Detection of Signals
—Estimation of Signal Parameters

4.1 INTRODUCTION

In Chapter 2 we formulated the detection and estimation problems in the classical context. In order to provide background for several areas, we first examined a reasonably general problem. Then, in Section 2.6 of Chapter 2, we investigated the more precise results that were available in the general Gaussian case.

In Chapter 3 we developed techniques for representing continuous processes by sets of numbers. The particular representation that we considered in detail was appropriate primarily for Gaussian processes.

We now want to use these representations to extend the results of the classical theory to the case in which the observations consist of continuous waveforms.

4.1.1 Models

The problems of interest to us in this chapter may be divided into two categories. The first is the detection problem which arises in three broad areas: digital communications, radar/sonar, and pattern recognition and classification. The second is the signal parameter estimation problem which also arises in these three areas.

Detection. The conventional model of a simple digital communication system is shown in Fig. 4.1. The source puts out a binary digit (either 0 or 1) every T seconds. The most straightforward system would transmit either \( \sqrt{E_0} s_0(t) \) or \( \sqrt{E_1} s_1(t) \) during each interval. In a typical space communication system an attenuated version of the transmitted signal would be received with negligible distortion. The received signal consists of \( \sqrt{E_0} s_0(t) \) or \( \sqrt{E_1} s_1(t) \) plus an additive noise component.

The characterization of the noise depends on the particular application. One source, always present, is thermal noise in the receiver front end. This
noise can be modeled as a sample function from a Gaussian random process. As we proceed to more complicated models, we shall encounter other sources of interference that may turn out to be more important than the thermal noise. In many cases we can redesign the system to eliminate these other interference effects almost entirely. Then the thermal noise will be the disturbance that limits the system performance. In most systems the spectrum of the thermal noise is flat over the frequency range of interest, and we may characterize it in terms of a spectral height of $N_0/2$ joules. An alternate characterization commonly used is effective noise temperature $T_e$ (e.g., Valley and Wallman [1] or Davenport and Root [2], Chapter 10). The two are related simply by

$$N_0 = kT_e,$$

where $k$ is Boltzmann’s constant, $1.38 \times 10^{-23}$ joule/°K and $T_e$ is the effective noise temperature, °K.

Thus in this particular case we could categorize the receiver design as a problem of detecting one of two known signals in the presence of additive white Gaussian noise.

If we look into a possible system in more detail, a typical transmitter could be as shown in Fig. 4.2. The transmitter has an oscillator with nominal center frequency of $\omega_0$. It is biphase modulated according to whether the source output is 1 (0°) or 0 (180°). The oscillator’s instantaneous phase varies slowly, and the receiver must include some auxiliary equipment to measure the oscillator phase. If the phase varies slowly enough, we shall see that accurate measurement is possible. If this is true, the problem may be modeled as above. If the measurement is not accurate, however, we must incorporate the phase uncertainty in our model.

A second type of communication system is the point-to-point ionospheric scatter system shown in Fig. 4.3 in which the transmitted signal is scattered by the layers in the ionosphere. In a typical system we can transmit a “one” by sending a sine wave of a given frequency and a “zero” by a sine wave of another frequency. The receiver signal may vary as shown in Fig. 4.4. Now, the receiver has a signal that fluctuates in amplitude and phase.

In the commonly-used frequency range most of the additive noise is Gaussian.

Corresponding problems are present in the radar context. A conventional pulsed radar transmits a signal as shown in Fig. 4.5. If a target is present, the sequence of pulses is reflected. As the target fluctuates, the amplitude and phase of the reflected pulses change. The returned signal consists of a sequence of pulses whose amplitude and phase are unknown. The problem
is to examine this sequence in the presence of receiver noise and decide whether a target is present.

There are obvious similarities between the two areas, but there are also some differences:

1. In a digital communication system the two types of error (say 1, when 0 was sent, and vice versa) are usually of equal importance. Furthermore, a signal may be present on both hypotheses. This gives a symmetry to the problem that can be exploited. In a radar/sonar system the two types of error are almost always of unequal importance. In addition, a signal is present only on one hypothesis. This means that the problem is generally nonsymmetric.

2. In a digital communication system the probability of error is usually an adequate measure of system performance. Normally, in radar/sonar a reasonably complete ROC is needed.

3. In a digital system we are sending a sequence of digits. Thus we can correct digit errors by putting some structure into the sequence. In the radar/sonar case this is not an available alternative.

In spite of these differences, a great many of the basic results will be useful for both areas.

**Estimation.** The second problem of interest is the estimation of signal parameters, which is encountered in both the communications and radar/sonar areas. We discuss a communication problem first.

Consider the analog message source shown in Fig. 4.6a. For simplicity we assume that it is a sample function from a bandlimited random process (2W cps: double-sided). We could then sample it every 1/2W seconds without losing any information. In other words, given these samples at the

---

Fig. 4.4 Signal component in time-varying channel.

Fig. 4.5 Signals in radar model: (a) transmitted sequence of rf pulses; (b) received sequence [amplified (time-shift not shown)].

Fig. 4.6 Analog message transmission.
receiver, we could reconstruct the message exactly (e.g. Nyquist, [4] or Problem 3.3.6). Every $T$ seconds ($T = 1/2W$) we transmit a signal that depends on the particular value $A_i$ at the last sampling time. In the system in Fig. 4.6b the amplitude of a sinusoid depends on the value of $A_i$. This system is referred to as a pulse-amplitude modulation system (PAM). In the system in Fig. 4.6c the frequency of the sinusoid depends on the sample value. This system is referred to as a pulse frequency modulation system (PFM). The signal is transmitted over a channel and is corrupted by noise (Fig. 4.7). The received signal in the $i$th interval is:

$$r(t) = s(t, A_i) + n(t), \quad T_i \leq t \leq T_{i+1}. \quad (2)$$

The purpose of the receiver is to estimate the values of the successive $A_i$ and use these estimates to reconstruct the message.

A typical radar system is shown in Fig. 4.8. In a conventional pulsed radar the transmitted signal is a sinusoid with a rectangular envelope.

$$s_i(t) = \sqrt{2E_i} \sin \omega_c t, \quad 0 \leq t \leq T_i, \quad (3a)$$

$$= 0, \quad \text{elsewhere.}$$

The returned signal is delayed by the round-trip time to the target. If the target is moving, there is Doppler shift. Finally there is a random amplitude and phase due to the target fluctuation. The received signal in the absence of noise is

$$s_i(t) = v \sqrt{2E_i} \sin [(\omega_c + \omega_D)(t - \tau) + \phi], \quad \tau \leq t \leq \tau + T, \quad (3b)$$

$$= 0, \quad \text{elsewhere.}$$

Here we estimate $\tau$ and $\omega_D$ (or, equivalently, target range and velocity). Once again, there are obvious similarities between the communication and radar/sonar problem. The basic differences are the following:

1. In the communications context $A_i$ is a random variable with a probability density that is usually known. In radar the range or velocity limits of interest are known. The parameters, however, are best treated as nonrandom variables (e.g., discussion in Section 2.4).

2. In the radar case the difficulty may be compounded by a lack of knowledge regarding the target's presence. Thus the detection and estimation problem may have to be combined.

3. In almost all radar problems a phase reference is not available.

Other models of interest will appear naturally in the course of our discussion.

### 4.1.2 Format

Because this chapter is long, it is important to understand the over-all structure. The basic approach has three steps:

1. The observation consists of a waveform $r(t)$. Thus the observation space may be infinite dimensional. Our first step is to map the received signal into some convenient decision or estimation space. This will reduce the problem to one studied in Chapter 2.

2. In the detection problem we then select decision regions and compute the ROC or $Pr(\epsilon)$. In the estimation problem we evaluate the variance or mean-square error.

3. We examine the results to see what they imply about system design and performance.

We carry out these steps for a sequence of models of increasing complexity (Fig. 4.9) and develop the detection and estimation problem in parallel. By emphasizing their parallel nature for the simple cases, we can save appreciable effort in the more complex cases by considering only one problem in the text and leaving the other as an exercise. We start with the simple models and then proceed to the more involved.
4.2 Detection and Estimation in White Gaussian Noise

<table>
<thead>
<tr>
<th>Channel</th>
<th>Signal detection</th>
<th>Signal parameter estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Additive white noise</td>
<td>Simple binary</td>
<td>Single parameter, linear</td>
</tr>
<tr>
<td>Additive colored noise</td>
<td>General binary</td>
<td>Single parameter, nonlinear</td>
</tr>
<tr>
<td>Simple random</td>
<td>$M$-ary</td>
<td>Multiple parameter</td>
</tr>
<tr>
<td>Multiple channels</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 4.9 Sequence of models.

A logical question is: if the problem is so simple, why is the chapter so long? This is a result of our efforts to determine how the model and its parameters affect the design and performance of the system. We feel that only by examining some representative problems in detail can we acquire an appreciation for the implications of the theory.

Before proceeding to the solution, a brief historical comment is in order. The mathematical groundwork for our approach to this problem was developed by Grenander [5]. The detection problem relating to optimum radar systems was developed at the M.I.T. Radiation Laboratory, (e.g., Lawson and Uhlenbeck [6]) in the early 1940's. Somewhat later Woodward and Davies [7, 8] approached the radar problem in a different way. The detection problem was formulated at about the same time in a manner similar to ours by both Peterson, Birdsall, and Fox [9] and Middleton and Van Meter [10], whereas the estimation problem was first done by Slepian [11]. Parallel results with a communications emphasis were developed by Kotelnikov [12, 13] in Russia. Books that deal almost exclusively with radar include Helstrom [14] and Wainstein and Zubakov [15]. Books that deal almost exclusively with communication include Kotelnikov [13], Harman [16], Baghdady (ed.) [17], Wozencraft and Jacobs [18], and Golomb et al. [19]. The last two parts of Middleton [47] cover a number of topics in both areas. By presenting the problems side by side we hope to emphasize their inherent similarities and contrast their differences.

4.2 DETECTION AND ESTIMATION IN WHITE GAUSSIAN NOISE

In this section we formulate and solve the detection and estimation problems for the case in which the interference is additive white Gaussian noise.

We consider the detection problem first: the simple binary case, the general binary case, and the $M$-ary case are discussed in that order. By using the concept of a sufficient statistic the optimum receiver structures are simply derived and the performances for a number of important cases are evaluated. Finally, we study the sensitivity of the optimum receiver to the detailed assumptions of our model.

As we have seen in the classical context, the decision and estimation problems are closely related; linear estimation will turn out to be essentially the same as simple binary detection. When we proceed to the nonlinear estimation problem, new issues will develop, both in specifying the estimator structure and in evaluating its performance.

4.2.1 Detection of Signals in Additive White Gaussian Noise

Simple Binary Detection. In the simplest binary decision problem the received signal under one hypothesis consists of a completely known signal, $\sqrt{E} s(t)$, corrupted by an additive zero-mean white Gaussian noise $w(t)$ with spectral height $N_0/2$; the received signal under the other hypothesis consists of the noise $w(t)$ alone. Thus

$$ r(t) = \sqrt{E} s(t) + w(t), \quad 0 \leq t \leq T; H_1, $$

$$ = w(t), \quad 0 \leq t \leq T; H_0. $$

(4)

For convenience we assume that

$$ \int_0^T s^2(t) \, dt = 1, \quad \text{(5)} $$

so that $E$ represents the received signal energy. The problem is to observe $r(t)$ over the interval $[0, T]$ and decide whether $H_0$ or $H_1$ is true. The criterion may be either Bayes or Neyman-Pearson.

The following ideas will enable us to solve this problem easily:

1. Our observation is a time-continuous random waveform. The first step is to reduce it to a set of random variables (possibly a countably infinite set).

2. One method is the series expansion of Chapter 3:

$$ r(t) = \lim_{K \to \infty} \sum_{k=1}^{K} r_k \phi_k(t); \quad 0 \leq t \leq T. $$

(6)

When $K = K'$, there are $K'$ coefficients in the series, $r_1, \ldots, r_{K'}$ which we could denote by the vector $r_K$. In our subsequent discussion we suppress the $K'$ subscript and denote the coefficients by $r$. 
3. In Chapter 2 we saw that if we transformed \( r \) into two independent vectors, \( f \) (the sufficient statistic) and \( y \), as shown in Fig. 4.10, our decision could be based only on \( l \), because the values of \( y \) did not depend on the hypothesis. The advantage of this technique was that it reduced the dimension of the decision space to that of \( l \). Because this is a binary problem we know that \( l \) will be one-dimensional.

Here the method is straightforward. If we choose the first orthonormal function to be \( s(t) \), the first coefficient in the decomposition is the Gaussian random variable,

\[
\begin{align*}
\tau_1 &= \int_0^T s(t) w(t) \, dt \triangleq w_1; H_0, \\
\tau_i &= \int_0^T s(t) [\sqrt{E} s(t) + w(t)] \, dt = \sqrt{E} + w_i; H_1, \\
\end{align*}
\]

The remaining \( \tau_i \) \( (i > 1) \) are Gaussian random variables which can be generated by using some arbitrary orthonormal set whose members are orthogonal to \( s(t) \).

\[
\begin{align*}
\tau_1 &= \int_0^T s(t) w(t) \, dt \triangleq w_1; H_0, \\
\tau_i &= \int_0^T \phi_i(t) [\sqrt{E} s(t) + w(t)] \, dt = w_i; H_1, \quad i \neq 1.
\end{align*}
\]

From Chapter 3 (44) we know that

\[
E(w_i w_j) = 0; \quad i \neq j.
\]

Because \( w_i \) and \( w_j \) are jointly Gaussian, they are statistically independent (see Property 3 on p. 184).

We see that only \( \tau_1 \) depends on which hypothesis is true. Further, all \( \tau_i \) \( (i > 1) \) are statistically independent of \( \tau_1 \). Thus \( \tau_1 \) is a sufficient statistic \( (\tau_1 = l) \). The other \( \tau_i \) correspond to \( y \). Because they will not affect the decision, there is no need to compute them.

Several equivalent receiver structures follow immediately. The structure in Fig. 4.11 is called a correlation receiver. It correlates the input \( r(t) \) with a stored replica of the signal \( s(t) \). The output is \( r_1 \), which is a sufficient statistic \( (r_1 = l) \) and is a Gaussian random variable. Once we have obtained \( r_1 \), the decision problem will be identical to the classical problem in Chapter 2 (specifically, Example 1 on pp. 27–28). We compare \( l \) to a threshold in order to make a decision.

An equivalent realization is shown in Fig. 4.12. The impulse response of the linear system is simply the signal reversed in time and shifted,

\[
h(\tau) = s(T - \tau).
\]

The output at time \( T \) is the desired statistic \( l \). This receiver is called a matched filter receiver. (It was first derived by North [20].) The two structures are mathematically identical; the choice of which structure to use depends solely on ease of realization.

Just as in Example 1 of Chapter 2, the sufficient statistic \( l \) is Gaussian under either hypothesis. Its mean and variance follow easily:

\[
\begin{align*}
E(l|H_1) &= E(r_1|H_1) = \sqrt{E}, \\
E(l|H_0) &= E(r_1|H_0) = 0, \\
\end{align*}
\]

\[
\begin{align*}
\text{Var}(l|H_0) &= \text{Var}(l|H_1) = \frac{N_0}{2}. \\
\end{align*}
\]

Thus we can use the results of Chapter 2, (64)–(68), with

\[
d = \left( \frac{2E}{N_0} \right)^{1/2}.
\]

\[
\begin{align*}
\text{Fig. 4.10} & \quad \text{Generation of sufficient statistics.} \\
\end{align*}
\]

\[
\begin{align*}
\text{Fig. 4.11} & \quad \text{Correlation receiver.} \\
\end{align*}
\]

\[
\begin{align*}
\text{Fig. 4.12} & \quad \text{Matched filter receiver.} \\
\end{align*}
\]
The curves in Figs. 2.9a and 2.9b of Chapter 2 are directly applicable and are reproduced as Figs. 4.13 and 4.14. We see that the performance depends only on the received signal energy $E$ and the noise spectral height $N_0$—the signal shape is not important. This is intuitively logical because the noise is the same along any coordinate.

The key to the simplicity in the solution was our ability to reduce an infinite dimensional observation space to a one-dimensional decision space by exploiting the idea of a sufficient statistic. Clearly, we should end up with the same receiver even if we do not recognize that a sufficient statistic is available. To demonstrate this we construct the likelihood ratio directly. Three observations lead us easily to the solution.

1. If we approximate $r(t)$ in terms of some finite set of numbers, $r_1, \ldots, r_K$, we have a problem in classical detection theory that we can solve.

2. If we choose the set $r_1, r_2, \ldots, r_K$ so that

$$P_{R_1, R_2, \ldots, R_K|H_i}(R_1, R_2, \ldots, R_K|H_i) = \prod_{i=1}^{K} p_{r_i|H_i}(R_i|H_i), \quad i = 0, 1,$$

that is, the observations are conditionally independent, we have an easy problem to solve.
3. Because we know that it requires an infinite set of numbers to represent \( r(t) \) completely, we want to get the solution in a convenient form so that we can let \( K \to \infty \).

We denote the approximation that uses \( K \) coefficients as \( r_K(t) \). Thus
\[
    r_K(t) = \sum_{i=1}^{K} r_i \phi_i(t), \quad 0 \leq t \leq T, \tag{13}
\]
where
\[
    r_i = \int_{0}^{T} r(t) \phi_i(t) \, dt, \quad i = 1, 2, \ldots, K, \tag{14}
\]
and the \( \phi_i(t) \) belong to an arbitrary complete orthonormal set of functions. Using (14), we see that under \( H_0 \)
\[
    r_i = \int_{0}^{T} w(t) \phi_i(t) \, dt = w_i, \tag{15}
\]
and under \( H_1 \)
\[
    r_i = \int_{0}^{T} \sqrt{E} s(t) \phi_i(t) \, dt + \int_{0}^{T} w(t) \phi_i(t) \, dt = s_i + w_i. \tag{16}
\]
The coefficients \( s_i \) correspond to an expansion of the signal
\[
    s_K(t) = \sum_{i=1}^{K} s_i \phi_i(t), \quad 0 \leq t \leq T, \tag{17}
\]
and
\[
    \sqrt{E} s(t) = \lim_{K \to \infty} s_K(t). \tag{18}
\]
The \( r_i \)'s are Gaussian with known statistics:
\[
    E(r_i|H_0) = 0, \\
    E(r_i|H_1) = s_i, \\
    \text{Var}(r_i|H_0) = \text{Var}(r_i|H_1) = \frac{N_0}{2}. \tag{19}
\]
Because the noise is "white," these coefficients are independent along any set of coordinates. The likelihood ratio is
\[
    \Lambda[r_K(t)] = \frac{p_{r|H_1}(R|H_1)}{p_{r|H_0}(R|H_0)} = \prod_{i=1}^{K} \frac{1}{\sqrt{2\pi N_0}} \exp \left( -\frac{1}{2} \left( \frac{R_i - s_i}{\sqrt{N_0}} \right)^2 \right). \tag{20}
\]
Taking the logarithm and canceling common terms, we have
\[
    \ln \Lambda[r_K(t)] = \frac{2}{N_0} \sum_{i=1}^{K} R_i s_i - \frac{1}{N_0} \sum_{i=1}^{K} s_i^2. \tag{21}
\]
The two sums are easily expressed as integrals. From Parseval's theorem,
\[
    \sum_{i=1}^{K} R_i s_i = \int_{0}^{T} r_K(t) s_K(t) \, dt
\]
and
\[
    \sum_{i=1}^{K} s_i^2 = \int_{0}^{T} s_K^2(t) \, dt, \tag{22}
\]
We now have the log likelihood ratio in a form in which it is convenient to pass to the limit:
\[
    \text{I.i.m. } \ln \Lambda[r_K(t)] \propto \ln \Lambda[r(t)] = t \int_{0}^{T} r(t) s(t) \, dt - E \frac{2}{N_0}. \tag{23}
\]
The first term is just the sufficient statistic we obtained before. The second term is a bias. The resulting likelihood ratio test is
\[
    \sqrt{E} \int_{0}^{T} r(t) s(t) \, dt \leq \ln \eta + \frac{E}{N_0}. \tag{24}
\]
(Recall from Chapter 2 that \( \eta \) is a constant which depends on the costs and a priori probabilities in a Bayes test and the desired \( P_F \) in a Neyman-Pearson test.) It is important to observe that even though the probability density \( p_{r|H_0}(r(t)|H_0) \) is not well defined for either hypothesis, the likelihood ratio is.

Before going on to more general problems it is important to emphasize the two separate features of the signal detection problem:

1. First we reduce the received waveform to a single number which is a point in a decision space. This operation is performed physically by a correlation operation and is invariant to the decision criterion that we plan to use. This invariance is important because it enables us to construct the waveform processor without committing ourselves to a particular criterion.

2. Once we have transformed the received waveform into the decision space we have only the essential features of the problem left to consider. Once we get to the decision space the problem is the same as that studied in Chapter 2. The actual received waveform is no longer important and all physical situations that lead to the same picture in a decision space are identical for our purposes. In our simple example we saw that all signals of equal energy map into the same point in the decision space. It is therefore obvious that the signal shape is unimportant.

The separation of these two parts of the problem leads to a clearer understanding of the fundamental issues.
General Binary Detection in White Gaussian Noise. The results for the simply binary problem extend easily to the general binary problem. Let
\[ r(t) = \sqrt{E_1} s_1(t) + w(t), \quad 0 \leq t \leq T; H_1, \]
\[ = \sqrt{E_0} s_0(t) + w(t), \quad 0 \leq t \leq T; H_0, \quad (25) \]
where \( s_0(t) \) and \( s_1(t) \) are normalized but are not necessarily orthogonal. We denote the correlation between the two signals as
\[ \rho \triangleq \int_0^T s_0(t) s_1(t) \, dt. \]
(Note that |\( \rho \)| \( \leq 1 \) because the signals are normalized.)

We choose our first two orthogonal functions as follows:
\[ \phi_1(t) = s_1(t), \quad 0 \leq t \leq T, \quad (26) \]
\[ \phi_2(t) = \frac{1}{\sqrt{1 - \rho^2}} [s_0(t) - \rho s_1(t)], \quad 0 \leq t \leq T. \quad (27) \]

We see that \( \phi_2(t) \) is obtained by subtracting out the component of \( s_0(t) \) that is correlated with \( \phi_1(t) \) and normalizing the result. The remaining \( \phi_i(t) \) consist of an arbitrary orthonormal set whose members are orthogonal to \( \phi_1(t) \) and \( \phi_2(t) \) and are chosen so that the entire set is complete. The coefficients are
\[ r_i = \int_0^T r(t) \phi_i(t) \, dt; \quad i = 1, 2, \ldots . \quad (28) \]
All of the \( r_i \) except \( r_1 \) and \( r_2 \) do not depend on which hypothesis is true and are statistically independent of \( r_1 \) and \( r_2 \). Thus a two-dimensional decision region, shown in Fig. 4.15a, is adequate. The mean value of \( r_i \) along each coordinate is
\[ E[r_i | H_0] = \sqrt{E_0} \int_0^T s_0(t) \phi_i(t) \, dt, \triangleq s_{0i}, \quad i = 1, 2, \ldots, H_0, \quad (29) \]
\[ E[r_i | H_1] = \sqrt{E_1} \int_0^T s_1(t) \phi_i(t) \, dt, \triangleq s_{1i}, \quad i = 1, 2, \ldots, H_1. \quad (30) \]
The likelihood ratio test follows directly from Section 2.6 (2.327)
\[ \ln \Lambda = -\frac{1}{N_0} \sum_{i=1}^2 (R_i - s_{1i})^2 + \frac{1}{N_0} \sum_{i=1}^2 (R_i - s_{0i})^2 \frac{H_1}{H_0} \ln \eta, \quad (31a) \]
\[ \ln \Lambda = -\frac{1}{N_0} |R - s_1|^2 + \frac{1}{N_0} |R - s_0|^2 \frac{H_1}{H_0} \ln \eta, \quad (31b) \]
or, canceling common terms and rearranging the result,
\[ R^T(s_1 - s_0) \frac{H_1}{H_0} \frac{N_0}{2} \ln \eta + \frac{1}{2} (|s_1|^2 - |s_0|^2). \quad (31c) \]
Thus only the product of \( R^T \) with the difference vector \( s_1 - s_0 \) is used to make a decision. Therefore the decision space is divided into two parts by a line perpendicular to \( s_1 - s_0 \) as shown in Fig. 4.15b. The noise components along the \( r_1 \) and \( r_2 \) axes are independent and identically distributed.

Now observe that we can transform the coordinates as shown in Fig. 4.15b. The noises along the new coordinates are still independent, but only the coefficient along the \( f \) coordinate depends on the hypothesis and the \( y \) coefficient may be disregarded. Therefore we can simplify our receiver by
generating \( I \) instead of \( r_1 \) and \( r_2 \). The function needed to generate the statistic is just the normalized version of the difference signal. Denote the difference signal by \( s_d(t) \):
\[
s_d(t) = \sqrt{E_1} s_1(t) - \sqrt{E_0} s_0(t). \tag{32}
\]
The normalized function is
\[
f_d(t) = \frac{\sqrt{E_1} s_1(t) - \sqrt{E_0} s_0(t)}{(E_1 - 2\rho \sqrt{E_0 E_1} - E_0)^{1/2}}. \tag{33}
\]
The receiver is shown in Fig. 4.16. (Note that this result could have been obtained directly by choosing \( f_d(t) \) as the first orthonormal function.)

Thus once again the binary problem reduces to a one-dimensional decision space. The statistic \( I \) is Gaussian:
\[
E(I | H_1) = \frac{E_1 - \sqrt{E_0 E_1} \rho}{(E_1 - 2\rho \sqrt{E_0 E_1} + E_0)^{1/2}} \tag{34}
\]
\[
E(I | H_0) = \frac{\sqrt{E_0 E_1} \rho - E_0}{(E_1 - 2\rho \sqrt{E_0 E_1} + E_0)^{1/2}}. \tag{35}
\]
The variance is \( N_0/2 \) as before. Thus
\[
d^2 = \frac{2}{N_0} (E_1 + E_0 - 2\rho \sqrt{E_0 E_1}). \tag{36}
\]

Observe that if we normalized our coordinate system so that noise variance was unity then \( d \) would be the distance between the two signals. The resulting probabilities are
\[
P_F = \text{erfc} \left( \frac{\ln \eta}{d} + \frac{d}{2} \right), \tag{37}
\]
\[
P_D = \text{erfc} \left( \frac{\ln \eta}{d} - \frac{d}{2} \right). \tag{38}
\]

[These equations are just (2.67) and (2.68).]

The best choice of signals follows easily. The performance index \( d \) is monotonically related to the distance between the two signals in the decision space. For fixed energies the best performance is obtained by making \( \rho = -1 \). In other words,
\[
s_0(t) = -s_1(t). \tag{39}
\]

Once again the signal shape is not important.

When the criterion is minimum probability of error (as would be the logical choice in a binary communication system) and the a priori probabilities of the two hypotheses are equal, the decision region boundary has a simple interpretation. It is the perpendicular bisector of the line connecting the signal points (Fig. 4.17). Thus the receiver under these circumstances can be interpreted as a minimum-distance receiver and the error probability is
\[
\Pr (\varepsilon) = \int_{d/2}^{\infty} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{x^2}{2} \right) dx = \text{erfc} \left( \frac{d}{2} \right). \tag{40}
\]

If, in addition, the signals have equal energy, the bisector goes through the origin and we are simply choosing the signal that is most correlated with \( r(t) \). This can be referred to as a "largest-of" receiver (Fig. 4.18).

The discussion can be extended to the \( M \)-ary problem in a straightforward manner.

**M-ary Detection in White Gaussian Noise.** Assume that there are \( M \)-hypotheses:
\[
r(t) = \sqrt{E} s_i(t) + w(t); \quad 0 \leq t \leq T; H_i. \tag{41}
\]
The $s_i(t)$ all have unit energy but may be correlated:

$$\int_0^T s_i(t) s_j(t) \, dt = \rho_{ij}, \quad i, j = 1, 2, \ldots, M. \tag{42}$$

This problem is analogous to the $M$-hypothesis problem in Chapter 2. We saw that the main difficulty for a likelihood ratio test with arbitrary costs was the specification of the boundaries of the decision regions. We shall devote our efforts to finding a suitable set of sufficient statistics and evaluating the minimum probability of error for some interesting cases.

First we construct a suitable coordinate system to find a decision space with the minimum possible dimensionality. The procedure is a simple extension of the method used for two dimensions. The first coordinate function is just the first signal. The second coordinate function is that component of the second signal which is linearly independent of the first and so on. We let

$$\phi_1(t) = s_1(t), \quad \phi_2(t) = (1 - \rho_{12}^2)^{-\frac{1}{2}}[s_2(t) - \rho_{12} s_1(t)]. \tag{43a}$$

To construct the third coordinate function we write

$$\phi_3(t) = c_3 [s_3 (t) - c_1 \phi_1 (t) - c_2 \phi_2 (t)], \tag{43b}$$

and find $c_1$ and $c_2$ by requiring orthogonality and $c_3$ by requiring $\phi_3(t)$ to be normalized. (This is called the Gram-Schmidt procedure and is developed in detail in Problem 4.2.7.) We proceed until one of two things happens:

1. $M$ orthonormal functions are obtained.
2. $N \lt M$ orthonormal functions are obtained and the remaining signals can be represented by linear combinations of these orthonormal functions. Thus the decision space will consist of at most $M$ dimensions and fewer if the signals are linearly dependent.†

† Observe that we are talking about algebraic dependence.

We then use this set of orthonormal functions to generate $N$ coefficients $(N \leq M)$

$$r_i \triangleq \int_0^T r(t) \phi_i (t) \, dt, \quad i = 1, 2, \ldots, N. \tag{44a}$$

These are statistically independent Gaussian random variables with variance $N_0/2$ whose means depend on which hypothesis is true,

$$E[r_i | H_i] \triangleq m_{ij}, \quad i = 1, \ldots, N, \quad j = 1, \ldots, M. \tag{44b}$$

The likelihood ratio test follows directly from our results in Chapter 2 (Problem No. 2.6.1). When the criterion is minimum $P_e$, we compute

$$l_j = \ln P_j - \frac{1}{N_0} \sum_{i=1}^N (R_i - m_{ij})^2, \quad j = 1, \ldots, M, \tag{45}$$

and choose the largest. (The modification for other cost assignments is given in Problem No. 2.3.2.)

Two examples illustrate these ideas.

**Example 1.** Let

$$s_i(t) = \left(\frac{2}{T}\right)^{\frac{1}{2}} \sin \left(\omega_i t + (i - 1) \frac{T}{2}\right), \quad 0 \leq t \leq T, \quad i = 1, 2, 3, 4, \tag{46}$$

and

$$w = \frac{2\pi n}{T}$$

($n$ is an arbitrary integer). We see that

$$\phi_1(t) = \left(\frac{2}{T}\right)^{\frac{1}{2}} \sin \omega_i t, \quad 0 \leq t \leq T, \tag{47}$$

and

$$\phi_2(t) = \left(\frac{2}{T}\right)^{\frac{1}{2}} \cos \omega_i t, \quad 0 \leq t \leq T.$$  

We see $s_3(t)$ and $s_4(t)$ are $-\phi_1(t)$ and $-\phi_2(t)$ respectively. Thus, in this case, $M = 4$ and $N = 2$. The decision space is shown in Fig. 4.19a. The decision regions follow easily when the criterion is minimum probability of error and the a priori probabilities are equal. Using the result in (45), we obtain the decision regions in Fig. 4.19b.

**Example 2.** Let

$$s_1(t) = \left(\frac{2}{T}\right)^{\frac{1}{2}} \sin \frac{2\pi n}{T} t, \quad 0 \leq t \leq T,$$

$$s_2(t) = \left(\frac{2}{T}\right)^{\frac{1}{2}} \sin \frac{4\pi n}{T} t, \quad 0 \leq t \leq T,$$

$$s_3(t) = \left(\frac{2}{T}\right)^{\frac{1}{2}} \sin \frac{6\pi n}{T} t, \quad 0 \leq t \leq T,$$

($n$ is an arbitrary integer) and

$$E_i = E, \quad i = 1, 2, 3.$$  

Now

$$\phi_i(t) = s_i(t). \tag{49}$$
Detection of Signals in Additive White Gaussian Noise: M-ary Detectors

Example 1 (continued). We assume that the hypotheses are equally likely. Now the problem is symmetrical. Thus it is sufficient to assume that $s_i(t)$ was transmitted and compute the resulting $Pr(\cdot)$. (Clearly, $Pr(\cdot) = Pr(\cdot|H_i)$, $i = 1, \ldots, 4$.) We also observe that the answer would be invariant to a $45^\circ$ rotation of the signal set because the noise is circularly symmetric.

Thus the problem of interest reduces to the simple diagram shown in Fig. 4.21. The $Pr(\cdot)$ is simply the probability that $r$ lies outside the first quadrant when $H_1$ is true.

Now $r_1$ and $r_2$ are independent Gaussian variables with identical means and variances:

$$E(r_1|H_i) = E(r_2|H_i) = \frac{E}{2}$$

and

$$\text{Var}(r_1|H_i) = \text{Var}(r_2|H_i) = \frac{N_o}{2}$$

The $Pr(\cdot)$ can be obtained by integrating $p_{r_1,r_2|h_i}(R_1, R_2|H_i)$ over the area outside the first quadrant. Equivalently, $Pr(\cdot)$ is the integral over the first quadrant subtracted from unity.

$$Pr(\cdot) = 1 - \left[ \frac{1}{2\pi N_0} \right] \exp \left[ -\frac{1}{2N_0} \left( \frac{R_1^2 + R_2^2}{2} - \frac{E^2}{2N_0} \right) \right]$$

Changing variables, we have

$$Pr(\cdot) = 1 - \left[ \frac{1}{2\pi N_0} \right] \exp \left[ -\frac{1}{2} \frac{E^2}{N_0} \right] = 1 - \left( \frac{E^2}{N_0} \right)^{1/2}$$

which is the desired result.

Another example of interest is a generalization of Example 2.

Example 3. Let us assume that

$$r(t) = \sqrt{E} s_i(t) + w(t), \quad 0 \leq t \leq T, \quad H_i, \quad i = 1, 2, \ldots, M$$

and

$$p_{r_1} = 0$$

and the hypotheses are equally likely. Because the energies are equal, it is convenient to
implement the LRT as a "greatest of" receiver as shown in Fig. 4.22. Once again the problem is symmetric, so we may assume $H_i$ is true. Then an error occurs if any $l_j > l_i : j \neq 1$, where

$$l_i \triangleq \int_0^\tau r(t) s_i(t) \, dt, \quad j = 1, 2, \ldots, M.$$ 

Thus

$$\Pr (\varepsilon) = \Pr (\varepsilon | H_i) = 1 - \Pr (\text{all } l_j < l_i : j \neq 1) \quad (55)$$

or, noting that the $l_i (j \neq 1)$ have the same density on $H_i$,

$$\Pr (\varepsilon) = 1 - \int_{-\infty}^\infty P_{\varepsilon | H_i} (L_i | H_i) \left( \int_{-\infty}^\infty P_{l_j | H_i} (L_j | H_i) \, dL_j \right)^{M-1} \, dL_i. \quad (56)$$

In this particular case the densities are

$$P_{\varepsilon | H_i} (L_i | H_i) = \frac{1}{\sqrt{\pi} N_0} \exp \left( -\frac{1}{2} \frac{(L_i - \sqrt{E})^2}{N_0/2} \right) \quad (57)$$

and

$$P_{l_j | H_i} (L_j | H_i) = \frac{1}{\sqrt{2\pi} N_0} \exp \left( -\frac{1}{2} \frac{L_j^2}{N_0/2} \right), \quad j \neq 1. \quad (58)$$

Substituting these densities into (56) and normalizing the variables, we obtain

$$\Pr (\varepsilon) = 1 - \int_{-\infty}^\infty dx \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \frac{(x - (2E/N_0)^{1/2})^2}{2} \right) \left( \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{y^2}{2} \right] \, dy \right)^{M-1}. \quad (59)$$

Unfortunately, we cannot integrate this analytically. Numerical results for certain

![Fig. 4.22](image)

``Largest of'' receiver.

values of $M$ and $E/N_0$ are tabulated in [21] and shown in Fig. 4.23. For some of our purposes an approximate analytic expression is more interesting. We derive a very simple bound. Some other useful bounds are derived in the problems. Looking at (55), we see that we could rewrite the $\Pr (\varepsilon)$ as

$$\Pr (\varepsilon) = \Pr (\text{any } l_j > l_i : j \neq 1), \quad (60)$$

$$\Pr (\varepsilon) = \Pr (l_j > l_i \text{ or } l_k > l_i \text{ or } \cdots \text{ or } l_M > l_i). \quad (61)$$
Now, several $l_i$ can be greater than $l_i$. (The events are not mutually exclusive.) Thus

$$\Pr (\varepsilon) = \Pr (l_2 > l_1) + \Pr (l_3 > l_1) + \cdots + \Pr (l_m > l_1),$$

(62)

$$\Pr (\varepsilon) \leq (M - 1) \left( \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{x^2}{2} \right) dy \right)^{M - 1}$$

but the term in the bracket is just the expression of the probability of error for two orthogonal signals. Using (36) with $\rho = 0$ and $E_1 = E_2 = E$ in (40), we have

$$\Pr (\varepsilon) \leq (M - 1) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{x^2}{2} \right) dx$$

(Equation 64 also follows directly from (63) by a change of variables.) We can further simplify this equation by using (2.71):

$$\Pr (\varepsilon) \leq \frac{(M - 1)}{\sqrt{2\pi} \cdot \sqrt{E/\eta_0}} \exp \left( \frac{E}{2\eta_0} \right)$$

(65)

We observe that the upper bound increases linearly with $M$. The bound on the Pr ($\varepsilon$) given by this expression is plotted in Fig. 4.23 for $M = 16$.

A related problem in which $M$ orthogonal signals arise is that of transmitting a sequence of binary digits.

Example 4. Sequence of Digits. Consider the simple digital system shown in Fig. 4.24, in which the source puts out a binary digit every $T$ seconds. The outputs 0 and 1 are equally likely. The available transmitter power is $P$. For simplicity we assume that we are using orthogonal signals. The following choices are available:

1. Transmit one of two orthogonal signals every $T$ seconds. The energy per signal is $PT$.
2. Transmit one of four orthogonal signals every $2T$ seconds. The energy per signal is $2PT$. For example, the encoder could use the mapping,

$$00 \rightarrow s_1(t),$$
$$01 \rightarrow s_2(t),$$
$$10 \rightarrow s_3(t),$$
$$11 \rightarrow s_4(t).$$

$$\int_0^{2T} s_i(t) n(t) dt = \delta_{ij}$$

Fig. 4.24 Digital communication system.

3. In general, we could transmit one of $M$ orthogonal signals every $T \log_2 M$ seconds. The energy per signal is $TP \log_2 M$. To compute the probability of error we use (59):

$$\Pr (\varepsilon) = 1 - \int_{-\infty}^{\infty} dx \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{x^2}{2} \right) \int_{-\infty}^{\infty} dy \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{y^2}{2} \right)

\times \left[ \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{y^2}{2} \right) dy \right]^{M - 1}.$$  

(66)

The results have been calculated numerically (19) and are plotted in Fig. 4.25. The behavior is quite interesting. Above a certain value of $PT/\eta_0$ the error probability decreases with increased $M$. Below this value the converse is true. It is instructive to investigate the behavior as $M \rightarrow \infty$. We obtain from (66), by a simple change of variables,

$$\lim_{M \rightarrow \infty} (1 - \Pr(\varepsilon)) = \int_{-\infty}^{\infty} dy \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \lim_{M \rightarrow \infty} \left[ \text{erf} \left( \frac{y + \frac{2P \log_2 M}{\eta_0}}{\sqrt{N_0}} \right) \right]^{M - 1}.$$  

(67)

Now consider the limit of the logarithm of the expression in the brace:

$$\lim_{M \rightarrow \infty} \frac{\text{erf} \left[ y + \frac{2P \log_2 M}{\eta_0} \right]}{(M - 1)^{1/2}}.$$  

(68)

Evaluating this limit by treating $M$ as a continuous variable and using L'Hôpital's rule, we find that (see Problem 4.2.15)

$$\lim_{M \rightarrow \infty} \text{erf} \left[ y + \frac{2P \log_2 M}{\eta_0} \right] = \begin{cases} -\infty, & PT \eta_0 < \ln 2, \\ 0, & PT \eta_0 > \ln 2. \end{cases}$$  

(69)

Thus, from the continuity of logarithm,

$$\lim_{M \rightarrow \infty} \Pr (\varepsilon) = \begin{cases} 0, & PT \eta_0 > \ln 2, \\ 1, & PT \eta_0 < \ln 2. \end{cases}$$  

(70)

Thus we see that there is a definite threshold effect. The value of $T$ is determined by how fast the source produces digits. Specifically, the rate in binary digits per second is

$$R = \frac{1}{2} \text{binary digits/sec}.$$  

(71)

Using orthogonal signals, we see that if

$$R < \frac{P}{\ln 2 \eta_0}$$  

(72)

the probability of error will go to zero. The obvious disadvantage is the bandwidth requirement. As $M \rightarrow \infty$, the transmitted bandwidth goes to $\infty$.

The result in (72) was derived for a particular set of signals. Shannon has shown (e.g., [22] or [23]) that the right-hand side is the bound on the
rate for error-free transmission for any communication scheme. This rate is referred to as the capacity of an infinite bandwidth, additive white Gaussian noise channel.

\[ C_w = \frac{1}{\ln 2} \frac{P}{N_0} \text{ bits/sec.} \] (73)

Shannon has also derived an expression for a bandlimited channel \((W_{ch}: \text{single-sided})\):

\[ C = W_{ch} \log_2 \left(1 + \frac{P}{W_{ch}N_0}\right) \] (74)

These capacity expressions are fundamental to the problem of sending sequences of digits. We shall not consider this problem, for an adequate discussion would take us too far afield. Suitable references are [18] and [66].

In this section we have derived the canonic receiver structures for the \(M\)-ary hypothesis problem in which the received signal under each hypothesis is a known signal plus additive white Gaussian noise. The simplicity resulted because we were always able to reduce an infinite dimensional observation space to a finite (\(\leq M\)) dimensional decision space.

In the problems we consider some of the implications of these results. Specific results derived in the problems include the following:

1. The probability of error for any set of \(M\) equally correlated signals can be expressed in terms of an equivalent set of \(M\) orthogonal signals (Problem 4.2.9).

2. The lowest value of uniform correlation is \((- (M - 1))^{-1}\) Signals with this property are optimum when there is no bandwidth restriction (Problems 4.2.9-4.2.12). They are referred to as Simplex signals.

3. For large \(M\), orthogonal signals are essentially optimum.

**Sensitivity.** Before leaving the problem of detection in the presence of white noise we shall discuss an important issue that is frequently overlooked. We have been studying the mathematical model of a physical system and have assumed that we know the quantities of interest such as \(s(t), E, \) and \(N_0\) exactly. In an actual system these quantities will vary from their nominal values. It is important to determine how the performance of the optimum receiver will vary when the nominal values are perturbed. If the performance is highly sensitive to small perturbations, the validity of the nominal performance calculation is questionable. We shall discuss sensitivity in the context of the simple binary detection problem.

The model for this problem is

\[ r(t) = \sqrt{E} s(t) + w(t), \quad 0 \leq t \leq T: H_1 \]
\[ r(t) = w(t), \quad 0 \leq t \leq T: H_0. \] (75)

**Fig. 4.25** Probability of decision error: \(M\) orthogonal signals, power constraint.
The receiver consists of a matched filter followed by a decision device. The impulse response of the matched filter depends on the shape of \( s(t) \). The energy and noise levels affect the decision level in the general Bayes case. (In the Neyman–Pearson case only the noise level affects the threshold setting.) There are several possible sensitivity analyses. Two of these are the following:

1. Assume that the actual signal energy and signal shape are identical to those in the model. Calculate the change in \( P_D \) and \( P_F \) due to a change in the white noise level.

2. Assume that the signal energy and the noise level are identical to those in the model. Calculate the change in \( P_D \) and \( P_F \) due to a change in the signal.

In both cases we can approach the problem by first finding the change in \( d \) due to the changes in the model and then seeing how \( P_D \) and \( P_F \) are affected by a change in \( d \). In this section we shall investigate the effect of an inaccurate knowledge of signal shape on the value of \( d \). The other questions mentioned above are left as an exercise. We assume that we have designed a filter that is matched to the assumed signal \( s(t) \),

\[
h(T - t) = s(t), \quad 0 \leq t \leq T, \tag{76}
\]

and that the received waveform on \( H_1 \) is

\[
r(t) = s_d(t) + w(t), \quad 0 \leq t \leq T, \tag{77}
\]

where \( s_d(t) \) is the actual signal received. There are two general methods of relating \( s_d(t) \) to \( s(t) \). We call the first the function-variation method.

**Function-Variation Method.** Let

\[
s_d(t) = \sqrt{E} s(t) + \sqrt{E_s} s_e(t), \quad 0 \leq t \leq T, \tag{78}
\]

where \( s_e(t) \) is a normalized waveform representing the inaccuracy. The energy in the error signal is constrained to equal \( E_s \).

The effect can be most easily studied by examining the decision space (more precisely an augmented decision space). To include all of \( s_d(t) \) in the decision space we think of adding another matched filter,

\[
\tilde{h}_d(T - t) = \tilde{s}_d(t) = \frac{s_d(t) - \rho_s s(t)}{\sqrt{1 - \rho_s^2}}, \quad 0 \leq t \leq T, \tag{79}
\]

where \( \rho_s \) is the correlation between \( s_d(t) \) and \( s(t) \). (Observe that we do not do this physically.) We now have a two-dimensional space. The effect of the constraint is clear. Any \( s_d(t) \) will lead to a point on the circle surrounding \( s \), as shown in Fig. 4.26. Observe that the decision still uses only the coordinate along \( s(t) \). The effect is obvious. The error signal that causes the largest performance degradation is

\[
s_e(t) = -s(t). \tag{80}
\]

Then

\[
d^2 = \frac{2}{N_0} (\sqrt{E} - \sqrt{E_s})^2. \tag{81}
\]

To state the result another way,

\[
\frac{\Delta d}{d} = \frac{\sqrt{2E_s/N_0}}{\sqrt{2E/N_0}} = \left( \frac{E_s}{E} \right)^{1/2} \tag{82}
\]

where

\[
\Delta d \doteq d_a - d. \tag{83}
\]

We see that small energy in the error signal implies a small change in performance. Thus the test is insensitive to small perturbations. The second method is called the parameter-variation method.

**Parameter-Variation Method.** This method can best be explained by an example. Let

\[
s(t) = \left( \frac{2}{T} \right)^{1/2} \sin \omega_d t, \quad 0 \leq t \leq T, \tag{84}
\]

be the nominal signal. The actual signal is

\[
s_a(t) = \left( \frac{2}{T} \right)^{1/2} \sin (\omega_d t + \theta), \quad 0 \leq t \leq T, \tag{85}
\]

which, for \( \theta = 0 \), corresponds to the nominal signal. The augmented
detailed assumption. In the sequel we shall encounter several examples of this sensitivity.

We now turn to the problem of linear estimation.

### 4.2.2 Linear Estimation

In Section 4.1.1 we formulated the problem of estimating signal parameters in the presence of additive noise. For the case of additive white noise the received waveform is

\[ r(t) = s(t, A) + w(t), \quad 0 \leq t \leq T, \quad (88a) \]

where \( w(t) \) is a sample function from a white Gaussian noise process with spectral height \( N_0/2 \). The parameter \( A \) is the quantity we wish to estimate. If it is a random parameter we will assume that the a priori density is known and use a Bayes estimation procedure. If it is a nonrandom variable we will use ML estimates. The function \( s(t, A) \) is a deterministic mapping of \( A \) into a time function. If \( s(t, A) \) is a linear mapping (in other words, superposition holds), we refer to the system using the signal as a linear signaling (or linear modulation) system. Furthermore, for the criterion of interest the estimator will turn out to be linear so we refer to the problem as a linear estimation problem. In this section we study linear estimation and in Section 4.2.3, nonlinear estimation. For linear modulation (88a) can always be written as

\[ r(t) = A \sqrt{E} s(t) + w(t), \quad 0 \leq t \leq T, \quad (88b) \]

where \( s(t) \) has unit energy.

We can solve the linear estimation problem easily by exploiting its similarity to the detection problem that we just solved. From Section 2.4 we know that the likelihood function is needed. We recall, however, that the problem is greatly simplified if we can find a sufficient statistic and work with it instead of the received waveform. If we compare (88b) and (4)–(7), it is clear that a sufficient statistic is \( r_1 \), where

\[ r_1 = \int_0^T r(t) s(t) \, dt. \quad (89) \]

Just as in Section 4.2.1, the probability density of \( r_1 \), given \( a = A \), is Gaussian:

\[ E(r_1 | A) = A \sqrt{E}, \]

\[ \text{Var}(r_1 | A) = \frac{N_0}{2}. \quad (90) \]

It is easy to verify that the coefficients along the other orthogonal functions
272  4.2 Detection and Estimation in White Gaussian Noise

[see (8)] are independent of \( a \). Thus the waveform problem reduces to the classical estimation problem (see pp. 58–59).

The logarithm of the likelihood function is

\[
\log(l(A)) = -\frac{1}{2} \frac{(R_1 - A \sqrt{E})^2}{N_0/2}.
\]

(91)

If \( A \) is a nonrandom variable, the ML estimate is the value of \( A \) at which this function is a maximum. Thus

\[
\hat{a}_{\text{ml}}(R_1) = \frac{R_1}{\sqrt{E}}.
\]

(92)

The receiver is shown in Fig. 4.28a. We see the estimate is unbiased.

If \( a \) is a random variable with a probability density \( p_a(A) \), then the MAP estimate is the value of \( A \) where

\[
\log(p_a(A)) = -\frac{1}{2} \frac{(R_1 - A \sqrt{E})^2}{N_0/2} + \log(p_a(A)),
\]

(93)

is a maximum. For the special case in which \( a \) is Gaussian, \( N(0, \sigma_a) \), the MAP estimate is easily obtained by differentiating \( l_p(A) \) and equating the result to zero:

\[
\frac{\partial l_p(A)}{\partial A} = \frac{R_1 - A \sqrt{E}}{N_0/2} \sqrt{E} - \frac{A}{\sigma_a^2} = 0
\]

and

\[
\hat{a}_{\text{map}}(R_1) = \frac{2E/N_0 - 1/\sigma_a^2}{2E/N_0 + 1/\sigma_a^2} \frac{R_1}{\sqrt{E}}.
\]

(95)

In both the ML and MAP cases it is easy to show that the result is the absolute maximum.

The MAP receiver is shown in Fig. 4.28b. Observe that the only difference between the two receivers is a gain. The normalized error variances follow easily: for MAP

\[
\frac{E[a_n^2]}{\sigma_a^2} = \sigma_{a_n}^2 \triangleq \frac{\sigma_{a_n}^2}{\sigma_a^2} = \left(1 + \frac{2\sigma_a^2E}{N_0}\right)^{-1} \quad \text{(MAP)}.
\]

(96)

The quantity \( \sigma_a^2E \) is the expected value of the received energy. For ML

\[
\sigma_{a_n}^2 \triangleq \frac{\sigma_{a_n}^2}{\sigma_a^2} = \left(\frac{2\sigma_a^2E}{N_0}\right)^{-1} \quad \text{(ML)}.
\]

(97)

Here \( \sigma_a^2 E \) is the actual value of the received energy. We see that the variance of the maximum likelihood estimate is the reciprocal of \( \sigma^2 \), the performance index of the simple binary problem. In both cases we see that the only way to decrease the mean-square error is to increase the energy-to-noise ratio. In many situations the available energy-to-noise ratio is not adequate to provide the desired accuracy. In these situations we try a nonlinear signaling scheme in an effort to achieve the desired accuracy. In the next section we discuss the nonlinear estimation.

Before leaving linear estimation, we should point out that the MAP estimate is also the Bayes estimate for a large class of criteria. Whenever \( a \) is Gaussian the a posteriori density is Gaussian and Properties 1 and 2 on pp. 60–61 are applicable. This invariance to criterion depends directly on the linear signaling model.

4.2.3 Nonlinear Estimation

The system in Fig. 4.7 illustrates a typical nonlinear estimation problem. The received signal is

\[
r(t) = s(t, A) + w(t), \quad 0 \leq t \leq T.
\]

(98)

From our results in the classical case we know that a sufficient statistic does not exist in general. As before, we can construct the likelihood

---

**Fig. 4.28** Optimum receiver, linear estimation: (a) ML receiver; (b) MAP receiver.
section of the text; for example, the received signals are corrupted by additive zero-mean Gaussian noise which is independent of the hypotheses.

Section P4.2 Additive White Gaussian Noise

### Binary Detection

**Problem P4.2.1.** Derive an expression for the probability of detection \( P_d \), in terms of \( d \) and \( P_F \), for the given signal in the additive white Gaussian noise detection problem. [See (37) and (38)].

**Problem P4.2.2.** In a binary FSK system one of two sinusoids of different frequencies is transmitted; for example,

\[
s_1(t) = f(t) \cos 2\pi f_c t, \quad 0 \leq t \leq T, \\
s_2(t) = f(t) \cos 2\pi (f_c + \Delta f) t, \quad 0 \leq t \leq T,
\]

where \( f_c \gg 1/T \) and \( \Delta f \). The correlation coefficient is

\[
\rho = \frac{\int_0^T f^2(t) \cos(2\pi \Delta f t) \, dt}{\sqrt{\int_0^T f^4(t) \, dt}}.
\]

The transmitted signal is corrupted by additive white Gaussian noise \( N_0/2 \).

1. Evaluate \( \rho \) for a rectangular pulse; that is,

\[
f(t) = \left( \frac{2E}{T} \right)^{1/2}, \quad 0 \leq t \leq T, \\
= 0, \quad \text{elsewhere.}
\]

Sketch the result as a function of \( \Delta f / T \).

2. Assume that we require \( P_F (\varepsilon) = 0.01 \). What value of \( E/N_0 \) is necessary to achieve this if \( \Delta f = \infty \)? Plot the increase in \( E/N_0 \) over this asymptotic value that is necessary to achieve the same \( P_F (\varepsilon) \) as a function of \( \Delta f / T \).

**Problem P4.2.3.** The risk involved in an experiment is

\[
R = C_x P_x P_0 + C_d P_d P_1.
\]

The applicable ROC is Fig. 2.9. You are given (a) \( C_x = 2 \); (b) \( C_d = 1 \); (c) \( P_1 \) may vary between 0 and 1. Sketch the line on the ROC that will minimize your maximum possible risk (i.e., assume \( P_1 \) is chosen to make \( R \) as large as possible. Your line should be a locus of the thresholds that will cause the maximum to be as small as possible)

**Problem P4.2.4.** Consider the linear feedback system shown below

\[
\begin{align*}
\text{Fig. P4.1}
\end{align*}
\]

The function \( x(t) \) is a known deterministic function that is zero for \( t < 0 \). Under \( H_F, A_1 = A_2 \). Under \( H_F, A_1 = A_0 \). The noise \( w(t) \) is a sample function from a white Gaussian process of spectral height \( N_0/2 \). We observe \( r(t) \) over the interval \((0, T)\). All initial conditions in the feedback system are zero.

1. Find the likelihood ratio test.
2. Find an expression for \( P_F \) and \( P_r \) for the special case in which \( x(t) = \delta(t) \) (an impulse) and \( T = \infty \).

**Problem P4.2.5.** Three commonly used methods for transmitting binary signals over an additive Gaussian noise channel are on-off keying (ASK), frequency-shift keying (FSK), and phase-shift keying (PSK):

\[
\begin{align*}
H_F: r(t) &= s_0(t) + w(t), \quad 0 \leq t \leq T, \\
H_1: r(t) &= s_1(t) + w(t), \quad 0 \leq t \leq T,
\end{align*}
\]

where \( w(t) \) is a sample function from a white Gaussian process of spectral height \( N_0/2 \). The signals for the three cases are as follows:

<table>
<thead>
<tr>
<th>( s_0(t) )</th>
<th>( s_1(t) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 0 )</td>
<td>( \sqrt{2E/T} ) sin ( \omega_1 t )</td>
</tr>
<tr>
<td>( \sqrt{2E/T} ) sin ( \omega_0 t )</td>
<td>( \sqrt{2E/T} ) sin ( \omega_0 t )</td>
</tr>
<tr>
<td>( \sqrt{2E/T} ) sin ( \omega_1 t )</td>
<td>( \sqrt{2E/T} ) sin ( \omega_0 t )</td>
</tr>
</tbody>
</table>

where \( \omega_0 = \omega_1 = 2\pi n/T \) for some non-zero integer \( n \) and \( \omega_0 = \omega_1 = 2\pi mT \) for some non-zero integer \( m \).

1. Derive appropriate signal spaces for the three techniques.
2. Find \( \Delta^2 \) and the resulting probability of error for the three schemes (assum e that the two hypotheses are equally likely).
3. Comment on the relative efficiency of the three schemes (a) with regard to utilization of transmitter energy, (b) with regard to ease of implementation.
4. Give an example in which the model of this problem does not accurately describe the actual physical situation.

**Problem P4.2.6. Suboptimum Receivers.** In this problem, we investigate the degradation in performance that results from using a filter other than the optimum receiver filter. A reasonable performance comparison is the increase in transmitted energy required to overcome the decrease in \( \Delta^2 \) that results from the mismatching. We would hope that for many practical cases the equipment simplification that results from using other than the matched filter is well worth the required increase in transmitted energy. The system of interest is shown in Fig. P4.2, in which

\[
\int_0^T r^2(t) \, dt = 1 \quad \text{and} \quad E[w(t)] w(t) = \frac{N_0}{2} \delta(t - \tau).
\]

\[
\begin{align*}
H_F: \sqrt{E}[s(t)] \\
H_0: 0 \\
0 \leq t \leq T
\end{align*}
\]

\[
\begin{align*}
h(t) \\
\text{Sample} \\
\text{at } t = T \\
\text{Decide}
\end{align*}
\]

\[
\begin{align*}
H_F \text{ or } H_0
\end{align*}
\]

**Sub-optimum receiver**
The received waveform is

$$H_1 r(t) = \sqrt{E} s(t) + w(t), \quad -\infty < t < \infty,$$

$$H_2 r(t) = w(t), \quad -\infty < t < \infty.$$ 

We know that

$$h_{opt}(t) = \begin{cases} s(T - t), & 0 \leq t \leq T, \\ 0, & \text{elsewhere}, \end{cases}$$

and

$$d_{opt}^2 = \frac{2E}{N_0}.$$

Suppose that

$$h(t) = e^{-a t} u_+(t), \quad -\infty < t < \infty,$$

$$s(t) = \left( \frac{1}{T} \right)^{\frac{1}{2}} 1 \leq t \leq T,$$

$$0, \quad \text{elsewhere}.$$

1. Choose the parameter \( a \) to maximize the output signal-to-noise ratio \( d^2 \).
2. Compute the resulting \( d^2 \) and compare with \( d_{opt}^2 \). How many decibels must the transmitter energy be increased to obtain the same performance?

**M-ary Signals**

**Problem 4.2.7. Gram-Schmidt.** In this problem we go through the details of the geometric representation of a set of \( M \) waveforms in terms of \( N (N \leq M) \) orthogonal signals.

Consider the \( M \) signals \( s_1(t), \ldots, s_M(t) \) which are either linearly independent or linearly dependent. If they are linearly independent, we can write (by definition)

$$\sum_{i=1}^{M} a_i s_i(t) = 0.$$

1. Show that if \( M \) signals are linearly dependent, then \( s_M(t) \) can be expressed in terms of \( s_{i}(t) ; i = 1, \ldots, M - 1 $$.
2. Continue this procedure until you obtain \( N \)-linearly independent signals and \( M-N \) signals expressed in terms of them. \( N \) is called the *dimension* of the signal set.
3. Carry out the details of the Gram-Schmidt procedure described on p. 258.

**Problem 4.2.8. Translation/Simplex Signals [18].** For maximum a posteriori reception the probability of error is not affected by a linear translation of the signals in the decision space; for example, the two decision spaces in Figs. P4.3a and P4.3b have the same \( \Pr (\epsilon) \). Clearly, the sets do not require the same energy. Denote the average energy in a signal set as

$$E \triangleq \sum_{i=1}^{M} \Pr (H_i) |s_i|^2 = \sum_{i=1}^{N} \Pr (H_i) |s_i|^2 = \sum_{i=1}^{M} \Pr (H_i) dt.$$ 

1. Find the linear translation that minimizes the average energy of the translated signal set; that is, minimize

$$E \triangleq \sum_{i=1}^{N} \Pr (H_i) |s_i - m|^2.$$

2. Explain the geometric meaning of the result in part 1.
3. Apply the result in part 1 to the case of \( M \) orthogonal equal-energy signals representing equally likely hypotheses. The resulting signals are called *Simplex signals*. Sketch the signal vectors for \( M = 2, 3, 4 \).
4. What is the energy required to transmit each signal in the Simplex set?
5. Discuss the energy reduction obtained in going from the orthogonal set to the Simplex set while keeping the same \( \Pr (\epsilon) \).

**Problem 4.2.9. Equally correlated signals.** Consider \( M \) equally correlated signals

$$E \int_{-\infty}^{\infty} s_i(t) s_j(t) dt = \begin{cases} E_i, & i = j, \\ 0, & i \neq j. \end{cases}$$

1. Prove

$$\frac{1}{M-1} \leq \rho \leq 1.$$

2. Verify that the left inequality is given by a Simplex set.
3. Prove an equally correlated set with energy \( E \) has the same \( \Pr (\epsilon) \) as an orthogonal set with energy \( E_{\text{ortho}} = E (1 - \rho) \).
4. Express the \( \Pr (\epsilon) \) of the Simplex set in terms of the \( \Pr (\epsilon) \) for the orthogonal set and \( M \).

**Problem 4.2.10. M Signals, Arbitrary Correlation.** Consider an \( M \)-ary system used to transmit equally likely messages. The signals have equal energy and may be correlated:

$$\rho_{ij} = \int_{-\infty}^{\infty} s_i(t) s_j(t) dt, \quad i, j = 1, 2, \ldots, M.$$

The channel adds white Gaussian noise with spectral height \( N_0/2 \). Thus

$$r(t) = \sqrt{E} s(t) + w(t), \quad 0 \leq t \leq T : H_i, \quad i = 1, \ldots, M.$$

1. Draw a block diagram of an optimum receiver containing \( M \) matched filters. What is the minimum number of matched filters that can be used?
2. Let \( \rho \) be the signal correlation matrix. The \( ij \) element is \( \rho_{ij} \). If \( \rho \) is non-singular, what is the dimension of the signal space?
3. Find an expression for \( \Pr (\epsilon | H_i) \), the probability of error, assuming \( H_i \) is true. Assume that \( \rho \) is non-singular.
4. Find an expression for \( \Pr (\epsilon) \).
5. Is this error expression valid for Simplex signals? (Is \( \rho \) singular?)
Find \( p_s(x) \).
5. Using the results in (4), we have
\[
1 - Pr(\epsilon) = \frac{1}{M} \exp \left( -\frac{E}{N_0} \right) \int_{-\infty}^{\infty} \exp \left( \frac{(2\pi/N_0)^{1/2} X}{2} \right) p_s(X) dX.
\]
Use \( p_s(X) \) from (4) to obtain the desired result.

**Problem 4.2.12 (continuation).**
1. Using the expression in (P.1) of Problem 4.2.11, show that \( \partial Pr(\epsilon) / \partial R_i > 0 \).
Does your derivation still hold if \( i \to i \) and \( f \to j \)?
2. Use the results of part 1 and Problem 4.2.9 to develop an intuitive argument that the Simplex set is locally optimum.
Comment: The proof of local optimality is contained in [70]. The proof of global optimality is contained in [71].

**Problem 4.2.13.** Consider the system in Problem 4.2.10. Define
\[ p_{\text{max}} = \max_{s} p_{s}. \]
1. Prove that \( Pr(\epsilon) \) on any signal set is less than the \( Pr(\epsilon) \) for a set of equally correlated signals with correlation equal to \( p_{\text{max}} \).
2. Express this in terms of the error probability for a set of orthogonal signals.
3. Show that the \( Pr(\epsilon) \) is upper bounded by
\[
Pr(\epsilon) \leq (M - 1)^{c} - \text{erfc}_{s} \left( \frac{E}{N_0} / (1 - p_{\text{max}}) \right)^{l}. \]

**Problem 4.2.14 [72].** Consider the system in Problem 4.2.10. Define
\[ d_i: \text{distance between the } i^{th} \text{ message point and the nearest neighbor}. \]
Observe
\[
d_i = \min_{j} 2 \sqrt{(1 - r_{ij}) E / N_0} \]
\[
d = \frac{1}{M} \sum_{i=1}^{M} d_i, \]
\[ d_{\text{min}} = \min_{i} d_i. \]
Prove
\[ \text{erfc}_{s}(d) \leq Pr(\epsilon) \leq (M - 1) \text{erfc}_{s}(d_{\text{min}}). \]
Note that this result extends to signals with unequal energies in an obvious manner.

**Problem 4.2.15.** In (68) of the text we used the limit
\[
\lim_{M \to \infty} \ln \left[ 1 + \left( \frac{M}{N_0} \right)^{1/2} \right] = 0.
\]
Use l'Hôpital's rule to verify the limits asserted in (69) and (70).

**Problem 4.2.16.** The error probability in (66) is the probability of error in deciding which signal was sent. Each signal corresponds to a sequence of digits; for example, if \( M = 8 \),
\[
000 \to s_0(t), \quad 100 \to s_1(t),
001 \to s_2(t), \quad 101 \to s_3(t),
010 \to s_4(t), \quad 110 \to s_5(t),
011 \to s_6(t), \quad 111 \to s_7(t).
\]
Therefore an error in the signal decision does not necessarily mean that all digits will be in error. Frequently the digit (or bit) error probability \( P_{e_b}(e) \) is the error of interest.

1. Verify that if an error is made any of the other \( M - 1 \) signals are equally likely to be chosen.
2. Verify that the expected number of bits in error, given a signal error is made, is
   \[
   \sum_{i=1}^{M} \left( \frac{\log_2 M}{i} \right) = \frac{(\log_2 M)M}{2(M-1)}
   \]
3. Verify that the bit error probability is
   \[
   P_{e_b}(e) = \frac{M}{2(M-1)} P_r(e).
   \]

4. Sketch the behavior of the bit error probability for \( M = 2, 4, \) and 8 (use Fig. 4.25).

**Problem 4.2.17. Bi-orthogonal Signals.** Prove that for a set of \( M \) bi-orthogonal signals with energy \( E \) and equally likely hypotheses the \( P_r(e) \) is

\[
P_r(e) = 1 - \int_{0}^{\infty} \exp \left( -\frac{1}{N_0} \right) \left[ \int_{x}^{\infty} \exp \left( -\frac{x^2}{N_0} \right) dx \right]^{M-1} dx.
\]

Verify that this \( P_r(e) \) approaches the error probability for orthogonal signals for large \( M \) and \( d^2 \). What is the advantage of the bi-orthogonal set?

**Problem 4.2.18.** Consider the following digital communication system. There are four equally probable hypotheses. The signals transmitted under the hypotheses are

\[
\begin{align*}
H_1: & \left( \frac{2E}{T} \right) A \sin \omega_1 t, & 0 \leq t \leq T, \\
H_2: & \left( \frac{2E}{T} \right) A \sin \omega_2 t, & 0 \leq t \leq T, \\
H_3: & \left( \frac{2E}{T} \right) A \sin \omega_3 t, & 0 \leq t \leq T, \\
H_4: & \left( \frac{2E}{T} \right) A \sin \omega_4 t, & 0 \leq t \leq T,
\end{align*}
\]

The signal is corrupted by additive Gaussian white noise \( w(t) \), \( (N_0/2) \).

1. Draw a block diagram of the minimum probability of error receiver and the decision space and compute the resulting probability of error.
2. How does the probability of error behave for large \( A^2/N_0 \)?

**Problem 4.2.19. M-ary ASK [72].** An ASK system is used to transmit equally likely messages

\[
s(t) = \sqrt{E} \phi_i(t), \quad i = 1, 2, \ldots, M,
\]

where

\[
\sqrt{E} = (l - 1) \Delta, \quad \int_{0}^{T} \phi_i(t) dt = 1.
\]

The received signal under the \( i \)th hypothesis is

\[
r(t) = s(t) + w(t), \quad 0 \leq t \leq T; \quad H_i, \quad i = 1, 2, \ldots, M,
\]

where \( w(t) \) is a white noise with spectral height \( N_0/2 \).

1. Draw a block diagram of the optimum receiver.
2. Draw the decision space and compute the \( P_r(e) \).
3. What is the average transmitted energy?

Note: \( \sum_{i=1}^{M} j^2 = (n - 1) n (2n - 1) / 6 \).

4. What translation of the signal set in the decision space would maintain the \( P_r(e) \) while minimizing the average transmitted energy?

**Problem 4.2.20 (continuation).** Use the sequence transmission model on pp. 264–265 with the ASK system in part 4 of Problem 4.2.19. Consider specifically the case in which \( M = 4 \). How should the digit sequence be mapped into signals to minimize the bit error probability? Compute the signal error probability and the bit error probability.

**Problem 4.2.21. M-ary PSK [72].** A communication system transmitter sends one of \( M \) messages over an additive white Gaussian noise channel (spectral height \( N_0/2 \)) using the signals

\[
s_i(t) = \left( \frac{2E}{T} \right)^{1/2} \cos \left( \frac{2 \pi}{M} t + \frac{2m_i}{M} \right), \quad 0 \leq t \leq T,
\]

elsewhere, \( i = 0, 1, 2, \ldots, M - 1 \),

where \( \pi \) is an integer. The messages are equally likely. This type of system is called an \( M \)-ary phase-shift-keyed (PSK) system.

1. Draw a block diagram of the optimum receiver. Use the minimum number of filters.
2. Draw the decision-space and decision lines for various \( M \).
3. Prove

\[
a \leq P_r(e) \leq 2a,
\]

where

\[
a = \text{erfc} \left( \frac{2E}{N_0} \sin \frac{\pi}{M} \right).
\]

**Problem 4.2.22 (continuation).** Optimum PSK [73]. The basic system is shown in Fig. 4.24. The possible signaling strategies are the following:

1. Use a binary PSK set with the energy in each signal equal to PT.
2. Use an M-ary PSK set with the energy in each signal equal to PT \( \log_2 M \).

Discuss how you would choose \( M \) to minimize the digit error probability. Compare bi-phase and four phase PSK on this basis.

**Problem 4.2.23 (continuation).** In the context of an M-ary PSK system discuss qualitatively the effect of an incorrect phase reference. In other words, the nominal signal
set is given in Problem 4.2.22 and the receiver is designed on that basis. The actual signal set, however, is

\[ s(t) = \begin{cases} \left( \frac{2E}{T} \right)^n \cos \left( \frac{2\pi n}{T} t + \frac{2\pi i}{M} + \theta \right), & 0 \leq t \leq T, \quad i = 1, 2, \ldots, M, \\ 0, & \text{elsewhere}, \quad n \text{ is an integer}, \end{cases} \]

where \( \theta \) is a random phase angle. How does the importance of a phase error change as \( M \) increases?

**Estimation**

**Problem 4.2.24. Bhattacharyya Bound.** Let

\[ r(t) = s(t, A) + w(t), \quad 0 \leq t \leq T, \]

where \( s(t, A) \) is differentiable \( k \) times with respect to \( A \). The noise has spectral height \( N_0/2 \).

1. Extend the Bhattacharyya bound technique developed in Problem 4.2.23 to the waveforms for the \( n = 2 \) case. Assume that \( A \) is nonrandom variable.
2. Repeat the case in which \( A \) is a Gaussian random variable; \( N(0, \sigma_A) \).
3. Extend the results in parts 1 and 2 to the case in which \( n = 3 \).

**Problem 4.2.25.** Consider the problem in Example 1 on p. 276. In addition to the unknown time of arrival, the pulse has an unknown amplitude. Thus

\[ r(t) = b s(t - a) + w(t), \quad -T \leq t \leq T, \]

where \( a \) is a uniformly distributed random variable (see Fig. 4.29b) and \( b \) is Gaussian, \( N(0, \sigma_b) \)

Draw a block diagram of a receiver to generate the joint MAP estimates, \( \hat{d}_{\text{MAP}} \)
and \( \hat{b}_{\text{MAP}} \).

**Problem 4.2.26.** The known signal \( s(t), 0 \leq t \leq T \), is transmitted over a channel with unknown nonnegative gain \( A \) and additive Gaussian noise \( n(t) \):

\[ \int_0^T s^2(t) \, dt = E, \]

\[ K_a(t, \tau) = \frac{N_0}{2} \delta(t - \tau), \]

1. What is the maximum likelihood estimate of \( A \)?
2. What is the bias in the estimate?
3. Is the estimate asymptotically unbiased?

**Problem 4.2.27.** Consider the stationary Poisson random process \( x(t) \). A typical sample function is shown in Fig. P4.4.

**Event times**

\[ \begin{array}{cccccc}
\hline
0 & x & x & x & x & T \\
\hline
\end{array} \]

*Fig. P4.4*

The probability of \( n \) events in any interval \( \tau \) is

\[ \Pr(n, \tau) = \frac{(k\tau)^n}{n!} e^{-k\tau}. \]

The parameter \( k \) of the process is an unknown nonrandom variable which we want to estimate. We observe \( x(t) \) over an interval \((0, T)\).

1. Is it necessary to record the event times or is it adequate to count the number of events that occur in the interval? Prove that \( n^* \), the number of events that occur \( n \times 10^{-2} \) in the interval \((0, T)\) is a sufficient statistic.
2. Find the Cramer-Rao inequality for any unbiased estimate of \( k \).
3. Find the maximum-likelihood estimate of \( k \). Call this estimate \( \hat{k} \).
4. Prove that \( \hat{k} \) is unbiased.
5. Find

\[ \text{Var}(\hat{k} - k). \]

6. Is the maximum-likelihood estimate efficient?

**Problem 4.2.28.** When a signal is transmitted through a particular medium, the amplitude of the output is inversely proportional to the murkiness of the medium. Before observation the output of the medium is corrupted by additive, white Gaussian noise. (Spectral height \( N_0/2 \), double-sided.) Thus

\[ r(t) = \frac{1}{M} f(t) + w(t), \quad 0 \leq t \leq T, \]

where \( f(t) \) is a known signal and

\[ \int_0^T f^2(t) \, dt = E. \]

We want to design an optimum Murky-Meter.

1. Assume that \( M \) is a nonrandom variable. Derive the block diagram of a system whose output is the maximum-likelihood estimate of \( M \) (denoted by \( \hat{m}_{\text{ML}} \)).
2. Now assume that \( M \) is a Gaussian random variable with zero mean and variance \( \sigma_m^2 \). Find the equation that specifies the maximum a posteriori estimate of \( M \) (denoted by \( \hat{m}_{\text{MAP}} \)).
3. Show that

\[ \hat{m}_{\text{MAP}} \rightarrow \hat{m}_{\text{ML}}, \]

\[ \sigma_{m}^{-2} \rightarrow \infty. \]

**Section 4.3 Nonwhite Additive Gaussian Noise**

**Mathematical Preliminaries**

**Problem 4.3.1. Reversibility.** Prove that \( h_{\text{w}}(t, u) \) [defined in (157)] is a reversible operation by demonstrating an \( h_{\text{w}}^{-1}(t, u) \) such that

\[ \int_{T_i}^{T_f} h_{\text{w}}(t, u) h_{\text{w}}^{-1}(u, z) \, du = \delta(t - z). \]

What restrictions on the noise are needed?
section of the text; for example, the received signals are corrupted by additive zero-mean Gaussian noise which is independent of the hypotheses.

Section P4.2 Additive White Gaussian Noise

Binary Detection

Problem 4.2.1. Derive an expression for the probability of detection for a binary detection problem, in terms of the received signal and the noise, for the Bernoulli process. [see (37) and (38)].

Problem 4.2.2. In a binary FSK system one of two sinusoids of different frequencies is transmitted; for example,

\[ s_1(t) = f(t) \cos 2\pi f_1 t, \quad 0 \leq t \leq T, \]
\[ s_2(t) = f(t) \cos 2\pi (f_0 + \Delta f) t, \quad 0 \leq t \leq T, \]

where \( f_0 \gg 1/T \) and \( \Delta f \). The correlation coefficient is

\[ \rho = \frac{\int_0^T f^2(t) \cos(2\pi \Delta f) dt}{\sqrt{\int_0^T f^2(t) dt \int_0^T \cos^2(2\pi \Delta f) dt}}. \]

The transmitted signal is corrupted by additive white Gaussian noise \( (N_0/2) \).

1. Evaluate \( \rho \) for a rectangular pulse; that is,

\[ f(t) = \begin{cases} \frac{2E}{T} & 0 \leq t \leq T, \\ 0, & \text{elsewhere.} \end{cases} \]

Sketch the result as a function of \( \Delta fT \).

2. Assume that we require \( Pr(e) = 0.01 \). What value of \( E/N_0 \) is necessary to achieve this if \( \Delta f = \infty \)? Plot the increase in \( E/N_0 \) over this asymptotic value that is necessary to achieve the same \( Pr(e) \) as a function of \( \Delta fT \).

Problem 4.2.3. The risk involved in an experiment is

\[ R = C_T P_T P_0 + C_M P_M P_1. \]

The applicable ROC is Fig. 2.9. You are given (a) \( C_T = 2 \); (b) \( C_m = 1 \); (c) \( P_T \) may vary between 0 and 1. Sketch the line on the ROC that will minimize your maximum possible risk (i.e., assume \( P_T \) is chosen to make \( R \) as large as possible. Your line should be a locus of the thresholds that will cause the maximum to be as small as possible).

Problem 4.2.4. Consider the linear feedback system shown below

\[ \begin{align*}
\text{Fig. P4.1} \\
\text{Problem 4.2.4}.
\end{align*} \]

The function \( x(t) \) is a known deterministic function that is zero for \( t < 0 \). Under \( H_1, A_1 = A_1 \). Under \( H_0, A_1 = A_0 \). The noise \( w(t) \) is a sample function from a white Gaussian process of spectral height \( N_0/2 \). We observe \( r(t) \) over the interval \( (0, T) \). All initial conditions in the feedback system are zero.

1. Find the likelihood ratio test.
2. Find an expression for \( P_D \) and \( P_F \) for the special case in which \( x(t) = \delta(t) \) (an impulse) and \( T = \infty \).

Problem 4.2.5. Three commonly used methods for transmitting binary signals over an additive Gaussian noise channel are on-off keying (ASK), frequency-shift keying (FSK), and phase-shift keying (PSK):

\[ H_0: r(t) = s_0(t) + w(t), \quad 0 \leq t \leq T, \]
\[ H_1: r(t) = s_1(t) + w(t), \quad 0 \leq t \leq T, \]

where \( w(t) \) is a sample function from a white Gaussian process of spectral height \( N_0/2 \). The signals for the three cases are as follows:

\[
\begin{array}{|c|c|c|}
\hline
& \text{ASK} & \text{FSK} & \text{PSK} \\
\hline
s_0(t) & 0 & \sqrt{2E/T} \sin \omega_0 t & \sqrt{2E/T} \sin \omega_0 t \\
\hline
s_1(t) & \sqrt{2E/T} \sin \omega_1 t & \sqrt{2E/T} \sin \omega_0 t & \sqrt{2E/T} \sin (\omega_0 t + \pi) \\
\hline
\end{array}
\]

where \( \omega_0 - \omega_1 = 2\pi n/T \) for some nonzero integer \( n \) and \( \omega_0 = 2\pi mT \) for some nonzero integer \( m \).

1. Draw appropriate signal spaces for the three techniques.
2. Find \( d^2 \) and the resulting probability of error for the three schemes (assume that the two hypotheses are equally likely).
3. Comment on the relative efficiency of the three schemes (a) with regard to utilization of transmitter energy, (b) with regard to ease of implementation.
4. Give an example in which the model of this problem does not accurately describe the actual physical situation.

Problem 4.2.6. Suboptimum Receivers. In this problem we investigate the degradation in performance that results from using a filter other than the optimum receiver. A reasonable performance comparison is the increase in transmitted energy required to overcome the decrease in \( d^2 \) that results from the mismatching. We would hope that for many practical cases the equipment simplification that results from using other than the matched filter is worth the required increase in transmitted energy. The system of interest is shown in Fig. P4.2, in which

\[ \int_0^T (n^2(t) dt = 1 \text{ and } E[w(t) w(t')] = \frac{N_0}{2} \delta(t - t'). \]

\[ \text{Fig. P4.2} \]

\[ \begin{align*}
H_1: \sqrt{E} & s(t) \\
H_0: 0 \\
0 \leq t \leq T
\end{align*} \]
The received waveform is
\[ H_1 r(t) = \sqrt{E} s(t) + w(t), \quad -\infty < t < \infty, \]
\[ H_2 r(t) = w(t), \quad -\infty < t < \infty. \]

We know that
\[ h_{opt}(t) = \begin{cases} s(T - t), & 0 \leq t \leq T, \\ 0, & \text{elsewhere}, \end{cases} \]
and
\[ d_{opt}^2 = \frac{2E}{N_0}. \]

Suppose that
\[ h(t) = e^{-alt}u(-t), \quad -\infty < t < \infty, \]
\[ s(t) = \left( \frac{1}{T} \right)^{1/2}, \quad 0 \leq t \leq T, \]
\[ 0, \quad \text{elsewhere}. \]

1. Choose the parameter \( a \) to maximize the output signal-to-noise ratio \( d^2 \).
2. Compute the resulting \( d^2 \) and compare with \( d_{opt}^2 \). How many decibels must the transmitter energy be increased to obtain the same performance?

**M-ARY SIGNALS**

**Problem 4.2.7. Gram-Schmidt.** In this problem we go through the details of the geometric representation of a set of \( M \) waveforms in terms of \( N(N \leq M) \) orthogonal signals.

Consider the \( M \) signals \( s_1(t), \ldots, s_M(t) \) which are either linearly independent or linearly dependent. If they are linearly dependent, we can write (by definition)
\[ \sum_{i=1}^{M} a_i s_i(t) = 0. \]

1. Show that if \( M \) signals are linearly dependent, then \( s_M(t) \) can be expressed in terms of \( s_i(t); i = 1, \ldots, M - 1 \).
2. Continue this procedure until you obtain \( N \)-linearly independent signals and \( M-N \) signals expressed in terms of them. \( N \) is called the dimension of the signal set.
3. Carry out the details of the Gram-Schmidt procedure described on p. 258.

**Problem 4.2.8. Translation/Simplex Signals** [18]. For maximum a posteriori reception the probability of error is not affected by a linear translation of the signals in the decision space; for example, the two decision spaces in Figs. P4.3a and P4.3b have the same Pr (e). Clearly, the sets do not require the same energy. Denote the average energy in a signal set as
\[ E \triangleq \sum_{i=1}^{M} \Pr(H_i) |s_i|^2 = \sum_{i=1}^{N} \Pr(H_i) E \int_{0}^{T} s_i^2(t) \, dt. \]

1. Find the linear translation that minimizes the average energy of the translated signal set; that is, minimize
\[ E \triangleq \sum_{i=1}^{M} \Pr(H_i) |s_i - m|^2. \]

2. Explain the geometric meaning of the result in part 1.
3. Apply the result in part 1 to the case of \( M \) orthogonal equal-energy signals representing equally likely hypotheses. The resulting signals are called Simplex signals. Sketch the signal vectors for \( M = 2, 3, 4 \).
4. What is the energy required to transmit each signal in the Simplex set?
5. Discuss the energy reduction obtained in going from the orthogonal set to the Simplex set while keeping the same Pr (e).

**Problem 4.2.9. Equally correlated signals.** Consider \( M \) equally correlated signals
\[ E \int_{0}^{T} s_i(t) s_j(t) \, dt = \left\{ \begin{array}{ll} E, & i = j, \\ \rho E, & i \neq j. \end{array} \right. \]

1. Prove
\[ -\frac{1}{M-1} \leq \rho \leq 1. \]

2. Verify that the left inequality is given by a Simplex set.
3. Prove that an equally-correlated set with energy \( E \) has the same Pr (e) as an orthogonal set with energy \( E_{orth} = E(1 - \rho) \).
4. Express the Pr (e) of the Simplex set in terms of the Pr (e) for the orthogonal set and \( M \).

**Problem 4.2.10. M Signals, Arbitrary Correlation.** Consider an \( M \)-ary system used to transmit equally likely messages. The signals have equal energy and may be correlated:
\[ \rho_{ij} = \int_{0}^{T} s_i(t) s_j(t) \, dt, \quad i, j = 1, 2, \ldots, M. \]

The channel adds white Gaussian noise with spectral height \( N_0/2 \). Thus
\[ r(t) = \sqrt{E} s_i(t) + w(t), \quad 0 \leq t \leq T; H_i, \quad i = 1, \ldots, M. \]

1. Draw a block diagram of an optimum receiver containing \( M \) matched filters. What is the minimum number of matched filters that can be used?
2. Let \( \rho \) be the signal correlation matrix. The \( ij \) element is \( \rho_{ij} \). If \( \rho \) is nonsingular, what is the dimension of the signal space?
3. Find an expression for \( \Pr(e|H_i) \), the probability of error, assuming \( H_i \) is true. Assume that \( \rho \) is nonsingular.
4. Find an expression for \( \Pr(e) \).
5. Is this error expression valid for Simplex signals? (Is \( \rho \) singular?)
Problem 4.2.11 (continuation). Error Probability [69]. In this problem we derive an alternate expression for the Pr (ε) for the system in Problem 4.2.10. The desired expression is

\[ 1 - \Pr (\varepsilon) = \frac{1}{M} \exp \left( -\frac{E}{N_0} \right) \int_{-\infty}^{\infty} \exp \left[ \frac{2E}{N_0} x \right] \times \left[ \frac{d}{dx} \int_{-\infty}^{x} \cdots \int_{-\infty}^{x} \exp \left( -\frac{1}{2} y^2 p r x \right) dy \right] dx. \]  

(P.1)

Develop the following steps:

1. Rewrite the receiver in terms of \( M \) orthonormal functions \( \phi_i(t) \). Define

\[ s_i(t) = \sum_{k=1}^{M} s_{ik} \phi_i(t), \quad i = 1, 2, \ldots, M, \]

\[ r(t) = \sum_{k=1}^{M} r_k \phi_k(t). \]

Verify that the optimum receiver forms the statistics

\[ l_i = \int_{0}^{T} r(t) s_i(t) dt = \sum_{k=1}^{M} s_{ik} R_k \]

and choose the greatest.

2. Assume that \( s_k(t) \) is transmitted. Show

\[ \Pr (\varepsilon|s_k) = \Pr (R \text{ in } Z_k) = \Pr \left( \sum_{j=1}^{M} s_{jk} R_k = \max_{j} \sum_{k=1}^{M} s_{jk} R_k \right). \]

3. Verify that

\[ \Pr(\varepsilon) = \frac{1}{M} \exp \left( -\frac{E}{N_0} \right) \sum_{k=1}^{M} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp \left\{ \frac{-(1/2N_0) \sum_{j=1}^{M} R_j^2}{(2\pi)^{M/2}} \right\} \times \exp \left[ \frac{2}{N_0} \max_{j} \sum_{k=1}^{M} R_k s_{jk} \right]. \]  

(P.2)

4. Define

\[ f(R) = \exp \left\{ \max_{j} \left[ \left( \frac{2}{EN_0} \right)^{1/2} \sum_{k=1}^{M} s_{jk} R_k \right] \right\} \]

and observe that (P.2) can be viewed as the expectation of \( f(R) \) over a set of statistically independent zero-mean Gaussian variables, \( R_k \), with variance \( N_0/2 \). To evaluate this expectation, define

\[ z_j = \left( \frac{2}{EN_0} \right)^{1/2} \sum_{k=1}^{M} s_{jk} R_k, \quad j = 1, 2, \ldots, M, \]

and

\[ z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_M \end{bmatrix}. \]

Find \( p_{\varepsilon}(z) \). Define

\[ x = \max_{j} z_j. \]

Find \( p_{\varepsilon}(x) \). Define

\[ 1 - \Pr (\varepsilon) = \frac{1}{M} \exp \left( -\frac{E}{N_0} \right) \int_{-\infty}^{\infty} \exp \left[ \frac{2E}{N_0} x \right] p_{\varepsilon}(X) \, dx. \]

Use \( p_{\varepsilon}(X) \) from (4) to obtain the desired result.

Problem 4.2.12 (continuation).

1. Using the expression in (P.1) of Problem 4.2.11, show that \( \theta \Pr (\varepsilon) / \theta d \theta > 0 \). Does your derivation still hold if \( 1 \to i \) and \( 2 \to j \)?

2. Use the results of part 1 and Problem 4.2.9 to develop an intuitive argument that the Simplex set is locally optimum.

Comment: The proof of local optimality is contained in [70]. The proof of global optimality is contained in [71].

Problem 4.2.13. Consider the system in Problem 4.2.10. Define

\[ p_{\text{max}} = \max_{1 \leq j \leq M} p_{\varepsilon}. \]

1. Prove that \( p_{\varepsilon} \) on any signal set is less than the \( p_{\varepsilon} \) for a set of equally correlated signals with correlation equal to \( p_{\text{max}}. \)

2. Express this in terms of the error probability for a set of orthogonal signals.

3. Show that the \( p_{\varepsilon} \) is upper bounded by

\[ p_{\varepsilon} \leq (M - 1) \left( \text{erfc} \left( \frac{E}{N_0} (1 - p_{\text{max}}) \right) \right). \]

Problem 4.2.14 [72]. Consider the system in Problem 4.2.10. Define

\[ d_i : \text{distance between the } \text{i} \text{th message point and the nearest neighbor.} \]

Observe

\[ d_i = \min \left[ 2 \sqrt{1 - p_{\varepsilon}} (E/N_0) \right], \]

\[ d = \frac{1}{M} \sum_{i=1}^{M} d_i, \]

\[ d_{\text{min}} = \min_{i} d_i. \]

Prove

\[ \text{erfc} \left( d \right) \leq p_{\varepsilon} \leq (M - 1) \text{erfc} \left( d_{\text{min}} \right). \]

Note that this result extends to signals with unequal energies in an obvious manner.

Problem 4.2.15. In (68) of the text we used the limit

\[ \lim_{M \to \infty} \text{erfc} \left[ x + \frac{2PT \log_2 M}{N_0} \right] = \frac{1}{1/(M - 1)}. \]

Use l'Hospital's rule to verify the limits asserted in (69) and (70).

Problem 4.2.16. The error probability in (66) is the probability of error in deciding which signal was sent. Each signal corresponds to a sequence of digits; for example, if \( M = 8 \),

\[
\begin{align*}
000 &\rightarrow s_0(t) & 100 &\rightarrow s_3(t) \\
001 &\rightarrow s_1(t) & 101 &\rightarrow s_4(t) \\
010 &\rightarrow s_2(t) & 110 &\rightarrow s_5(t) \\
011 &\rightarrow s_3(t) & 111 &\rightarrow s_6(t).
\end{align*}
\]
Therefore an error in the signal decision does not necessarily mean that all digits will be in error. Frequently the digit (or bit) error probability \( \Pr_e (e) \) is the error of interest.

1. Verify that if an error is made any of the other \( M - 1 \) signals are equally likely to be chosen.
2. Verify that the expected number of bits in error, given a signal error is made, is

\[
\sum_{i=1}^{\log_2 M} \left( \frac{\log_2 M}{\log_2 M} \right) = \frac{(\log_2 M)M}{2(M - 1)}
\]

3. Verify that the bit error probability is

\[
\Pr_e (e) = \frac{M}{2(M - 1)} \Pr (e).
\]

4. Sketch the behavior of the bit error probability for \( M = 2, 4, \) and \( 8 \) (use Fig. 4.25).

**Problem 4.1.17. Bi-orthogonal Signals.** Prove that for a set of \( M \) bi-orthogonal signals with energy \( E \) and equally likely hypotheses the \( \Pr (e) \) is

\[
\Pr (e) = 1 - \int_0^1 \frac{1}{\sqrt{\pi N_0}} \exp \left[ -\frac{1}{N_0} (x - \sqrt{E} y) \right] \left[ \int_x^1 \frac{1}{\sqrt{\pi N_0}} \exp \left( -\frac{y^2}{N_0} \right) dy \right]^{M-1} dx.
\]

Verify that this \( \Pr (e) \) approaches the error probability for orthogonal signals for large \( M \) and \( \sigma^2 \). What is the advantage of the bi-orthogonal set?

**Problem 4.1.18.** Consider the following digital communication system. There are four equally probable hypotheses. The signals transmitted under the hypotheses are

\[
H_1: \left( \frac{2}{T} \right)^{1/2} A \sin \omega_0 t, \quad 0 \leq t \leq T,
\]

\[
H_2: \left( \frac{2}{3} \right)^{1/2} A \sin \omega_0 t, \quad 0 \leq t \leq T,
\]

\[
H_3: -\left( \frac{2}{T} \right)^{1/2} A \sin \omega_0 t, \quad 0 \leq t \leq T,
\]

\[
H_4: -\left( \frac{2}{3} \right)^{1/2} A \sin \omega_0 t, \quad 0 \leq t \leq T.
\]

The signal is corrupted by additive Gaussian white noise \( w(t) \), \( (N_0/2) \).

1. Draw a block diagram of the minimum probability of error receiver and the decision space and compute the resulting probability of error.
2. How does the probability of error behave for large \( \Delta T/N_0 \)?

**Problem 4.1.19. M-ary ASK [72].** An ASK system is used to transmit equally likely messages

\[
s_i(t) = \sqrt{E_i} \phi_i(t), \quad i = 1, 2, \ldots, M,
\]

where

\[
\sqrt{E_i} = (i - 1) \Delta, \quad \int_0^T \phi_i^2(t) dt = 1.
\]

The received signal under the \( i \)th hypothesis is

\[
r(t) = s_i(t) + w(t), \quad 0 \leq t \leq T: H_i, \quad i = 1, 2, \ldots, M,
\]

where \( w(t) \) is a white noise with spectral height \( N_0/2 \).

1. Draw a block diagram of the optimum receiver. Use the minimum number of filters.
2. Draw the decision space and decision lines for various \( M \).
3. Prove

\[
\alpha \leq \Pr (e) \leq 2 \alpha,
\]

where

\[
\alpha = \text{erfc} \left( \left( \frac{2E_i}{N_0} \right)^{1/2} \sin \frac{T}{M} \right).
\]

**Problem 4.2.20 (continuation). Optimum PSK [73].** The basic system is shown in Fig. 4.24. The possible signaling strategies are the following:

1. Use a binary PSK set with the energy in each signal equal to \( PT \).
2. Use an \( M \)-ary PSK set with the energy in each signal equal to \( PT \log M \).

Discuss how you would choose \( M \) to minimize the digit error probability. Compare bi-phase and four phase PSK on this basis.

**Problem 4.2.23 (continuation).** In the context of an \( M \)-ary PSK system discuss qualitatively the effect of an incorrect phase reference. In other words, the nominal signal