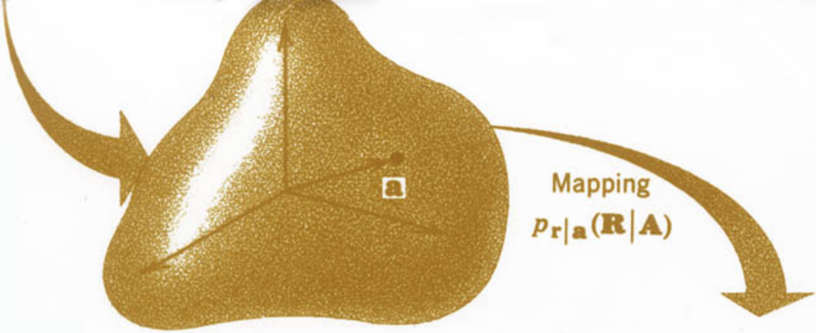

DETECTION, ESTIMATION, AND MODULATION THEORY

Part I

HARRY L. VAN TREES

Detection, Estimation, and Modulation Theory

This page intentionally left blank

Detection, Estimation, and Modulation Theory

Part I. Detection, Estimation, and Linear Modulation Theory

HARRY L. VAN TREES
George Mason University

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.
New York • Chichester • Weinheim • Brisbane • Singapore • Toronto

This text is printed on acid-free paper. ♾

Published simultaneously in Canada.

For ordering and customer service, call 1-800-CALL-WILEY.

Library of Congress Cataloging in Publication Data is available.

ISBN 0-471-09517-6

Printed in the United States of America

10 9 8 7 6 5 4 3

To Diane

and Stephen, Mark, Kathleen, Patricia,
Eileen, Harry, and Julia

and the next generation—
Brittany, Erin, Thomas, Elizabeth, Emily,
Dillon, Bryan, Julia, Robert, Margaret,
Peter, Emma, Sarah, Harry, Rebecca,
and Molly

This page intentionally left blank

Preface for Paperback Edition

In 1968, Part I of *Detection, Estimation, and Modulation Theory* [VT68] was published. It turned out to be a reasonably successful book that has been widely used by several generations of engineers. There were thirty printings, but the last printing was in 1996. Volumes II and III ([VT71a], [VT71b]) were published in 1971 and focused on specific application areas such as analog modulation, Gaussian signals and noise, and the radar–sonar problem. Volume II had a short life span due to the shift from analog modulation to digital modulation. Volume III is still widely used as a reference and as a supplementary text. In a moment of youthful optimism, I indicated in the the Preface to Volume III and in Chapter III-14 that a short monograph on optimum array processing would be published in 1971. The bibliography lists it as a reference, *Optimum Array Processing,* Wiley, 1971, which has been subsequently cited by several authors. After a 30-year delay, *Optimum Array Processing,* Part IV of *Detection, Estimation, and Modulation Theory* will be published this year.

A few comments on my career may help explain the long delay. In 1972, MIT loaned me to the Defense Communication Agency in Washington, D.C. where I spent three years as the Chief Scientist and the Associate Director of Technology. At the end of the tour, I decided, for personal reasons, to stay in the Washington, D.C. area. I spent three years as an Assistant Vice-President at COMSAT where my group did the advanced planning for the INTELSAT satellites. In 1978, I became the Chief Scientist of the United States Air Force. In 1979, Dr. Gerald Dinneen, the former Director of Lincoln Laboratories, was serving as Assistant Secretary of Defense for C3I. He asked me to become his Principal Deputy and I spent two years in that position. In 1981, I joined M/A-COM Linkabit. Linkabit is the company that Irwin Jacobs and Andrew Viterbi had started in 1969 and sold to M/A-COM in 1979. I started an Eastern operation which grew to about 200 people in three years. After Irwin and Andy left M/A-COM and started Qualcomm, I was responsible for the government operations in San Diego as well as Washington, D.C. In 1988, M/A-COM sold the division. At that point I decided to return to the academic world.

I joined George Mason University in September of 1988. One of my priorities was to finish the book on optimum array processing. However, I found that I needed to build up a research center in order to attract young research-oriented faculty and doctoral students. The process took about six years. The Center for Excellence in Command, Control, Communications, and Intelligence has been very successful and has generated over $300 million in research funding during its existence. During this growth period, I spent some time on array processing but a concentrated effort was not possible. In 1995, I started a serious effort to write the Array Processing book.

Throughout the *Optimum Array Processing* text there are references to Parts I and III of *Detection, Estimation, and Modulation Theory*. The referenced material is available in several other books, but I am most familiar with my own work. Wiley agreed to publish Part I and III in paperback so the material will be readily available. In addition to providing background for Part IV, Part I is still useful as a text for a graduate course in Detection and Estimation Theory. Part III is suitable for a second level graduate course dealing with more specialized topics.

In the 30-year period, there has been a dramatic change in the signal processing area. Advances in computational capability have allowed the implementation of complex algorithms that were only of theoretical interest in the past. In many applications, algorithms can be implemented that reach the theoretical bounds.

The advances in computational capability have also changed how the material is taught. In Parts I and III, there is an emphasis on compact analytical solutions to problems. In Part IV, there is a much greater emphasis on efficient iterative solutions and simulations. All of the material in parts I and III is still relevant. The books use continuous time processes but the transition to discrete time processes is straightforward. Integrals that were difficult to do analytically can be done easily in Matlab®. The various detection and estimation algorithms can be simulated and their performance compared to the theoretical bounds. We still use most of the problems in the text but supplement them with problems that require Matlab® solutions.

We hope that a new generation of students and readers find these reprinted editions to be useful.

HARRY L. VAN TREES

Fairfax, Virginia
June 2001

Preface

The area of detection and estimation theory that we shall study in this book represents a combination of the classical techniques of statistical inference and the random process characterization of communication, radar, sonar, and other modern data processing systems. The two major areas of statistical inference are decision theory and estimation theory. In the first case we observe an output that has a random character and decide which of two possible causes produced it. This type of problem was studied in the middle of the eighteenth century by Thomas Bayes [1]. In the estimation theory case the output is related to the value of some parameter of interest, and we try to estimate the value of this parameter. Work in this area was published by Legendre [2] and Gauss [3] in the early nineteenth century. Significant contributions to the classical theory that we use as background were developed by Fisher [4] and Neyman and Pearson [5] more than 30 years ago. In 1941 and 1942 Kolmogoroff [6] and Wiener [7] applied statistical techniques to the solution of the optimum linear filtering problem. Since that time the application of statistical techniques to the synthesis and analysis of all types of systems has grown rapidly. The application of these techniques and the resulting implications are the subject of this book.

This book and the subsequent volume, Detection, Estimation, and Modulation Theory, Part II, are based on notes prepared for a course entitled "Detection, Estimation, and Modulation Theory," which is taught as a second-level graduate course at M.I.T. My original interest in the material grew out of my research activities in the area of analog modulation theory. A preliminary version of the material that deals with modulation theory was used as a text for a summer course presented at M.I.T. in 1964. It turned out that our viewpoint on modulation theory could best be understood by an audience with a clear understanding of modern detection and estimation theory. At that time there was no suitable text available to cover the material of interest and emphasize the points that I felt were important, so I started writing notes. It was clear that in order to present the material to graduate students in a reasonable amount of time it would be necessary to develop a unified presentation of the three topics: detection, estimation, and modulation theory, and exploit the fundamental ideas that connected them. As the development proceeded, it grew in size until the material that was originally intended to be background for modulation theory occupies the entire contents of this book. The original material on modulation theory starts at the beginning of the second book. Collectively, the two books provide a unified coverage of the three topics and their application to many important physical problems.

For the last three years I have presented successively revised versions of the material in my course. The audience consists typically of 40 to 50 students who have completed a graduate course in random processes which covered most of the material in Davenport and Root [8]. In general, they have a good understanding of random process theory and a fair amount of practice with the routine manipulation required to solve problems. In addition, many of them are interested in doing research in this general area or closely related areas. This interest provides a great deal of motivation which I exploit by requiring them to develop many of the important ideas as problems. It is for this audience that the book is primarily intended. The appendix contains a detailed outline of the course.

On the other hand, many practicing engineers deal with systems that have been or should have been designed and analyzed with the techniques developed in this book. I have attempted to make the book useful to them. An earlier version was used successfully as a text for an in-plant course for graduate engineers.

From the standpoint of specific background little advanced material is required. A knowledge of elementary probability theory and second moment characterization of random processes is assumed. Some familiarity with matrix theory and linear algebra is helpful but certainly not necessary. The level of mathematical rigor is low, although in most sections the results could be rigorously proved by simply being more careful in our derivations. We have adopted this approach in order not to obscure the important ideas with a lot of detail and to make the material readable for the kind of engineering audience that will find it useful. Fortunately, in almost all cases we can verify that our answers are intuitively logical. It is worthwhile to observe that this ability to check our answers intuitively would be necessary even if our derivations were rigorous, because our ultimate objective is to obtain an answer that corresponds to some physical system of interest. It is easy to find physical problems in which a plausible mathematical model and correct mathematics lead to an unrealistic answer for the original problem.

We have several idiosyncrasies that it might be appropriate to mention. In general, we look at a problem in a fair amount of detail. Many times we look at the same problem in several different ways in order to gain a better understanding of the meaning of the result. Teaching students a number of ways of doing things helps them to be more flexible in their approach to new problems. A second feature is the necessity for the reader to solve problems to understand the material fully. Throughout the course and the book we emphasize the development of an ability to work problems. At the end of each chapter are problems that range from routine manipulations to significant extensions of the material in the text. In many cases they are equivalent to journal articles currently being published. Only by working a fair number of them is it possible to appreciate the significance and generality of the results. Solutions for an individual problem will be supplied on request, and a book containing solutions to about one third of the problems is available to faculty members teaching the course. We are continually generating new problems in conjunction with the course and will send them to anyone who is using the book as a course text. A third issue is the abundance of block diagrams, outlines, and pictures. The diagrams are included because most engineers (including myself) are more at home with these items than with the corresponding equations.

One problem always encountered is the amount of notation needed to cover the large range of subjects. We have tried to choose the notation in a logical manner and to make it mnemonic. All the notation is summarized in the glossary at the end of the book. We have tried to make our list of references as complete as possible and to acknowledge any ideas due to other people.

A number of people have contributed in many ways and it is a pleasure to acknowledge them. Professors W. B. Davenport and W. M. Siebert have provided continual encouragement and technical comments on the various chapters. Professors Estil Hoversten and Donald Snyder of the M.I.T. faculty and Lewis Collins, Arthur Baggeroer, and Michael Austin, three of my doctoral students, have carefully read and criticized the various chapters. Their suggestions have improved the manuscript appreciably. In addition, Baggeroer and Collins contributed a number of the problems in the various chapters and Baggeroer did the programming necessary for many of the graphical results. Lt. David Wright read and criticized Chapter 2. L. A. Frasco and H. D. Goldfein, two of my teaching assistants, worked all of the problems in the book. Dr. Howard Yudkin of Lincoln Laboratory read the entire manuscript and offered a number of important criticisms. In addition, various graduate students taking the course have made suggestions which have been incorporated. Most of the final draft was typed by Miss Aina Sils. Her patience with the innumerable changes is

sincerely appreciated. Several other secretaries, including Mrs. Jarmila Hrbek, Mrs. Joan Bauer, and Miss Camille Tortorici, typed sections of the various drafts.

As pointed out earlier, the books are an outgrowth of my research interests. This research is a continuing effort, and I shall be glad to send our current work to people working in this area on a regular reciprocal basis. My early work in modulation theory was supported by Lincoln Laboratory as a summer employee and consultant in groups directed by Dr. Herbert Sherman and Dr. Barney Reiffen. My research at M.I.T. was partly supported by the Joint Services and the National Aeronautics and Space Administration under the auspices of the Research Laboratory of Electronics. This support is gratefully acknowledged.

Harry L. Van Trees

Cambridge, Massachusetts
October, 1967.

REFERENCES

[1] Thomas Bayes, "An Essay Towards Solving a Problem in the Doctrine of Chances," *Phil. Trans*, **53**, 370–418 (1764).
[2] A. M. Legendre, *Nouvelles Méthodes pour La Détermination ces Orbites des Comètes*, Paris, 1806.
[3] K. F. Gauss, *Theory of Motion of the Heavenly Bodies Moving About the Sun in Conic Sections*, reprinted by Dover, New York, 1963.
[4] R. A. Fisher, "Theory of Statistical Estimation," *Proc. Cambridge Philos. Soc.*, **22**, 700 (1925).
[5] J. Neyman and E. S. Pearson, "On the Problem of the Most Efficient Tests of Statistical Hypotheses," *Phil. Trans. Roy. Soc. London*, A **231**, 289, (1933).
[6] A. Kolmogoroff, "Interpolation and Extrapolation von Stationären Zufälligen Folgen," *Bull. Acad. Sci. USSR, Ser. Math.* **5**, 1941.
[7] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, Tech. Press of M.I.T. and Wiley, New York, 1949 (originally published as a classified report in 1942).
[8] W. B. Davenport and W. L. Root, *Random Signals and Noise*, McGraw-Hill, New York, 1958.

Contents

1

Introduction

In these two books, we shall study three areas of statistical theory, which we have labeled detection theory, estimation theory, and modulation theory. The goal is to develop these theories in a common mathematical framework and to demonstrate how they can be used to solve a wealth of practical problems in many diverse physical situations.

In this chapter we present three outlines of the material. The first is a topical outline in which we develop a qualitative understanding of the three areas by examining some typical problems of interest. The second is a logical outline in which we explore the various methods of attacking the problems. The third is a chronological outline in which we explain the structure of the books.

1.1 TOPICAL OUTLINE

An easy way to explain what is meant by detection theory is to examine several physical situations that lead to detection theory problems.

A simple digital communication system is shown in Fig. 1.1. The source puts out a binary digit every T seconds. Our object is to transmit this sequence of digits to some other location. The channel available for transmitting the sequence depends on the particular situation. Typically, it could be a telephone line, a radio link, or an acoustical channel. For

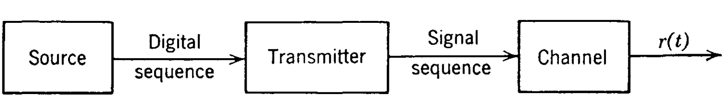

Fig. 1.1 Digital communication system.

purposes of illustration, we shall consider a radio link. In order to transmit the information, we must put it into a form suitable for propagating over the channel. A straightforward method would be to build a device that generates a sine wave,

$$s_1(t) = \sin \omega_1 t, \tag{1}$$

for T seconds if the source generated a "one" in the preceding interval, and a sine wave of a different frequency,

$$s_0(t) = \sin \omega_0 t, \tag{2}$$

for T seconds if the source generated a "zero" in the preceding interval. The frequencies are chosen so that the signals $s_0(t)$ and $s_1(t)$ will propagate over the particular radio link of concern. The output of the device is fed into an antenna and transmitted over the channel. Typical source and transmitted signal sequences are shown in Fig. 1.2. In the simplest kind of channel the signal sequence arrives at the receiving antenna attenuated but essentially undistorted. To process the received signal we pass it through the antenna and some stages of rf-amplification, in the course of which a thermal noise $n(t)$ is added to the message sequence. Thus in any T-second interval we have available a waveform $r(t)$ in which

$$r(t) = s_1(t) + n(t), \qquad 0 \le t \le T, \tag{3}$$

if $s_1(t)$ was transmitted, and

$$r(t) = s_0(t) + n(t), \qquad 0 \le t \le T, \tag{4}$$

if $s_0(t)$ was transmitted. We are now faced with the problem of deciding which of the two possible signals was transmitted. We label the device that does this a decision device. It is simply a processor that observes $r(t)$ and guesses whether $s_1(t)$ or $s_0(t)$ was sent according to some set of rules. This is equivalent to guessing what the source output was in the preceding interval. We refer to designing and evaluating the processor as a detection

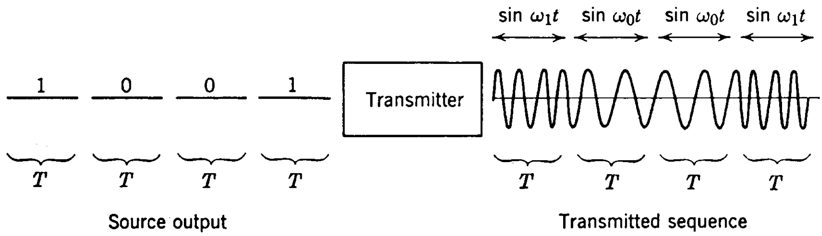

Fig. 1.2 Typical sequences.

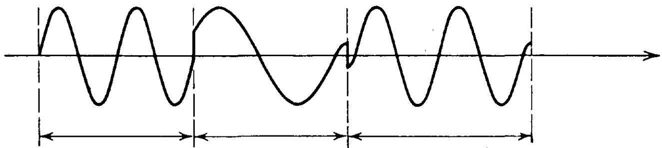

Fig. 1.3 Sequence with phase shifts.

theory problem. In this particular case the only possible source of error in making a decision is the additive noise. If it were not present, the input would be completely known and we could make decisions without errors. We denote this type of problem as the *known signal in noise problem.* It corresponds to the lowest level (i.e., simplest) of the detection problems of interest.

An example of the next level of detection problem is shown in Fig. 1.3. The oscillators used to generate $s_1(t)$ and $s_0(t)$ in the preceding example have a phase drift. Therefore in a particular T-second interval the received signal corresponding to a "one" is

$$r(t) = \sin(\omega_1 t + \theta_1) + n(t), \qquad 0 \le t \le T, \tag{5}$$

and the received signal corresponding to a "zero" is

$$r(t) = \sin(\omega_0 t + \theta_0) + n(t), \qquad 0 \le t \le T, \tag{6}$$

where θ_0 and θ_1 are unknown constant phase angles. Thus even in the absence of noise the input waveform is not completely known. In a practical system the receiver may include auxiliary equipment to measure the oscillator phase. If the phase varies slowly enough, we shall see that essentially perfect measurement is possible. If this is true, the problem is the same as above. However, if the measurement is not perfect, we must incorporate the signal uncertainty in our model.

A corresponding problem arises in the radar and sonar areas. A conventional radar transmits a pulse at some frequency ω_c with a rectangular envelope:

$$s_t(t) = \sin \omega_c t, \qquad 0 \le t \le T. \tag{7}$$

If a target is present, the pulse is reflected. Even the simplest target will introduce an attenuation and phase shift in the transmitted signal. Thus the signal available for processing in the interval of interest is

$$\begin{aligned} r(t) &= V_r \sin[\omega_c(t - \tau) + \theta_r] + n(t), && \tau \le t \le \tau + T, \\ &= n(t), && 0 \le t < \tau, \tau + T < t < \infty, \end{aligned} \tag{8}$$

if a target is present and

$$r(t) = n(t), \qquad 0 \leq t < \infty, \tag{9}$$

if a target is absent. We see that in the absence of noise the signal still contains three unknown quantities: V_r, the amplitude, θ_r, the phase, and τ, the round-trip travel time to the target.

These two examples represent the second level of detection problems. We classify them as *signal with unknown parameters in noise problems.*

Detection problems of a third level appear in several areas. In a passive sonar detection system the receiver listens for noise generated by enemy vessels. The engines, propellers, and other elements in the vessel generate acoustical signals that travel through the ocean to the hydrophones in the detection system. This composite signal can best be characterized as a sample function from a random process. In addition, the hydrophone generates self-noise and picks up sea noise. Thus a suitable model for the detection problem might be

$$r(t) = s_{\Omega}(t) + n(t) \tag{10}$$

if the target is present and

$$r(t) = n(t) \tag{11}$$

if it is not. In the absence of noise the signal is a sample function from a random process (indicated by the subscript Ω).

In the communications field a large number of systems employ channels in which randomness is inherent. Typical systems are tropospheric scatter links, orbiting dipole links, and chaff systems. A common technique is to transmit one of two signals separated in frequency. (We denote these frequencies as ω_1 and ω_0.) The resulting received signal is

$$r(t) = s_{\Omega_1}(t) + n(t) \tag{12}$$

if $s_1(t)$ was transmitted and

$$r(t) = s_{\Omega_0}(t) + n(t) \tag{13}$$

if $s_0(t)$ was transmitted. Here $s_{\Omega_1}(t)$ is a sample function from a random process centered at ω_1, and $s_{\Omega_0}(t)$ is a sample function from a random process centered at ω_0. These examples are characterized by the lack of any deterministic signal component. Any decision procedure that we design will have to be based on the difference in the statistical properties of the two random processes from which $s_{\Omega_0}(t)$ and $s_{\Omega_1}(t)$ are obtained. This is the third level of detection problem and is referred to as a *random signal in noise problem.*

In our examination of representative examples we have seen that detection theory problems are characterized by the fact that we must decide which of several alternatives is true. There were only two alternatives in the examples cited; therefore we refer to them as binary detection problems. Later we will encounter problems in which there are M alternatives available (the M-ary detection problem). Our hierarchy of detection problems is presented graphically in Fig. 1.4.

There is a parallel set of problems in the estimation theory area. A simple example is given in Fig. 1.5, in which the source puts out an analog message $a(t)$ (Fig. 1.5*a*). To transmit the message we first sample it every T seconds. Then, every T seconds we transmit a signal that contains

	Detection theory
Level 1. Known signals in noise	1. Synchronous digital communication 2. Pattern recognition problems
Level 2. Signals with unknown parameters in noise	1. Conventional pulsed radar or sonar, target detection 2. Target classification (orientation of target unknown) 3. Digital communication systems without phase reference 4. Digital communication over slowly-fading channels
Level 3. Random signals in noise	1. Digital communication over scatter link, orbiting dipole channel, or chaff link 2. Passive sonar 3. Seismic detection system 4. Radio astronomy (detection of noise sources)

Fig. 1.4 Detection theory hierarchy.

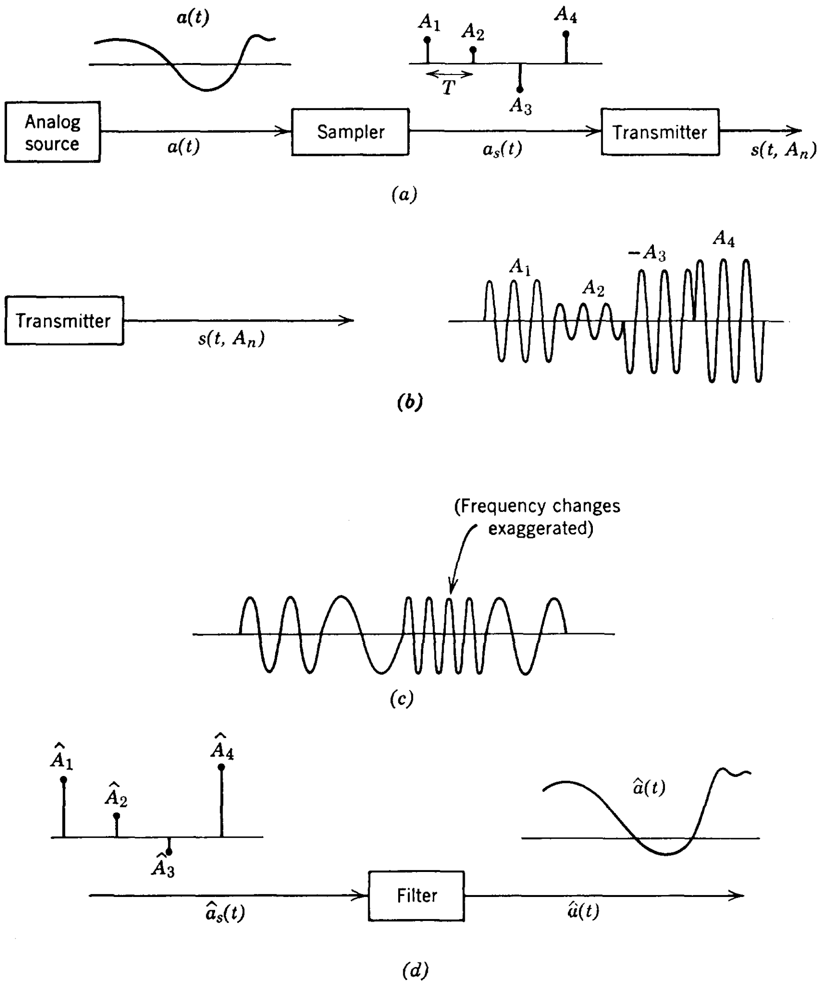

Fig. 1.5 **(*a*) Sampling an analog source; (*b*) pulse-amplitude modulation; (*c*) pulse-frequency modulation; (*d*) waveform reconstruction.**

a parameter which is uniquely related to the last sample value. In Fig. 1.5*b* the signal is a sinusoid whose amplitude depends on the last sample. Thus, if the sample at time nT is A_n, the signal in the interval $[nT, (n + 1)T]$ is

$$s(t, A_n) = A_n \sin \omega_c t, \qquad nT \le t \le (n + 1)T. \tag{14}$$

A system of this type is called a pulse amplitude modulation (PAM) system. In Fig. 1.5*c* the signal is a sinusoid whose frequency in the interval differs from the reference frequency ω_c by an amount proportional to the preceding sample value,

$$s(t, A_n) = \sin(\omega_c t + A_n t), \qquad nT \le t \le (n+1)T. \tag{15}$$

A system of this type is called a pulse frequency modulation (PFM) system. Once again there is additive noise. The received waveform, given that A_n was the sample value, is

$$r(t) = s(t, A_n) + n(t), \qquad nT \le t \le (n+1)T. \tag{16}$$

During each interval the receiver tries to estimate A_n. We denote these estimates as $\hat{A}_n$. Over a period of time we obtain a sequence of estimates, as shown in Fig. 1.5*d*, which is passed into a device whose output is an estimate of the original message $a(t)$. If $a(t)$ is a band-limited signal, the device is just an ideal low-pass filter. For other cases it is more involved.

If, however, the parameters in this example were known and the noise were absent, the received signal would be completely known. We refer to problems in this category as *known signal in noise problems.* If we assume that the mapping from A_n to $s(t, A_n)$ in the transmitter has an inverse, we see that if the noise were not present we could determine A_n unambiguously. (Clearly, if we were allowed to design the transmitter, we should always choose a mapping with an inverse.) The *known signal in noise problem* is the first level of the estimation problem hierarchy.

Returning to the area of radar, we consider a somewhat different problem. We assume that we know a target is present but do not know its range or velocity. Then the received signal is

$$\begin{aligned} r(t) &= V_r \sin[(\omega_c + \omega_d)(t - \tau) + \theta_r] + n(t), && \tau \le t \le \tau + T, \\ &= n(t), && 0 \le t < \tau, \tau + T < t < \infty, \end{aligned} \tag{17}$$

where ω_d denotes a Doppler shift caused by the target's motion. We want to estimate τ and ω_d. Now, even if the noise were absent and τ and ω_d were known, the signal would still contain the unknown parameters V_r and θ_r. This is a typical second-level estimation problem. As in detection theory, we refer to problems in this category as *signal with unknown parameters in noise problems.*

At the third level the signal component is a random process whose statistical characteristics contain parameters we want to estimate. The received signal is of the form

$$r(t) = s_\Omega(t, A) + n(t), \tag{18}$$

where $s_\Omega(t, A)$ is a sample function from a random process. In a simple case it might be a stationary process with the narrow-band spectrum shown in Fig. 1.6. The shape of the spectrum is known but the center frequency

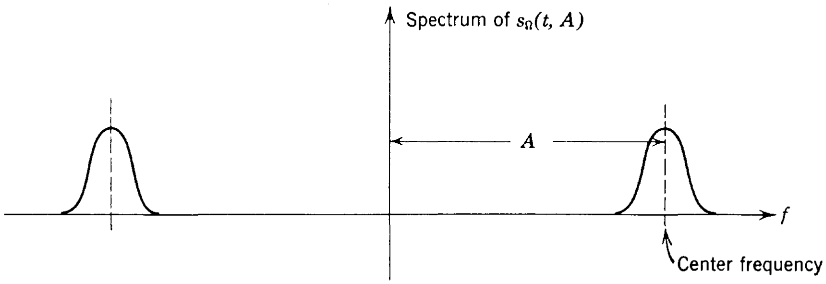

Fig. 1.6 Spectrum of random signal.

	Estimation Theory
Level 1. Known signals in noise	1. PAM, PFM, and PPM communication systems with phase synchronization 2. Inaccuracies in inertial systems (e.g., drift angle measurement)
Level 2. Signals with unknown parameters in noise	1. Range, velocity, or angle measurement in radar/sonar problems 2. Discrete time, continuous amplitude communication system (with unknown amplitude or phase in channel)
Level 3. Random signals in noise	1. Power spectrum parameter estimation 2. Range or Doppler spread target parameters in radar/sonar problem 3. Velocity measurement in radio astronomy 4. Target parameter estimation: passive sonar 5. Ground mapping radars

Fig. 1.7 Estimation theory hierarchy.

is not. The receiver must observe $r(t)$ and, using the statistical properties of $s_\Omega(t, A)$ and $n(t)$, estimate the value of A. This particular example could arise in either radio astronomy or passive sonar. The general class of problem in which the signal containing the parameters is a sample function from a random process is referred to as the *random signal in noise problem.* The hierarchy of estimation theory problems is shown in Fig. 1.7.

We note that there appears to be considerable parallelism in the detection and estimation theory problems. We shall frequently exploit these parallels to reduce the work, but there is a basic difference that should be emphasized. In binary detection the receiver is either "right" or "wrong." In the estimation of a continuous parameter the receiver will seldom be exactly right, but it can try to be close most of the time. This difference will be reflected in the manner in which we judge system performance.

The third area of interest is frequently referred to as modulation theory. We shall see shortly that this term is too narrow for the actual problems. Once again a simple example is useful. In Fig. 1.8 we show an analog message source whose output might typically be music or speech. To convey the message over the channel, we transform it by using a modulation scheme to get it into a form suitable for propagation. The transmitted signal is a continuous waveform that depends on $a(t)$ in some deterministic manner. In Fig. 1.8 it is an amplitude modulated waveform:

$$s[t, a(t)] = [1 + ma(t)] \sin(\omega_c t). \tag{19}$$

(This is conventional double-sideband AM with modulation index m.) In Fig. 1.8*c* the transmitted signal is a frequency modulated (FM) waveform:

$$s[t, a(t)] = \sin\left[\omega_c t + \int_{-\infty}^{t} a(u)\, du\right]. \tag{20}$$

When noise is added the received signal is

$$r(t) = s[t, a(t)] + n(t). \tag{21}$$

Now the receiver must observe $r(t)$ and put out a continuous estimate of the message $a(t)$, as shown in Fig. 1.8. This particular example is a first-level modulation problem, for if $n(t)$ were absent and $a(t)$ were known the received signal would be completely known. Once again we describe it as a *known signal in noise problem.*

Another type of physical situation in which we want to estimate a continuous function is shown in Fig. 1.9. The channel is a time-invariant linear system whose impulse response $h(\tau)$ is unknown. To estimate the impulse response we transmit a known signal $x(t)$. The received signal is

$$r(t) = \int_0^\infty h(\tau)\, x(t - \tau)\, d\tau + n(t). \tag{22}$$

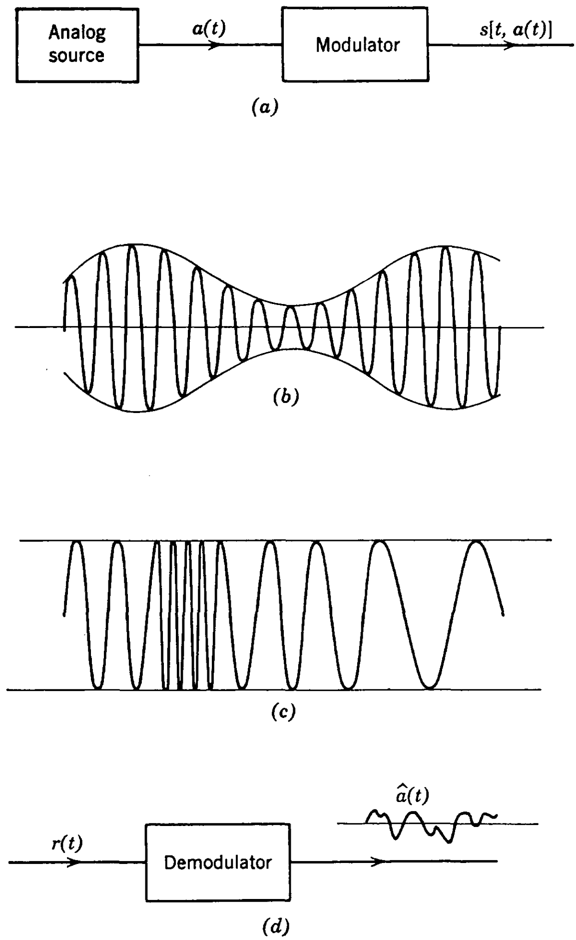

Fig. 1.8 A modulation theory example: (*a*) analog transmission system; (*b*) amplitude modulated signal; (*c*) frequency modulated signal; (*d*) demodulator.

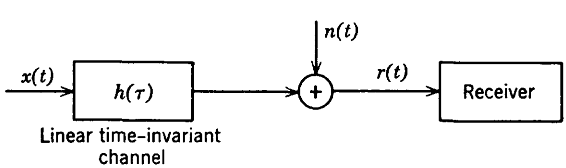

Fig. 1.9 Channel measurement.

The receiver observes $r(t)$ and tries to estimate $h(\tau)$. This particular example could best be described as a continuous estimation problem. Many other problems of interest in which we shall try to estimate a continuous waveform will be encountered. For convenience, we shall use the term *modulation theory* for this category, even though the term continuous waveform estimation might be more descriptive.

The other levels of the modulation theory problem follow by direct analogy. In the amplitude modulation system shown in Fig. 1.8*b* the receiver frequently does not know the phase angle of the carrier. In this case a suitable model is

$$r(t) = (1 + ma(t)) \sin (\omega_c t + \theta) + n(t), \tag{23}$$

	Modulation Theory (Continuous waveform estimation)
1. Known signals in noise	1. Conventional communication systems such as AM (DSB-AM, SSB), FM, and PM with phase synchronization 2. Optimum filter theory 3. Optimum feedback systems 4. Channel measurement 5. Orbital estimation for satellites 6. Signal estimation in seismic and sonar classification systems 7. Synchronization in digital systems
2. Signals with unknown parameters in noise	1. Conventional communication systems without phase synchronization 2. Estimation of channel characteristics when phase of input signal is unknown
3. Random signals in noise	1. Analog communication over randomly varying channels 2. Estimation of statistics of time-varying processes 3. Estimation of plant characteristics

Fig. 1.10 Modulation theory hierarchy.

where θ is an unknown parameter. This is an example of a *signal with unknown parameter problem* in the modulation theory area.

A simple example of a third-level problem (*random signal in noise*) is one in which we transmit a frequency-modulated signal over a radio link whose gain and phase characteristics are time-varying. We shall find that if we transmit the signal in (20) over this channel the received waveform will be

$$r(t) = V(t) \sin \left[\omega_c t + \int_{-\infty}^{t} a(u)\, du + \theta(t)\right] + n(t), \tag{24}$$

where $V(t)$ and $\theta(t)$ are sample functions from random processes. Thus, even if $a(u)$ were known and the noise $n(t)$ were absent, the received signal would still be a random process. An over-all outline of the problems of interest to us appears in Fig. 1.10. Additional examples included in the table to indicate the breadth of the problems that fit into the outline are discussed in more detail in the text.

Now that we have outlined the areas of interest it is appropriate to determine how to go about solving them.

1.2 POSSIBLE APPROACHES

From the examples we have discussed it is obvious that an inherent feature of all the problems is randomness of source, channel, or noise (often all three). Thus our approach must be statistical in nature. Even assuming that we are using a statistical model, there are many different ways to approach the problem. We can divide the possible approaches into two categories, which we denote as "structured" and "nonstructured." Some simple examples will illustrate what we mean by a structured approach.

Example 1. The input to a linear time-invariant system is $r(t)$:

$$\begin{aligned} r(t) &= s(t) + w(t) \qquad 0 \le t \le T, \\ &= 0, \qquad \text{elsewhere.} \end{aligned} \tag{25}$$

The impulse response of the system is $h(\tau)$. The signal $s(t)$ is a known function with energy E_s,

$$E_s = \int_0^T s^2(t)\, dt, \tag{26}$$

and $w(t)$ is a sample function from a zero-mean random process with a covariance function:

$$K_w(t, u) = \frac{N_0}{2}\, \delta(t-u). \tag{27}$$

We are concerned with the output of the system at time T. The output due to the signal is a deterministic quantity:

$$s_o(T) = \int_0^T h(\tau)\, s(T - \tau)\, d\tau. \tag{28}$$

The output due to the noise is a random variable:

$$n_o(T) = \int_0^T h(\tau)\, n(T - \tau)\, d\tau. \tag{29}$$

We can define the output signal-to-noise ratio at time T as

$$\frac{S}{N} \triangleq \frac{s_o^2(T)}{E[n_o^2(T)]}, \tag{30}$$

where $E(\cdot)$ denotes expectation.

Substituting (28) and (29) into (30), we obtain

$$\frac{S}{N} = \frac{\left[\int_0^T h(\tau)\, s(T - \tau)\, d\tau\right]^2}{E\left[\iint_0^T h(\tau)\, h(u)\, n(T - \tau)\, n(T - u)\, d\tau\, du\right]}. \tag{31}$$

By bringing the expectation inside the integral, using (27), and performing the integration with respect to u, we have

$$\frac{S}{N} = \frac{\left[\int_0^T h(\tau)\, s(T - \tau)\, d\tau\right]^2}{N_0/2 \int_0^T h^2(\tau)\, d\tau}. \tag{32}$$

The problem of interest is to choose $h(\tau)$ to maximize the signal-to-noise ratio. The solution follows easily, but it is not important for our present discussion. (See Problem 3.3.1.)

This example illustrates the three essential features of the structured approach to a statistical optimization problem:

Structure. The processor was required to be a linear time-invariant filter. We wanted to choose the best system in this class. Systems that were not in this class (e.g., nonlinear or time-varying) were not allowed.

Criterion. In this case we wanted to maximize a quantity that we called the signal-to-noise ratio.

Information. To write the expression for S/N we had to know the signal shape and the covariance function of the noise process.

If we knew more about the process (e.g., its first-order probability density), we could not use it, and if we knew less, we could not solve the problem. Clearly, if we changed the criterion, the information required might be different. For example, to maximize x

$$x = \frac{s_o^4(T)}{E[n_o^4(T)]}, \tag{33}$$

the covariance function of the noise process would not be adequate. Alternatively, if we changed the structure, the information required might