Binary Signal Detection in AWGN

1 Examples of Signal Sets for Binary Data Transmission

In an $M$-ary data transmission system there is a collection $\{s_i | 0 \leq i < M\}$ of $M$ signals, which are also referred to as waveform. Information is conveyed to the receiver by transmitting signals from this collection. For a binary data transmission, the collection has only two signals.

For a binary baseband communication system, each of the two signals has its energy concentrated at low frequencies (below some frequency $W$). In order to send a sequence of binary digit, a corresponding sequence of waveforms is transmitted to the receiver. If the binary digits are generated at the rate of one digit every $T$ units of time, the waveform must be transmitted at a rate of $1/T$. If the waveforms are time limited and of duration $T$, the transmitted signal consists of a sequence of nonoverlapping waveforms in consecutive time intervals of duration $T$.

The collection of waveform that is available to the transmitter is known as the signal set. A binary signal set consists of two waveforms $s_0(t)$ and $s_1(t)$. If the waveforms have duration $T$, then for both $i = 0$ and $i = 1$,

$$s_i(t) = 0 \text{ for } t < 0 \text{ and } t > T. \quad (1)$$

Let $a_0, a_1, a_{N-1}$ be a sequence of binary digits (i.e., for each $n$, $a_n = 0$ or $a_n = 1$). In order to send this sequence of binary digits to a receiver, a sequence of waveforms is transmitted. The composite signal $x(t)$ formed by the sequence of waveforms can be written as

$$x(t) = \sum_{n=0}^{N-1} s_{a_n}(t - nT); \quad (2)$$

that is, $a_n$ is sent by transmitting the waveform $s_{a_n}(t-nT)$. We will discuss the above representation in detail in Chapter 4.

For such waveforms, the composite signal $x(t)$, given by (2), satisfies

$$x(t) = s_{a_n}(t - nT), nT \leq t < (n+1)T. \quad (3)$$

For example, if the binary sequence 010 is to be sent, then

$$x(t) = s_0(t), \quad 0 \leq t < T$$

$$x(t) = s_1(t), \quad T \leq t < 2T$$

$$x(t) = s_0(t), \quad 2T \leq t < 3T.$$
where \( A \) is a positive number representing the amplitude of the signal. The signal set, given by 1-(a), is defined

\[
s_0(t) = 0, \text{ for all } t \\
s_1(t) = s(t)
\]

This signal set is one form of on-off signalling. The data bit 1 is represented by the presence of the pulse (on), and the data bit 0 is represented by its absence (off). The signal set is also an example of an orthogonal signal set. We say that signals \( s_0(t) \) and \( s_1(t) \) are orthogonal on the interval \([0, T]\) if

\[
\int_0^T s_0(t)s_1(t)dt = 0. \tag{5}
\]

The second example is obtained by letting

\[
s_0(t) = -s(t) \\
s_1(t) = s(t)
\]

as shown in Figure 1-(b). In this example, a positive pulse represents a 1 and a negative pulse represents a 0. If the pulse shapes are identical and the pulses have opposite polarity, as in this example, the resulting signal set is referred to as an antipodal signal set. In other words, a signal set \( \{s_0(t), s_1(t)\} \) is an antipodal signal set and the two signals are said to be antipodal if \( s_1(t) = -s_0(t) \) for all \( t \).

The third example is given by

\[
s_0(t) = \begin{cases} A & 0 \leq t < T/2 \\ 0 & \text{otherwise} \end{cases} \tag{6}
\]

\[
s_1(t) = \begin{cases} A & T/2 \leq t < T \\ 0 & \text{otherwise} \end{cases} \tag{7}
\]

By observation, we know that these two signals satisfy (5). Thus they are two orthogonal signals, which is referred to as orthogonal signalling.

Note that all signal sets shown in Figure 1 are time-limited signals that start at \( t = 0 \) and have duration \( T \), which is referred to as being time-limited to the interval \([0, T]\). Also notice that each of the signals has finite energy.

## 2 Detection

The receiver in a digital communication system cannot observe the transmitted signal. Instead, it observes a signal that is only statistically related to the transmitted signal. Based on the observation, the receiver makes inferences about the binary sequence that was sent. At present, we consider binary decisions only. Such a receiver is often referred to as a hard-decision receiver. Other types of decisions are possible, such as including the option of allowing the receiver to declare it does not know the value of a digit, but such an option is not allowed for the performance analysis given in this note.
Figure 1: Three examples of baseband signal sets
For binary decisions, there are only two possible outcomes: The receiver’s decision is correct or it is wrong. If the receiver makes the wrong decision, we say it has made an error. Our goal is to design the receiver in a way that minimizes the probability that it makes an error.

A receiver consists of one or more time-invariant linear (LTI) filters followed by one sampler (or more samplers) and a decision device. A sampler and a decision device are used in all receivers. Here channel noise is additive Gaussian white noise (AWGN). The structure of a receiver for detecting binary signals in an AWGN channel is shown in Figure 2.

2.1 The General Model

The signals considered are given by

\[
\begin{align*}
    s_0(t) &= \phi_0(t) \\
    s_1(t) &= \phi_1(t)
\end{align*}
\]

where \(\{\phi_0(t), \phi_1(t)\}\) is a binary signal set of the type described in Section 1, which are finite-energy, time limited signals of duration \(T\). The general model of the receiver structure for detecting binary signals in AWGN is shown in Figure 2. The received signal \(r(t)\) is the sum of the AWGN process \(N(t)\) and the signal \(s_i(t)\) where \(i = 0\) if the binary digit 0 is sent and \(i = 1\) if the binary digit 1 is sent. The filter shown in the Figure 2 is a LTI filter with impulse response \(h(t)\). The output of this filter, which is denoted by \(Y(t)\), is sampled at time \(T_0\). The output \(Y(T_0)\) of the sampler is then compare with a threshold \(\alpha\) in order to make a decision between the two alternatives 0 and 1. A detailed description follows.

First, we consider the channel noise process. Here we consider memoryless channel, i.e., the signal is only corrupted by the AWGN process \(n(t)\). The AWGN process \(n(t)\) does not depend on the transmitted signal. This channel is called an additive white Gaussian noise (AWGN) channel. The essential features AWGN channel model are as follows. The output of the channel is equal to the sum of the input signal and the channel noise. The channel noise a white Gaussian process, has zero mean, and power spectral density \(N_0/2\), and it independent of the input to the channel.

Although we speak of "channel" noise, it must be remembered that much of noise originates in the receiver itself. This noise is the thermal noise due to random motion of the conduction electrons in the resistive components of the receiver. Channel model includes the RF portion (if it applies) of the receiver, or receiver front end as it is sometimes called; thus, the thermal noise is viewed as part of the channel. Our concerns of the design of the communication systems is with the demodulation of the received signal, and this demodulation takes place after the signal is converted to a lower
frequency. Consequently, the received signal in this model is not the RF signal. In this chapter, the baseband signal (or a PCM signal) is considered to be the received signal.

The received signal is \( r(t) \), which is a random process. This is the input signal of the filter, i.e.,
\[
r(t) = s_i(t) \ast \delta(t) + n(t) = s_i(t) + n(t).
\] (8)

This process has a signal component \( s_i(t) \) and a noise component \( n(t) \). Since the filter is linear, its output \( Y(t) \) can be written as the sum of a signal component and a noise component. The signal component of \( Y(t) \) is just the convolution of the signal \( s_i(t) \) with the impulse \( h(t) \). Let \( s_{io}(t) \) and \( n_o(t) \) to represent the signal component and noise component, respectively, of the output of the filter. Then
\[
Y(t) = s_{io}(t) + n_o(t)
\] (9)
where the index \( i \) denotes the binary digit that is sent, and
\[
s_{io}(t) = s_i(t) \ast h(t) = \int_{-\infty}^{\infty} s_i(t - \tau)h(\tau)d\tau
\] (10)
and
\[
n_o(t) = n(t) \ast h(t) = \int_{-\infty}^{\infty} n(t - \tau)h(\tau)d\tau.
\] (11)

The noise component \( n(t) \) of the input gives rise to a noise component \( n_o(t) \) at the output of the filter. The properties of the noise process \( n_0(t) \) can be obtained by applying the methods in Chapter 2, which will given later.

The filter is followed by a sampler. The sampler is basically a switch that briefly at time \( T_0 \). The value of \( T_0 \) is arbitrary for now, we will determine it later by considering the optimal sampling time. The output of the sampler is the random variable \( Y(T_0) \), which is the input to the threshold device (see the Figure 1 in Section 2 of Chapter 3 in the slides).

In the threshold device, \( Y(T_0) \) is compared with a threshold level \( \alpha \). The output \( \hat{s} \) of the threshold device is determined by whether \( Y(T_0) > \alpha \) or \( Y(T_0) \leq \alpha \). The decision that the binary digit 0 was transmitted corresponds to \( \hat{s} = 0 \) as the output of the threshold device, and the decision that 1 was sent corresponds to \( \hat{s} = 1 \). For a fixed value \( \alpha \), there are only two nontrivial decision rules that are possible with this type of threshold device. One is to decide 0 was sent if \( Y(T_0) \leq \alpha \) and 1 was sent if \( Y(T_0) > \alpha \). The other just opposite: Decide 1 is sent if \( Y(T_0) \leq \alpha \) and 0 was sent if \( Y(T_0) > \alpha \).

The choice as to which of these two rules should be used depends on the relative values \( s_{0o}(T_0) \) and \( s_{1o}(T_0) \). Since the signal \( s_{io}(t) \) is the convolution of the signal \( s_i(t) \) with the impulse response \( h(t) \), the value of \( s_{io}(T_0) \) depends on \( s_i(t) \), \( h(t) \), and the sampling time \( T_0 \). If the following inequality is true:
\[
s_{1o}(T_0) > s_{0o}(T_0)
\]
then the decision rule that should be used is to decide 1 was sent (i.e., \( \hat{s} = 1 \)) if \( Y(T_0) > \alpha \) and 0 was sent (i.e., \( \hat{s} = 0 \)) if \( Y(T_0) \leq \alpha \).

This conclude the description of the components of the communication system shown in Figure 2. The next step is to analyze its performance in terms of optimal criteria in the sense that minimizes the probability of error.
2.2 The Filter Output

The output of the filter is the random process \( Y(t) \), which is given by (9). For the AWGN channel, \( n(t) \) is a white Gaussian noise process. Thus, \( R_n(\tau) = (N_0/2)\delta(\tau) \) (see Chapter 2). Applying the result in Chapter 2, we have

\[
S_{n_o}(f) = |H(f)|^2 S_n(f)
\]

where \( H(f) \) is the transfer function for the LTI filter \( h(t) \), \( S_n(f) \) is the spectral density for the channel noise process, and \( S_{n_o}(f) \) is the spectral density for the output process \( n_o(t) \). If needed, the autocorrelation function for \( n_o(t) \) can be computed from its spectral density by taking the inverse Fourier transform, but in many problems this is not required. Since \( n(t) \) is white, we obtain

\[
S_{n_o}(f) = \frac{1}{2} N_0 |H(f)|^2
\]

where \( N_0/2 \) is the spectral density for the noise process \( n(t) \).

Note that \( n_o(t) \) is still Gaussian. In order to determine the error probabilities, we really do not need a complete characterization of the random process \( Y(t) \). All that matters is the values of this process at the sampling time \( T_0 \), and what is actually required is a characterization of the random variable \( Y(T_0) \). The decision made by the receiver is based solely on the value of \( Y(T_0) \) since this is the only input to the threshold device.

2.3 The Input to the Threshold Device

The input to the threshold device is the random variable \( Y(T_0) \), which is the sum of the deterministic quantity \( s_o(T_0) \) and the random variable \( n_o(T_0) \). Because the receiver’s decision is based entirely on the random variable \( Y(T_0) \), this random variable is called the decision variable. The decision variable contains all of the information from the receiver input \( Y(t) \) that is actually used in the making decision.

Clearly, for any system of interest, \( n_o(T_0) \) does not depend on which signal is sent. This follows from the definition of the additive Gaussian noise channel: The process \( n(t) \) is independent of the channel input for such a channel. Consequently, the process \( n_o(t) \) is independent of the transmitted signal. It follows that the random variable \( n_o(T_0) \) does not depend on \( t \).

Suppose that 0 is sent. The decision variable is then given by

\[
Y(T_0) = s_o(T_0) + n_o(T_0).
\]

For convenience, we define a random process

\[
Y_0(t) = s_0o(t) + n_o(t).
\]  \hspace{1cm} (12)

When 0 is sent, \( Y(t) \) is equal to \( Y_0(t) \) for all \( t \). The random variable \( Y_0(T_0) \) is Gaussian because it is the sum of a deterministic value \( s_0o(T_0) \) and a Gaussian random variable \( n_o(T_0) \). The mean of \( Y_0(T_0) \), denoted by \( \mu_0(T_0) \), is given by

\[
\mu_0(T_0) = E[Y_0(T_0)] = s_0o(T_0) + E[n_o(T_0)].
\]

Since \( E[n_o(T_0)] = E[n(T_0)]H(0) \) and the mean of \( n(t) \) is zero, then the mean of the output noise process is zero, i.e., \( E[n_o(T_0)] = 0 \). Therefore,

\[
\mu_0(T_0) = s_0o(T_0) = s_0(T_0) * h(T_0) = \int_{-\infty}^{\infty} s_0(T_0 - \tau) h(\tau) d\tau.
\]  \hspace{1cm} (13)
The variance of \( Y_0(T_0) \) is given by
\[
\text{Var}(Y_0(T_0)) = E[(Y_0(T_0) - \mu_0(T_0))^2] = E[(Y_0(T_0) - s_{0o}(T_0))^2] = E[n_o(T_0)] = R_{n_o}(0).
\]

Next, suppose that 1 is sent. The decision variable is then the random variable
\[
Y(T_0) = Y_1(T_0)
\]
where \( Y_1(t) \) is the random process defined by
\[
Y_1(t) = s_{1o}(t) + n_o(t)
\]
The random variable \( Y_1(T_0) \) is Gaussian with mean \( \mu_1(T_0) \), given by
\[
\mu_1(T_0) = E[Y_1(T_0)] = s_{1o}(T_0) + E[n_o(T_0)] = s_{1o}(T_0) = \int_{-\infty}^{\infty} s_1(T_0 - \tau) h(\tau) d\tau.
\]
Similarly, we can obtain that the variance of \( Y_1(T_0) \), given as follows
\[
\text{Var}(Y_1(T_0)) = E[(Y_1(T_0) - \mu_1(T_0))^2] = E[(Y_1(T_0) - s_{1o}(T_0))^2] = E[n_o(T_0)] = R_{n_o}(0).
\]
Thus we obtain that
\[
\text{Var}(Y_0(T_0)) = \text{Var}(Y_1(T_0)) = R_{n_o}(0).
\]
If we set \( \sigma^2 = R_{n_o}(0) \), then \( \text{Var}(Y_i(T_0)) = \sigma^2 \) for both \( i = 0 \) and \( i = 1 \). In other words, the two decision variables have the same variance which is independent of which signal is sent.

One conclusion that can be drawn from this is that the only difference between two decision variables \( Y_0(T_0) \) and \( Y_1(T_0) \) is the mean. Both of these are Gaussian random variables; hence, they are completely characterized by their means and variances. However, they have the same variance \( \sigma^2 \). Thus, the decision device at the output of the sampler must discriminate between two random variables that differ only in their mean values. If 0 is sent, the mean is \( \mu_0(T_0) = s_{0o}(T_0) \), but if 1 is sent, the mean is \( \mu_1(T_0) = s_{1o}(T_0) \). Intuitively, we expect that the ability of the decision device to discriminate between these two cases should depend on the difference of the means \( \mu_1(T_0) - \mu_0(T_0) \). Notice that
\[
\mu_1(T_0) - \mu_0(T_0) = s_{1o}(T_0) - s_{0o}(T_0)
\]
\[
= \int_{-\infty}^{\infty} [s_1(T_0 - \tau) - s_0(T_0 - \tau)] h(\tau) d\tau, \text{ or equivalently,}
\]
\[
= \int_{-\infty}^{\infty} [s_1(\tau) - s_0(\tau)] h(T_0 - \tau) d\tau.
\]
Thus the difference of the means is a function of the impulse response of the filter and the difference of the two signals \( s_0(t) \) and \( s_1(t) \). The performance of the communication system is optimized by
use of signals that are as different as possible and a filter with an impulse response that accentuates this difference as much as possible. The selection of the signals and the filter to optimize performance is discussed later.

Example 1

Assume that the signal selection is antipodal signaling, i.e., \(s_0(t) = -s(t)\) and \(s_1(t) = s(t)\) where \(s(t) = A > 0\) for \(0 \leq t < T\), otherwise \(s(t) = 0\). The impulse response \(h(t)\) is given by \(h(t) = s(T - t)\) (it is called a matched filter). If 0 is sent, the input to the threshold device is the random variable \(Y(T_0) = Y_0(T_0) = s_{0\theta}(T_0) + n_\theta(T_0)\) which is Gaussian with mean \(s_{0\theta}(T_0)\) and variance \(\sigma^2 = \frac{1}{2}N_0T\). If 1 is sent, the input to the threshold device is the random variable \(Y(T_0) = Y_1(T_0) = s_{1\theta}(T_0) + n_\theta(T_0)\) which is Gaussian with mean \(s_{1\theta}(T_0)\) and variance \(\sigma^2\). The difference in the means is

\[
\mu_1(T_0) - \mu_0(T_0) = s_{1\theta}(T_0) - [-s_{1\theta}(T_0)] = 2s_{1\theta}(T_0)
\]

Since \(h(t) = s(T - t)\), we have

\[
\mu_1(T_0) - \mu_0(T_0) = 2s_{1\theta}(T_0) = 2s(T_0) \ast h(T_0)
\]

\[
= \begin{cases} 
2A^2T_0 & 0 \leq T_0 \leq T \\
2A^2(T - T_0) & T < T_0 \leq 2T \\
0 & \text{otherwise}
\end{cases}
\]

The difference is maximized if the sampling time \(T_0\) is equal to \(T\).

For certain filters, the transfer function \(H(f)\) is easier to work with than the impulse response. In general, we have

\[
\sigma^2 = \frac{1}{2}N_0 \int_{-\infty}^{\infty} |H(f)|^2 df.
\]

By Parseval’s theory, we have

\[
\sigma^2 = \frac{1}{2}N_0 \int_{-\infty}^{\infty} h(t)^2 dt.
\]

Thus, for the AWGN channel, \(\sigma^2\) can be evaluated by integrating either the square of the magnitude of the transfer function or the square of the impulse response, which is equal to the energy of the impulse response.

2.4 The Error Probabilities

Having characterized the decision variable \(Y(T_0)\), we can now give analytical expressions for the error probabilities \(P(e|i), i = 0, 1\), where \(P(e|0)\) denotes the probability that the decision made by the receiver is wrong when 0 is sent (i.e., when \(s_0(t)\) is transmitted), and \(P(e|1)\) denotes the probability that the decision made by the receiver is wrong when 1 is sent.
The receiver is wrong when 1 is sent if and only if the input to the threshold device is less than or equal to the threshold \( \alpha \). Under this condition that 1 is sent, the input \( Y(T_0) \) to the threshold device is the random variable \( Y_0(T_0) \). Let \( X = Y(T_0) \) and \( f_X(x|i) \) be the pdf of \( X \) given that digit \( i \) is sent, \( i = 0, 1 \). We denote the \( X \) being Gaussian with mean \( \mu \) and variance \( \sigma^2 \) by \( N(\mu, \sigma^2) \). From the previous discussions, we have \( f_X(x|i) \) is equal to the pdf of a Gaussian random variable with mean \( s_{io}(T_0) \) and variance \( \sigma^2 \), i.e., the pdf of a Gaussian variables with \( N(s_{io}(T_0), \sigma^2) \) for \( i = 0, 1 \). Thus

\[
P(e|1) = P\{Y(T_0) \leq \alpha \mid 1\} = P\{Y_1(T_0) \leq \alpha\} = \Phi((\alpha - s_{io}(T_0))/\sigma) = Q(s_{io}(T_0) - \alpha)/\sigma).
\]

Similarly, the probability of error when 0 is sent is given by

\[
P(e|0) = P\{Y(T_0) > \alpha \mid 0\} = P\{Y_0(T_0) > \alpha\} = 1 - P\{Y_0(T_0) \leq \alpha\} = 1 - \Phi((\alpha - s_{io}(T_0))/\sigma) = Q((\alpha - s_{io}(T_0))/\sigma).
\]

By taking advantage of the fact that \((-1)^i\) is equal to 1 if \( i = 0 \) and is \(-1\) if \( i = 1 \), we can summarize the key result as follows. If the transmitted signals are \( s_0(t) \) and \( s_1(t) \), and if \( \mu_i(T_0) = s_{io}(T_0) \) is equal to the mean of the signal component of the output of the filter \( h(t) \), sampled at time \( T_0 : 0 < T_0 \leq T \), then probability of error given the signal \( s_i(t) \) is transmitted is

\[
P(e|i) = Q((-1)^i(s_{io}(T_0) - \alpha)/\sigma)
\]

where \( \alpha \) is the threshold, \( T_0 \) is the sampling time, and \( \sigma \) is the standard deviation of the output of the filter.

**Example 2. On-Off Signals**

Consider a binary communication system with the signals \( s_1(t) = s(t) \) and \( s_0(t) = 0 \) where \( s(t) = A > 0 \) for \( 0 \leq t < T \) and otherwise \( s(t) = 0 \). The receiver filter has impulse response \( h(t) = s(T-t) \), i.e., the filter is the matched filter, and the sampling time is \( T_0 = T \). The output signal \( s_{1o}(t) \) is given by

\[
s_{1o}(t) = \begin{cases} \frac{A^2}{2}t & 0 \leq t \leq T \\ \frac{A^2}{2}(T-t) & T < t \leq 2T \\ 0 & \text{otherwise} \end{cases}
\]

and the output signal \( s_{0o}(t) \) is identically zero. It follows that \( \mu_1(T) = s_{1o}(T) = A^2T \) and \( \mu_0(T) = 0 \). For an AWGN channel with psd \( N_0/2 \) and threshold \( \alpha \), the error probabilities obtained by (15) are

\[
P(e|1) = Q((A^2T - \alpha)/\sigma) \quad P(e|0) = Q(\alpha/\sigma).
\]
Note that \( \int_{-\infty}^{\infty} h(t)^2 dt = A^2 T \), the energy of the signal since \( h(t) = s(T - t) \), denoted as \( E \). Thus the variance \( \sigma^2 = \frac{1}{2} N_0 A^2 T = \frac{1}{2} N_0 E \), and the error probability expressions become

\[
P(e|1) = Q\left( \frac{(A^2 T - \alpha)}{\sqrt{(1/2)N_0E}} \right)
\]
\[
P(e|0) = Q\left( \frac{\alpha}{\sqrt{(1/2)N_0E}} \right).
\]

The typical selection for the threshold is to be in the range \( 0 < \alpha < A^2 T \). Usually, it takes \( \alpha = A^2 T/2 \). Under this value of \( \alpha \), note that \( A^2 T - \alpha = A^2 T/2 = E/2 \), then we have

\[
P(e|1) = Q\left( \sqrt{\frac{E}{2N_0}} \right)
\]
\[
P(e|0) = Q\left( \frac{(E/2)}{\sqrt{(1/2)N_0E}} \right) = Q\left( \sqrt{\frac{E}{2N_0}} \right).
\]

The average probability of error is defined as

\[
P(e) = P\{ 1 \text { is sent}\} P(e|1) + P\{ 0 \text { is sent}\} P(e|0).
\]

If the two signals are transmitted equally likely, i.e., \( P\{ 1 \text { is sent}\} = P\{ 0 \text { is sent}\} = 1/2 \), then we have

\[
P(e) = \frac{1}{2} Q\left( \sqrt{\frac{E}{2N_0}} \right) + \frac{1}{2} Q\left( \sqrt{\frac{E}{2N_0}} \right) = Q\left( \sqrt{\frac{E}{2N_0}} \right) = \frac{1}{2} \text{erfc} \left( \frac{1}{2} \sqrt{\frac{E}{N_0}} \right).
\]

Question: (a) For the case of the sampling time \( T_0 = T/2 \), derive the average error probability. For \( h(t) = p_T(t) \) where \( p_T(t) = 1 \) for \( 0 \leq t \leq T \) and otherwise \( p_T(t) = 0 \), and the sampling time \( T_0 = T \), derive the average error probability.

### 3 Optimization of the Threshold

Note that \( Q(x) \) is a decreasing function of \( x \). From (15), if all of the parameters of the signals and the noise are fixed, the error probabilities \( P(e|1) \) and \( P(e|0) \) are functions of the threshold \( \alpha \). From the individual expressions for \( P(e|1) \) and \( P(e|0) \), it is clear that one of them is an increasing function of \( \alpha \) and the other is a decreasing function of \( \alpha \). For an AWGN channel, the optimal threshold in the sense of minimizing the error probabilities, is given by

\[
\alpha = [\mu_0(T_0) + \mu_1(T_0)]/2
\]

which is referred to as the minimax threshold in the literature. In this case, the error probability is given by

\[
P_m(e) = P(e|1) = P(e|0) = Q\left( \frac{[\mu_1(T_0) - \mu_0(T_0)]/(2\sigma)}{\sqrt{1/2)N_0E}} \right).
\]

(16)
4 The Matched Filter for the AWGN Channel

Because $Q(x)$ is a decreasing function of $x$, $P(e|i)$ is minimized by maximizing the quantity $[\mu_1(T_0) - \mu_0(T_0)]/(2\sigma)$. Because $\mu_1(T_0) - \mu_0(T_0)$ depends on the signals, but not on the noise, and $\sigma$ depends on the noise spectral density, but not on the signals, the quantity of interest can be thought of as a signal-to-noise ratio. It is convenient to define

$$ (SNR)_o = \frac{[\mu_1(T_0) - \mu_0(T_0)]}{2\sigma}, $$

so that (16) becomes

$$ P_m(e) = Q(SNR_o). $$

Keep in mind that $SNR_o$ depends on the filter impulse response $h(t)$, the sampling time $T_0$, and the signal set $\{s_0(t), s_1(t)\}$. Note that

$$ \mu_i(T_0) = (s_i * h)(T_0), \quad i = 0, 1 $$

and for the AWGN channel, recall that the variance $\sigma^2$ can be written as

$$ \sigma^2 = \frac{N_0}{2} \int_{-\infty}^{\infty} h^2(t)dt $$

where $N_0/2$ is the psd of the noise.

These results are valid for any LTI filter. The goal of this section is to find the optimum filter; that is, we wish to find the filter that gives the smallest values of the error probability $P_m(e)$. Here we use the threshold $\alpha = \frac{[\mu_1(T_0) + \mu_0(T_0)]}{2}$, which is referred to as the minmax threshold in the literature. Thus the signal-to-noise ratio is a function of the filter impulse response, substituting (18) to the SNR representation, we have

$$ (SNR)_o = \frac{(s_1 * h)(T_0) - (s_0 * h)(T_0)}{\sqrt{2N_0||h||}} $$

where

$$ ||h|| = \left\{ \int_{-\infty}^{\infty} h^2(t)dt \right\}^{1/2} = \left\{ \int_{-\infty}^{\infty} H^2(f)df \right\}^{1/2}. $$

Let us define

$$ g(t) = s_1(t) - s_0(t) $$

whose Fourier spectral is given by $G(f)$. Then maximizing $SNR_o$ is equivalent to maximize $||g * h(T_0)||^2$. Let $V^*(f) = H(f)e^{j2\pi fT_0}$. The Schwarz inequality (see Appendix G) is applied to the maximization of the signal-to-noise ratio. This gives, in the notation of frequency domain,

$$ (SNR)_o^2 = \frac{\left| \int_{-\infty}^{\infty} G(f)V^*(f)df \right|^2}{2N_0||h||^2} \leq \frac{\int_{-\infty}^{\infty} |G(f)|^2df \int_{-\infty}^{\infty} |V^*(f)|^2df}{2N_0||h||^2}.$$
where the equality is true if and only
\[ G(f) = k' V(f) \]
or equivalently,
\[ H(f) = k G^*(f) e^{-j 2\pi f T_0} \]  
(21)
where \( k \) is an arbitrary constant.

We conclude that the optimum filter has impulse response
\[ h(t) = \lambda g(T_0 - t) = \lambda [s_1(T_0 - t) - s_0(T_0 - t)] \]  
(22)
for some choice of \( \lambda \). It follows that for a proper choice of \( c \) (the gain constant is unimportant, we typically let \( c = 1/2 \) (or \( \lambda = 1 \)),
\[ (SNR)_{\text{max}} = \frac{c ||h||^2}{\sqrt{2 N_0 ||h||}} = \frac{||g(T_0 - t)/2||}{\sqrt{N_0/2}}. \]
The filter given by (22) is called the match filter for the AWGN channel. For the remainder of this section, we assume that
\[ g_{T_0}(t) = g(T_0 - t)/2 = \frac{1}{2} [s_1(T_0 - t) - s_0(T_0 - t)]. \]
The signal-to-noise ratio for the matched filter receiver can be expressed in terms of more fundamental parameters of the signal set \( \{s_0(t), s_1(t)\} \). We start with
\[ (SNR)_o = \frac{||g_{T_0}||}{\sqrt{N_0/2}} \]
and then use the fact that
\[ ||g_{T_0}||^2 = \int_{-\infty}^{\infty} |g_{T_0}(t)|^2 dt = \frac{1}{4} \int_{-\infty}^{\infty} [s_1(T_0 - t) - s_0(T_0 - t)]^2 dt \]
\[ = \frac{1}{4} \int_{-\infty}^{\infty} [s_1^2(u) + s_0^2(u)] du - \frac{1}{2} \int_{-\infty}^{\infty} s_1(u)s_0(u) du \]
\[ = \frac{1}{4} (E_0 + E_1) - \frac{1}{2} \rho \]
where \( \rho \) is the integral in the second term (the inner product of signals \( s_0 \) and \( s_1 \)). Letting \( E = (E_0 + E_1)/2 \) and \( r = \rho/E \), we can write \( ||g_{T_0}||^2 \) as
\[ ||g_{T_0}||^2 = E(1 - r)/2. \]
The signal-to-noise ratio for a receiver with the matched filter is therefore given by
\[ (SNR)_o = \{E(1 - r)/N_0\}^{1/2}. \]  
(23)
The parameter \( E \) is called the average energy for the signal set, and \( r \) is the correlation coefficient (or normalized correlation) for the two signals \( s_0 \) and \( s_1 \).

The first observation to be made about (23) is that the signal-to-noise ratio does not depend on the sampling time \( T_0 \) if the matched filter is used. The intuitive reason for this is that the matched filter, as defined by
\[ h(t) = c[s_1(T_0 - t) - s_0(T_0 - t)]. \]
automatically compensates for any changes in the sampling time. It should also noted that, because
the noise is WSS, the variance of the output noise is independent of the sampling time. (This is
true for any LTI filter.)

Another observation for (23) is that the signal-to-noise ratio and hence the error probabilities
depend on two signal parameters only. These are the average energy in the two signals \( s_0 \) and \( s_1 \)
and the inner product of \( s_0 \) and \( s_1 \). The detailed structure of the signals are unimportant.

**Example 3.** Antipodal signals have equal energy, so \( E = E_0 = E_1 \). We drop the subscripts and
denote the energy by \( E \). The inner product for antipodal signals is

\[
\rho = \int_{-\infty}^{\infty} s_1(t)s_0(t)dt = -\int_{-\infty}^{\infty} [s_1(t)]^2dt = -E_1.
\]

The correlation coefficient for antipodal signals is therefore \( r = -1 \). The resulting signal-to-noise
ratio is

\[
(SNR)_o = \sqrt{2E/N_0},
\]

and the error probabilities for the minimax threshold are given by

\[
P(e|1) = P(e|0) = Q(\sqrt{2E/N_0}).
\]