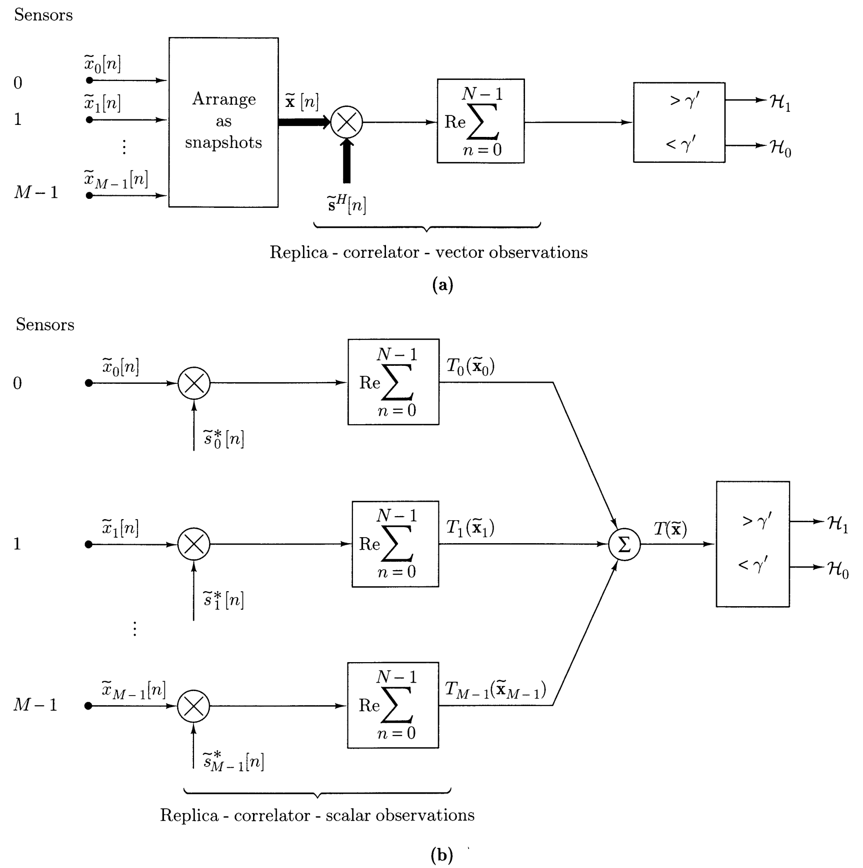


Figure 13.5. Replica-correlator for array data (a) Temporal ordering (b) Spatial ordering.

this detector in two different ways. As written above, we correlate the snapshots $\tilde{\mathbf{x}}[n]$ with the known signal snapshots $\tilde{\mathbf{s}}[n]$. Alternatively,

$$\begin{aligned} T(\tilde{\mathbf{x}}) &= \operatorname{Re}\left(\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\tilde{x}_m[n]\tilde{s}_m^*[n]\right) \\ &= \operatorname{Re}\left(\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\tilde{x}_m[n]\tilde{s}_m^*[n]\right) \\ &= \operatorname{Re}\left(\sum_{m=0}^{M-1}\tilde{\mathbf{s}}_m^H\tilde{\mathbf{x}}_m\right) \\ &= \sum_{m=0}^{M-1}T_m(\tilde{\mathbf{x}}_m) \end{aligned}$$

where $T_m(\tilde{\mathbf{x}}_m) = \operatorname{Re}(\tilde{\mathbf{s}}_m^H\tilde{\mathbf{x}}_m)$ is the replica-correlator output of the mth sensor. Hence, the detector correlates the output of each sensor with the known signal at that sensor and then sums the results over all sensors. This is because the output of each sensor is independent of all the others. This is shown in Figure 13.5b. In essence, because the data set is a 2-D array, we can correlate over columns first and then rows or vice-versa (see Problem 13.15).

The detection performance is easily found from the scalar case. From (13.5) we have

$$T_m(\tilde{\mathbf{x}}_m) \sim \begin{cases} \mathcal{N}(0, \sigma^2\tilde{\mathbf{s}}_m^H\tilde{\mathbf{s}}_m/2) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(\tilde{\mathbf{s}}_m^H\tilde{\mathbf{s}}_m, \sigma^2\tilde{\mathbf{s}}_m^H\tilde{\mathbf{s}}_m/2) & \text{under } \mathcal{H}_1. \end{cases}$$

Utilizing the independence of the $T_m(\tilde{\mathbf{x}}_m)$'s, it follows that

$$T(\tilde{\mathbf{x}}) = \sum_{m=0}^{M-1}T_m(\tilde{\mathbf{x}}_m) \sim \begin{cases} \mathcal{N}\left(0, \sum_{m=0}^{M-1}\sigma^2\tilde{\mathbf{s}}_m^H\tilde{\mathbf{s}}_m/2\right) & \text{under } \mathcal{H}_0 \\ \mathcal{N}\left(\sum_{m=0}^{M-1}\tilde{\mathbf{s}}_m^H\tilde{\mathbf{s}}_m, \sum_{m=0}^{M-1}\sigma^2\tilde{\mathbf{s}}_m^H\tilde{\mathbf{s}}_m/2\right) & \text{under } \mathcal{H}_1. \end{cases}$$

But $\sum_{m=0}^{M-1}\tilde{\mathbf{s}}_m^H\tilde{\mathbf{s}}_m = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1}|\tilde{s}_m[n]|^2 = \mathcal{E}$, which is the total signal energy, so that

$$\begin{aligned} P_{FA} &= Q\left(\frac{\gamma'}{\sqrt{\sigma^2\mathcal{E}/2}}\right) \\ P_D &= Q\left(\frac{\gamma' - \mathcal{E}}{\sqrt{\sigma^2\mathcal{E}/2}}\right) \end{aligned}$$

or

$$P_D = Q\left(Q^{-1}(P_{FA}) - \sqrt{d^2}\right)$$

where $d^2 = 2\mathcal{E}/\sigma^2$. It is clear that the performance will increase as the signal energy at each sensor increases and also as the number of sensors increases.

13.6.2 Known Deterministic Signal and General Noise Covariance

Now we use (13.10) so that the NP detector decides $\mathcal{H}_1$ if

$$T(\tilde{\mathbf{x}}) = \operatorname{Re}(\tilde{\mathbf{s}}^H\mathbf{C}^{-1}\tilde{\mathbf{x}}) > \gamma'. \tag{13.27}$$

Of course if $\mathbf{C} = \sigma^2\mathbf{I}$, then we have the previous case. The detection performance has already been given by (13.12). The only difference now is in the definitions of $\tilde{\mathbf{s}}$, $\tilde{\mathbf{x}}$, and $\mathbf{C}$. From (13.12) it becomes

$$\begin{aligned} P_{FA} &= Q\left(\frac{\gamma'}{\sqrt{\tilde{\mathbf{s}}^H\mathbf{C}^{-1}\tilde{\mathbf{s}}/2}}\right) \\ P_D &= Q\left(\frac{\gamma' - \tilde{\mathbf{s}}^H\mathbf{C}^{-1}\tilde{\mathbf{s}}}{\sqrt{\tilde{\mathbf{s}}^H\mathbf{C}^{-1}\tilde{\mathbf{s}}/2}}\right) \end{aligned}$$

or

$$P_D = Q\left(Q^{-1}(P_{FA}) - \sqrt{d^2}\right) \tag{13.28}$$

where $d^2 = 2\tilde{\mathbf{s}}^H\mathbf{C}^{-1}\tilde{\mathbf{s}}$.

13.6.3 Known Deterministic Signal in Temporally Uncorrelated Noise

If the noise is uncorrelated from temporal snapshot to snapshot but correlated between sensors at the same time instant, then $\mathbf{C}[i,j] = \mathbf{0}$ for $i \neq j$. Note that $\mathbf{C}[i,i]$ is not diagonal in general. Then,

$$\mathbf{C}^{-1} = \begin{bmatrix} \mathbf{C}^{-1}[0,0] & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^{-1}[1,1] & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}^{-1}[N-1,N-1] \end{bmatrix}$$

and (13.27) reduces to

$$T(\tilde{\mathbf{x}}) = \operatorname{Re}\left(\sum_{n=0}^{N-1}\tilde{\mathbf{s}}^H[n]\mathbf{C}^{-1}[n,n]\tilde{\mathbf{x}}[n]\right) \tag{13.29}$$

and the performance follows from (13.28) with $d^2 = 2\sum_{n=0}^{N-1}\tilde{\mathbf{s}}^H[n]\mathbf{C}^{-1}[n,n]\tilde{\mathbf{s}}[n]$.

13.6.4 Known Deterministic Signal in Spatially Uncorrelated Noise

If the noise processes are uncorrelated from sensor to sensor, but each noise process is temporally correlated (at any one sensor), then using a spatial ordering we have

$$\underline{\mathbf{C}}^{-1} = \begin{bmatrix} \mathbf{C}_{00}^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{11}^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}_{M-1,M-1}^{-1} \end{bmatrix}.$$

The NP detector becomes from (13.27) to decide $\mathcal{H}_1$ if

$$\begin{aligned} T(\tilde{\mathbf{x}}) &= \operatorname{Re}\left(\underline{\tilde{\mathbf{s}}}^H \underline{\mathbf{C}}^{-1} \underline{\tilde{\mathbf{x}}}\right) \\ &= \operatorname{Re}\left(\sum_{m=0}^{M-1} \tilde{\mathbf{s}}_m^H \mathbf{C}_{mm}^{-1} \tilde{\mathbf{x}}_m\right) \qquad (13.30) \\ &= \sum_{m=0}^{M-1} T_m(\tilde{\mathbf{x}}_m) > \gamma' \end{aligned}$$

where $T_m(\tilde{\mathbf{x}}_m) = \operatorname{Re}(\tilde{\mathbf{s}}_m^H \mathbf{C}_{mm}^{-1} \tilde{\mathbf{x}}_m)$ is the generalized matched filter output or replica-correlator output of the mth sensor. The detection performance follows from (13.28) with $d^2 = 2\sum_{m=0}^{M-1} \tilde{\mathbf{s}}_m^H \mathbf{C}_{mm}^{-1} \tilde{\mathbf{s}}_m$.

13.6.5 Random Signal in CWGN

We now turn our attention to the detection of a random signal. The detection problem is

$$\begin{aligned} \mathcal{H}_0 &: \tilde{\mathbf{x}}[n] = \tilde{\mathbf{w}}[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 &: \tilde{\mathbf{x}}[n] = \tilde{\mathbf{s}}[n] + \tilde{\mathbf{w}}[n] & n = 0, 1, \ldots, N-1 \end{aligned}$$

where $\tilde{\mathbf{s}}[n]$ is a complex Gaussian random process with zero mean and covariance matrix $\mathbf{C}_{\tilde{s}} = E(\tilde{\mathbf{s}}\tilde{\mathbf{s}}^H)$ and $\tilde{\mathbf{w}}[n]$ is a vector CWGN process with known variance σ^2, and is independent of $\tilde{\mathbf{s}}[n]$. Then, (13.14) applies directly so that we decide $\mathcal{H}_1$ if

$$T(\tilde{\mathbf{x}}) = \tilde{\mathbf{x}}^H \hat{\tilde{\mathbf{s}}} > \gamma' \qquad (13.31)$$

where

$$\hat{\tilde{\mathbf{s}}} = \mathbf{C}_{\tilde{s}}(\mathbf{C}_{\tilde{s}} + \sigma^2\mathbf{I})^{-1}\tilde{\mathbf{x}}.$$

Some special cases are of interest. If the signal is spatially and temporally uncorrelated (or white), then $\mathbf{C}_{\tilde{s}} = \sigma_{\tilde{s}}^2\mathbf{I}$ and we have

$$\hat{\tilde{\mathbf{s}}} = \frac{\sigma_{\tilde{s}}^2}{\sigma_{\tilde{s}}^2 + \sigma^2}\tilde{\mathbf{x}}$$

and

$$\begin{aligned} T(\tilde{\mathbf{x}}) &= \frac{\sigma_{\tilde{s}}^2}{\sigma_{\tilde{s}}^2 + \sigma^2}\tilde{\mathbf{x}}^H\tilde{\mathbf{x}} \\ &= \frac{\sigma_{\tilde{s}}^2}{\sigma_{\tilde{s}}^2 + \sigma^2}\sum_{m=0}^{M-1} \tilde{\mathbf{x}}_m^H\tilde{\mathbf{x}}_m \end{aligned}$$

using a spatial ordering. Then, we decide $\mathcal{H}_1$ if

$$T'(\tilde{\mathbf{x}}) = \sum_{m=0}^{M-1} T_m(\tilde{\mathbf{x}}_m) > \gamma''$$

where $T_m(\tilde{\mathbf{x}}_m) = \sum_{n=0}^{N-1} |x_m[n]|^2$ is the energy out of the mth sensor. This is the sum of the energy test statistics from each sensor (see Example 5.1).

If the signal is only temporally uncorrelated, then since

$$\mathbf{C}_{\tilde{s}} = \begin{bmatrix} \mathbf{C}_{\tilde{s}}[0,0] & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{\tilde{s}}[1,1] & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}_{\tilde{s}}[N-1,N-1] \end{bmatrix}$$

and letting $\hat{\tilde{\mathbf{s}}} = [\hat{\tilde{\mathbf{s}}}[0]\,\hat{\tilde{\mathbf{s}}}[1]\ldots\hat{\tilde{\mathbf{s}}}[N-1]]^T$, it follows that

$$\hat{\tilde{\mathbf{s}}}[n] = \mathbf{C}_{\tilde{s}}[n,n](\mathbf{C}_{\tilde{s}}[n,n] + \sigma^2\mathbf{I})^{-1}\tilde{\mathbf{x}}[n]$$

and from (13.31)

$$T(\tilde{\mathbf{x}}) = \sum_{n=0}^{N-1} \tilde{\mathbf{x}}^H[n]\hat{\tilde{\mathbf{s}}}[n]$$

which is shown in Figure 13.6a.

If the signal is spatially uncorrelated, then since

$$\underline{\mathbf{C}}_{\tilde{s}} = \begin{bmatrix} \mathbf{C}_{00} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{11} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}_{M-1,M-1} \end{bmatrix}$$

is block-diagonal and letting $\hat{\underline{\tilde{\mathbf{s}}}} = [\hat{\tilde{\mathbf{s}}}_0\,\hat{\tilde{\mathbf{s}}}_1\ldots\hat{\tilde{\mathbf{s}}}_{M-1}]^T$, it follows that

$$\hat{\tilde{\mathbf{s}}}_m = \mathbf{C}_{mm}(\mathbf{C}_{mm} + \sigma^2\mathbf{I})^{-1}\tilde{\mathbf{x}}_m$$

and from (13.31)

$$T(\tilde{\mathbf{x}}) = \sum_{m=0}^{M-1} \tilde{\mathbf{x}}_m^H\hat{\tilde{\mathbf{s}}}_m$$

which is shown in Figure 13.6b.

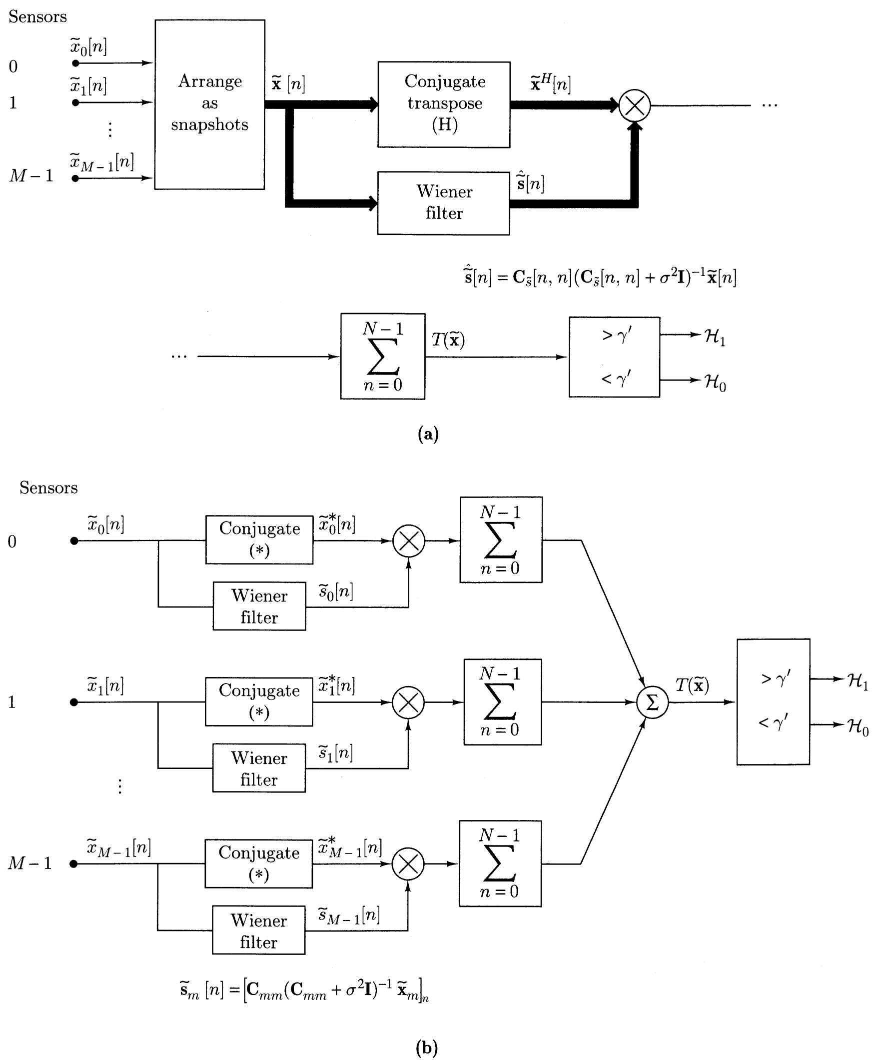


Figure 13.6. Estimator-correlator for array data (a) Temporally uncorrelated signal (b) Spatially uncorrelated signal.

13.6.6 Deterministic Signal with Unknown Parameters in CWGN

Lastly, we examine the case of a deterministic signal with unknown parametes in CWGN. We assume that the signal obeys the classical linear model so that the results of Section 13.4.1 can be applied. The detection problem is

$$\begin{array}{lll} \mathcal{H}_0 : \tilde{\mathbf{x}}[n] = \tilde{\mathbf{w}}[n] & n = 0,1,\ldots,N-1 \\ \mathcal{H}_1 : \tilde{\mathbf{x}}[n] = \sum_{i=1}^{p} \mathbf{h}_i[n]\theta_i + \tilde{\mathbf{w}}[n] & n = 0,1,\ldots,N-1 \end{array} \tag{13.32}$$

where the deterministic signal is $\tilde{\mathbf{s}}[n] = \sum_{i=1}^{p} \mathbf{h}_i[n]\theta_i$, and $\tilde{\mathbf{w}}[n]$ is vector CWGN with known variance σ^2. The $M \times 1$ observation vectors $\mathbf{h}_i[n]$ are complex and known and the signal parameters θ_i are assumed to be complex and unknown. As an example of the signal modeling, consider the case of two complex sinusoids that are received at M sensors. Each sinusoidal signal received at a sensor is identical except that it is delayed with respect to the other sensors. Let $n_m(\beta_1), n_m(\beta_2)$ be the delay in samples of the first and second sinusoid, respectively, at the mth sensor. The arguments β_1, β_2 indicate the direction of arrival of the propagating sinusoids. Then, the signal received at the mth sensor can be expressed as

$$\tilde{s}_m[n] = A\exp\left[j(2\pi f_1(n - n_m(\beta_1)) + \phi)\right] + B\exp\left[j(2\pi f_2(n - n_m(\beta_2)) + \psi)\right]$$

where, as usual, $\tilde{s}_m[n]$ is the nth temporal sample received at the mth sensor. It is assumed that $\{A, \phi, B, \psi\}$ are unknown and that the frequencies f_1, f_2 and delays $n_m(\beta_1), n_m(\beta_2)$ are known. Arranging the samples as snapshots, we have the linear signal model of (13.32)

$$\begin{aligned} \tilde{\mathbf{s}}[n] &= \underbrace{\begin{bmatrix} \exp[j2\pi f_1(n - n_0(\beta_1))] \\ \exp[j2\pi f_1(n - n_1(\beta_1))] \\ \vdots \\ \exp[j2\pi f_1(n - n_{M-1}(\beta_1))] \end{bmatrix}}_{\mathbf{h}_1[n]} \underbrace{A\exp(j\phi)}_{\theta_1} \\ &+ \underbrace{\begin{bmatrix} \exp[j2\pi f_2(n - n_0(\beta_2))] \\ \exp[j2\pi f_2(n - n_1(\beta_2))] \\ \vdots \\ \exp[j2\pi f_2(n - n_{M-1}(\beta_2))] \end{bmatrix}}_{\mathbf{h}_2[n]} \underbrace{B\exp(j\psi)}_{\theta_2}. \end{aligned}$$

This model will be used again in Section 13.8.

We can now express the signal in (13.32) in the form of the classical linear model by using a temporal ordering as follows.

$$\tilde{\mathbf{s}} = \begin{bmatrix} \tilde{\mathbf{s}}[0] \\ \tilde{\mathbf{s}}[1] \\ \vdots \\ \tilde{\mathbf{s}}[N-1] \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{p} \mathbf{h}_i[0]\theta_i \\ \sum_{i=1}^{p} \mathbf{h}_i[1]\theta_i \\ \vdots \\ \sum_{i=1}^{p} \mathbf{h}_i[N-1]\theta_i \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} \mathbf{h}_1[0] & \mathbf{h}_2[0] & \dots & \mathbf{h}_p[0] \\ \mathbf{h}_1[1] & \mathbf{h}_2[1] & \dots & \mathbf{h}_p[1] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{h}_1[N-1] & \mathbf{h}_2[N-1] & \dots & \mathbf{h}_p[N-1] \end{bmatrix}}_{\mathbf{H}} \boldsymbol{\theta}.$$

Note that the observation matrix $\mathbf{H}$ is $MN \times p$ since each $\mathbf{h}_i[n]$ is $M \times 1$.

Now using the linear model results of (13.19), we can implement a GLRT to decide $\mathcal{H}_1$ if

$$T(\tilde{\mathbf{x}}) = \frac{\hat{\boldsymbol{\theta}}_1^H \mathbf{H}^H \mathbf{H} \hat{\boldsymbol{\theta}}_1}{\sigma^2/2} > \gamma' \tag{13.33}$$

where $\hat{\boldsymbol{\theta}}_1 = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\tilde{\mathbf{x}}$. The test statistic can be simplified by letting $\hat{\tilde{\mathbf{s}}} = \mathbf{H}\hat{\boldsymbol{\theta}}_1$ so that

$$\begin{aligned} T(\tilde{\mathbf{x}}) &= \frac{\tilde{\mathbf{x}}^H\mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\tilde{\mathbf{x}}}{\sigma^2/2} = \frac{\hat{\tilde{\mathbf{s}}}^H\tilde{\mathbf{x}}}{\sigma^2/2} \\ &= \frac{\sum_{n=0}^{N-1} \hat{\tilde{\mathbf{s}}}^H[n]\tilde{\mathbf{x}}[n]}{\sigma^2/2} \end{aligned} \tag{13.34}$$

using a temporal ordering or equivalently

$$T(\underline{\tilde{\mathbf{x}}}) = \frac{\sum_{m=0}^{M-1} \hat{\tilde{\mathbf{s}}}_m^H \tilde{\mathbf{x}}_m}{\sigma^2/2} \tag{13.35}$$

using a spatial ordering. The detection performance becomes from (13.20)

$$P_{FA} = Q_{\chi^2_{2p}}(\gamma')$$

$$P_D = Q_{\chi'^2_{2p}(\lambda)}(\gamma') \tag{13.36}$$

where

$$\lambda = \frac{\boldsymbol{\theta}_1^H\mathbf{H}^H\mathbf{H}\boldsymbol{\theta}_1}{\sigma^2/2}. \tag{13.37}$$

13.7 Estimator-Correlator for Large Data Records

As we discussed in Chapter 5, it is possible to implement an estimator-correlator in the frequency domain if the data record length is large enough or as $N \to \infty$. This followed from an approximate eigendecomposition and subsequent diagonalization of the signal covariance matrix. In the vector case a similar diagonalization results in a transformed signal covariance matrix that is *block-diagonal.* Before motivating this result we need to define the ACF and PSD for a *vector* or *multichannel* WSS random process. Recall that the covariance matrix of $\tilde{\mathbf{x}}$ is

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}[0,0] & \mathbf{C}[0,1] & \dots & \mathbf{C}[0,N-1] \\ \mathbf{C}[1,0] & \mathbf{C}[1,1] & \dots & \mathbf{C}[1,N-1] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}[N-1,0] & \mathbf{C}[N-1,1] & \dots & \mathbf{C}[N-1,N-1] \end{bmatrix}$$

where $\mathbf{C}[i,j] = E(\tilde{\mathbf{x}}[i]\tilde{\mathbf{x}}^H[j])$. If the multichannel time series $\tilde{\mathbf{x}}[n]$ is WSS, then the covariance matrix $\mathbf{C}[i,j]$ only depends on the time lag between the samples. In this case

$$\begin{aligned} \mathbf{C}[i,j] &= E(\tilde{\mathbf{x}}[i]\tilde{\mathbf{x}}^H[j]) = E(\tilde{\mathbf{x}}[i-j]\tilde{\mathbf{x}}^H[0]) \\ &= E(\tilde{\mathbf{x}}^*[0]\tilde{\mathbf{x}}^T[i-j])^T. \end{aligned}$$

We will define $E(\tilde{\mathbf{x}}^*[0]\tilde{\mathbf{x}}^T[k])$ as the multichannel ACF at lag k or more generally [Kay 1988], the multichannel ACF is defined as the $M \times M$ matrix

$$\mathbf{R}_{\tilde{x}\tilde{x}}[k] = E(\tilde{\mathbf{x}}^*[n]\tilde{\mathbf{x}}^T[n+k]).$$

Thus, $\mathbf{C}[i,j] = \mathbf{R}^T_{\tilde{x}\tilde{x}}[i-j]$ and the covariance matrix becomes

$$\mathbf{C} = \begin{bmatrix} \mathbf{R}^T_{\tilde{x}\tilde{x}}[0] & \mathbf{R}^T_{\tilde{x}\tilde{x}}[-1] & \dots & \mathbf{R}^T_{\tilde{x}\tilde{x}}[-(N-1)] \\ \mathbf{R}^T_{\tilde{x}\tilde{x}}[1] & \mathbf{R}^T_{\tilde{x}\tilde{x}}[0] & \dots & \mathbf{R}^T_{\tilde{x}\tilde{x}}[-(N-2)] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}^T_{\tilde{x}\tilde{x}}[N-1] & \mathbf{R}^T_{\tilde{x}\tilde{x}}[N-2] & \dots & \mathbf{R}^T_{\tilde{x}\tilde{x}}[0] \end{bmatrix}. \tag{13.38}$$

Note that the $M \times M$ matrix along any NW-SE diagonal is the same (termed a *block-Toeplitz* matrix). If we define $r_{ij}[k] = E(\tilde{x}_i^*[n]\tilde{x}_j[n+k])$ for $i = 0, 1, \dots, M-1; j =$

$0, 1, \ldots, M-1$, then $r_{ij}[k]$ is the *cross-correlation function* between sensors i and j at lag k. Then the multichannel ACF at lag k is

$$\mathbf{R}_{\tilde{x}\tilde{x}}[k] = \begin{bmatrix} r_{00}[k] & r_{01}[k] & \ldots & r_{0,M-1}[k] \\ r_{10}[k] & r_{11}[k] & \ldots & r_{1,M-1}[k] \\ \vdots & \vdots & \ddots & \vdots \\ r_{M-1,0}[k] & r_{M-1,1}[k] & \ldots & r_{M-1,M-1}[k] \end{bmatrix}.$$

The Fourier transform of $\mathbf{R}_{\tilde{x}\tilde{x}}[k]$, which is defined as the Fourier transform of each element of $\mathbf{R}_{\tilde{x}\tilde{x}}[k]$, yields

$$\begin{aligned} \mathbf{P}_{\tilde{x}\tilde{x}}(f) &= \sum_{k=-\infty}^{\infty} \mathbf{R}_{\tilde{x}\tilde{x}}[k] \exp(-j2\pi f k) \\ &= \begin{bmatrix} P_{00}(f) & P_{01}(f) & \ldots & P_{0,M-1}(f) \\ P_{10}(f) & P_{11}(f) & \ldots & P_{1,M-1}(f) \\ \vdots & \vdots & \ddots & \vdots \\ P_{M-1,0}(f) & P_{M-1,1}(f) & \ldots & P_{M-1,M-1}(f) \end{bmatrix} \end{aligned}$$

where

$$P_{ij}(f) = \sum_{k=-\infty}^{\infty} r_{ij}[k] \exp(-j2\pi f k)$$

is the cross-PSD between sensors i and j. The terms on the main diagonal are the auto-PSDs at the sensors or just the usual PSD. The matrix $\mathbf{P}_{\tilde{x}\tilde{x}}(f)$ is termed the *cross-spectral matrix* (CSM). It can be shown to be Hermitian and positive definite (see Problem 13.19).

We now illustrate how the covariance matrix $\mathbf{C}$ can be approximately block-diagonalized, extending the results in Chapter 2. To see this consider a multichannel process whose ACF satisfies $\mathbf{R}_{\tilde{x}\tilde{x}}[k] = \mathbf{0}$ for $k \geq 2$ (termed a *multichannel* MA(1) process). Then from (13.38)

$$\mathbf{C} = \begin{bmatrix} \mathbf{R}^T_{\tilde{x}\tilde{x}}[0] & \mathbf{R}^T_{\tilde{x}\tilde{x}}[-1] & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{R}^T_{\tilde{x}\tilde{x}}[1] & \mathbf{R}^T_{\tilde{x}\tilde{x}}[0] & \mathbf{R}^T_{\tilde{x}\tilde{x}}[-1] & \mathbf{0} & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{R}^T_{\tilde{x}\tilde{x}}[1] & \mathbf{R}^T_{\tilde{x}\tilde{x}}[0] & \mathbf{R}^T_{\tilde{x}\tilde{x}}[-1] \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{R}^T_{\tilde{x}\tilde{x}}[1] & \mathbf{R}^T_{\tilde{x}\tilde{x}}[0] \end{bmatrix}.$$

Now define a multichannel sinusoidal matrix as

$$\mathbf{V}_i = \frac{1}{\sqrt{N}} \begin{bmatrix} \mathbf{I}_M \\ \mathbf{I}_M \exp(j2\pi f_i) \\ \vdots \\ \mathbf{I}_M \exp[j2\pi f_i(N-1)] \end{bmatrix}$$

where $f_i = i/N$ for $i = 0, 1, \ldots, N-1$ and hence $\mathbf{V}_i$ has dimension $MN \times M$, and $\mathbf{I}_M$ is the $M \times M$ identity matrix. Then, evaluating $\mathbf{C}\mathbf{V}_i$ we have

$$\begin{aligned} &\mathbf{C}\mathbf{V}_i \\ &= \frac{1}{\sqrt{N}} \begin{bmatrix} \mathbf{R}^T_{\tilde{x}\tilde{x}}[0] + \mathbf{R}^T_{\tilde{x}\tilde{x}}[-1]\exp(j2\pi f_i) \\ \mathbf{R}^T_{\tilde{x}\tilde{x}}[1] + \mathbf{R}^T_{\tilde{x}\tilde{x}}[0]\exp(j2\pi f_i) + \mathbf{R}^T_{\tilde{x}\tilde{x}}[-1]\exp(j4\pi f_i) \\ \vdots \\ \mathbf{R}^T_{\tilde{x}\tilde{x}}[1]\exp[j2\pi f_i(N-2)] + \mathbf{R}^T_{\tilde{x}\tilde{x}}[0]\exp[j2\pi f_i(N-1)] \end{bmatrix} \\ &= \frac{1}{\sqrt{N}} \begin{bmatrix} [\mathbf{R}^T_{\tilde{x}\tilde{x}}[0] + \mathbf{R}^T_{\tilde{x}\tilde{x}}[-1]\exp(j2\pi f_i)]1 \\ [\mathbf{R}^T_{\tilde{x}\tilde{x}}[1]\exp(-j2\pi f_i) + \mathbf{R}^T_{\tilde{x}\tilde{x}}[0] + \mathbf{R}^T_{\tilde{x}\tilde{x}}[-1]\exp(j2\pi f_i)]\exp(j2\pi f_i) \\ \vdots \\ [\mathbf{R}^T_{\tilde{x}\tilde{x}}[1]\exp(-j2\pi f_i) + \mathbf{R}^T_{\tilde{x}\tilde{x}}[0]]\exp[j2\pi f_i(N-1)] \end{bmatrix} \\ &\approx \frac{1}{\sqrt{N}} \begin{bmatrix} \mathbf{I}_M \mathbf{P}^T_{\tilde{x}\tilde{x}}(f_i) \\ \mathbf{I}_M \exp(j2\pi f_i)\mathbf{P}^T_{\tilde{x}\tilde{x}}(f_i) \\ \vdots \\ \mathbf{I}_M \exp[j2\pi f_i(N-1)]\mathbf{P}^T_{\tilde{x}\tilde{x}}(f_i) \end{bmatrix} \\ &= \frac{1}{\sqrt{N}} \begin{bmatrix} \mathbf{I}_M \\ \mathbf{I}_M \exp(j2\pi f_i) \\ \vdots \\ \mathbf{I}_M \exp[j2\pi f_i(N-1)] \end{bmatrix} \mathbf{P}^T_{\tilde{x}\tilde{x}}(f_i) = \mathbf{V}_i \mathbf{P}^T_{\tilde{x}\tilde{x}}(f_i) \end{aligned}$$

since

$$\mathbf{P}_{\tilde{x}\tilde{x}}(f) = \mathbf{R}_{\tilde{x}\tilde{x}}[-1]\exp(j2\pi f) + \mathbf{R}_{\tilde{x}\tilde{x}}[0] + \mathbf{R}_{\tilde{x}\tilde{x}}[1]\exp(-j2\pi f)$$

is the CSM. Now letting $i = 0, 1, \ldots, N-1$, we have

$$\mathbf{C}\underbrace{[\,\mathbf{V}_0 \;\; \mathbf{V}_1 \;\; \ldots \;\; \mathbf{V}_{N-1}\,]}_{\mathbf{V}} =$$

$$\underbrace{[\,\mathbf{V}_0 \;\; \mathbf{V}_1 \;\; \ldots \;\; \mathbf{V}_{N-1}\,]}_{\mathbf{V}} \underbrace{\begin{bmatrix} \mathbf{P}^T_{\tilde{x}\tilde{x}}(f_0) & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{P}^T_{\tilde{x}\tilde{x}}(f_1) & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{P}^T_{\tilde{x}\tilde{x}}(f_{N-1}) \end{bmatrix}}_{\mathbf{P}^T_{\tilde{x}}}. \tag{13.39}$$

But $\mathbf{V}$ is a unitary matrix (which is to say that $\mathbf{V}^H = \mathbf{V}^{-1}$) since

$$\mathbf{V}^H\mathbf{V} = \begin{bmatrix} \mathbf{V}_0^H\mathbf{V}_0 & \mathbf{V}_0^H\mathbf{V}_1 & \cdots & \mathbf{V}_0^H\mathbf{V}_{N-1} \\ \mathbf{V}_1^H\mathbf{V}_0 & \mathbf{V}_1^H\mathbf{V}_1 & \cdots & \mathbf{V}_1^H\mathbf{V}_{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{V}_0^H\mathbf{V}_{N-1} & \mathbf{V}_1^H\mathbf{V}_{N-1} & \cdots & \mathbf{V}_{N-1}^H\mathbf{V}_{N-1} \end{bmatrix}$$

and

$$\mathbf{V}_k^H\mathbf{V}_l = \frac{1}{N}\sum_{n=0}^{N-1}\mathbf{I}_M\exp(-j2\pi f_k n)\mathbf{I}_M\exp(j2\pi f_l n) = \mathbf{I}_M\delta_{kl}.$$

As a result, we have that

$$\mathbf{V}^H\mathbf{C}\mathbf{V} = \mathbf{P}_{\tilde{x}}^T \tag{13.40}$$

and thus $\mathbf{C}$ has been approximately *block-diagonalized*. If the $M \times M$ blocks of $\mathbf{P}_{\tilde{x}}^T$ are diagonal or if

$$\mathbf{P}_{\tilde{x}\tilde{x}}^T(f_i) = \mathbf{P}_{\tilde{x}\tilde{x}}(f_i) = \begin{bmatrix} P_{00}(f_i) & 0 & \cdots & 0 \\ 0 & P_{11}(f_i) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & P_{M-1,M-1}(f_i) \end{bmatrix}$$

for $i = 0, 1, \ldots, N-1$, then $\mathbf{C}$ is approximately *diagonalized*. The latter case will occur if the cross-PSDs are zero or if $P_{ij}(f) = 0$ for $i \neq j$. In this case the WSS random processes are uncorrelated from sensor to sensor.

We now are able to approximate the estimator-correlator using a WSS assumption for the signal process. The economies in doing so result from the block-diagonal form of $\mathbf{C}_{\tilde{s}}$ that allow a specialization of the estimator-correlator of Section 13.6.5. Let $\mathbf{P}_{\tilde{s}}$ be the block-diagonal matrix defined in (13.39) with $\mathbf{P}_{\tilde{s}\tilde{s}}(f_i)$ replacing $\mathbf{P}_{\tilde{x}\tilde{x}}(f_i)$. Then, from (13.40) we have for a multichannel WSS signal process and a large data record that

$$\mathbf{V}^H\mathbf{C}_{\tilde{s}}\mathbf{V} = \mathbf{P}_{\tilde{s}}^T$$

where $\mathbf{V}$ is a unitary matrix. As a result, the MMSE estimator of $\tilde{\mathbf{s}}$ is

$$\begin{aligned} \hat{\tilde{\mathbf{s}}} &= \mathbf{C}_{\tilde{s}}(\mathbf{C}_{\tilde{s}} + \sigma^2\mathbf{I})^{-1}\tilde{\mathbf{x}} \\ &= \mathbf{V}\mathbf{P}_{\tilde{s}}^T\mathbf{V}^H(\mathbf{V}\mathbf{P}_{\tilde{s}}^T\mathbf{V}^H + \sigma^2\mathbf{I})^{-1}\tilde{\mathbf{x}} \\ &= \mathbf{V}\mathbf{P}_{\tilde{s}}^T\mathbf{V}^H[\mathbf{V}(\mathbf{P}_{\tilde{s}}^T + \sigma^2\mathbf{I})\mathbf{V}^H]^{-1}\tilde{\mathbf{x}} \\ &= \mathbf{V}\mathbf{P}_{\tilde{s}}^T(\mathbf{P}_{\tilde{s}}^T + \sigma^2\mathbf{I})^{-1}\mathbf{V}^H\tilde{\mathbf{x}} \end{aligned}$$

or

$$\mathbf{V}^H\hat{\tilde{\mathbf{s}}} = \mathbf{P}_{\tilde{s}}^T(\mathbf{P}_{\tilde{s}}^T + \sigma^2\mathbf{I})^{-1}\mathbf{V}^H\tilde{\mathbf{x}}.$$

Due to the block-diagonal property of $\mathbf{P}_{\tilde{s}}^T$ we have

$$\mathbf{P}_{\tilde{s}}^T(\mathbf{P}_{\tilde{s}}^T + \sigma^2\mathbf{I})^{-1} =$$

$$\begin{bmatrix} \mathbf{P}_{\tilde{s}\tilde{s}}^T(f_0)(\mathbf{P}_{\tilde{s}\tilde{s}}^T(f_0) + \sigma^2\mathbf{I})^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{\tilde{s}\tilde{s}}^T(f_1)(\mathbf{P}_{\tilde{s}\tilde{s}}^T(f_1) + \sigma^2\mathbf{I})^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{P}_{\tilde{s}\tilde{s}}^T(f_{N-1})(\mathbf{P}_{\tilde{s}\tilde{s}}^T(f_{N-1}) + \sigma^2\mathbf{I})^{-1} \end{bmatrix}.$$

But

$$\begin{aligned} \mathbf{V}^H\tilde{\mathbf{x}} &= \frac{1}{\sqrt{N}}\begin{bmatrix} \mathbf{V}_0^H\tilde{\mathbf{x}} \\ \mathbf{V}_1^H\tilde{\mathbf{x}} \\ \vdots \\ \mathbf{V}_{N-1}^H\tilde{\mathbf{x}} \end{bmatrix} \\ &= \frac{1}{\sqrt{N}}\begin{bmatrix} \sum_{n=0}^{N-1}\tilde{\mathbf{x}}[n]\exp(-j2\pi f_0 n) \\ \sum_{n=0}^{N-1}\tilde{\mathbf{x}}[n]\exp(-j2\pi f_1 n) \\ \vdots \\ \sum_{n=0}^{N-1}\tilde{\mathbf{x}}[n]\exp(-j2\pi f_{N-1} n) \end{bmatrix} \\ &= \frac{1}{\sqrt{N}}\begin{bmatrix} \mathbf{X}(f_0) \\ \mathbf{X}(f_1) \\ \vdots \\ \mathbf{X}(f_{N-1}) \end{bmatrix} \end{aligned}$$

where $\mathbf{X}(f) = \sum_{n=0}^{N-1}\tilde{\mathbf{x}}[n]\exp(-j2\pi f n)$ is the Fourier transform of $\tilde{\mathbf{x}}[n]$ for $n = 0, 1, \ldots, N-1$, and letting $\hat{\mathbf{S}}(f)$ be the Fourier transform of $\hat{\mathbf{s}}[n]$, it follows that

$$\mathbf{V}^H\hat{\tilde{\mathbf{s}}} = \frac{1}{\sqrt{N}}\begin{bmatrix} \hat{\mathbf{S}}(f_0) \\ \hat{\mathbf{S}}(f_1) \\ \vdots \\ \hat{\mathbf{S}}(f_{N-1}) \end{bmatrix}.$$

Combining these results we have that

$$\hat{\mathbf{S}}(f_i) = \mathbf{P}_{\tilde{s}\tilde{s}}^T(f_i)(\mathbf{P}_{\tilde{s}\tilde{s}}^T(f_i) + \sigma^2\mathbf{I})^{-1}\mathbf{X}(f_i) \tag{13.41}$$

for $i = 0, 1, \ldots, N-1$. This is the multichannel Wiener filter estimator of the signal Fourier transform at frequency f_i (see Section 5.5 for the scalar case). The block-diagonal nature of $\mathbf{P}_{\tilde{s}}^T$ has resulted in a decoupling in frequency. Finally, we decide $\mathcal{H}_1$ if

$$\begin{aligned} T(\tilde{\mathbf{x}}) &= \tilde{\mathbf{x}}^H \hat{\tilde{\mathbf{s}}} = \tilde{\mathbf{x}}^H \mathbf{V}\mathbf{V}^H \hat{\tilde{\mathbf{s}}} \\ &= (\mathbf{V}^H \tilde{\mathbf{x}})^H \mathbf{V}^H \hat{\tilde{\mathbf{s}}} \\ &= \frac{1}{N} \sum_{i=0}^{N-1} \mathbf{X}^H(f_i)\hat{\mathbf{S}}(f_i) > \gamma' \end{aligned} \tag{13.42}$$

or if

$$T(\tilde{\mathbf{x}}) = \frac{1}{N} \sum_{i=0}^{N-1} \mathbf{X}^H(f_i)\mathbf{P}_{\tilde{s}\tilde{s}}^T(f_i)(\mathbf{P}_{\tilde{s}\tilde{s}}^T(f_i) + \sigma^2\mathbf{I})^{-1}\mathbf{X}(f_i) > \gamma'. \tag{13.43}$$

If we let $X_m(f)$ be the Fourier transform of the N samples at the mth sensor or $X_m(f) = [\mathbf{X}(f)]_m$ and similarly $\hat{S}_m(f) = [\hat{\mathbf{S}}(f)]_m$, we have from (13.42)

$$T(\tilde{\mathbf{x}}) = \frac{1}{N} \sum_{i=0}^{N-1} \left(\sum_{m=0}^{M-1} X_m^*(f_i)\hat{S}_m(f_i) \right) \tag{13.44}$$

and is shown in Figure 13.7. It is seen that an estimator-correlator operates on the sensor data for a single frequency bin. Then, the estimator-correlator outputs for each frequency bin are averaged together. Note from (13.41), however, that in general $\hat{S}_m(f_i)$ depends on the data at all of the sensors, so that a complete decoupling has not been obtained. Only if the sensor output processes are uncorrelated from sensor to sensor, so that $\mathbf{P}_{\tilde{s}\tilde{s}}(f_i)$ is diagonal, can we claim this. In this instance, then, we have from (13.41) that

$$\hat{S}_m(f_i) = [\hat{\mathbf{S}}(f_i)]_m = \frac{P_{mm}(f_i)}{P_{mm}(f_i) + \sigma^2} X_m(f_i)$$

where $P_{mm}(f)$ is the auto-PSD of the signal at the mth sensor. It follows from (13.44) that

$$\begin{aligned} T(\tilde{\mathbf{x}}) &= \sum_{m=0}^{M-1} \sum_{i=0}^{N-1} \frac{P_{mm}(f_i)}{P_{mm}(f_i) + \sigma^2} \frac{1}{N} |X_m(f_i)|^2 \\ &= \sum_{m=0}^{M-1} \sum_{i=0}^{N-1} \frac{P_{mm}(f_i)}{P_{mm}(f_i) + \sigma^2} I_m(f_i) \\ &= \sum_{m=0}^{M-1} T_m(\tilde{\mathbf{x}}_m) \end{aligned} \tag{13.45}$$

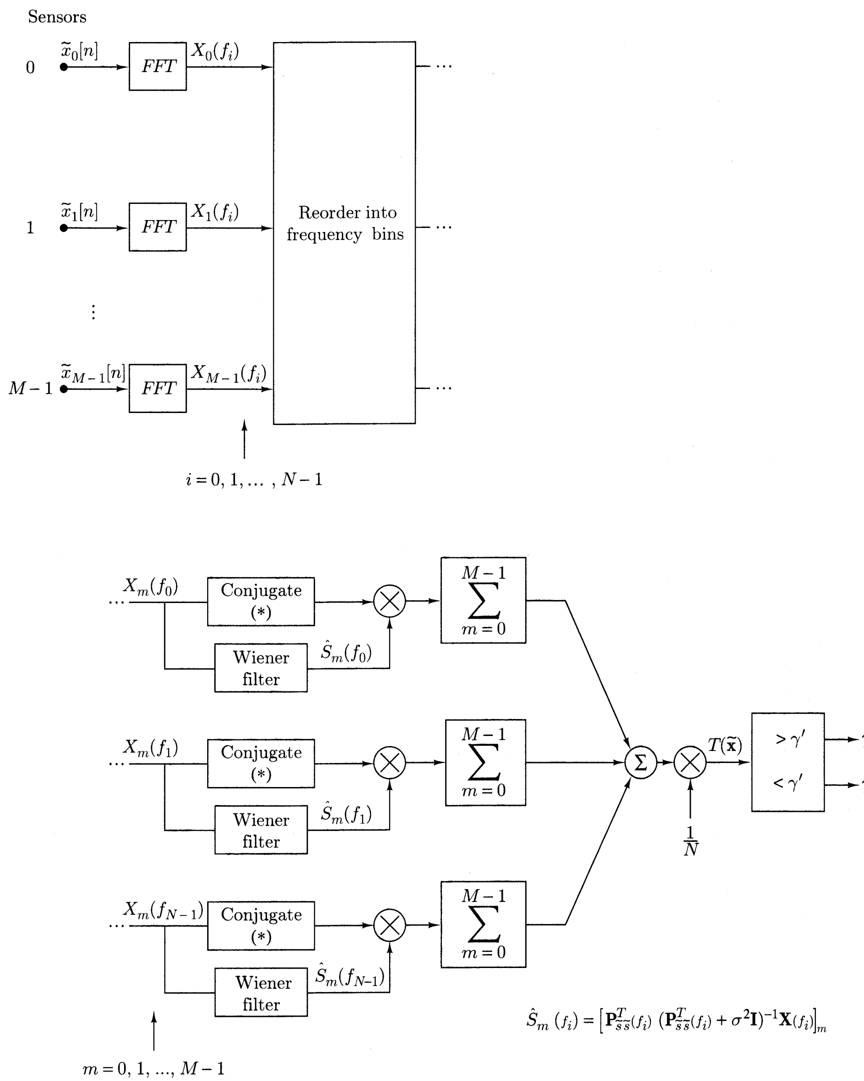


Figure 13.7. Multichannel estimator-correlator for WSS signal process.

where $T_m(\tilde{\mathbf{x}}_m)$ is the estimator-correlator based on a single sensor, which in turn only depends upon the periodogram $I_m(f)$ for that sensor. (The reader may wish to compare (13.45) to (5.27)). In essence we have achieved a decorrelation in time by applying a Fourier transform but only a decorrelation in space if the sensor outputs are also uncorrelated. In typical array processing problems the signal is strongly correlated in space. In fact, this is the principal reason to use multiple sensors – to exploit this correlation.

An important special case occurs when $\mathbf{P}_{\tilde{s}\tilde{s}}(f)$ is Toeplitz or when $P_{ij}(f)$ depends only on $j-i$. This will result from the dependence on $r_{ij}[k]$ on only $j-i$. Then, the data are WSS in space. An example is a line array in which the distance between adjacent sensors is the same (see Problem 13.21). Then, $\mathbf{P}_{\tilde{s}\tilde{s}}(f)$ can be approximately diagonalized by a *spatial* Fourier transform. Alternatively, a 2-D Fourier transform may be used to approximately diagonalize $\mathbf{C}_{\tilde{s}}$. This follows because $\mathbf{C}_{\tilde{s}}$ is now *doubly-block Toeplitz*, which is to say that it is block-Toeplitz and each block is Toeplitz (see Problem 13.22). An example of this case is explored further in the next section.

13.8 Signal Processing Examples

We now utilize some of the previous results in designing detectors for the array processing problem. A major simplification that can now be made stems from the assumption that the signal originates at some point in space (either radiated from a source or reflected by a target) and propagates to the array of sensors. As such, the form of the signal is the same at all sensors except for the delay due to the difference in propagation times. The signal source is assumed to be in the far field so that the signal is constant along a plane, i.e., along a planar wavefront. We will consider an array of sensors with arbitrary spacings and then specialize each detector to a uniformly spaced line array, which is a common geometrical configuration. The direction of propagation of the signal is assumed to be in a plane. The problem that we will address is to detect a signal, either deterministic with unknown parameters or random with known PDF, embedded in noise that is complex Gaussian and uncorrelated in time and space (from sensor to sensor) with variance σ^2. The problem is depicted in Figure 13.8. The source emits a real bandpass signal $s(t)$, which arrives at the mth sensor τ_m seconds later. As a result, the signal at the mth sensor is $s_m(t) = s(t-\tau_m)$. As is customary in processing bandpass signals, the complex envelope of $s_m(t)$ is extracted (see [Kay-I 1993, pp. 495–496]). The complex envelope of a bandpass signal $s(t)$ is the complex low-pass signal $\tilde{s}(t)$ defined in the relationship

$$s(t) = 2\mathrm{Re}(\tilde{s}(t)\exp(j2\pi F_0 t))$$

where F_0 is the center frequency of the bandpass signal in Hz. At the mth sensor we have

$$s(t-\tau_m) \;=\; 2\mathrm{Re}(\tilde{s}(t-\tau_m)\exp(j2\pi F_0(t-\tau_m)))$$

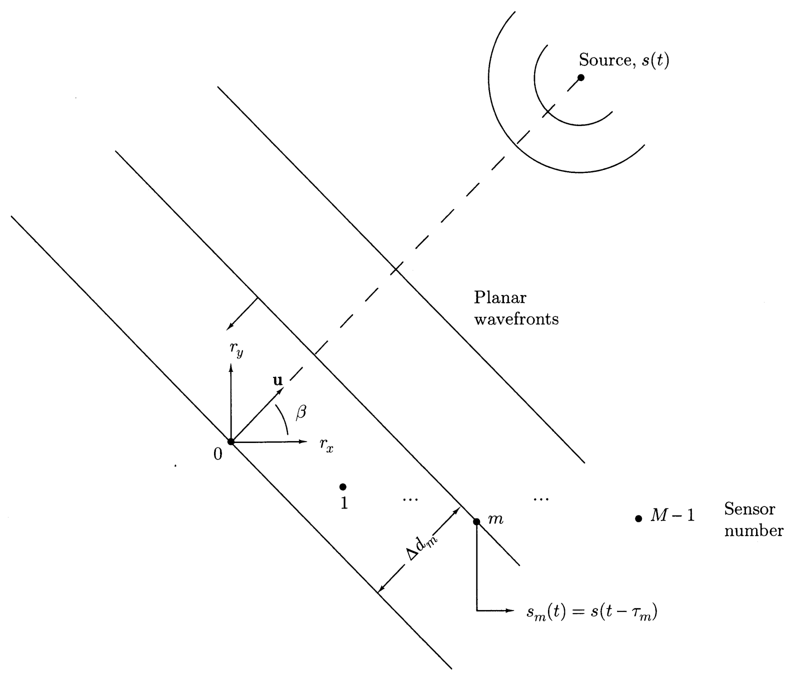


Figure 13.8. Array processing problem.

$$= \;2\mathrm{Re}(\tilde{s}(t-\tau_m)\exp(-j2\pi F_0\tau_m)\exp(j2\pi F_0 t)).$$

The complex envelope at the mth sensor is seen to be

$$\tilde{s}_m(t) = \tilde{s}(t-\tau_m)\exp(-j2\pi F_0\tau_m). \tag{13.46}$$

Based on the positions of the sensors, we can relate τ_m to τ_0 as follows. Let $\tau_m = \tau_0 + \Delta\tau_m$, where $\Delta\tau_m$ is the additional time required for the signal to propagate from the zeroth to the mth sensor. The additional delay is just the distance between the zeroth and mth sensors *along the direction of propagation* divided by the propagation speed c or

$$\begin{aligned}\Delta\tau_m &= \frac{\Delta d_m}{c} \\ &= -\frac{(\mathbf{r}_m - \mathbf{r}_0)^T\mathbf{u}}{c}\end{aligned}$$

$$= -\frac{\mathbf{r}_m^T\mathbf{u}}{c}$$

where $\mathbf{r}_m$ is the position of the mth sensor ($\mathbf{r}_0 = [0\,0]^T$) and $\mathbf{u} = [\cos\beta\ \sin\beta]^T$ is a unit vector opposite to the direction of propagation. Now, from (13.46) the received complex envelope signal at the mth sensor is

$$\tilde{s}_m(t) = \tilde{s}(t - \tau_0 + \mathbf{r}_m^T\mathbf{u}/c)\exp\left[-j2\pi F_0(\tau_0 - \mathbf{r}_m^T\mathbf{u}/c)\right]. \tag{13.47}$$

Note that the delay depends not only on the sensor location but also on the source location in azimuth or β. We now examine two cases of interest. The first case assumes that the signal is deterministic with unknown parameters, typical of an active sonar or radar system, while the second case assumes that the signal is a random process that is WSS with known PSD, typical of a broadband passive sonar system.

13.8.1 Active Sonar/Radar

We assume a sinusoidal transmitted signal with frequency F_0 Hz. The reflected signal is then also sinusoidal but with frequency F_0+F_D Hz, due to the Doppler shift of a moving target. The signal at the point of reflection is then $s(t) = B\cos[2\pi(F_0+F_D)t+\psi)]$ with a corresponding complex envelope $\tilde{s}(t) = (B/2)\exp[j(2\pi F_D t+\psi)]$. Hence, from (13.47) the complex envelope at the mth sensor is

$$\tilde{s}_m(t) = (B/2)\exp\left[j(2\pi F_D(t-\tau_0+\mathbf{r}_m^T\mathbf{u}/c)+\psi)\right]\exp\left[-j2\pi F_0(\tau_0 - \mathbf{r}_m^T\mathbf{u}/c)\right].$$

If we sample the complex envelope at $t_n = n\Delta$, then we have at the mth sensor the signal $\tilde{s}_m[n] = \tilde{s}_m(n\Delta)$. Now let $F_D\Delta = f_D$ and $F_0\Delta = f_0$ be the frequencies in cycles/sample (or a fraction of the sampling frequency) and also define $n_m(\beta) = -\mathbf{r}_m^T\mathbf{u}/(c\Delta)$ as the additional propagation delay to the mth sensor in samples (assumed to be an integer). Then, we have

$$\tilde{s}_m[n] = (B/2)\exp\left[j(2\pi f_D n+\phi)\right]\exp\left[-j2\pi(f_0+f_D)n_m(\beta)\right]$$

where $\phi = -2\pi(F_0+F_D)\tau_0+\psi$. Letting $B/2 = A$ and $f_1 = f_0+f_D$ we have finally that

$$\tilde{s}_m[n] = A\exp\left[j(2\pi f_D n+\phi)\right]\exp\left[-j2\pi f_1 n_m(\beta)\right]. \tag{13.48}$$

It is seen that the received signal at the mth sensor is sinusoidal in time with amplitude A, frequency f_D, and phase ϕ, and additionally there is a *differential phase* $-2\pi f_1 n_m(\beta)$ that depends on the sensor number. Of course, the latter is due to the additional propagation time to the mth sensor. We will assume that the signal length is N samples. Then the time samples for which the signal is present at the mth sensor are

$$n = \tau_0/\Delta + n_m(\beta), \tau_0/\Delta + n_m(\beta)+1, \tau_0/\Delta + n_m(\beta) + N - 1$$

since the delay in samples to the mth sensor is $(\tau_0 - \mathbf{r}_m^T\mathbf{u}/c)/\Delta$. In order to observe the signal at all of the sensors, our data window length must be the signal length plus the additional delay between the first arriving signal and the last arriving signal or the time interval $[\tau_0/\Delta + n_m(\beta)_{\min}, \tau_0/\Delta + n_m(\beta)_{\max} + N - 1]$. If $n_m(\beta)_{\max} - n_m(\beta)_{\min}$ is small relative to N, then the observation interval $[\tau_0/\Delta + n_m(\beta)_{\min}, \tau_0/\Delta + n_m(\beta)_{\min} + N - 1]$ will include most of the signal samples at each sensor. Making this assumption and referencing the first sample to $\tau_0/\Delta + n_m(\beta)_{\min}$, we have from (13.48) our signal model

$$\tilde{s}_m[n] = A\exp\left[j(2\pi(f_D n - f_1 n_m(\beta)) + \phi)\right] \qquad n = 0,1,\ldots,N-1 \tag{13.49}$$

and for $m = 0,1,\ldots,M-1$. It is assumed that the only unknown parameters are A and ϕ. For this to be valid we must know f_D, the Doppler frequency, β, the arrival angle, and τ_0/Δ, the arrival time of the signal at the zeroth sensor. In practice, these are seldom known, and so the GLRT statistic that we will derive will need to be maximized over these additional parameters. Assuming that the noise is CWGN in time and space or all the samples are uncorrelated and have variance σ^2, we can apply the GLRT of Section 13.6.6. We need only obtain $\mathbf{H}$ based on (13.49). As in Section 13.6.6, we arrange the signal samples in temporal order as

$$\tilde{\mathbf{s}} = \begin{bmatrix} \tilde{\mathbf{s}}[0] \\ \vdots \\ \tilde{\mathbf{s}}[N-1] \end{bmatrix} = \underbrace{\begin{bmatrix} \exp[-j2\pi f_1 n_0(\beta)] \\ \vdots \\ \exp[-j2\pi f_1 n_{M-1}(\beta)] \\ \cdots\cdots\cdots\cdots \\ \vdots \\ \cdots\cdots\cdots\cdots \\ \exp[j2\pi(f_D(N-1) - f_1 n_0(\beta))] \\ \vdots \\ \exp[j2\pi(f_D(N-1) - f_1 n_{M-1}(\beta))] \end{bmatrix}}_{\mathbf{H}} \underbrace{A\exp(j\phi)}_{\boldsymbol{\theta}}$$

where $\mathbf{H}$ has dimension $MN \times 1$. Noting that $\mathbf{H}^H\mathbf{H} = MN$, we have from (13.33) that the GLRT decides $\mathcal{H}_1$ if

$$T(\tilde{\mathbf{x}}) = \frac{MN\hat{\boldsymbol{\theta}}_1^H\hat{\boldsymbol{\theta}}_1}{\sigma^2/2} > \gamma'.$$

But $\boldsymbol{\theta}$ is a complex scalar and thus

$$T(\tilde{\mathbf{x}}) = \frac{MN|\hat{\theta}_1|^2}{\sigma^2/2}$$

where

$$\hat{\theta}_1 = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\tilde{\mathbf{x}} = \mathbf{H}^H\tilde{\mathbf{x}}/(MN)$$

and

$$\mathbf{H}^H\tilde{\mathbf{x}} = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} \tilde{x}_m[n]\exp[-j2\pi(f_D n - f_1 n_m(\beta))].$$

Then

$$T(\tilde{\mathbf{x}}) = \frac{1}{MN\sigma^2/2}\left|\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} \tilde{x}_m[n]\exp[-j2\pi(f_D n - f_1 n_m(\beta))]\right|^2. \tag{13.50}$$

The test statistic can be expressed in a slightly more intuitive form by letting

$$\tilde{x}_B[n] = \frac{1}{M}\sum_{m=0}^{M-1} \tilde{x}_m[n]\exp[j2\pi f_1 n_m(\beta)] \tag{13.51}$$

so that

$$T(\tilde{\mathbf{x}}) = \frac{M}{\sigma^2/2}\frac{1}{N}\left|\sum_{n=0}^{N-1} \tilde{x}_B[n]\exp(-j2\pi f_D n)\right|^2. \tag{13.52}$$

Now we observe that the test statistic is just the scaled periodogram evaluated at $f = f_D$ of the $\tilde{x}_B[n]$ data. To discern the effect of the combining operation of (13.51), we note that the received signal at the mth sensor is

$$\tilde{s}_m[n] = \exp[-j2\pi f_1 n_m(\beta)]A\exp[j(2\pi f_D n + \phi)]$$

which is a sinusoid whose phase varies according to the sensor number. Of course, this phase difference is due to the disparity in propagation delays. Now if $\tilde{x}_m[n] = \tilde{s}_m[n] + \tilde{w}_m[n]$, we have

$$\begin{aligned}\tilde{x}_B[n] &= \frac{1}{M}\sum_{m=0}^{M-1} \tilde{s}_m[n]\exp[j2\pi f_1 n_m(\beta)] + \frac{1}{M}\sum_{m=0}^{M-1} \tilde{w}_m[n]\exp[j2\pi f_1 n_m(\beta)] \\ &= A\exp[j(2\pi f_D n + \phi)] + \frac{1}{M}\sum_{m=0}^{M-1} \tilde{w}_m[n]\exp[j2\pi f_1 n_m(\beta)]. \end{aligned} \tag{13.53}$$

The effect of this combining operation is to phase all of the sensor output signals before adding them, in effect, nullifying the phase differences due to propagation. This processing step is known as *beamforming*. It can be shown to be a spatial filtering operation or equivalently a matched filter applied to the spatial data [Knight, Pridham, and Kay 1981]. The utility is that the signals are added in phase while the noise is not. As a result, the SNR is enhanced by the spatial averaging. This improvement is called the *array gain* and is analogous to the processing gain for temporal filtering (see Chapter 4). To determine the improvement define the array gain (AG) as the ratio of the output SNR η_{out} to the input SNR η_{in} or

$$\text{AG} = 10\log_{10}\frac{\eta_{\text{out}}}{\eta_{\text{in}}} \quad \text{dB}.$$

The input SNR is $\eta_{\text{min}} = A^2/\sigma^2$ while at the output of the beamformer the SNR is from (13.53)

$$\eta_{\text{out}} = \frac{A^2}{\text{var}\left(\frac{1}{M}\sum_{m=0}^{M-1}\tilde{w}_m[n]\exp(j2\pi f_1 n_m(\beta))\right)}.$$

Since $\tilde{w}_m[n]$ is IID in m (being CWGN in space), we have that

$$\begin{aligned}\text{var}\left(\frac{1}{M}\sum_{m=0}^{M-1}\tilde{w}_m[n]\exp(j2\pi f_1 n_m(\beta))\right) &= \frac{1}{M^2}\sum_{m=0}^{M-1}\text{var}\left[\tilde{w}_m[n]\exp(j2\pi f_1 n_m(\beta))\right] \\ &= \frac{1}{M^2}\sum_{m=0}^{M-1} E(|\tilde{w}_m[n]|^2) = \sigma^2/M\end{aligned}$$

so that AG $= 10\log_{10} M$ dB. The effect of beamforming is to improve the SNR by a factor of M, where M is the number of sensors. The entire detector is shown in Figure 13.9. In order to implement beamforming we require knowledge of the arrival angle β of the signal. Note that the beamformer effects spatial processing, while the periodogram or quadrature matched filter effects temporal processing.

The performance of the detector follows from (13.36) and (13.37) with $p = 1$ as

$$\begin{aligned}P_{FA} &= Q_{\chi_2^2}(\gamma') = \exp(-\gamma'/2) \\ P_D &= Q_{\chi_2'^2(\lambda)}(\gamma')\end{aligned}$$

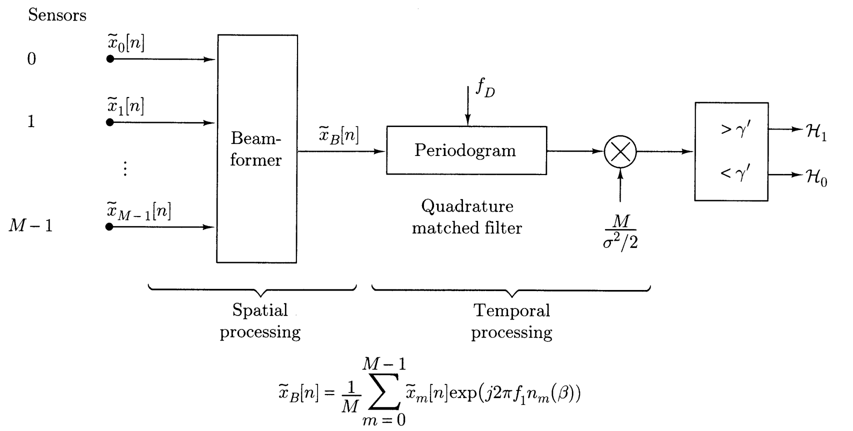


Figure 13.9. Active sonar/radar array detector for sinusoidal signal.

where

$$\begin{aligned}\lambda &= \frac{\boldsymbol{\theta}_1^H \mathbf{H}^H \mathbf{H} \boldsymbol{\theta}_1}{\sigma^2/2} = \frac{\tilde{\mathbf{s}}^H \tilde{\mathbf{s}}}{\sigma^2/2} \\ &= \frac{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} |\tilde{s}_m[n]|^2}{\sigma^2/2} = \frac{MNA^2}{\sigma^2/2}.\end{aligned}$$

Note that the noncentrality parameter is increased by MN or $10\log_{10} M + 10\log_{10} N$ dB over the single sample case. The first term is the array gain while the second term is the processing gain.

We next specialize the results to the uniformly spaced line array shown in Figure 13.10. The only difference from the more general case is that the delays can be explicitly found as

$$n_m(\beta) = -m\frac{d}{c\Delta}\cos\beta$$

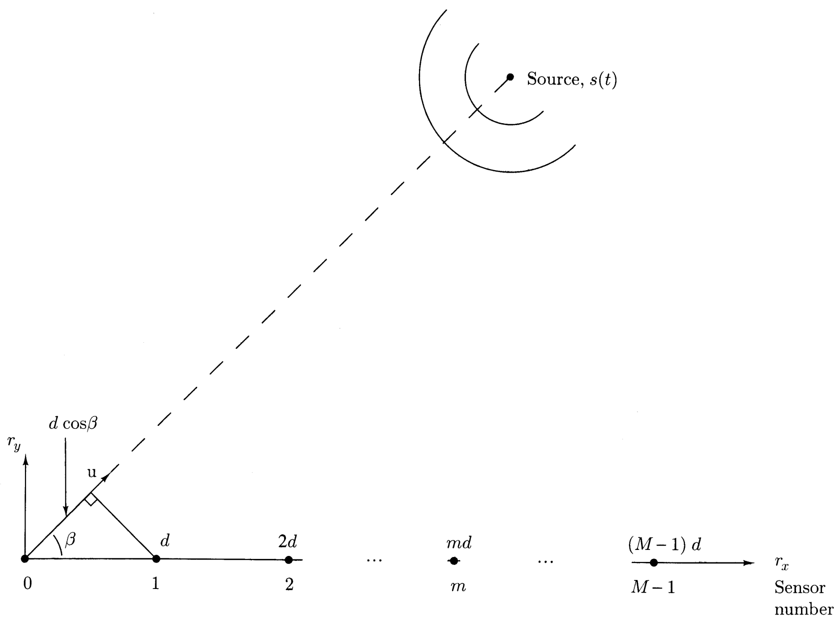


Figure 13.10. Geometry for uniformly spaced line array.

where we assume that the delays are integer multiples of Δ so that $n_m(\beta)$ is an integer. The received signal then becomes from (13.49)

$$\tilde{s}_m[n] = A\exp\left[j(2\pi(f_D n + f_1\frac{d}{c\Delta}\cos\beta\, m) + \phi)\right]$$

which is a 2-D sinusoid with temporal frequency f_D and spatial frequency $f_s = f_1 d/(c\Delta)\cos\beta$. The detector of (13.50) reduces to

$$T(\tilde{\mathbf{x}}) = \frac{1}{\sigma^2/2}\frac{1}{NM}\left|\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\tilde{x}_m[n]\exp\left[-j2\pi(f_s m + f_D n)\right]\right|^2. \tag{13.54}$$

The test statistic is a scaled 2-D periodogram or quadrature matched filter evaluated at the known temporal frequency f_D and known spatial frequency f_s. In practice, f_D and β are unknown and thus, f_s and f_D needed in (13.54) are unknown. The GLRT then picks the peak of a normalized 2-D periodogram as the detection statistic. In this case a 2-D FFT can be used to efficiently evaluate the periodogram. Lastly, since the arrival time is also unknown, the entire detector must process sequential blocks of data in time, which are usually overlapped. In this case the increase in P_{FA} will be approximately linear with the number of frequency bins, arrival angles, and blocks used (see Example 7.5). The detection probability will be the same as for the known arrival time, arrival angle, and frequency case. See also [Knight, Pridham, and Kay 1981] for additional practical considerations.

13.8.2 Broadband Passive Sonar

We now assume that the complex envelope of the emitted signal is a broadband complex WSS Gaussian random process. The noise is assumed to be CWGN as before. As such we would expect an estimator-correlator structure for the detector. To simplify the derivation and results we further assume that the data record length is large enough so that we can use the asymptotic approximation of Section 13.7. This will hold whenever the data record length is much larger than the correlation time of the signal, which is the effective length of the signal ACF. The emitted signal is modeled as a bandpass zero mean real Gaussian random process with center frequency F_0. We will denote the complex envelope of the emitted signal as $\tilde{s}[n]$, and note that it will be a zero mean complex Gaussian random process, which is assumed to be WSS. The PSD of $\tilde{s}[n]$ is *assumed known* and is given by $P_{\tilde{s}\tilde{s}}(f)$. To apply the results of Section 13.7 we need only determine the CSM of the signal or $\mathbf{P}_{\tilde{s}\tilde{s}}(f)$. Then, from (13.43) we decide $\mathcal{H}_1$ if

$$T(\tilde{\mathbf{x}}) = \frac{1}{N}\sum_{i=0}^{N-1}\mathbf{X}^H(f_i)\mathbf{P}_{\tilde{s}\tilde{s}}^T(f_i)(\mathbf{P}_{\tilde{s}\tilde{s}}^T(f_i) + \sigma^2\mathbf{I})^{-1}\mathbf{X}(f_i) > \gamma'$$

where

$$[\mathbf{X}(f_i)]_m = \sum_{n=0}^{N-1} \tilde{x}_m[n] \exp(-j2\pi f_i n)$$

is the discrete Fourier transform of the data at the output of the mth sensor. To find the CSM we use

$$[\mathbf{P}_{\tilde{s}\tilde{s}}(f)]_{lm} = P_{lm}(f) = \mathcal{F}\{r_{lm}[k]\}$$

where $r_{lm}[k] = E(\tilde{s}_l^*[n]\tilde{s}_m[n+k])$ is the cross-correlation between the lth and mth sensors. The received complex envelope signal at the mth sensor is from (13.47)

$$\begin{aligned}\tilde{s}_m[n] &= \tilde{s}_m(n\Delta)\\ &= \tilde{s}(n\Delta - \tau_0 - n_m(\beta)\Delta)\exp\left[-j2\pi F_0(\tau_0 + n_m(\beta)\Delta)\right]\end{aligned}$$

where we have used $n_m(\beta) = -\mathbf{r}_m^T\mathbf{u}/(c\Delta)$. Since $\tilde{s}[n] = \tilde{s}(n\Delta)$, we have

$$\tilde{s}_m[n] = \tilde{s}[n - n_m(\beta) - \tau_0/\Delta]\exp\left[-j2\pi F_0(\tau_0 + n_m(\beta)\Delta)\right].$$

The observation interval should be chosen so that the first arriving signal and the last arriving signal are included in the data set. Also, the data record length, after alignment to compensate for the time delays, should be as large as possible. In practice, it is limited by the stationarity of the signal process. The ACF becomes

$$\begin{aligned}r_{lm}[k]\\ &= E(\tilde{s}^*[n - n_l(\beta) - \tau_0/\Delta]\tilde{s}[n + k - n_m(\beta) - \tau_0/\Delta])\\ &\qquad \cdot \exp\left[j2\pi F_0\Delta(n_l(\beta) - n_m(\beta))\right]\\ &= r_{\tilde{s}\tilde{s}}[k + n_l(\beta) - n_m(\beta)]\exp\left[j2\pi F_0\Delta(n_l(\beta) - n_m(\beta))\right]\end{aligned}$$

and therefore

$$\begin{aligned}[\mathbf{P}_{\tilde{s}\tilde{s}}(f)]_{lm} &= \mathcal{F}\{r_{lm}[k]\}\\ &= P_{\tilde{s}\tilde{s}}(f)\exp\left[j2\pi(f + F_0\Delta)(n_l(\beta) - n_m(\beta))\right].\end{aligned}$$

Hence, in matrix form we have that

$$\mathbf{P}_{\tilde{s}\tilde{s}}(f) = P_{\tilde{s}\tilde{s}}(f)\mathbf{e}^*(\beta)\mathbf{e}^T(\beta)$$

where $\mathbf{e}(\beta) = [1 \ \exp[-j2\pi(f+F_0\Delta)n_1(\beta)] \ldots \exp[-j2\pi(f+F_0\Delta)n_{M-1}(\beta)]]^T$ (recall that $n_0(\beta) = 0$) or that

$$\mathbf{P}_{\tilde{s}\tilde{s}}^T(f) = P_{\tilde{s}\tilde{s}}(f)\mathbf{e}(\beta)\mathbf{e}^H(\beta).$$

Note that the CSM has rank one, allowing us to easily invert $\mathbf{P}_{\tilde{s}\tilde{s}}^T(f) + \sigma^2\mathbf{I}$ using Woodbury's identity. Using (13.16) and letting $\mathbf{e}_i(\beta)$ denote $\mathbf{e}(\beta)$ for $f = f_i$, it follows that

$$\begin{aligned}(\mathbf{P}_{\tilde{s}\tilde{s}}^T(f_i) + \sigma^2\mathbf{I})^{-1} &= (\sigma^2\mathbf{I} + P_{\tilde{s}\tilde{s}}(f_i)\mathbf{e}_i(\beta)\mathbf{e}_i^H(\beta))^{-1}\\ &= \frac{1}{\sigma^2}\mathbf{I} - \frac{1}{\sigma^4}\frac{P_{\tilde{s}\tilde{s}}(f_i)\mathbf{e}_i(\beta)\mathbf{e}_i^H(\beta)}{1 + MP_{\tilde{s}\tilde{s}}(f_i)/\sigma^2}\end{aligned}$$

and

$$\begin{aligned}\mathbf{P}_{\tilde{s}\tilde{s}}^T(f_i)(\mathbf{P}_{\tilde{s}\tilde{s}}^T(f_i) + \sigma^2\mathbf{I})^{-1} &= P_{\tilde{s}\tilde{s}}(f_i)\mathbf{e}_i(\beta)\mathbf{e}_i^H(\beta)\left[\frac{1}{\sigma^2}\mathbf{I} - \frac{1}{\sigma^4}\frac{P_{\tilde{s}\tilde{s}}(f_i)\mathbf{e}_i(\beta)\mathbf{e}_i^H(\beta)}{1 + MP_{\tilde{s}\tilde{s}}(f_i)/\sigma^2}\right]\\ &= \left(1 - \frac{MP_{\tilde{s}\tilde{s}}(f_i)/\sigma^2}{1 + MP_{\tilde{s}\tilde{s}}(f_i)/\sigma^2}\right)\frac{P_{\tilde{s}\tilde{s}}(f_i)}{\sigma^2}\mathbf{e}_i(\beta)\mathbf{e}_i^H(\beta)\\ &= \frac{P_{\tilde{s}\tilde{s}}(f_i)}{P_{\tilde{s}\tilde{s}}(f_i) + \sigma^2/M}\frac{1}{M}\mathbf{e}_i(\beta)\mathbf{e}_i^H(\beta).\end{aligned}$$

The MMSE estimator of the signal is from (13.41)

$$\hat{\mathbf{S}}(f_i) = \frac{P_{\tilde{s}\tilde{s}}(f_i)}{P_{\tilde{s}\tilde{s}}(f_i) + \sigma^2/M}\frac{\mathbf{e}_i^H(\beta)\mathbf{X}(f_i)}{M}\mathbf{e}_i(\beta) \qquad (13.55)$$

or for the mth sensor

$$\begin{aligned}\hat{S}_m(f_i) &= \frac{P_{\tilde{s}\tilde{s}}(f_i)}{P_{\tilde{s}\tilde{s}}(f_i) + \sigma^2/M}\left(\underbrace{\frac{1}{M}\sum_{m=0}^{M-1} X_m(f_i)\exp[j2\pi(F_0\Delta + f_i)n_m(\beta)]}_{\hat{S}_B(f_i)}\right)\\ &\qquad \cdot \exp[-j2\pi(F_0\Delta + f_i)n_m(\beta)].\end{aligned} \qquad (13.56)$$

This is the MMSE estimator for the Fourier transform of the complex envelope or low-pass signal at $f = f_i$ for the mth sensor. Note that it consists of a narrowband *frequency domain beamformer* that forms $\hat{S}_B(f_i)$ by properly phasing and then averaging all sinusoidal components at $f = f_i$. This is necessary to compensate for the phase delay caused by the $n_m(\beta)$ time delay at each sensor. Next, the proper phasing is included via the $\exp[-j2\pi(F_0\Delta + f_i)n_m(\beta)]$ factor. Finally, the term

$$H(f_i) = \frac{P_{\tilde{s}\tilde{s}}(f_i)}{P_{\tilde{s}\tilde{s}}(f_i) + \sigma^2/M}$$

is a temporal Wiener filter necessary to filter out as much of the CWGN as possible. After beamforming, the noise power is reduced by a factor of M due to the array gain and hence the σ^2/M term in the Wiener filter. The overall detector becomes from (13.43)

$$\begin{aligned} T(\tilde{\mathbf{x}}) &= \frac{1}{N}\sum_{i=0}^{N-1}\mathbf{X}^H(f_i)\hat{\mathbf{S}}(f_i) &(13.57)\\ &= \frac{1}{N}\sum_{i=0}^{N-1}\left(\sum_{m=0}^{M-1}X_m^*(f_i)\hat{S}_m(f_i)\right). &(13.58)\end{aligned}$$

Effectively, the use of frequency samples has decoupled the detection problem into one that computes the estimator-correlator based on a narrowband signal $X_m^*(f_i)\hat{S}_m(f_i)$ (for which we have only one sample at a given frequency), sums over all sensors according to $T_i(\tilde{\mathbf{x}}) = \sum_{m=0}^{M-1} X_m^*(f_i)\hat{S}_m(f_i)$, and averages the results over frequency. The entire detector is shown in Figure 13.11. An alternative representation is from (13.57) and (13.55)

$$\begin{aligned} T(\tilde{\mathbf{x}}) &= \frac{1}{N}\sum_{i=0}^{N-1}\mathbf{X}^H(f_i)H(f_i)\mathbf{e}_i(\beta)\frac{\mathbf{e}_i^H(\beta)}{M}\mathbf{X}(f_i)\\ &= \frac{1}{NM}\sum_{i=0}^{N-1}H(f_i)|\mathbf{e}_i^H(\beta)\mathbf{X}(f_i)|^2\\ &= \frac{M}{N}\sum_{i=0}^{N-1}H(f_i)\left|\frac{1}{M}\sum_{m=0}^{M-1}X_m(f_i)\exp[j2\pi(F_0\Delta+f_i)n_m(\beta)]\right|^2 \end{aligned}$$

(13.59)

which is seen to consist of a frequency domain beamformer, squarer, Wiener filter, and frequency averager [Van Trees 1966].

In the special case of a uniformly spaced line array we have that $n_m(\beta) = -md/(c\Delta)\cos\beta$. The beamformer becomes from (13.56)

$$\begin{aligned} \hat{S}_B(f_i) &= \frac{1}{M}\sum_{m=0}^{M-1}X_m(f_i)\exp[-j2\pi(F_0\Delta+f_i)d/(c\Delta)\cos\beta\, m]\\ &= \frac{1}{M}\sum_{m=0}^{M-1}X_m(f_i)\exp(-j2\pi f_s m)\\ &= \frac{1}{M}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\tilde{x}_m[n]\exp[-j2\pi(f_s m+f_i n)] &(13.60)\end{aligned}$$

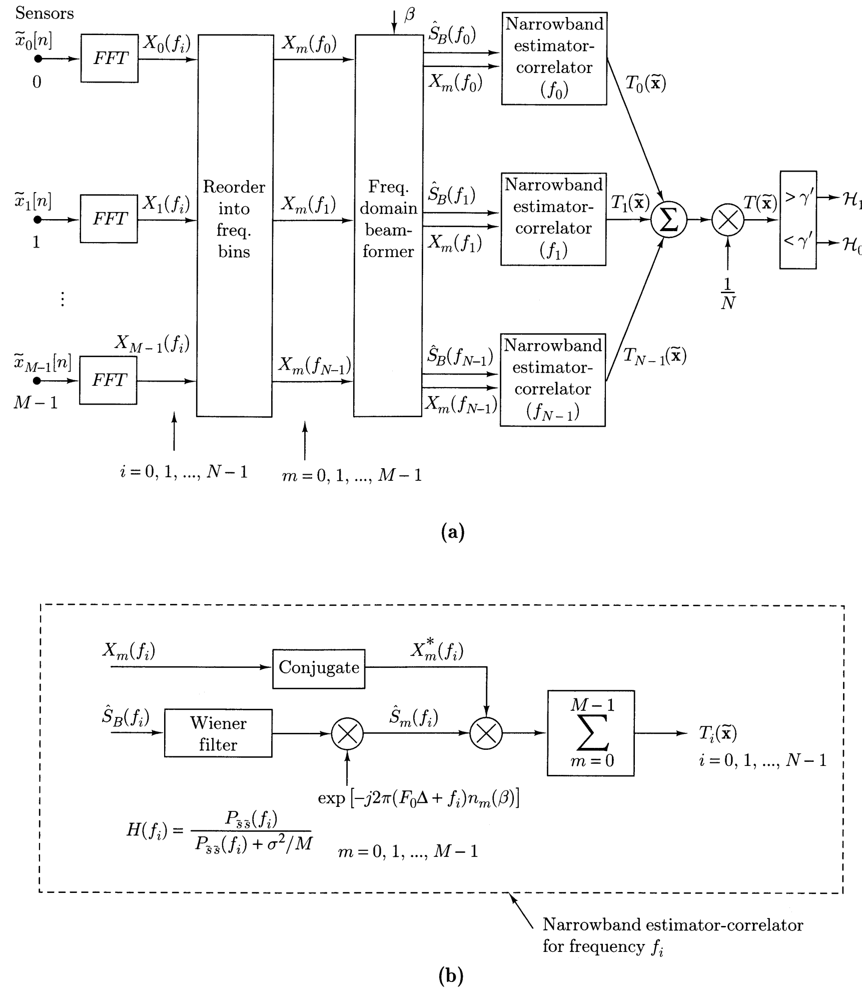


Figure 13.11. Broadband passive sonar array detector.

which is the scaled 2-D discrete Fourier transform of the 2-D data set $\tilde{x}_m[n]$ with $f_s = (F_0\Delta + f_i)d/(c\Delta)\cos\beta$ being the spatial frequency and f_i being the temporal frequency.

In practice, F_0 and β are unknown so that the GLRT must maximize over these parameters. Also, the length of the observation interval N is determined by the assumed stationarity of the source. See also [Knight, Pridham, and Kay 1981] for additional practical considerations.

References

Graybill, F.A., *Matrices with Applications in Statistics*, Wadsworth, Pacific Grove, Ca., 1969.

Hannan, E.J. *Multiple Time Series*, J. Wiley, New York, 1970.

Kay, S.M., *Modern Spectral Estimation: Theory and Application*, Prentice-Hall, Englewood Cliffs, N.J., 1988.

Knight, W.S., R.G. Pridham, S.M. Kay, "Digital Signal Processing for Sonar," *Proc. IEEE*, Vol. 69, pp. 1451–1506, Nov. 1981.

Robinson, E.A., *Multichannel Time Series Analysis with Digital Computer Programs*, Holden-Day, San Francisco, Ca. 1967.

Van Trees, H.L., "A Unified Theory for Optimum Array Processing," Report No. 4160866, Arthur D. Little, Inc., Cambridge, Mass., August 1966.

Van Trees, H.L., *Detection, Estimation, and Modulation Theory*, Vol. III, J. Wiley, New York, 1971.

Problems

13.1 A complex sinusoid $\tilde{s}[n] = \tilde{A}\exp(j2\pi f_0 n)$ is to be detected in CWGN with known variance σ^2 based on $\tilde{x}[n]$ for $n = 0, 1, \ldots, N-1$. If $\tilde{A}$ and f_0 are known, determine the NP detector (including the threshold) and its performance.

13.2 Express (13.3) in matched filter form by letting

$$T(\tilde{\mathbf{x}}) = \operatorname{Re}\left.\left(\sum_{k=0}^{n} \tilde{h}[n-k]\tilde{x}[k]\right)\right|_{n=N-1}$$

and determining $\tilde{h}[n]$.

13.3 In this problem we rederive the results of Example 13.1 using a real data formulation. To do so, let the complex observations $\tilde{x}[n] = u[n] + jv[n]$ for $n = 0, 1, \ldots, N-1$ be arranged in a real vector as $\mathbf{x} = [\mathbf{u}^T\,\mathbf{v}^T]^T$, where $\mathbf{u} = [u[0]\,u[1]\ldots u[N-1]]^T$ and $\mathbf{v} = [v[0]\,v[1]\ldots v[N-1]]^T$. Find the NP test statistic for the detection problem stated in Example 13.1 by using an LRT based on the real data vector $\mathbf{x}$. Then, show that the statistics are identical. Hint: The data vector $\mathbf{x}$ will have a real multivariate PDF under $\mathcal{H}_0$ and $\mathcal{H}_1$.

13.4 Let $\tilde{x}[0] = A\exp(j\phi) + \tilde{w}[0]$, where A, ϕ are known constants and $\tilde{w}[0] \sim \mathcal{CN}(0, \sigma^2)$. Consider the PDF for the 2×1 real random vector $[\operatorname{Re}(\tilde{x}[0])\,\operatorname{Im}(\tilde{x}[0])]^T$. Plot the PDF contours under $\mathcal{H}_0$ for which $A = 0$, and also under $\mathcal{H}_1$ for the values $A = 1$, $\phi = 0$ and $A = 1$, $\phi = \pi/2$. Explain why the value of ϕ does not affect the detection performance of an NP detector.

13.5 Verify (13.11) and (13.12) by finding the moments of $\tilde{z} = \tilde{\mathbf{s}}^H\mathbf{C}^{-1}\tilde{\mathbf{s}}$.

13.6 For a complex Gaussian random signal with mean zero and known covariance matrix $\mathbf{C}_{\tilde{s}} = \sigma_{\tilde{s}}^2\mathbf{I}$ embedded in CWGN with known variance σ^2, find the NP detection statistic. Explain your results.

13.7 Show that

a. $(\mathbf{AB})^H = \mathbf{B}^H\mathbf{A}^H$

b. $\mathbf{A} + \mathbf{B}$ is Hermitian if $\mathbf{A}$ and $\mathbf{B}$ are Hermitian

c. $\mathbf{A}^{-1}$ is Hermitian if $\mathbf{A}$ is Hermitian

by using the properties of matrix transposition. Then, show that $\mathbf{C}_{\tilde{s}}(\mathbf{C}_{\tilde{s}} + \sigma^2\mathbf{I})^{-1}$ is a Hermitian matrix. Hint: $\mathbf{C}_{\tilde{s}}$ and $(\mathbf{C}_{\tilde{s}} + \sigma^2\mathbf{I})^{-1}$ commute.

13.8 In this problem we determine the canonical detector for the complex estimator-correlator and its performance, extending the results of Section 5.3. Since $\mathbf{C}_{\tilde{s}}$ is a Hermitian and positive semidefinite matrix, it can be diagonalized as $\mathbf{V}^H\mathbf{C}_{\tilde{s}}\mathbf{V} = \mathbf{\Lambda}_{\tilde{s}}$. The $N \times N$ matrix $\mathbf{V}$ is the complex modal matrix or $\mathbf{V} = [\mathbf{v}_0\,\mathbf{v}_1\ldots\mathbf{v}_{N-1}]$, where $\mathbf{v}_i$ is the ith $N \times 1$ eigenvector, and the $N \times N$ matrix $\mathbf{\Lambda}_{\tilde{s}}$ is diagonal with $[n, n]$ element $\lambda_{\tilde{s}_n}$, where $\lambda_{\tilde{s}_n}$ is the nth eigenvalue. Since $\mathbf{C}_{\tilde{s}}$ is Hermitian, it follows that the eigenvectors can be chosen to be orthogonal so that $\mathbf{V}^H = \mathbf{V}^{-1}$, and also that the eigenvalues are real. Furthermore, because of the positive semidefinite property of $\mathbf{C}_{\tilde{s}}$, the eigenvalues are all non-negative. We further assume that the eigenvalues are distinct. Show that the estimator-correlator of (13.14) may be expressed in canonical form as

$$T(\tilde{\mathbf{x}}) = \sum_{n=0}^{N-1} \frac{\lambda_{\tilde{s}_n}}{\lambda_{\tilde{s}_n} + \sigma^2}|\tilde{y}[n]|^2$$

where $\tilde{\mathbf{y}} = \mathbf{V}^H\tilde{\mathbf{x}}$. Next, to find the detection performance note that $|\tilde{y}[n]|^2$ is a scaled χ_2^2 random variable and $\{\tilde{y}[0], \tilde{y}[1]\ldots, \tilde{y}[N-1]\}$ are independent random variables. Find P_{FA} and P_D by using the result that if $y = \sum_{n=0}^{N-1} \beta_n x_n$, where the β_n's are distinct with $\beta_n > 0$, and the x_n's are IID with PDF χ_2^2, then the PDF of y is

$$p(y) = \begin{cases} \sum_{n=0}^{N-1} C_n \frac{1}{2\beta_n} \exp\left[-\frac{y}{2\beta_n}\right] & y > 0 \\ 0 & y < 0 \end{cases}$$

where

$$C_n = \prod_{\substack{i=0 \\ i \neq n}}^{N-1} \frac{1}{1 - \beta_i/\beta_n}.$$

You should be able to show that

$$P_{FA} = \sum_{n=0}^{N-1} A_n \exp(-\gamma'/\alpha_n)$$

where $\alpha_n = \lambda_{\tilde{s}_n}\sigma^2/(\lambda_{\tilde{s}_n} + \sigma^2)$ and

$$A_n = \prod_{\substack{i=0 \\ i \neq n}}^{N-1} \frac{1}{1 - \alpha_i/\alpha_n}$$

and

$$P_D = \sum_{n=0}^{N-1} B_n \exp(-\gamma'/\lambda_{\tilde{s}_n})$$

where

$$B_n = \prod_{\substack{i=0 \\ i \neq n}}^{N-1} \frac{1}{1 - \lambda_{\tilde{s}_i}/\lambda_{\tilde{s}_n}}.$$

13.9 In Example 13.2 let $\tilde{h}[n] = 1$ so that $\tilde{s}[n]$ is a complex random DC level with $\tilde{A} \sim \mathcal{CN}(0, \sigma_A^2)$. Determine the NP detector and explain its operation. Also, find the detection performance.

13.10 Repeat Problem 13.9 but with $\tilde{h}[n] = \exp(j2\pi f_0 n)$. Discuss the effect of the signal, i.e, $\tilde{h}[n]$, on detection performance.

13.11 We wish to detect the complex deterministic DC level $\tilde{A}$ in CWGN with known variance σ^2 based on the data set $\tilde{x}[n]$ for $n = 0, 1, \ldots, N-1$. If $\tilde{A}$ is unknown, find the GLRT, including the threshold.

13.12 We wish to detect a complex sinusoid $\tilde{A}\exp(j2\pi f_0 n)$ of unknown deterministic amplitude $\tilde{A}$ and known frequency f_0 embedded in CWGN with known variance σ^2. Assuming the observations $\tilde{x}[n]$ for $n = 0, 1, \ldots, N-1$, find the GLRT, including the threshold. How does the detector change if f_0 is unknown?

13.13 We wish to detect the signal $\tilde{s}[n] = \tilde{A}_1\tilde{\phi}_1[n] + \tilde{A}_2\tilde{\phi}_2[n]$, where $\tilde{A}_1, \tilde{A}_2$ are deterministic and unknown and $\tilde{\phi}_1[n], \tilde{\phi}_2[n]$ are known orthonormal basis signals, in CWGN with known variance σ^2. Find the GLRT test statistic based on the observations $\tilde{x}[n]$ for $n = 0, 1, \ldots, N-1$. Next, specialize your results to the case when $\tilde{\phi}_1[n] = (1/\sqrt{N})\exp(j2\pi f_1 n)$ for $f_1 = k/N$ and $\tilde{\phi}_2[n] = (1/\sqrt{N})\exp(j2\pi f_2 n)$,where k, l are distinct integers with $k - l \neq rN$ for r an integer. Hint: Orthonormality means that $\sum_{n=0}^{N-1} \tilde{\phi}_i^*[n]\tilde{\phi}_j[n] = \delta_{ij}$.

13.14 The output of an array of sensors is observed. There are $M = 2$ sensors and $N = 3$ samples, with $\{1, 2, 3\}$ being observed at the output of the first sensor and $\{4, 5, 6\}$ being observed at the output of the second sensor. Find $\tilde{\mathbf{x}}[n]$, $\tilde{\mathbf{x}}_m$ and then $\tilde{\mathbf{x}}$, $\underline{\tilde{\mathbf{x}}}$.

13.15 We define the correlation (inner product) between two $N \times N$ matrices $\mathbf{A}$, $\mathbf{B}$ as

$$(\mathbf{A}, \mathbf{B}) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} [\mathbf{A}]_{ij}^* [\mathbf{B}]_{ij} = \text{tr}(\mathbf{A}^H \mathbf{B}).$$

Show that this correlation may be computed by correlating over columns first and then rows or vice-versa. To do so, let $\mathbf{A}$ and $\mathbf{B}$ be written in column form as $\mathbf{A} = [\mathbf{a}_0 \ldots \mathbf{a}_{N-1}]$ and $\mathbf{B} = [\mathbf{b}_0 \ldots \mathbf{b}_{N-1}]$, and also in row form as

$$\mathbf{A} = \begin{bmatrix} \mathbf{c}_0^T \\ \vdots \\ \mathbf{c}_{N-1}^T \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{d}_0^T \\ \vdots \\ \mathbf{d}_{N-1}^T \end{bmatrix}.$$

13.16 A sinusoidal random process is observed at the output of an array as

$$\tilde{x}_m[n] = \tilde{A}\exp\left[j(2\pi(f_0 m + f_1 n) + \phi)\right]$$

where $\tilde{A}$ is deterministic and ϕ is a random variable with $\phi \sim \mathcal{U}[0, 2\pi]$. Show that the cross-correlation between sensors m and m' is

$$r_{mm'}[k] = |\tilde{A}|^2 \exp\left[j2\pi(f_0(m' - m) + f_1 k)\right].$$

13.17 For the cross-correlation given in Problem 13.16, explicitly find $\mathbf{R}_{\tilde{x}\tilde{x}}[k]$, $\mathbf{P}_{\tilde{x}\tilde{x}}(f)$, and $\mathbf{C}$, if $M = 2$ and $N = 3$.

13.18 If $E(\tilde{x}_m[l]\tilde{x}^*_{m'}[n]) = 0$ for $m \neq m'$, find $\mathbf{R}_{\tilde{x}\tilde{x}}[k]$ and $\mathbf{P}_{\tilde{x}\tilde{x}}(f)$ for $M = 2$ and $N = 3$. Is $\mathbf{C}$ as given by (13.38) diagonal? What happens if we use a spatial ordering?

13.19 Prove that the CSM is Hermitian and positive semidefinite. Hint: To show that it is positive semidefinite let $\tilde{y}[n] = \tilde{\boldsymbol{\xi}}^T\tilde{\mathbf{x}}[n]$ and determine the ACF and PSD of $\tilde{y}[n]$.

13.20 In (13.45) show that as $N \to \infty$

$$T(\tilde{\mathbf{x}}) \to N\sum_{m=0}^{M-1}\int_{-\frac{1}{2}}^{\frac{1}{2}}\frac{P_{mm}(f)}{P_{mm}(f)+\sigma^2}I_m(f)df$$

and explain.

13.21 The continuous-time output of a uniformly spaced line array as shown in Figure 13.10 is converted to its complex envelope and then sampled. The resultant complex discrete-time process can be thought of as a sampling of the 2-D continuous process $\tilde{x}(r_x, t)$, where r_x denotes the distance along the x axis and t denotes time. With this interpretation we have that $\tilde{x}_m[n] = \tilde{x}(md, n\Delta)$ for $m = 0, 1, \ldots, M-1$; $n = 0, 1, \ldots, N-1$. If $\tilde{x}(r_x, t)$ is WSS in space and time, then we can define a *sampled spatial-temporal* ACF as

$$r[h,k] = E(\tilde{x}^*(md, n\Delta)\tilde{x}((m+h)d, (n+k)\Delta).$$

With this definition we can rewrite the cross-correlation function between sensors as

$$r_{ij}[k] = E(\tilde{x}^*(id, n\Delta)\tilde{x}(jd, (n+k)\Delta) = r[j-i, k].$$

Find $\mathbf{R}_{\tilde{x}\tilde{x}}[k]$ if $r_{ij}[k] = r[i-j, k]$. What special form does $\mathbf{R}_{\tilde{x}\tilde{x}}[k]$ have?

13.22 Find $\mathbf{C}_{\tilde{s}}$ if the spatial-temporal ACF is that defined in Problem 13.21 for $M = 2$ and $N = 3$. Explain why it is doubly block-Toeplitz (matrix is Toeplitz in blocks and each block is Toeplitz). Is it Toeplitz?

13.23 For the uniformly spaced line array shown in Figure 13.10, assume that $\beta = \pi/2$ (termed a *broadside* arrival). Show that the GLRT of (13.50) reduces to

$$T(\tilde{\mathbf{x}}) = \frac{M}{N\sigma^2/2}\left|\frac{1}{M}\sum_{m=0}^{M-1}X_m(f_D)\right|^2$$

where $X_m(f) = \sum_{n=0}^{N-1}\tilde{x}_m[n]\exp(-j2\pi fn)$ and explain.

13.24 For the uniformly spaced line array shown in Figure 13.10 assume that $\beta = \pi/2$. Show that the estimator-correlator of (13.59) reduces to

$$T(\tilde{\mathbf{x}}) = \frac{M}{N}\sum_{i=0}^{N-1}H(f_i)\left|\hat{S}_B(f_i)\right|^2$$

where

$$\hat{S}_B(f_i) = \frac{1}{M}\sum_{m=0}^{M-1}X_m(f_i)$$

and explain.

Appendix 13A

PDF of GLRT for Complex Linear Model

To find the PDF of the GLRT statistic (13.19), we first note that $\hat{\boldsymbol{\theta}}_1$ is a complex Gaussian random vector, since it is a linear transformation of the complex Gaussian random vector $\tilde{\mathbf{x}}$. Hence, it can be shown (see [Kay-I 1993, Example 15.9]) that $\hat{\boldsymbol{\theta}}_1 \sim \mathcal{CN}(\boldsymbol{\theta}_1, \sigma^2(\mathbf{H}^H\mathbf{H})^{-1})$. As a result, we have from (13.19) that $T(\tilde{\mathbf{x}}) = 2\hat{\boldsymbol{\theta}}_1^H \mathbf{C}_{\hat{\theta}_1}^{-1}\hat{\boldsymbol{\theta}}_1$, where $\mathbf{C}_{\hat{\theta}_1} = \sigma^2(\mathbf{H}^H\mathbf{H})^{-1}$ is the covariance matrix of $\hat{\boldsymbol{\theta}}_1$. We next let $\boldsymbol{\theta}_1 = \boldsymbol{\alpha} + j\boldsymbol{\beta}$ and $\hat{\boldsymbol{\theta}}_1 = \hat{\boldsymbol{\alpha}} + j\hat{\boldsymbol{\beta}}$ and define the corresponding $2p \times 1$ real vectors $\boldsymbol{\xi} = [\boldsymbol{\alpha}^T \, \boldsymbol{\beta}^T]^T$ and $\hat{\boldsymbol{\xi}} = [\hat{\boldsymbol{\alpha}}^T \, \hat{\boldsymbol{\beta}}^T]^T$. Then, the test statistic is equivalent to (see [Kay-I 1993, Appendix 15A])

$$T(\tilde{\mathbf{x}}) = \hat{\boldsymbol{\xi}}^T \mathbf{C}_{\hat{\xi}}^{-1} \hat{\boldsymbol{\xi}}$$

where $\mathbf{C}_{\hat{\xi}}$ is the $2p \times 2p$ real covariance matrix of the real estimator $\hat{\boldsymbol{\xi}}$. Since $\hat{\boldsymbol{\theta}}_1 \sim \mathcal{CN}(\boldsymbol{\theta}_1, \mathbf{C}_{\hat{\theta}_1})$, it follows that (see [Kay-I 1993, Chapter 15])

$$\hat{\boldsymbol{\xi}} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix}, \mathbf{C}_{\hat{\xi}}\right).$$

We next use the theorem that if $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$, where $\mathbf{x}$ is a real $N \times 1$ random vector, then $\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x} \sim \chi'^2_N(\lambda)$, where $\lambda = \boldsymbol{\mu}^T\mathbf{C}^{-1}\boldsymbol{\mu}$ (see Chapter 2). Then we have

$$T(\tilde{\mathbf{x}}) \sim \begin{cases} \chi^2_{2p} & \text{under } \mathcal{H}_0 \\ \chi'^2_{2p}(\lambda) & \text{under } \mathcal{H}_1 \end{cases}$$

where

$$\lambda = \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix}^T \mathbf{C}_{\hat{\xi}}^{-1} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix}.$$

But

$$\begin{aligned} \lambda &= \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix}^T \mathbf{C}_{\hat{\xi}}^{-1} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} \\ &= 2\boldsymbol{\theta}_1^H \mathbf{C}_{\hat{\theta}_1}^{-1} \boldsymbol{\theta}_1 \\ &= \frac{\boldsymbol{\theta}_1^H \mathbf{H}^H \mathbf{H} \boldsymbol{\theta}_1}{\sigma^2/2}. \end{aligned}$$

Appendix 1

Review of Important Concepts

A1.1 Linear and Matrix Algebra

Important results from linear and matrix algebra theory are reviewed in this section. In the discussions to follow it is assumed that the reader already has some familiarity with these topics. The specific concepts to be described are used heavily throughout the book. For a more comprehensive treatment the reader is referred to the books [Noble and Daniel 1977] and [Graybill 1969]. All matrices and vectors are assumed to be real.

A1.1.1 Definitions

Consider an $m \times n$ matrix $\mathbf{A}$ with elements a_{ij}, $i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$. A shorthand notation for describing $\mathbf{A}$ is

$$[\mathbf{A}]_{ij} = a_{ij}.$$

The *transpose* of $\mathbf{A}$, which is denoted by $\mathbf{A}^T$, is defined as the $n \times m$ matrix with elements a_{ji} or

$$[\mathbf{A}^T]_{ij} = a_{ji}.$$

A *square* matrix is one for which $m = n$. A square matrix is *symmetric* if $\mathbf{A}^T = \mathbf{A}$.

The *rank* of a matrix is the number of linearly independent rows or columns, whichever is less. The *inverse* of a square $n \times n$ matrix is the square $n \times n$ matrix $\mathbf{A}^{-1}$ for which

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$$

where $\mathbf{I}$ is the $n \times n$ identity matrix. The inverse will exist if and only if the rank of $\mathbf{A}$ is n. If the inverse does not exist, then $\mathbf{A}$ is *singular*.

The *determinant* of a square $n \times n$ matrix is denoted by $\det(\mathbf{A})$. It is computed as

$$\det(\mathbf{A}) = \sum_{j=1}^{n} a_{ij} C_{ij}$$

where

$$C_{ij} = (-1)^{i+j} M_{ij}.$$

M_{ij} is the determinant of the submatrix of $\mathbf{A}$ obtained by deleting the ith row and jth column and is termed the *minor* of a_{ij}. C_{ij} is the *cofactor* of a_{ij}. Note that any choice of i for $i = 1, 2, \ldots, n$ will yield the same value for $\det(\mathbf{A})$.

A *quadratic form* Q is defined as

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j.$$

In defining the quadratic form, it is assumed that $a_{ji} = a_{ij}$. This entails no loss in generality, since any quadratic function may be expressed in this manner. Q may also be expressed as

$$Q = \mathbf{x}^T \mathbf{A} \mathbf{x}$$

where $\mathbf{x} = [x_1\, x_2 \ldots x_n]^T$ and $\mathbf{A}$ is a square $n \times n$ matrix with $a_{ji} = a_{ij}$ or $\mathbf{A}$ is a symmetric matrix.

A square $n \times n$ matrix $\mathbf{A}$ is *positive semidefinite* if $\mathbf{A}$ is symmetric and

$$\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$$

for all real $\mathbf{x} \neq \mathbf{0}$. If the quadratic form is strictly positive, then $\mathbf{A}$ is *positive definite.* When referring to a matrix as positive definite or positive semidefinite, it is always assumed that the matrix is symmetric.

The *trace* of a square $n \times n$ matrix is the sum of its diagonal elements or

$$\operatorname{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}.$$

A *partitioned* $m \times n$ matrix $\mathbf{A}$ is one that is expressed in terms of its submatrices. An example is the 2×2 partitioning

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}.$$

Each "element" $\mathbf{A}_{ij}$ is a submatrix of $\mathbf{A}$. The dimensions of the partitions are given as

$$\begin{bmatrix} k \times l & k \times (n-l) \\ (m-k) \times l & (m-k) \times (n-l) \end{bmatrix}.$$

A1.1.2 Special Matrices

A *diagonal* matrix is a square $n \times n$ matrix with $a_{ij} = 0$ for $i \neq j$ or all elements off the principal diagonal are zero. A diagonal matrix appears as

$$\mathbf{A} = \begin{bmatrix} a_{11} & 0 & \ldots & 0 \\ 0 & a_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_{nn} \end{bmatrix}.$$

A diagonal matrix will sometimes be denoted by $\operatorname{diag}(a_{11}, a_{22}, \ldots, a_{nn})$. The inverse of a diagonal matrix is found by simply inverting each element on the principal diagonal.

A generalization of the diagonal matrix is the square $n \times n$ *block-diagonal* matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{A}_{kk} \end{bmatrix}$$

in which all submatrices $\mathbf{A}_{ii}$ are square and the other submatrices are identically zero. The dimensions of the submatrices need not be identical. For instance, if $k = 2$, $\mathbf{A}_{11}$ might have dimension 2×2 while $\mathbf{A}_{22}$ might be a scalar. If all $\mathbf{A}_{ii}$ are nonsingular, then the inverse is easily found as

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}_{11}^{-1} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22}^{-1} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{A}_{kk}^{-1} \end{bmatrix}.$$

Also, the determinant is

$$\det(\mathbf{A}) = \prod_{i=1}^{k} \det(\mathbf{A}_{ii}).$$

A square $n \times n$ matrix is *orthogonal* if

$$\mathbf{A}^{-1} = \mathbf{A}^T.$$

For a matrix to be orthogonal the columns (and rows) must be orthonormal or if

$$\mathbf{A} = [\; \mathbf{a}_1 \;\; \mathbf{a}_2 \;\; \ldots \;\; \mathbf{a}_n \;]$$

where $\mathbf{a}_i$ denotes the ith column, the conditions

$$\mathbf{a}_i^T \mathbf{a}_j = \begin{cases} 0 & \text{for } i \neq j \\ 1 & \text{for } i = j \end{cases}$$

must be satisfied. An important example of an orthogonal matrix arises in modeling of data by a sum of harmonically related sinusoids or by a discrete Fourier series. As an example, for n even

$$\mathbf{A} = \frac{1}{\sqrt{\frac{n}{2}}}\begin{bmatrix} \frac{1}{\sqrt{2}} & 1 & \cdots & \frac{1}{\sqrt{2}} & 0 & \cdots & 0 \\ \frac{1}{\sqrt{2}} & \cos\frac{2\pi}{n} & \cdots & \frac{1}{\sqrt{2}}\cos\frac{2\pi(\frac{n}{2})}{n} & \sin\frac{2\pi}{n} & \cdots & \sin\frac{2\pi(\frac{n}{2}-1)}{n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\sqrt{2}} & \cos\frac{2\pi(n-1)}{n} & \cdots & \frac{1}{\sqrt{2}}\cos\frac{2\pi\frac{n}{2}(n-1)}{n} & \sin\frac{2\pi(n-1)}{n} & \cdots & \sin\frac{2\pi(\frac{n}{2}-1)(n-1)}{n} \end{bmatrix}$$

is an orthogonal matrix. This follows from the orthogonality relationship for $i, j = 0, 1, \ldots, n/2$

$$\sum_{k=0}^{n-1} \cos\frac{2\pi k i}{n}\cos\frac{2\pi k j}{n} = \begin{cases} 0 & i \neq j \\ \frac{n}{2} & i = j = 1, 2, \ldots, \frac{n}{2}-1 \\ n & i = j = 0, \frac{n}{2} \end{cases}$$

and for $i, j = 1, 2, \ldots, n/2 - 1$

$$\sum_{k=0}^{n-1} \sin\frac{2\pi k i}{n}\sin\frac{2\pi k j}{n} = \frac{n}{2}\delta_{ij}$$

and finally for $i = 0, 1, \ldots, n/2; j = 1, 2, \ldots, n/2 - 1$

$$\sum_{k=0}^{n-1} \cos\frac{2\pi k i}{n}\sin\frac{2\pi k j}{n} = 0.$$

These orthogonality relationships may be proven by expressing the sines and cosines in terms of complex exponentials and using the result

$$\sum_{k=0}^{n-1} \exp\left(j\frac{2\pi}{n}kl\right) = n\delta_{l0}$$

for $l = 0, 1, \ldots, n-1$ [Oppenheim and Schafer 1975].

An *idempotent* matrix is a square $n \times n$ matrix that satisfies

$$\mathbf{A}^2 = \mathbf{A}.$$

This condition implies that $\mathbf{A}^l = \mathbf{A}$ for $l \geq 1$. An example is the *projection matrix*

$$\mathbf{A} = \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$$

where $\mathbf{H}$ is an $m \times n$ full rank matrix with $m > n$.

A square $n \times n$ *Toeplitz* matrix is defined as

$$[\mathbf{A}]_{ij} = a_{i-j}$$

or

$$\mathbf{A} = \begin{bmatrix} a_0 & a_{-1} & a_{-2} & \cdots & a_{-(n-1)} \\ a_1 & a_0 & a_{-1} & \cdots & a_{-(n-2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n-1} & a_{n-2} & a_{n-3} & \cdots & a_0 \end{bmatrix}. \tag{A1.1}$$

Each element along a northwest-southeast diagonal is the same. If in addition, $a_{-k} = a_k$, then $\mathbf{A}$ is *symmetric Toeplitz.*

A1.1.3 Matrix Manipulation and Formulas

Some useful formulas for the algebraic manipulation of matrices are summarized in this section. For $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$ the following relationships are useful.

$$\begin{aligned} (\mathbf{AB})^T &= \mathbf{B}^T\mathbf{A}^T \\ (\mathbf{A}^T)^{-1} &= (\mathbf{A}^{-1})^T \\ (\mathbf{AB})^{-1} &= \mathbf{B}^{-1}\mathbf{A}^{-1} \\ \det(\mathbf{A}^T) &= \det(\mathbf{A}) \\ \det(c\mathbf{A}) &= c^n\det(\mathbf{A}) \quad (c \text{ a scalar}) \\ \det(\mathbf{AB}) &= \det(\mathbf{A})\det(\mathbf{B}) \\ \det(\mathbf{A}^{-1}) &= \frac{1}{\det(\mathbf{A})} \\ \operatorname{tr}(\mathbf{AB}) &= \operatorname{tr}(\mathbf{BA}) \\ \operatorname{tr}(\mathbf{A}^T\mathbf{B}) &= \sum_{i=1}^{n}\sum_{j=1}^{n}[\mathbf{A}]_{ij}[\mathbf{B}]_{ij}. \end{aligned}$$

In differentiating linear and quadratic forms, the following formulas for the gradient are useful.

$$\begin{aligned} \frac{\partial \mathbf{b}^T\mathbf{x}}{\partial \mathbf{x}} &= \mathbf{b} \\ \frac{\partial \mathbf{x}^T\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} &= 2\mathbf{A}\mathbf{x} \end{aligned}$$

where it is assumed that $\mathbf{A}$ is a symmetric matrix. Also, for vectors $\mathbf{x}$ and $\mathbf{y}$ we have

$$\mathbf{y}^T\mathbf{x} = \operatorname{tr}(\mathbf{x}\mathbf{y}^T).$$

It is frequently necessary to determine the inverse of a matrix analytically. To do so, one can make use of the following formula. The inverse of a square $n \times n$ matrix is

$$\mathbf{A}^{-1} = \frac{\mathbf{C}^T}{\det(\mathbf{A})}$$

where $\mathbf{C}$ is the square $n \times n$ matrix of cofactors of $\mathbf{A}$. The cofactor matrix is defined by

$$[\mathbf{C}]_{ij} = (-1)^{i+j} M_{ij}$$

where M_{ij} is the minor of a_{ij} obtained by deleting the ith row and jth column of $\mathbf{A}$.

Another formula that is quite useful is the *matrix inversion lemma*

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{DA}^{-1}$$

where it is assumed that $\mathbf{A}$ is $n \times n$, $\mathbf{B}$ is $n \times m$, $\mathbf{C}$ is $m \times m$, and $\mathbf{D}$ is $m \times n$, and that the indicated inverses exist. A special case known as *Woodbury's identity* results for $\mathbf{B}$ an $n \times 1$ column vector $\mathbf{u}$, $\mathbf{C}$ a scalar of unity, and $\mathbf{D}$ a $1 \times n$ row vector $\mathbf{u}^T$. Then,

$$(\mathbf{A} + \mathbf{uu}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{uu}^T\mathbf{A}^{-1}}{1 + \mathbf{u}^T\mathbf{A}^{-1}\mathbf{u}}.$$

Partitioned matrices may be manipulated according to the usual rules of matrix algebra by considering each submatrix as an element. For multiplication of partitioned matrices the submatrices that are multiplied together must be conformable. As an illustration, for 2×2 partitioned matrices

$$\begin{aligned}\mathbf{AB} &= \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{bmatrix}.\end{aligned}$$

The transposition of a partitioned matrix is formed by transposing the submatrices of the matrix and applying T to each submatrix. For a 2×2 partitioned matrix

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^T = \begin{bmatrix} \mathbf{A}_{11}^T & \mathbf{A}_{21}^T \\ \mathbf{A}_{12}^T & \mathbf{A}_{22}^T \end{bmatrix}.$$

The extension of these properties to arbitrary partitioning is straightforward. Determination of the inverses and determinants of partitioned matrices is facilitated by employing the following formulas. Let $\mathbf{A}$ be a square $n \times n$ matrix partitioned as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} = \begin{bmatrix} k \times k & k \times (n-k) \\ (n-k) \times k & (n-k) \times (n-k) \end{bmatrix}.$$

Then,

$$\mathbf{A}^{-1} = \begin{bmatrix} (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} & -(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1} \end{bmatrix}$$

$$\begin{aligned}\det(\mathbf{A}) &= \det(\mathbf{A}_{22})\det(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}) \\ &= \det(\mathbf{A}_{11})\det(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})\end{aligned}$$

where the inverses of $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are assumed to exist.

A1.1.4 Theorems

Some important theorems used throughout the text are summarized in this section.

1. A square $n \times n$ matrix $\mathbf{A}$ is invertible (nonsingular) if and only if its columns (or rows) are linearly independent or, equivalently, if its determinant is nonzero. In such a case, $\mathbf{A}$ is *full rank*. Otherwise, it is singular.

2. A square $n \times n$ matrix $\mathbf{A}$ is positive definite if and only if

- **a.** it can be written as

$$\mathbf{A} = \mathbf{CC}^T \tag{A1.2}$$

where $\mathbf{C}$ is also $n \times n$ and is full rank and hence invertible, or

- **b.** the principal minors are all positive. (The ith principal minor is the determinant of the submatrix formed by deleting all rows and columns with an index greater than i.) If $\mathbf{A}$ can be written as in (A1.2), but $\mathbf{C}$ is not full rank or the principal minors are only nonnegative, then $\mathbf{A}$ is positive semidefinite.

3. If $\mathbf{A}$ is positive definite, then the inverse exists and may be found from (A1.2) as $\mathbf{A}^{-1} = (\mathbf{C}^{-1})^T(\mathbf{C}^{-1})$.

4. Let $\mathbf{A}$ be positive definite. If $\mathbf{B}$ is an $m \times n$ matrix of full rank with $m \leq n$, then $\mathbf{BAB}^T$ is also positive definite.

5. If $\mathbf{A}$ is positive definite (positive semidefinite), then

- **a.** the diagonal elements are positive (nonnegative)
- **b.** the determinant of $\mathbf{A}$, which is a principal minor, is positive (nonnegative).

A1.1.5 Eigendecompostion of Matrices

An *eigenvector* of a square $n \times n$ matrix $\mathbf{A}$ is an $n \times 1$ vector $\mathbf{v}$ satisfying

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v} \tag{A1.3}$$

for some scalar λ, which may be complex. λ is the *eigenvalue* of $\mathbf{A}$ corresponding to the eigenvector $\mathbf{v}$. It is assumed that the eigenvector is normalized to have unit length or $\mathbf{v}^T\mathbf{v} = 1$. If $\mathbf{A}$ is symmetric, then one can always find n linearly independent eigenvectors, although they will not in general be unique. An example is the identity matrix for which any vector is an eigenvector with eigenvalue 1. If $\mathbf{A}$ is symmetric, then the eigenvectors corresponding to distinct eigenvalues are orthonormal or $\mathbf{v}_i^T\mathbf{v}_j = \delta_{ij}$, and the eigenvalues are real. If, furthermore, the matrix is positive definite (positive semidefinite), then the eigenvalues are positive (nonnegative). For a positive semidefinite matrix the rank is equal to the number of nonzero eigenvalues.

The defining relation of (A1.3) can also be written as

$$\mathbf{A}\,[\ \mathbf{v}_1\ \ \mathbf{v}_2\ \ \ldots\ \ \mathbf{v}_n\ ] = [\ \lambda_1\mathbf{v}_1\ \ \lambda_2\mathbf{v}_2\ \ \ldots\ \ \lambda_n\mathbf{v}_n\ ]$$

or

$$\mathbf{A}\mathbf{V} = \mathbf{V}\mathbf{\Lambda} \tag{A1.4}$$

where

$$\begin{aligned} \mathbf{V} &= [\ \mathbf{v}_1\ \ \mathbf{v}_2\ \ \ldots\ \ \mathbf{v}_n\ ] \\ \mathbf{\Lambda} &= \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n). \end{aligned}$$

If $\mathbf{A}$ is symmetric so that the eigenvectors corresponding to distinct eigenvalues are orthonormal and the remaining eigenvectors are chosen to yield an orthonormal eigenvector set, then $\mathbf{V}$ is an orthogonal matrix. The matrix $\mathbf{V}$ is termed the *modal matrix*. As such, its inverse is $\mathbf{V}^T$, so that (A1.4) becomes

$$\begin{aligned} \mathbf{A} &= \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T \\ &= \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^T. \end{aligned}$$

Also, the inverse is easily determined as

$$\begin{aligned} \mathbf{A}^{-1} &= \mathbf{V}^{T^{-1}}\mathbf{\Lambda}^{-1}\mathbf{V}^{-1} \\ &= \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{V}^T \\ &= \sum_{i=1}^{n} \frac{1}{\lambda_i}\mathbf{v}_i\mathbf{v}_i^T. \end{aligned}$$

A final useful relationship follows from (A1.4) as

$$\begin{aligned} \det(\mathbf{A}) &= \det(\mathbf{V})\det(\mathbf{\Lambda})\det(\mathbf{V}^{-1}) \\ &= \det(\mathbf{\Lambda}) \\ &= \prod_{i=1}^{n} \lambda_i. \end{aligned}$$

It should be noted that the modal matrix $\mathbf{V}$, which diagonalizes a symmetric matrix, can be employed to decorrelate a set of random variables. Assume that $\mathbf{x}$ is a vector random variable with zero mean and covariance matrix $\mathbf{C}_x$. Then $\mathbf{y} = \mathbf{V}^T\mathbf{x}$ for $\mathbf{V}$ the modal matrix of $\mathbf{C}_x$ is a vector random variable with zero mean and *diagonal* covariance matrix. This follows from (A1.4) since the covariance matrix of $\mathbf{y}$ is

$$\begin{aligned} \mathbf{C}_y &= E(\mathbf{y}\mathbf{y}^T) = E(\mathbf{V}^T\mathbf{x}\mathbf{x}^T\mathbf{V}) \\ &= \mathbf{V}^T E(\mathbf{x}\mathbf{x}^T)\mathbf{V} = \mathbf{V}^T\mathbf{C}_x\mathbf{V} \\ &= \mathbf{\Lambda} \end{aligned}$$

where $\mathbf{\Lambda}$ is a diagonal matrix.

A1.1.6 Inequalities

The Cauchy-Schwarz inequality can be used to simplify maximization problems and provide explicit solutions. For two vectors $\mathbf{x}$ and $\mathbf{y}$ it asserts that

$$(\mathbf{y}^T\mathbf{x})^2 \leq (\mathbf{y}^T\mathbf{y})(\mathbf{x}^T\mathbf{x})$$

with equality if and only if $\mathbf{y} = c\mathbf{x}$ for c an arbitrary constant. As applied to functions $g(x)$ and $h(x)$, which we assume to be complex functions of a real variable for generality, it takes the form

$$\left| \int g(x)h(x)dx \right|^2 \leq \int |g(x)|^2 dx \int |h(x)|^2 dx$$

with equality if and only if $g(x) = ch^*(x)$ for c an arbitrary complex constant.

A1.2 Random Processes and Time Series Modeling

An assumption is made that the reader already has some familiarity with basic random process theory. This section serves as a review of these topics. For those readers needing a more extensive treatment the text [Papoulis 1965] on probability and random processes is recommended. For a discussion of time series modeling see [Kay 1988].

A1.2.1 Random Process Characterization

A *discrete random process* $x[n]$ is a sequence of random variables defined for every integer n. If the discrete random process is wide sense stationary (WSS), then it has a *mean*

$$E(x[n]) = \mu_x$$

which does not depend on n, and an *autocorrelation function* (ACF)

$$r_{xx}[k] = E(x[n]x[n+k]) \tag{A1.5}$$

which depends only on the lag k between the two samples and not their absolute positions. Also, the *autocovariance function* is defined as

$$c_{xx}[k] = E\left[(x[n]-\mu_x)(x[n+k]-\mu_x)\right] = r_{xx}[k] - \mu_x^2.$$

In a similar manner, two jointly WSS random processes $x[n]$ and $y[n]$ have a *cross-correlation function* (CCF)

$$r_{xy}[k] = E(x[n]y[n+k])$$

and a *cross-covariance function*

$$c_{xy}[k] = E\left[(x[n]-\mu_x)(y[n+k]-\mu_y)\right] = r_{xy}[k] - \mu_x\mu_y.$$

Some useful properties of the ACF and CCF are

$$\begin{aligned} r_{xx}[0] &\geq |r_{xx}[k]| \\ r_{xx}[-k] &= r_{xx}[k] \\ r_{xy}[-k] &= r_{yx}[k]. \end{aligned}$$

Note that $r_{xx}[0]$ is positive, which follows from (A1.5).

The z transforms of the ACF and CCF defined as

$$\begin{aligned} \mathcal{P}_{xx}(z) &= \sum_{k=-\infty}^{\infty} r_{xx}[k]z^{-k} \\ \mathcal{P}_{xy}(z) &= \sum_{k=-\infty}^{\infty} r_{xy}[k]z^{-k} \end{aligned}$$

lead to the definition of the power spectral density (PSD). When evaluated on the unit circle, $\mathcal{P}_{xx}(z)$ and $\mathcal{P}_{xy}(z)$ become the *auto-PSD*, $P_{xx}(f) = \mathcal{P}_{xx}(\exp[j2\pi f])$, and *cross-PSD*, $P_{xy}(f) = \mathcal{P}_{xy}(\exp[j2\pi f])$, or explicitly

$$P_{xx}(f) = \sum_{k=-\infty}^{\infty} r_{xx}[k]\exp(-j2\pi fk) \tag{A1.6}$$

$$P_{xy}(f) = \sum_{k=-\infty}^{\infty} r_{xy}[k]\exp(-j2\pi fk). \tag{A1.7}$$

It also follows from the definition of the cross-PSD and the property $r_{yx}[k] = r_{xy}[-k]$ that

$$P_{yx}(f) = P_{xy}^*(f).$$

The auto-PSD describes the distribution in frequency of the power of $x[n]$ and as such is real and nonnegative. The cross-PSD, on the other hand, is in general complex. The magnitude of the cross-PSD describes whether frequency components in $x[n]$ are associated with large or small amplitudes at the same frequency in $y[n]$, and the phase of the cross-PSD indicates the phase lag or lead of $x[n]$ with respect to $y[n]$ for a given frequency component. Note that both spectral densities are periodic with period 1. The frequency interval $-1/2 \leq f \leq 1/2$ will be considered the fundamental period. When there is no confusion, $P_{xx}(f)$ will simply be referred to as the *power spectral density*.

A process that will frequently be encountered is discrete white noise. It is defined as a zero mean process having an ACF

$$r_{xx}[k] = \sigma^2\delta[k]$$

where $\delta[k]$ is the discrete impulse function. This says that each sample is uncorrelated with all the others. Using (A1.6), the PSD becomes

$$P_{xx}(f) = \sigma^2$$

and is observed to be completely flat with frequency. Alternatively, white noise is composed of equi-power contributions from all frequencies.

For a linear shift invariant (LSI) system with impulse response $h[n]$ and with a WSS random process input, various relationships between the correlations and spectral density functions of the input process $x[n]$ and output process $y[n]$ hold. The correlation relationships are

$$\begin{aligned} r_{xy}[k] &= h[k] \star r_{xx}[k] = \sum_{l=-\infty}^{\infty} h[l]r_{xx}[k-l] \\ r_{yx}[k] &= h[-k] \star r_{xx}[k] = \sum_{l=-\infty}^{\infty} h[-l]r_{xx}[k-l] \\ r_{yy}[k] &= h[k] \star r_{yx}[k] = h[k] \star h[-k] \star r_{xx}[k] \\ &= \sum_{m=-\infty}^{\infty} h[k-m] \sum_{l=-\infty}^{\infty} h[-l]r_{xx}[m-l] \end{aligned}$$

where $\star$ denotes convolution. Denoting the system function by $\mathcal{H}(z) = \sum_{n=-\infty}^{\infty} h[n]z^{-n}$, the following relationships for the PSDs follow from these correlation properties.

$$\begin{aligned}\mathcal{P}_{xy}(z) &= \mathcal{H}(z)\mathcal{P}_{xx}(z)\\ \mathcal{P}_{yx}(z) &= \mathcal{H}(1/z)\mathcal{P}_{xx}(z)\\ \mathcal{P}_{yy}(z) &= \mathcal{H}(z)\mathcal{H}(1/z)\mathcal{P}_{xx}(z).\end{aligned}$$

In particular, letting $H(f) = \mathcal{H}[\exp(j2\pi f)]$ be the frequency response of the LSI system results in

$$\begin{aligned}P_{xy}(f) &= H(f)P_{xx}(f)\\ P_{yx}(f) &= H^*(f)P_{xx}(f)\\ P_{yy}(f) &= |H(f)|^2 P_{xx}(f).\end{aligned}$$

For the special case of a white noise input process the output PSD becomes

$$P_{yy}(f) = |H(f)|^2\sigma^2 \tag{A1.8}$$

since $P_{xx}(f) = \sigma^2$. This will form the basis of the time series models to be described shortly.

A1.2.2 Gaussian Random Process

A Gaussian random process is one for which any samples $x[n_0], x[n_1], \ldots, x[n_{N-1}]$ are jointly distributed according to a multivariate Gaussian PDF. If the samples are taken at successive times to generate the vector $\mathbf{x} = [x[0]\, x[1] \ldots x[N-1]]^T$, then assuming a zero mean WSS random process, the covariance matrix takes on the form

$$\mathbf{C}_x = \begin{bmatrix} r_{xx}[0] & r_{xx}[1] & \ldots & r_{xx}[N-1]\\ r_{xx}[1] & r_{xx}[0] & \ldots & r_{xx}[N-2]\\ \vdots & \vdots & \ddots & \vdots\\ r_{xx}[N-1] & r_{xx}[N-2] & \ldots & r_{xx}[0]\end{bmatrix}.$$

The covariance matrix, or more appropriately the autocorrelation matrix, has the special symmetric Toeplitz structure of (A1.1) with $a_k = a_{-k}$.

An important Gaussian random process is the white process. As discussed previously, the ACF for a white process is a discrete delta function. In light of the definition of a Gaussian random process a *white Gaussian random process* $x[n]$ with mean zero and variance σ^2 is one for which

$$x[n] \sim N(0, \sigma^2) \qquad -\infty < n < \infty$$

and

$$r_{xx}[m-n] = E(x[n]x[m]) = 0 \qquad m \neq n.$$

Because of the Gaussian assumption, any two samples are statistically independent.

A1.2.3 Time Series Models

A useful class of time series models consists of the rational transfer function models. The time series $x[n]$ is modeled as the output of a LSI filter with frequency response $H(f)$ excited at the input by white noise $u[n]$ with variance σ_u^2. As such, its PSD follows from (A1.8) as

$$P_{xx}(f) = |H(f)|^2\sigma_u^2.$$

The first time series model is termed an *autoregressive* (AR) process, which has the time domain representation

$$x[n] = -\sum_{k=1}^{p} a[k]x[n-k] + u[n].$$

It is said to be an AR process of order p and is denoted by AR(p). The AR parameters consist of the filter coefficients $\{a[1], a[2], \ldots, a[p]\}$ and the driving white noise variance σ_u^2. Since the frequency response is

$$H(f) = \frac{1}{1 + \sum_{k=1}^{p} a[k]\exp(-j2\pi fk)}$$

the AR PSD is

$$P_{xx}(f) = \frac{\sigma_u^2}{\left|1 + \sum_{k=1}^{p} a[k]\exp(-j2\pi fk)\right|^2}.$$

It can be shown that the ACF satisfies the recursive difference equation

$$r_{xx}[k] = \begin{cases} -\sum_{l=1}^{p} a[l]r_{xx}[k-l] & k \geq 1\\ -\sum_{l=1}^{p} a[l]r_{xx}[l] + \sigma_u^2 & k = 0.\end{cases}$$

In matrix form this becomes for $k = 1, 2, \ldots, p$

$$\begin{bmatrix} r_{xx}[0] & r_{xx}[1] & \ldots & r_{xx}[p-1]\\ r_{xx}[1] & r_{xx}[0] & \ldots & r_{xx}[p-2]\\ \vdots & \vdots & \ddots & \vdots\\ r_{xx}[p-1] & r_{xx}[p-2] & \ldots & r_{xx}[0]\end{bmatrix} \begin{bmatrix} a[1]\\ a[2]\\ \vdots\\ a[p]\end{bmatrix} = -\begin{bmatrix} r_{xx}[1]\\ r_{xx}[2]\\ \vdots\\ r_{xx}[p]\end{bmatrix}$$

and

$$\sigma_u^2 = r_{xx}[0] + \sum_{k=1}^{p} a[k]r_{xx}[k].$$

Given the ACF samples $r_{xx}[k]$ for $k = 0, 1, \ldots, p$, the AR parameters may be determined by solving the set of p linear equations. These equations are termed the *Yule-Walker equations.* As an example, for an AR(1) process

$$r_{xx}[k] = -a[1]r_{xx}[k-1] \qquad k \geq 1$$

which can be solved to yield

$$r_{xx}[k] = r_{xx}[0]\,(-a[1])^k \qquad k \geq 0$$

and from

$$\sigma_u^2 = r_{xx}[0] + a[1]r_{xx}[1]$$

we can solve for $r_{xx}[0]$ to produce the ACF of an AR(1) process

$$r_{xx}[k] = \frac{\sigma_u^2}{1 - a^2[1]}\,(-a[1])^{|k|}.$$

The corresponding PSD is

$$P_{xx}(f) = \frac{\sigma_u^2}{|1 + a[1]\exp(-j2\pi f)|^2}.$$

The AR(1) PSD will have most of its power at lower frequencies if $a[1] < 0$ and at higher frequencies if $a[1] > 0$. The system function for the AR(1) process is

$$\mathcal{H}(z) = \frac{1}{1 + a[1]z^{-1}}$$

and has a pole at $z = -a[1]$. Hence, for a stable process we must have $|a[1]| < 1$.

While the AR process is generated as the output of a filter having only poles, the *moving average* (MA) process is formed by passing white noise through a filter whose system function has only zeros. The time domain representation of a MA(q) process is

$$x[n] = u[n] + \sum_{k=1}^{q} b[k]u[n-k].$$

Since the filter frequency response is

$$H(f) = 1 + \sum_{k=1}^{q} b[k]\exp(-j2\pi fk)$$

its PSD is

$$P_{xx}(f) = \left|1 + \sum_{k=1}^{q} b[k]\exp(-j2\pi fk)\right|^2 \sigma_u^2.$$

The MA(q) process has the ACF

$$r_{xx}[k] = \begin{cases} \sigma_u^2 \displaystyle\sum_{l=0}^{q-|k|} b[l]b[l+|k|] & |k| \leq q \\ 0 & |k| > q. \end{cases}$$

The system function is

$$\mathcal{H}(z) = 1 + \sum_{k=1}^{q} b[k]z^{-k}$$

and is seen to consist of zeros only. Although not required for stability, it is usually assumed that the zeros satisfy $|z_i| < 1$. This is because z_i and $1/z_i^*$ can both result in the same PSD, resulting in a problem of identifiability of the process parameters.

References

Graybill, F.A., *Introduction to Matrices with Application in Statistics*, Wadsworth, Belmont, Ca., 1969.

Kay, S., *Modern Spectral Estimation: Theory and Application*, Prentice-Hall, Englewood Cliffs, N.J., 1988.

Noble, B., J.W. Daniel, *Applied Linear Algebra*, Prentice-Hall, Englewood Cliffs, N.J., 1977.

Oppenheim, A.V., R.W. Schafer, *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, N.J., 1975.

Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1965.

Appendix 2

Glossary of Symbols and Abbreviations (Vols. I & II)

Symbols

(Boldface characters denote vectors or matrices. All others are scalars.)

$*$	complex conjugate
$\star$	convolution
$\hat{}$	denotes estimator
$\check{}$	denotes estimator
$\tilde{}$	denotes complex quantity
$\sim$	denotes *is distributed according to*
$\stackrel{a}{\sim}$	denotes *is asymptotically distributed according to*
$\arg\max_{\theta} g(\theta)$	denotes the value of θ that maximizes $g(\theta)$
$[\mathbf{A}]_{ij}$	ijth element of $\mathbf{A}$
$[\mathbf{b}]_i$	ith element of $\mathbf{b}$
$\mathrm{Bmse}(\hat{\theta})$	Bayesian mean square error of $\hat{\theta}$
χ^2_n	chi-squared distribution with n degrees of freedom
$\chi'^2_n(\lambda)$	noncentral chi-squared distribution with n degrees of freedom and noncentrality parameter λ
$\mathrm{cov}(x, y)$	covariance of x and y
$\mathbf{C}_x$ or $\mathbf{C}_{xx}$	covariance matrix of $\mathbf{x}$
$\mathbf{C}_{xy}$	covariance matrix of $\mathbf{x}$ and $\mathbf{y}$
$\mathbf{C}_{y\|x}$	covariance matrix of $\mathbf{y}$ with respect to PDF of $\mathbf{y}$ conditioned on $\mathbf{x}$
$\mathcal{CN}(\tilde{\mu}, \sigma^2)$	complex normal distribution with mean $\tilde{\mu}$ and variance σ^2
$\mathcal{CN}(\tilde{\boldsymbol{\mu}}, \mathbf{C})$	multivariate complex normal distribution with mean $\tilde{\boldsymbol{\mu}}$ and covariance $\mathbf{C}$
d^2	deflection coefficient

$\delta(t)$	Dirac delta function
$\delta[n]$	discrete-time impulse sequence
δ_{ij}	Kronecker delta
Δ	time sampling interval
$\det(\mathbf{A})$	determinant of matrix $\mathbf{A}$
$\operatorname{diag}(\cdots)$	diagonal matrix with elements $\cdots$ on main diagonal
$\mathbf{e}_i$	natural unit vector in ith direction
E	expected value
E_x	expected value with respect to PDF of $\mathbf{x}$
$E(x;\mathcal{H}_i)$	expected value of x assuming $\mathcal{H}_i$ true
$E_{x\|\theta}$ or $E(x\|\theta)$	conditional expected value with respect to PDF of x conditioned on θ
$E(x\|\mathcal{H}_i)$	expected value of x conditioned on $\mathcal{H}_i$ being true
$\mathcal{E}$	energy
$\bar{\mathcal{E}}$	average energy
η	signal-to-noise-ratio
f	discrete-time frequency
F	continuous-time frequency
$\mathcal{F}$	Fourier transform
$\mathcal{F}^{-1}$	inverse Fourier transform
$F_{m,n}$	F distribution with m numerator and n denominator degrees of freedom
$F'_{m,n}(\lambda)$	noncentral F distribution with m numerator and n denominator degrees of freedom and noncentrality parameter λ
γ (γ', γ'', etc.)	threshold
$\Gamma(x)$	Gamma function
^{H}	conjugate transpose
$\mathbf{H}$	observation matrix
(x,y)	inner product of x and y
$i(\theta)$	Fisher information for single data sample and scalar θ
$I(\theta)$	Fisher information for scalar θ
$\mathbf{I}$ or $\mathbf{I}_n$	identity matrix or identity matrix of dimension $n \times n$
$\mathbf{I}(\boldsymbol{\theta})$	Fisher information matrix for vector $\boldsymbol{\theta}$
$\mathbf{I}_{\theta_r\theta_s}(\boldsymbol{\theta})$	Partition of $\mathbf{I}(\boldsymbol{\theta})$ of dimension $r \times s$
$I(f)$	periodogram
$I_x(f)$	periodogram based on $x[n]$
$\operatorname{Im}(\ )$	imaginary part of
j	$\sqrt{-1}$
$l(\mathbf{x})$	log-likelihood ratio
$L(\mathbf{x})$	likelihood ratio
$L_G(\mathbf{x})$	generalized likelihood ratio
$\operatorname{mse}(\hat{\theta})$	mean square error of $\hat{\theta}$ (classical)
$\mathbf{M}_{\hat{\theta}}$	mean square error matrix of $\hat{\boldsymbol{\theta}}$
μ	mean

n	sequence index
N	length of observed data set
$\mathcal{N}(\mu,\sigma^2)$	normal distribution with mean μ and variance σ^2
$\mathcal{N}(\boldsymbol{\mu},\mathbf{C})$	multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\mathbf{C}$
$\|\|x\|\|$	norm of x
$\mathbf{1}$	vector of all ones
$p(x)$ or $p_x(x)$	probability density function (PDF) of x
$p(\mathbf{x};\theta)$	PDF of $\mathbf{x}$ with θ as a parameter
$p(\mathbf{x};\mathcal{H}_i)$	PDF of $\mathbf{x}$ when $\mathcal{H}_i$ is true
$p(\mathbf{x};\theta,\mathcal{H}_i)$	PDF of $\mathbf{x}$ with θ as a parameter when $\mathcal{H}_i$ is true
$p(\mathbf{x}\|\theta)$	conditional PDF of $\mathbf{x}$ conditioned on θ
$p(\mathbf{x}\|\mathcal{H}_i)$	conditional PDF of $\mathbf{x}$ conditioned on $\mathcal{H}_i$ being true
$p(\mathbf{x}\|\theta,\mathcal{H}_i)$	conditional PDF of $\mathbf{x}$ conditioned on θ and $\mathcal{H}_i$ being true
$p(\mathbf{x};\theta\|\mathcal{H}_i)$	conditional PDF of $\mathbf{x}$ conditioned on $\mathcal{H}_i$ being true and with θ as a parameter
P_D	probability of detection
P_e	probability of error
P_{FA}	probability of false alarm
$P(\mathcal{H}_i)$	prior probability of $\mathcal{H}_i$
$P(\mathcal{H}_i\|\mathbf{x})$	posterior probability of $\mathcal{H}_i$
$\mathbf{P}$	projection matrix
$\mathbf{P}^{\perp}$	orthogonal projection matrix
$\frac{\partial}{\partial\mathbf{x}}$	gradient vector with respect to $\mathbf{x}$
$\frac{\partial^2}{\partial\mathbf{x}\partial\mathbf{x}^T}$	Hessian matrix with respect to $\mathbf{x}$
$\Pr\{\ \}$	probability
$P_{xx}(f)$	power spectral density of discrete-time process $x[n]$
$P_{xy}(f)$	cross-spectral density of discrete-time processes $x[n]$ and $y[n]$
$P_{xx}(F)$	power spectral density of continuous-time process $x(t)$
$\Phi(x)$	cumulative distribution function of $\mathcal{N}(0,1)$ random variable
ρ	correlation coefficient
ρ_s	signal correlation coefficient
$Q(x)$	probability that a $\mathcal{N}(0,1)$ random variable exceeds x
$Q^{-1}(u)$	value of $N(0,1)$ random variable which is exceeded with probability of u
$Q_{\chi^2_n}(x)$	probability that a χ^2_n random variable exceeds x
$Q_{\chi'^2_n(\lambda)}(x)$	probability that a $\chi'^2_n(\lambda)$ random variable exceeds x
$Q_{F_{m,n}}(x)$	probability that an $F_{m,n}$ random variable exceeds x
$Q_{F'_{m,n}(\lambda)}(x)$	probability that an $F'_{m,n}(\lambda)$ random variable exceeds x
$r_{xx}[k]$	autocorrelation function of discrete-time process $x[n]$
$r_{xx}(\tau)$	autocorrelation function of continuous-time process $x(t)$
$r_{xy}[k]$	cross-correlation function of discrete-time processes $x[n]$ and $y[n]$
$r_{xy}(\tau)$	cross-correlation function of continuous-time processes $x(t)$ and $y(t)$

$\mathbf{R}_{xx}$	autocorrelation matrix of $\mathbf{x}$
$\mathrm{Re}(\ )$	real part
σ^2	variance
$s[n]$	discrete-time signal
$\mathbf{s}$	vector of signal samples
$s(t)$	continuous-time signal
$\mathrm{sgn}(x)$	signum function ($=1$ for $x > 0$ and $=-1$ for $x < 0$)
$T(\mathbf{x})$	test statistic
$T_{\mathrm{R}}(\mathbf{x})$	Rao test statistic
$T_{\mathrm{W}}(\mathbf{x})$	Wald test statistic
t	continuous time
$\mathrm{tr}(\mathbf{A})$	trace of matrix $\mathbf{A}$
θ ($\boldsymbol{\theta}$)	unknown parameter (vector)
$\hat{\theta}$ ($\hat{\boldsymbol{\theta}}$)	estimator of θ ($\boldsymbol{\theta}$)
^{T}	transpose
$\mathcal{U}[a,b]$	uniform distribution over the interval $[a,b]$
$\mathrm{var}(x)$	variance of x
$\mathrm{var}(x;\mathcal{H}_i)$	variance of x assuming $\mathcal{H}_i$ true
$\mathrm{var}(x\|\theta)$	variance of conditional PDF or of $p(x\|\theta)$
$\mathrm{var}(x\|\mathcal{H}_i)$	variance of x conditioned on $\mathcal{H}_i$ being true
$\mathbf{V}$	modal matrix
$w[n]$	observation noise sequence
$\mathbf{w}$	vector of noise samples
$w(t)$	continuous-time noise
$x[n]$	observed discrete-time data
$\mathbf{x}$	vector of data samples
$x(t)$	observed continuous-time waveform
$\bar{x}$	sample mean of x
$\mathcal{Z}$	z transform
$\mathcal{Z}^{-1}$	inverse z transform
$\mathbf{0}$	vector or matrix of all zeros

Abbreviations

ACF	autocorrelation function
AG	array gain
ANC	adaptive noise canceler
AR	autoregressive
AR(p)	autoregressive process of order p
ARMA	autoregressive moving average
BLUE	best linear unbiased estimator
CCF	cross-correlation function
CDF	cumulative distribution function
CFAR	constant false alarm rate
CRLB	Cramer-Rao lower bound
CSM	cross-spectral matrix
CWGN	complex white Gaussian noise
DC	constant level (direct current)
DFT	discrete Fourier transform
EM	expectation-maximization
ENR	energy-to-noise ratio
FFT	fast Fourier transform
FIR	finite impulse response
GLRT	generalized likelihood ratio test
IID	independent and identically distributed
IIR	infinite impulse response
LLR	log-likelihood ratio
LRT	likelihood ratio test
LMMSE	linear minimum mean square error
LMP	locally most powerful
LO	locally optimum
LPC	linear predictive coding
LS	least squares
LSE	least squares estimator
LSI	linear shift invariant
MA	moving average
MAP	maximum a posteriori
ML	maximum likelihood
MLE	maximum likelihood estimator
MMSE	minimum mean square error
MSE	mean square error
MVDR	minimum variance distortionless response
MVU	minimum variance unbiased
NP	Neyman-Pearson
OOK	on-off keyed
PDF	probability density function
PG	processing gain
PRN	pseudorandom noise
PSD	power spectral density
RBLS	Rao-Blackwell-Lehmann-Scheffe
SNR	signal-to-noise ratio
TDL	tapped delay line (same as FIR)
TDOA	time difference of arrival
2-D	two-dimensional
WGN	white Gaussian noise
WSS	wide sense stationary

Index