An Introduction to Statistical Signal Processing With Applications

M. D. SRINATH

P. K. RAJASEKARAN

An Introduction to Statistical Signal Processing With Applications

M. D. SRINATH

Southern Methodist University
Dallas, Texas

P. K. RAJASEKARAN

Teledyne Brown Engineering
Huntsville, Alabama

A WILEY-INTERSCIENCE PUBLICATION

JOHN WILEY & SONS
New York • Chichester • Brisbane • Toronto

DS16/17

Library of Congress Cataloging in Publication Data:

Srinath, M D 1935–

An introduction to statistical signal processing with applications.

"A Wiley-Interscience publication."
Bibliography: p.
Includes Index.
1. Signal processing. 2. Signal theory (Telecommunication) 3. Estimation theory. I. Rajasekaran, P. K., 1942– joint author. II. Title.

TK5102.5.S68 621.38'0432 78-15417
ISBN 0-471-04404-0

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

PREFACE

Signal processing may broadly be considered to involve the recovery of information from physical observations. In the presence of random disturbances, we must resort to statistical techniques for the recovery of the information. Applications of statistical signal processing arise in a wide variety of areas such as communications, radio location of objects, seismic signal processing, and computer-assisted medical diagnosis.

This book is intended to serve as a first-year graduate level text in statistical signal processing and aims at covering certain basic techniques in the processing of stochastic signals and illustrating their use in a few specific applications. The motivation for writing this book arose from the instructional experience of the authors at Southern Methodist University (M.D.S.), Dallas, Texas, the Indian Institute of Technology (P.K.R.), Madras, India, and the Indian Institute of Science (M.D.S.), Bangalore, India, where they have been involved in teaching graduate level courses in statistical communication theory, radar systems, system identification, pattern recognition, stochastic control, and biomedical signal processing. An important realization of this experience was that an essentially common background in certain techniques for statistical signal processing is required for these apparently different courses. In particular, the topics of detection and estimation theory constitute a common foundation in many of these courses. Most available texts provide a treatment of these two topics in varying detail, but usually cover only one of the application areas in depth while making only a passing reference to the others. It was felt that there was a need for a text that presented these two basic topics in a clear and concise fashion and illustrated their use in several areas. The particular areas chosen for coverage in this text are communications, radar, pattern recognition, and system identification. These four areas are by no means exhaustive but were selected because of their generality and widespread use. Many of the techniques in the text can be either applied directly or extended to other application areas such as speech or image processing. It is expected that first-year graduate students as well as practicing engineers in different disciplines will use the text to acquire a background suitable for advanced study and a better understanding of the fields in which they are involved.

The background required to use this book is a course on systems theory and a course on stochastic processes. The mathematical rigor takes a middle path so as not to obscure intuitive reasoning. The first chapter presents an introduction and overview of the text. Chapter 2 is intended to serve as a quick review of certain results in systems theory and stochastic processes that are useful in later developments. The basic background in detection and estimation theory that is required for the latter applications chapters is provided in Chapters 3 through 6. The applications chapters, Chapters 7 through 10, can be covered selectively and in any order, depending on the interests of the class. For a class with strong interests in communication, a good suggestion is to review Chapter 2 quickly, follow up with Chapters 3 through 6 on theory, and emphasize Chapter 7 on communication systems and Chapter 8 on radar systems. On the other hand, for a class with strong interests in pattern recognition or control, Chapter 9 may be emphasized instead. Many exercises are included at the end of each chapter, serving the dual purpose of obtaining a better understanding of the text material, and presenting new applications or results not covered in the text.

We wish to express our gratitude to our graduate students and the reviewers of the manuscript for their comments which were extremely helpful in bringing this book to publication. Much of the contribution of one of the authors (P.K.R.) was accomplished during his stay at the Instituto Nacional de Astrofisica, Optica y Electronica (INAOE), Puebla, Mexico. He is grateful to Dr. G. Haro, Director-General and other authorities of INAOE for providing him the necessary facilities.

M. D. Srinath
P. K. Rajasekaran

Dallas, Texas
Huntsville, Alabama
March 1978

CONTENTS

CHAPTER 1

INTRODUCTION

Statistical signal processing has application in a wide variety of human activity. Such applications range from processing of seismic signals to computer-assisted medical diagnosis and treatment, tracking of objects in space, and traffic control.

Signal processing generally involves the recovery of information from physical observations. The processing required is relatively simple if the observation contains the information explicitly, and any interference present is exactly described. Often, the physical characteristics and limitations of the devices used for observation, and/or the media through which the information is observed or communicated, make this impossible. In fact, the interference is usually random in nature, and can only be described in terms of its average properties or statistics. The processing of such an observation in order to recover information can be termed *statistical signal processing*. This book is devoted to this topic and some applications.

To illustrate the type of problems that arise in statistical signal processing, let us consider the radio location of a flying object (target) in space. We can do this by beaming a packet of electromagnetic energy in the direction of the target and by observing the reflected electromagnetic wave. We have two problems to consider at this point. The primary problem is to decide whether there is an object present at all (the detection problem). If we decide that the object is present, we may desire to determine certain parameters associated with the object, say its range or velocity (the estimation problem). If there is no interference and if the reflected wave is not distorted through the transmission medium, the solution is straightforward. We monitor the reflected signal and observe the time delay τ between the transmitted and reflected waves, by noting the time at which a peak occurs. If the target is not present, there will be no peak. If a target is present, we can estimate its range as $R = \frac{\tau c}{2}$, where c is the velocity of propagation of the electromagnetic wave.

In the presence of interference (noise), the solution is no longer simple. The interference could be due to distortions through the transmission media, or thermal noise in the measurement device. The effect of the interference is to mask the peak that we are monitoring. We may get a spurious peak when no target is present or we may not be able to discern a peak when a target is present. In either case, the presence of noise can cause erroneous decisions. Our problem is to monitor the signal for a certain length of time, and decide about the presence of the object. This is the detection problem and falls into the general topic of statistical decision making.

If we decide that a target is present and seek to determine its range by observing the delay, we may still have difficulties because of the interference causing the peak to appear at the wrong time. We then have the problem of recovering the information (target range) from the "noisy" observation and is the estimation problem referred to earlier.

These two problems of detection and estimation arise in many other areas besides the radio location problem and are basic to all statistical signal processing techniques. Similar problems are encountered in areas such as communication, pattern recognition, and system identification. In analog communication systems, the transmitted message usually undergoes distortion during transmission. This distortion is usually characterized at the receiver as noise in the observations. In many cases, the message, for example, an audio or visual signal, may be modeled as a stochastic signal. The problem of recovering the message at the receiver can then be formulated as one of estimating a random signal in the presence of random noise.

In a digital communication system, the message is encoded into a sequence of binary digits (more generally into a sequence of several symbols). Typically these digits (bits), represented by a 1 or 0 are transmitted by sending suitably chosen pulses which again are distorted during transmission. The effect of this distortion is that, at the receiver it is no longer possible to determine which waveform was transmitted. We can again model the distortion during transmission as random noise in the receiver. The problem is one of deciding on the basis of noisy observations, whether the transmitted waveform corresponds to 1 or 0.

Recent work in efficient speech transmission involves characterizing the speech waveform by means of certain parameters that describe its spectrum. These parameters are then transmitted to the receiver which synthesizes the speech waveform from a knowledge of these paramters. The problem of extracting the parameters is an identification problem and is essentially one of estimating the parameters in a suitably chosen model for the speech waveform. These parameters undergo distortion during transmission so that at the receiver we have the problem of recovering (estimating) the parameters from noisy observations.

Pattern recognition systems also provide examples of systems involving signal processing. Consider that we want to design an automatic machine to distinguish between two handwritten characters, say "a" and "b". The characteristics of these letters will vary with the person writing them. Typically, a single specimen can be characterized as a sample from a population with a known statistical description. The machine is to be designed to distinguish between the two classes or categories—the letters "a" and "b"—when a sample is presented to it. Often the statistics of the population will be unknown and will have to be determined from samples from each of the two classes. Another pattern classification example arises in electroencephalograph(EEG) analysis. The EEG is a recording of electrical brain signals and may be used in a variety of clinical purposes. A typical application is in the determination of the sleep state of a patient. The EEG signals are corrupted by noise introduced by the measuring device so that the characteristic features of the EEG will have to be estimated in order to determine the sleep state. Other applications involve the study of evoked potentials or response to stimuli and can be used for determining sensory perception. The evoked response is superimposed on the normal EEG, which in this case may be treated as interference. The useful information from the evoked response is obtained in terms of certain characteristic lines in the power spectrum which can be used for diagnosis and study of any abnormalties.

The preceding has served to point out that the problems of detection and estimation arise either seperately or jointly in a wide variety of applications. While there is a seeming dichotomy between the two, in fact, there is an underlying similarity in the structure of the two problems, which facilitates the solution of many signal-processing problems. The aim of this book is to develop the fundamentals of detection and estimation theory as the basis of statistical signal processing and illustrate the application of the concepts and techniques in various areas. The application areas chosen are communications, radar systems, pattern recognition, and system identification. These areas are by no means exhaustive, but have been chosen because they are general enough to be of interest to a fairly wide audience.

1.1 ORGANIZATION

The book is addressed to the first-year graduate student in the disciplines of electrical and systems engineering. It also will serve the purpose of a graduate engineer in industry who is exposed to a wide variety of signal-processing problems but has not had an opportunity to obtain a cohesive background. Typically the student of this test may be expected to have had a prior course in systems and a course in probability and stochastic processes. The exposure to systems should preferably have included state variable and transform techniques for both continuous-time and discrete-time systems.

The book essentially consists of two parts. The first part consisting of Chapters 2 through 6 provides the basic framework of detection and estimation theory, while the second part which consists of Chapters 7 through 9 covers applications to the four areas mentioned earlier. Specifically our coverage is as follows. In Chapter 2, we provide a quick review of basic concepts from systems theory and stochastic processes which is useful in the developments in the later chapters. In particular we discuss Gauss–Markov models for signal generation and the complex representation of band-limited signals.

Chapter 3 is concerned with the first class of statistical signal-processing problems, namely, decision problems. In this chapter we discuss classical decision theory using statistical hypotheses testing methods. Various decision criteria are introduced and decision rules obtained. The performance of the decision rules are also investigated. The sequential decision test of Wald is also included for the sake of completeness. These concepts are extended in Chapter 4 to the detection of waveforms observed in noise. The results are primarily derived in terms of a binary hypotheses testing problem in which we are required to detect which one of two known waveforms is present in the observations. The correlation receiver and the matched filter are obtained for optimum processing when the interference is additive white noise. Receiver structures for signals observed in colored noise are derived. The chapter concludes with a brief section on the detection of signals with unknown parameters.

Chapter 5 deals with the problem of estimating parameters from observations that are corrupted by noise. The Bayes' estimators for parameters are obtained. Bounds on the performance of these estimators are presented. The concept of a reproducing density is introduced and applied to the estimation of parameters of probability density functions.

Chapter 6 considers the minimum-mean-square estimation of signals observed in noise. For stationary processes, the classical Wiener filter is derived for both continuous-time and discrete-time processes. The Kalman filter which provides an attractive recursive solution to the estimation problem in nonstationary processes is discussed. Estimation techniques for certain nonlinear signal models are also considered.

After developing the powerful tools of detection and estimation theory in these four chapters, in Chapter 7 we consider applications to communication systems. We first consider digital communication systems and develop optimum receiver structures for a variety of digital transmission schemes. We consider the performance of these receivers in various situations, including fading. Techniques of synchronization are presented and the problem of intersymbol interference discussed. The second part of this chapter considers demodulator structures for demodulation of analog modulated signals. Two different techniques, one using the maximum a posteriori estimation technique and the other using the minimum-mean-square estimation technique are used in deriving optimum demodulator structures.

In Chapter 8, we study applications to radar systems. The models for radar targets are derived using the complex representation of signals developed in Chapter 2. In addition to studying the detection of fluctuating and nonfluctuating targets, the problem of estimating such radar parameters as delay and Doppler frequency is investigated, and the concept of signal design to improve the accuracy in estimating these parameters is presented. The chapter concludes with an application of the Kalman filter for dynamic target tracking and maneuver detection.

Chapter 9 considers applications to pattern classification and system identification. The problem of pattern classification is posed as a hypothesis testing problem by associating the various pattern classes with hypotheses. The problem of learning the parameters of the associated density functions under both supervised and unsupervised conditions is discussed. The important problem of extracting the features of a pattern is also presented.

The second part of this chapter discusses the identification of systems based on input and output measurements. We first consider the identification of the parameters of an autoregressive (AR) or autoregressive-moving-average (ARMA) model of a stochastic process. Identification in signal models using state augmentation techniques is discussed. The maximum likelihood technique for identification in linear time-invariant models is also presented.

Three appendices are included. Appendix A provides a brief review of the bilateral Laplace and z-transforms which are used widely in the text. Appendix B summarizes pertinent features of certain optimization techniques which are used in Chapter 4 while Appendix C discusses vector and matrix operations which are useful. In particular, a useful matrix lemma is presented in this appendix.

While Chapters 3 through 6 are essential for an understanding of the remainder of the text, the text is structured such that the applications chapters can be covered selectively and in any order depending on the interest and background of the intended audience. The exercises at the end of each chapter serve a dual purpose—to better understand the material covered and to present new applications or results not covered in the text. While each chapter contains several references, a supplementary bibliography of about 200 references which contain many pertinent contributions, is provided at the end of the text.

CHAPTER 2

SIGNALS AND SYSTEMS

2.1 INTRODUCTION

In the previous chapter we have presented several examples in which statistical signal processing plays an important role. The role of the signal processor is to extract information from the observed signals and present it in a useful form. The proper design of the processor requires an appropriate description or characterization of the signals as well as of the systems generating these signals. For example, in radar tracking, we would like to know the location and velocity of the object being tracked. The signal of interest in this case is the radar return which is usually corrupted by various noises such as thermal noise in the processor. A description of the signal can be given in terms of the equations of motion of the object being tracked while the noise is usually modeled as a sample function of a stochastic process.

In this chapter we briefly review models for the generation of random signals and discuss some representations of these signals. We will find these results to be of use in our discussions in later chapters.

2.2 SYSTEM THEORY [1,2]

By a *system* we mean a model of a physical object or a collection of physical objects. A model is essentially an idealization of a physical object which retains the salient features of the system, but is easier to study. In this section we discuss some concepts and techniques which are useful in studying the behavior of systems.

A given physical system may give rise to many different models and the choice of a model depends on the use to which it is to be put. As an example, in studies of the trajectory of a satellite, we may use a particle as the model. In maneuvering, however, a rigid body model would be more appropriate.

Once a model has been determined, the next step is to obtain a mathematical representation of it. This requires the selection of pertinent variables, reference directions and coordinate axes and the application of appropriate physical laws. The equations that describe a system may assume many forms; they may be algebraic equations, integral or differential equations, and so on. The choice of a particular representation again depends on the application. Once the mathematical description of a system is obtained, the next step is to evaluate the behavior of the system and to modify the system if the behavior is not acceptable. The evaluation may be carried out by observing the general properties of the system, or by determining specific responses to typical inputs.

In obtaining a mathematical model we define certain quantities (variables) as the inputs to the system and certain other quantities as the outputs. For example, in an electric motor, we may consider the armature voltage or the field current as inputs and the shaft speed as the output. In general the inputs and output variables will be functions of time. If the output at any instant t, is dependent only on the value of the inputs at that instant, the system is a *zero-memory* system. Most systems, however, have memory. That is the output at t_1 depends not only on the inputs applied at t_1 but also on the inputs applied prior to and/or after t_1. Such systems are called dynamical systems. For dynamical systems, if an input $u(t)$ is applied for $t < t_1$, the output $y(t)$ for $t > t_1$ cannot in general, be uniquely determined, unless $u(t)$ is known for $t < t_1$. If two inputs $u_1(t)$ and $u_2(t)$ which are identical for $t > t_1$ but distinct for $t < t_1$ are applied to the system, the output $y(t)$ for $t > t_1$ will be different for the two inputs. For an input to give rise to a unique output, the system must be *initially relaxed* or *at rest*. That is if an input is applied for $t > t_1$, there must be no stored energy in the system prior to t_1. For systems initially relaxed at $t = -\infty$, we can write the relation between the input and the output as

$$y(t) = \mathrm{T}\{u(t)\} \tag{2.2.1}$$

where T is an operator or function describing this relationship.

We consider only systems in which the input–output relation is a differential equation (differential systems) or a difference equation (discrete systems). The system is linear or nonlinear depending on whether these equations are linear or nonlinear.

All physical systems are *causal* or *nonanticipative*. That is, if the system is initially relaxed, there can be no output prior to an input being applied. More precisely, if $[u(t), y(t)]$ and $[u_1(t), y_1(t)]$ correspond to two input–output pairs, the system is causal if $u(t) = u_1(t)$ for $t < t_1$ implies that $y(t) = y_1(t)$ for $t < t_1$. In some of our applications, however, we will have occasion to use noncausal models. An example is the ideal lowpass filter.

Linear Systems

The system described by Eq. (2.2.1) is said to be linear if, for any two arbitrary inputs $u_1(t)$ and $u_2(t)$ and any two scalars α_1 and α_2

$$\mathbf{T}\{\alpha_1 u_1(t) + \alpha_2 u_2(t)\} = \alpha_1 \mathbf{T}\{u_1(t)\} + \alpha_2 \mathbf{T}\{u_2(t)\} \tag{2.2.2}$$

It easily follows from Eq. (2.2.2) that for any finite n

$$\mathbf{T}\left\{\sum_{i=1}^{n} \alpha_i u_i(t)\right\} = \sum_{i=1}^{n} \alpha_i \mathbf{T}\{u_i(t)\} \tag{2.2.3}$$

We will assume that Eq. (2.2.3) holds also for infinite sums and integrals. In regard to input–output relationships, the terms *linear system* and *linear operator* are equivalent. Equation (2.2.2) is also referred to as the principle of superposition. We note that for the principle to hold, the system must be initially relaxed. That is, all initial conditions in the system are zero.

A *linear differential system* is characterized in terms of its input–output relation by a linear differential equation.

A system is *time invariant* or *stationary* if its parameters do not vary with time. For example the differential equation characterizing the system will be invariant with respect to a shift in the time-axis. That is, if the response to an input $u(t)$ is $y(t)$, then the response to the input $u(t)$ delayed by τ equals $y(t-\tau)$:

$$\mathbf{T}\{u(t-\tau)\} = y(t-\tau) \tag{2.2.4}$$

Impulse Response and the Convolution Integral

A linear system can be characterized in terms of its response to certain elementary inputs. One such input is the impulse function. Let us denote by $h(t, \tau)$ the response of the system at time t to a unit impulse applied at time τ. We note that $h(t, \tau)$ denotes a family of functions indexed by the parameter τ.

The output $y(t)$ to an arbitrary input $u(t)$ is then obtained from the convolution integral as

$$y(t) = \int_{-\infty}^{\infty} h(t, \tau)u(\tau)\, d\tau \tag{2.2.5}$$

We say that $y(t)$ is obtained as the convolution of the two functions $h(t, \tau)$ and $u(t)$ and write

$$y(t) = h(t, \tau) * u(t) \tag{2.2.6}$$

where $*$ refers to the convolution operation of Eq. (2.2.5).

For a causal system, the response is zero prior to an input being applied. Thus for a causal system we can write

$$h(t, \tau) = 0 \qquad t < \tau \tag{2.2.7}$$

It follows that the output of a linear causal system is given by

$$y(t) = \int_{-\infty}^{t} h(t, \tau)u(\tau)\, d\tau \tag{2.2.8}$$

If the input is zero prior to a finite time t_0, Eq. (2.2.8) reduces to

$$y(t) = \int_{t_0}^{t} h(t, \tau)u(\tau)\, d\tau \tag{2.2.9}$$

For time-invariant systems the convolution integral, Eq. (2.2.5), becomes

$$y(t) = \int_{-\infty}^{\infty} h(t-\tau)u(\tau)\, d\tau = \int_{-\infty}^{\infty} h(\tau)u(t-\tau)\, d\tau \tag{2.2.10}$$

If the system is causal, then $h(t) = 0$ for $t < 0$ and the output is given by

$$y(t) = \int_{-\infty}^{t} h(t-\tau)u(\tau)\, d\tau = \int_{0}^{\infty} h(\tau)u(t-\tau)\, d\tau \tag{2.2.11}$$

We will refer to any function $f(t)$ of t which is zero for negative values of t as a *causal function*. Similarly if $f(t) = 0$ for $t > 0$, we will refer to $f(t)$ as an *anticausal* function. If

$$f(t) = \begin{cases} f_a(t) & t < 0 \\ f_c(t) & t \geqslant 0 \end{cases} \tag{2.2.12}$$

we will refer to $f_c(t)$ as the *causal part* of $f(t)$ and $f_a(t)$ as the *anticausal* part of $f(t)$.