Optimum Receivers for the Additive White Gaussian Noise Channel
Table of Contents

- Optimum Receiver for Signals Corrupted by Additive White Gaussian Noise
  - Correlation Demodulator
  - Matched-Filter Demodulator
  - The Optimum Detector
  - Comparison of Digital Modulation Methods
We assume that the transmitter sends digital information by use of \( M \) signals waveforms \( \{s_m(t)=1,2,\cdots,M\} \). Each waveform is transmitted within the symbol interval of duration \( T \), i.e. \( 0 \leq t \leq T \).

The channel is assumed to corrupt the signal by the addition of white Gaussian noise, as shown in the following figure:

\[
r(t) = s_m(t) + n(t), \quad 0 \leq t \leq T
\]

where \( n(t) \) denotes a sample function of AWGN process with power spectral density \( \Phi_{nn}(f)=\frac{1}{2}N_0 \) W/Hz.
Optimum Receiver for Signals Corrupted by Additive White Gaussian Noise

◊ Our object is to design a receiver that is **optimum** in the sense that it minimizes the probability of making an error.

◊ It is convenient to subdivide the receiver into two parts—the *signal demodulator* and the *detector*.

![Diagram](image)

◊ The function of the *signal demodulator* is to convert the received waveform $r(t)$ into an $N$-dimensional vector $\mathbf{r} = [r_1, r_2, \ldots, r_N]$ where $N$ is the dimension of the transmitted signal waveform.

◊ The function of the *detector* is to decide which of the $M$ possible signal waveforms was transmitted based on the vector $\mathbf{r}$. 
Two realizations of the signal demodulator are described in the following section:

- One is based on the use of *signal correlators*.
- The second is based on the use of *matched filters*.

The optimum detector that follows the signal demodulator is designed to minimize the probability of error.
Correlation Demodulator

◊ We describe a *correlation demodulation* that decomposes the receiver signal and the noise into $N$-dimensional vectors.

◊ In other words, the signal and the noise are expanded into a series of linearly weighted orthonormal basis functions $\{f_n(t)\}$.

◊ It is assumed that the $N$ basis function $\{f_n(t)\}$ span the signal space, so every one of the possible transmitted signals of the set $\{s_m(t) = 1 \leq m \leq M\}$ can be represented as a linear combination of $\{f_n(t)\}$.
Suppose the receiver signal \( r(t) \) is passed through a parallel bank of \( N \) basis functions \( \{f_n(t)\} \), as shown in the following figure:

\[
\int_0^T r(t) f_k(t) dt = \int_0^T [s_m(t) + n(t)] f_k(t) dt
\]

\[
\Rightarrow \quad r_k = s_{mk} + n_k, \quad k = 1, 2, \ldots, N
\]

\[
\begin{align*}
s_{mk} &= \int_0^T s_m(t) f_k(t) dt, \quad k = 1, 2, \ldots, N \\
n_k &= \int_0^T n(t) f_k(t) dt, \quad k = 1, 2, \ldots, N
\end{align*}
\]

The signal is now represented by the vector \( s_m \) with components \( s_{mk}, k=1,2,\cdots N \). Their values depend on which of the \( M \) signals was transmitted.
From the above development, it follows that the correlator output \( \{r_k\} \) conditioned on the \( m \)th signal being transmitted are Gaussian random variables with mean

\[
E(r_k) = E(s_{mk} + n_k) = s_{mk}
\]

and equal variance

\[
\sigma_r^2 = \sigma_n^2 = \frac{1}{2} N_0
\]

Since the noise components \( \{n_k\} \) are uncorrelated Gaussian random variables, they are also statistically independent. As a consequence, the correlator outputs \( \{r_k\} \) conditioned on the \( m \)th signal being transmitted are statistically independent Gaussian variables.
The conditional probability density functions of the random variables $\mathbf{r} = [r_1 \ r_2 \ \cdots \ r_N]$ are:

$$p(\mathbf{r} \mid s_m) = \prod_{k=1}^{N} p(r_k \mid s_{mk}), \quad m = 1, 2, \ldots, M \quad ----(A)$$

$$p(r_k \mid s_{mk}) = \frac{1}{\sqrt{\pi N_0}} \exp \left[ - \frac{(r_k - s_{mk})^2}{N_0} \right], \quad k = 1, 2, \ldots, N \quad ----(B)$$

By substituting Equation (B) into Equation (A), we obtain the joint conditional PDFs

$$p(\mathbf{r} \mid s_m) = \frac{1}{(\pi N_0)^{N/2}} \exp \left[ - \sum_{k=1}^{N} \frac{(r_k - s_{mk})^2}{N_0} \right], \quad m = 1, 2, \ldots, M$$
Instead of using a bank of $N$ correlators to generate the variables $\{r_k\}$, we may use a bank of $N$ linear filters. To be specific, let us suppose that the impulse responses of the $N$ filters are:

$$h_k(t) = f_k(T - t), \quad 0 \leq t \leq T$$

where $\{f_k(t)\}$ are the $N$ basis functions and $h_k(t)=0$ outside of the interval $0 \leq t \leq T$.

The outputs of these filters are:

$$y_k(t) = r(t) * h_k(t)$$

$$= \int_0^t r(\tau) h_k(t - \tau) d\tau$$

$$= \int_0^t r(\tau) f_k(T - t + \tau) d\tau, \quad k = 1, 2, \ldots, N$$
Matched-Filter Demodulator

- If we sample the outputs of the filters at \( t = T \), we obtain
  \[
y_k(T) = \int_0^T r(\tau) f_k(\tau) d\tau = r_k, \quad k = 1, 2, \ldots, N
  \]

- A filter whose impulse response \( h(t) = s(T - t) \), where \( s(t) \) is assumed to be confined to the time interval \( 0 \leq t \leq T \), is called a matched filter to the signal \( s(t) \).

- An example of a signal and its matched filter are shown in the following figure.
The response of \( h(t) = s(T - t) \) to the signal \( s(t) \) is:

\[
y(t) = s(t) \ast h(t) = \int_0^t s(\tau) h(t - \tau) d\tau = \int_0^t s(\tau) s(T - t + \tau) d\tau
\]

which is the time-autocorrelation function of the signal \( s(t) \).

Note that the autocorrelation function \( y(t) \) is an even function of \( t \), which attains a peak at \( t = T \).
Matched-Filter Demodulator

- Matched filter demodulator that generates the observed variables \( \{ r_k \} \)
Properties of the matched filter.

If a signal $s(t)$ is corrupted by AWGN, the filter with an impulse response matched to $s(t)$ maximizes the output signal-to-noise ratio (SNR).

Proof:

Let us assume the receiver signal $r(t)$ consists of the signal $s(t)$ and AWGN $n(t)$ which has zero-mean and $\Phi_{nn}(f) = \frac{1}{2} N_0 \text{W/Hz}$.

Suppose the signal $r(t)$ is passed through a filter with impulse response $h(t)$, $0 \leq t \leq T$, and its output is sampled at time $t=T$. The output signal of the filter is:

$$y(t) = s(t) * h(t) = \int_0^t r(\tau)h(t-\tau)d\tau$$

$$= \int_0^t s(\tau)h(t-\tau)d\tau + \int_0^t n(\tau)h(t-\tau)d\tau$$
Matched-Filter Demodulator

◊ Proof: (cont.)

◊ At the sampling instant $t=T$:

$$y(T) = \int_{0}^{T} s(\tau) h(T - \tau) d\tau + \int_{0}^{T} n(\tau) h(T - \tau) d\tau$$

$$= y_s(T) + y_n(T)$$

◊ This problem is to select the filter impulse response that maximizes the output SNR$_0$ defined as:

$$\text{SNR}_0 = \frac{E[y_s^2(T)]}{E[y_n^2(T)]}$$

$$E[y_n^2(T)] = \int_{0}^{T} \int_{0}^{T} E[n(\tau)n(t)] h(T - \tau) h(T - t) dt d\tau$$

$$= \frac{1}{2} N_0 \int_{0}^{T} \int_{0}^{T} \delta(t - \tau) h(T - \tau) h(T - t) dt d\tau = \frac{1}{2} N_0 \int_{0}^{T} h^2(T - t) dt$$
Matched-Filter Demodulator

Proof: (cont.)

- By substituting for $y_s(T)$ and $E\left[ y_n^2(T) \right]$ into $\text{SNR}_0$.

$$\tau' = T - \tau$$

$$\text{SNR}_0 = \frac{\left[ \int_0^T s(\tau)h(T-\tau)d\tau \right]^2}{\frac{1}{2} N_0 \int_0^T h^2(T-t)dt} = \frac{\left[ \int_0^T h(\tau')s(T-\tau')d\tau' \right]^2}{\frac{1}{2} N_0 \int_0^T h^2(T-t)dt}$$

- Denominator of the SNR depends on the energy in $h(t)$.
- The maximum output SNR over $h(t)$ is obtained by maximizing the numerator subject to the constraint that the denominator is held constant.
Matched-Filter Demodulator

- Proof: (cont.)
  - **Cauchy-Schwarz inequality**: if \( g_1(t) \) and \( g_2(t) \) are finite-energy signals, then
    \[
    \left[ \int_{-\infty}^{\infty} g_1(t) g_2(t) dt \right]^2 \leq \int_{-\infty}^{\infty} g_1^2(t) dt \int_{-\infty}^{\infty} g_2^2(t) dt
    \]
    with equality when \( g_1(t) = C g_2(t) \) for any arbitrary constant \( C \).
  - If we set \( g_1(t) = h_1(t) \) and \( g_2(t) = s(T-t) \), it is clear that the SNR is maximized when \( h(t) = C s(T-t) \).
Proof: (cont.)

The output (maximum) SNR obtained with the matched filter is:

\[
\text{SNR}_0 = \left( \frac{1}{2N_0} \int_{0}^{T} h^2(T-t)dt \right) = \frac{1}{2N_0} \int_{0}^{T} \left( \int_{0}^{T} s(\tau)Cs(T-(T-\tau))d\tau \right)dt
\]

\[
= \frac{2}{N_0} \int_{0}^{T} s^2(t)dt = \frac{2\mathcal{E}}{N_0}
\]

Note that the output SNR from the matched filter depends on the energy of the waveform \(s(t)\) but not on the detailed characteristics of \(s(t)\).
The Optimum Detector

◊ Our goal is to design a signal detector that makes a decision on the transmitted signal in each signal interval based on the observation of the vector \( r \) in each interval such that the probability of a correct decision is maximized.

◊ We assume that there is no memory in signals transmitted in successive signal intervals.

◊ We consider a decision rule based on the computation of the posterior probabilities defined as

\[
P(s_m|r) = P(\text{signal } s_m \text{ was transmitted} | r), \quad m=1,2,\ldots,M.
\]

◊ The decision criterion is based on selecting the signal corresponding to the maximum of the set of posterior probabilities \( \{ P(s_m|r) \} \). This decision criterion is called the maximum a posterior probability (MAP) criterion.
Using Bayes’ rule, the posterior probabilities may be expressed as

\[ P(s_m | r) = \frac{p(r | s_m)P(s_m)}{p(r)} \quad \text{---(A)} \]

where \( P(s_m) \) is the *a priori probability* of the \( m \)th signal being transmitted.

The denominator of (A), which is independent of which signal is transmitted, may be expressed as

\[ p(r) = \sum_{m=1}^{M} p(r | s_m)P(s_m) \]

Some simplification occurs in the MAP criterion when the \( M \) signal are equally probable a priori, i.e., \( P(s_m) = 1/M \).

The decision rule based on finding the signal that maximizes \( P(s_m|r) \) is equivalent to finding the signal that maximizes \( P(r|s_m) \).
The Optimum Detector

- The conditional PDF $P(r|s_m)$ or any monotonic function of it is usually called the likelihood function.
- The decision criterion based on the maximum of $P(r|s_m)$ over the $M$ signals is called maximum-likelihood (ML) criterion.
- We observe that a detector based on the MAP criterion and one that is based on the ML criterion make the same decisions as long as a priori probabilities $P(s_m)$ are all equal.
- In the case of an AWGN channel, the likelihood function $p(r|s_m)$ is given by:

$$p(r|s_m) = \frac{1}{(\pi N_0)^N/2} \exp \left[ -\sum_{k=1}^{N} \frac{(r_k - s_{mk})^2}{N_0} \right], \quad m = 1, 2, \ldots, M$$

$$\ln p(r \mid s_m) = -\frac{1}{2} N \ln(\pi N_0) - \frac{1}{N_0} \sum_{k=1}^{N} (r_k - s_{mk})^2$$
The Optimum Detector

- The maximum of \( \ln p(r|s_m) \) over \( s_m \) is equivalent to finding the signal \( s_m \) that minimizes the **Euclidean distance**:

\[
D(r, s_m) = \sum_{k=1}^{N} (r_k - s_{mk})^2
\]

- We called \( D(r, s_m) \), \( m=1,2,\ldots,M \), the **distance metrics**.

- Hence, for the AWGN channel, the decision rule based on the ML criterion reduces to finding the signal \( s_m \) that is closest in distance to the receiver signal vector \( r \).

- We shall refer to this decision rule as **minimum distance detection**.
Example:

Consider the case of binary PAM signals in which the two possible signal points are $s_1 = -s_2 = \sqrt{\varepsilon_b}$, where $\varepsilon_b$ is the energy per bit. The priori probabilities are $P(s_1) = p$ and $P(s_2) = 1 - p$. Let us determine the metrics for the optimum MAP detector when the transmitted signal is corrupted with AWGN.

The receiver signal vector for binary PAM is:

$$r = \pm \sqrt{\varepsilon_b} + y_n(T)$$

where $y_n(T)$ is a zero mean Gaussian random variable with variance $\sigma_n^2 = \frac{1}{2} N_0$. 

The Optimum Detector

Example: (cont.)

◊ The conditional PDF $P(r|s_m)$ for two signals are

$$p(r \mid s_1) = \frac{1}{\sqrt{2\pi\sigma_n}} \exp \left[ -\frac{(r - \sqrt{\mathcal{E}_b})^2}{2\sigma_n^2} \right]$$

$$p(r \mid s_2) = \frac{1}{\sqrt{2\pi\sigma_n}} \exp \left[ -\frac{(r + \sqrt{\mathcal{E}_n})^2}{2\sigma_n^2} \right]$$

◊ Then the metrics $PM(r,s_1)$ and $PM(r,s_2)$ are

$$PM(r,s_1) = p \cdot p(r \mid s_1) = \frac{p}{\sqrt{2\pi\sigma_n}} \exp \left[ -\frac{(r - \sqrt{\mathcal{E}_b})^2}{2\sigma_n^2} \right]$$

$$PM(r,s_2) = (1-p) \cdot p(r \mid s_2) = \frac{1-p}{\sqrt{2\pi\sigma_n}} \exp \left[ -\frac{(r + \sqrt{\mathcal{E}_b})^2}{2\sigma_n^2} \right]$$
The Optimum Detector

Example: (cont.)

- If $PM(r,s_1) > PM(r,s_2)$, we select $s_1$ as the transmitted signal; otherwise, we select $s_2$. This decision rule may be expressed as:

$$\frac{PM(r,s_1)}{PM(r,s_2)} \begin{cases} s_1 & \geq 1 \\ s_2 & < 1 \end{cases}$$

$$\frac{PM(r,s_1)}{PM(r,s_2)} = \frac{p}{1-p} \exp \left[ \frac{(r + \sqrt{\epsilon_b})^2 - (r - \sqrt{\epsilon_b})^2}{2\sigma_n^2} \right]$$

$$\frac{(r + \sqrt{\epsilon_b})^2 - (r - \sqrt{\epsilon_b})^2}{2\sigma_n^2} \begin{cases} s_1 & \geq \ln \frac{1-p}{p} \\ s_2 & < \ln \frac{1-p}{p} \end{cases}$$

$$\sqrt{\epsilon_b} r \begin{cases} s_1 & \geq \frac{1}{2\sigma_n^2} \ln \frac{1-p}{p} = \frac{1}{4} N_0 \ln \frac{1-p}{p} \\ s_2 & < \frac{1}{2\sigma_n^2} \ln \frac{1-p}{p} \end{cases}$$
The Optimum Detector

Example: (cont.)

- The threshold is \( \frac{1}{4} N_0 \ln \frac{1-p}{p} \), denoted by \( \tau_h \), divides the real line into two regions, say \( R_1 \) and \( R_2 \), where \( R_1 \) consists of the set of points that are greater than \( \tau_h \) and \( R_2 \) consists of the set of points that are less than \( \tau_h \).

- If \( r \sqrt{\mathcal{E}_b} > \tau_h \), the decision is made that \( s_1 \) was transmitted.

- If \( r \sqrt{\mathcal{E}_b} < \tau_h \), the decision is made that \( s_2 \) was transmitted.
The Optimum Detector

Example: (cont.)

- The threshold $\tau_h$ depends on $N_0$ and $p$. If $p=1/2$, $\tau_h=0$.
- If $p>1/2$, the signal point $s_1$ is more probable and, hence, $\tau_h<0$. In this case, the region $R_1$ is larger than $R_2$, so that $s_1$ is more likely to be selected than $s_2$.
- The average probability of error is minimized.
- It is interesting to note that in the case of unequal priori probabilities, it is necessary to know not only the values of the priori probabilities but also the value of the power spectral density $N_0$, or equivalently, the noise-to-signal ratio, in order to compute the threshold.
- When $p=1/2$, the threshold is zero, and knowledge of $N_0$ is not required by the detector.
A compact and meaningful comparison of modulation methods is one based on the normalized data rate $R/W$ (bits per second per hertz of bandwidth) versus the SNR per bit ($\varepsilon_b/N_0$) required to achieve a given error probability.

In the case of PAM, QAM, and PSK, increasing $M$ results in a higher bit-to-bandwidth ratio $R/W$. 