Optimal Receivers for the AWGN Channel

Transmitter Model:

Consider a generic M-ary communication system where signal $s_m(t)$ is used to convey symbol $m \in \mathcal{M}$, where $\mathcal{M} = \{0, 1, \ldots, M-1\}$ is the symbol alphabet. The set of signals, $\{s_m(t) \mid m \in \mathcal{M}\}$, can be represented with K orthonormal basis signals $\{\phi_k(t) \mid k = 0, 1, \ldots, K-1\}$, with

$$s_m(t) = \sum_{k=0}^{K-1} s_{m,k} \phi_k(t)$$

where the weights are

$$s_{m,k} = \langle s_m(t), \phi_k(t) \rangle = \int_0^T s_m(t) \phi_k(t)\, dt$$

for all $m \in \mathcal{M}$ and $k \in \{0, 1, \ldots, K-1\}$.

Note: The Gram-Schmidt procedure not only describes how to find the basis signals and the weights, but also proves that a set of basis signals exists for any set of finite-energy data signals.

Additive White Gaussian Noise (AWGN) Channel Model:

Suppose symbol $m \in \mathcal{M}$ was transmitted. The received signal is represented by

$$r_c(t) = s_m(t) + w_c(t)$$

where

$s_m(t) =$ transmitted data signal

$w_c(t) =$ additive white Gaussian noise signal

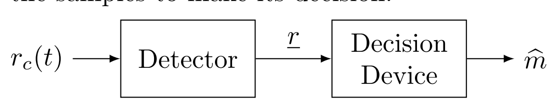
- stationary random process
- Gaussian distribution
- zero mean $\Rightarrow$ $\mu_w = \mathbf{E}[w_c(t)] = 0$
- white noise $\Rightarrow$ $\phi_w(\tau) = \mathbf{E}[w_c(t) w_c(t+\tau)] = \frac{\mathcal{N}_0}{2} \delta(\tau)$

Receiver:

The purpose of the receiver is to determine the transmitted symbol, m, based on observations of $r_c(t)$. Because of uncertainty introduced by the noise it is impossible to guarantee that the receiver will be able to correctly determine the transmitted symbol.

Optimal Receiver: An *optimal receiver* is one that is designed to minimize the probability that a decision error occurs. There exists no other receiver structure that can provide a lower probability of error.

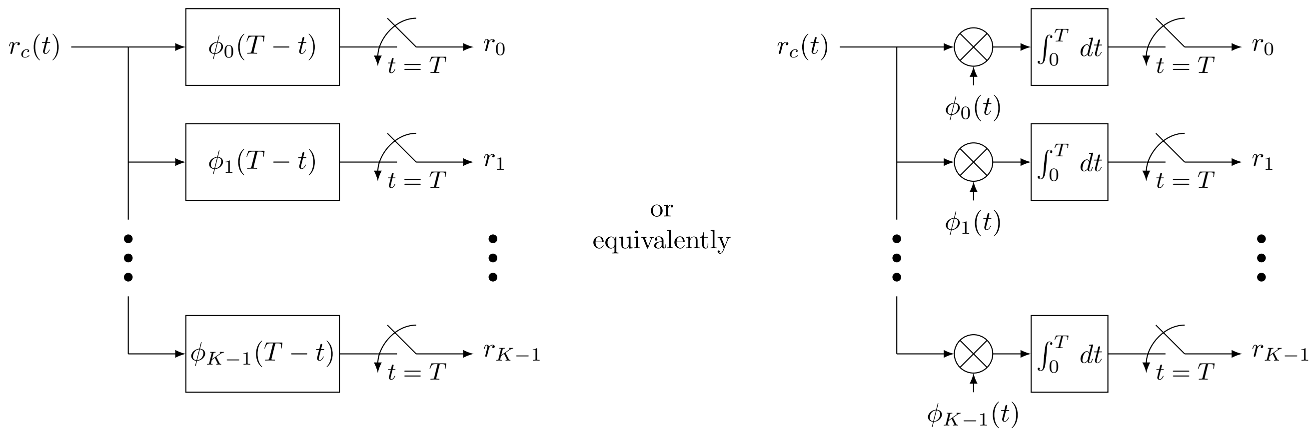
The optimal receiver can be separated into two stages, a detector, which filters and samples the received signal, and a decision device, which uses the samples to make its decision.



Detector – extracts a set of "sufficient statistics" from $r_c(t)$.
Decision Device – attempts to determine the transmitted symbol, m, based on $\underline{r} = [r_0\ r_1\ \cdots\ r_{K-1}]$.

Optimal Detector:

- filters and samples the received signal.
- optimal detector is composed of a bank of K matched filters (or correlators).
- filters are matched to the basis signals, $\{\phi_k(t) \mid k = 0, 1, \ldots, K-1\}$.



For $k \in \{0, 1, \ldots, K-1\}$, the received samples are

$$\begin{aligned} r_k &= \int_0^T r_c(t)\phi_k(t)\,dt \qquad\qquad [\text{but } r_c(t) = s_m(t) + w_c(t)] \\ &= \int_0^T s_m(t)\phi_k(t)\,dt + \int_0^T w_c(t)\phi_k(t)\,dt \\ &= s_{m,k} + w_k\ , \end{aligned}$$

where

$$w_k = \int_0^T w_c(t)\phi_k(t)\,dt$$

represents a sample of noise, and $\{s_{m,k}\}$ are the weights for signal $s_m(t)$. That is,

$$s_{m,k} = \langle s_m(t), \phi_k(t)\rangle = \int_0^T s_m(t)\phi_k(t)\,dt$$

The samples $\underline{r}$ represent the projection of the received signal onto the signal space defined by $\{\phi_k(t) | k = 0, 1, \ldots, K-1\}$.

Properties of w_k:

- Since $w_c(t)$ is a Gaussian random process, w_k is a Gaussian random variable.
- Mean:

$$\mathbf{E}[w_k] = \mathbf{E}\left[\int_0^T w_c(t)\phi_k(t)\,dt\right] = \int_0^T \mathbf{E}[w_c(t)]\,\phi_k(t)\,dt = 0\ .$$

- Covariance:

$$\begin{aligned} \mathbf{E}[w_k w_l] &= \mathbf{E}\left[\int_0^T w_c(t_1)\phi_k(t_1)\,dt_1 \int_0^T w_c(t_2)\phi_l(t_2)\,dt_2\right] \\ &= \int_0^T \int_0^T \mathbf{E}[w_c(t_1)w_c(t_2)]\,\phi_k(t_1)\phi_l(t_2)\,dt_1\,dt_2 \\ &= \int_0^T \int_0^T \frac{\mathcal{N}_0}{2}\delta(t_2 - t_1)\phi_k(t_1)\phi_l(t_2)\,dt_1\,dt_2 \\ &= \frac{\mathcal{N}_0}{2}\int_0^T \phi_k(t_2)\phi_l(t_2)\,dt_2 \\ &= \begin{cases} \mathcal{N}_0/2, & \text{if } k = l \\ 0, & \text{if } k \neq l \end{cases} \\ &= \frac{\mathcal{N}_0}{2}\delta_{l-k}\ . \end{aligned}$$

Properties of r_k:

– Mean:

$$\mathbf{E}[r_k] = \mathbf{E}[s_{m,k} + w_k] \;=\; s_{m,k} \;+\; \mathbf{E}[w_k] \;=\; s_{m,k} \;.$$

– Covariance:

$$\begin{aligned}
\mathbf{E}\left[\Big(r_k - \mathbf{E}[r_k]\Big)\Big(r_l - \mathbf{E}[r_l]\Big)\right] &= \mathbf{E}[w_k w_l] \\
&= \frac{\mathcal{N}_0}{2}\delta_{l-k} \\
&= \begin{cases} \mathcal{N}_0/2, & \text{if } k = l \\ 0, & \text{if } k \neq l \end{cases} \;.
\end{aligned}$$

– Distribution: $\{r_k\}$ are a set of independent Gaussian random variables, with $r_k \sim N(s_{m,k}, \mathcal{N}_0/2)$.

Residue:

In general, we cannot perfectly reconstruct $r_c(t)$ from the samples $\underline{r}$, so by sampling the filter outputs, some information has been lost. This lost information is the residual error,

$$r_e(t) = r_c(t) \;-\; \sum_{k=0}^{K-1} r_k \phi_k(t) \;.$$

However, $r_e(t)$ contains no relevant information to help in determining m.

Proof:

$$\begin{aligned}
r_e(t) &= s_m(t) + w_c(t) - \sum_{k=0}^{K-1} [s_{m,k} + w_k]\,\phi_k(t) \\
&= s_m(t) - \sum_{k=0}^{K-1} s_{m,k}\phi_k(t) + w_c(t) - \sum_{k=0}^{K-1} w_k\phi_k(t) \\
&= s_m(t) - s_m(t) + w_c(t) - \sum_{k=0}^{K-1} w_k\phi_k(t) \\
&= w_c(t) - \sum_{k=0}^{K-1} w_k\phi_k(t) \\
&= w_e(t) \;,
\end{aligned}$$

where

$$w_e(t) = w_c(t) - \sum_{k=0}^{K-1} w_k\phi_k(t) \;.$$

Since $w_c(t)$ and w_k are not based on m, $w_e(t)$ has the same value regardless of the transmitted signal. Therefore it will be of no direct assistance in determining m. However, $w_e(t)$ may provide some information about w_k, which could be used indirectly to determine m. But

$$\begin{aligned}
\mathbf{E}[w_k w_e(t)] &= \mathbf{E}\left[w_k\left(w_c(t) - \sum_{l=0}^{K-1} w_l\phi_l(t)\right)\right] \\
&= \mathbf{E}[w_k w_c(t)] - \mathbf{E}\left[w_k \sum_{l=0}^{K-1} w_l\phi_l(t)\right] \\
&= \mathbf{E}\left[\int_0^T w_c(\tau)\phi_k(\tau)\,d\tau\; w_c(t)\right] - \sum_{l=0}^{K-1}\mathbf{E}[w_k w_l]\,\phi_l(t) \\
&= \int_0^T \mathbf{E}[w_c(\tau)w_c(t)]\,\phi_k(\tau)\,d\tau - \sum_{l=0}^{K-1}\frac{\mathcal{N}_0}{2}\delta_{l-k}\phi_l(t) \\
&= \int_0^T \frac{\mathcal{N}_0}{2}\delta(t-\tau)\phi_k(\tau)\,d\tau - \frac{\mathcal{N}_0}{2}\phi_k(t) \\
&= \frac{\mathcal{N}_0}{2}\phi_k(t) - \frac{\mathcal{N}_0}{2}\phi_k(t) \\
&= 0 \;.
\end{aligned}$$

- Therefore, $w_e(t)$ and w_k are uncorrelated for all $t \in [0,T]$ and $k \in \{0, 1, \ldots, K-1\}$.
- Therefore, $w_e(t)$ is independent of w_k for all $t \in [0,T]$ and $k \in \{0, 1, \ldots, K-1\}$.
- Therefore, $w_e(t)$ contains no information about w_k.
- Therefore, knowledge of $w_e(t)$ is of no assistance in determining m.
- The samples $\underline{r}$ are a set of sufficient statistics for determining m. There is no additional information in $r_c(t)$ that is relevant.

Optimal Decision Device

The decision device must make a decision about which symbol was transmitted based on the received observations, $\underline{r}$. An optimal decision device is one that makes this decision in such a manner that the probability of a symbol error is minimized. Let $\widehat{m}$ be the decision made by the device.

Def^n: The *a priori probability distribution* is the probability distribution of the transmitted symbols before any data has been received. It is denoted by $\Pr\{m \text{ sent}\}$. Typically, each symbol is equally likely to have been transmitted, so $\Pr\{m \text{ sent}\} = 1/M$.

Def^n: The *a posteriori probability distribution* (APP) is the probability distribution of the transmitted symbols after the received signal has been observed. It is denoted by $\Pr\{m \text{ sent} \,|\, \underline{r} \text{ received}\}$.

Def^n: The conditional probability density function $f_{\underline{r}}(\underline{r} \,|\, m \text{ sent})$ is the pdf of observing $\underline{r}$ at the output of the detector, given that symbol m was transmitted. This is referred to as the *likelihood function*.

Maximum A Posteriori Probability (MAP) Decision Rule

To minimize the probability of an error, the decision device must maximize the probability that its decision is correct. It chooses $\widehat{m} = m$, for the value of m with the largest *a posteriori* probability. That is, choose $\widehat{m} = m$ if

$$\Pr\{m \text{ sent} \,|\, \underline{r} \text{ received}\} \;\geq\; \Pr\{l \text{ sent} \,|\, \underline{r} \text{ received}\} \;\text{ for all } l \neq m \;,$$

or equivalently

$$\widehat{m} = \arg\max_m \Pr\{m \text{ sent} \,|\, \underline{r} \text{ received}\} \;.$$

This is known as the *maximum a posteriori probability* (MAP) decision rule.

Example: Consider a system where one of $M = 4$ possible values could have been transmitted. Suppose, based on the received signal, the receiver calculates the following APP's

$$\begin{aligned}
\Pr\{0 \text{ sent} \,|\, \underline{r} \text{ received}\} &= 0.2\\
\Pr\{1 \text{ sent} \,|\, \underline{r} \text{ received}\} &= 0.1\\
\Pr\{2 \text{ sent} \,|\, \underline{r} \text{ received}\} &= 0.4\\
\Pr\{3 \text{ sent} \,|\, \underline{r} \text{ received}\} &= 0.3
\end{aligned}$$

According to the MAP decision rule, the receiver would chose $\widehat{m} = 2$, since it is most likely to have been transmitted based on the observations of the received signal.

Note: The probability of error in this case is 0.6, but any other choice for $\widehat{m}$ would lead to a higher probability of error.

The APP's can be calculated from the *likelihood function*, $f_{\underline{r}}(\underline{r} \,|\, m \text{ sent})$, with

$$\Pr\{m \text{ sent} \,|\, \underline{r} \text{ received}\} = \frac{f_{\underline{r}}(\underline{r} \,|\, m \text{ sent})\Pr\{m \text{ sent}\}}{f_{\underline{r}}(\underline{r})} = \frac{f_{\underline{r}}(\underline{r} \,|\, m \text{ sent}) \Pr\{m \text{ sent}\}}{\sum_{m'=0}^{M-1} f_{\underline{r}}(\underline{r} \,|\, m' \text{ sent}) \Pr\{m' \text{ sent}\}}$$

For the AWGN channel, since the components of $\underline{r} = [r_0\, r_1\, \cdots\, r_{K-1}]$ are independent, and each r_k has a Gaussian distribution with a mean of $s_{m,k}$ and a variance of $\mathcal{N}_0/2$, the likelihood function is

$$\begin{aligned}
f_{\underline{r}}(\underline{r} \,|\, m \text{ sent}) &= \prod_{k=0}^{K-1} f_{r_k}(r_k | m \text{ sent})\\
&= \prod_{k=0}^{K-1} \frac{1}{\sqrt{2\pi(\mathcal{N}_0/2)}} \exp\left\{-\frac{(r_k - s_{m,k})^2}{2(\mathcal{N}_0/2)}\right\}\\
&= \frac{1}{(\sqrt{\pi\mathcal{N}_0})^K} \exp\left\{-\frac{1}{\mathcal{N}_0}\sum_{k=0}^{K-1}(r_k - s_{m,k})^2\right\} \;.
\end{aligned}$$

Maximum Likelihood (ML) Decision Rule

Under certain conditions, the MAP decision rule can be simplified. Usually, all the symbols are equally likely to be transmitted, so the *a priori* probabilities $\Pr\{m \text{ sent}\} = 1/M$, so the MAP decision rule can be expressed as:

$$\widehat{m} = \arg\max_m \frac{f_{\underline{r}}(\underline{r} \mid m \text{ sent})1/M}{f_{\underline{r}}(\underline{r})}$$

or

$$\widehat{m} = \arg\max_m f_{\underline{r}}(\underline{r} \mid m \text{ sent}) \ .$$

This is know as the maximum likelihood (ML) decision rule. Note that the ML decision rule is equivalent to the MAP decision rule if the *a priori* probabilities are all equal.

Simplifications to the ML Decision Rule

Using the expression given above for the likelihood function, the ML decision rule can then be expressed as:

$$\widehat{m} = \arg\max_m \frac{1}{(\sqrt{\pi \mathcal{N}_0})^K} \exp\left\{-\frac{1}{\mathcal{N}_0}\sum_{k=0}^{K-1}(r_k - s_{m,k})^2\right\}$$

or

$$\widehat{m} = \arg\max_m \exp\left\{-\frac{1}{\mathcal{N}_0}\sum_{k=0}^{K-1}(r_k - s_{m,k})^2\right\}$$

or (by taking the log)

$$\widehat{m} = \arg\max_m \left(-\frac{1}{\mathcal{N}_0}\sum_{k=0}^{K-1}(r_k - s_{m,k})^2\right)$$

or

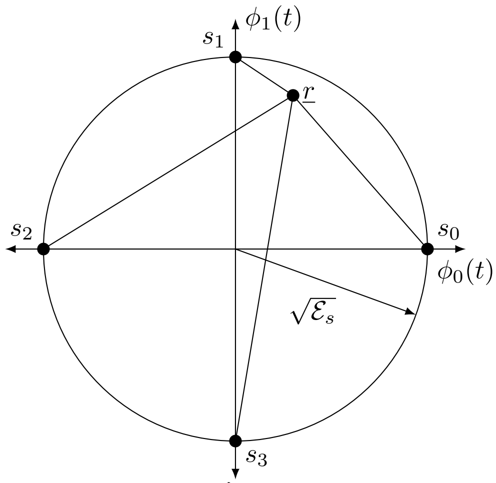
$$\widehat{m} = \arg\min_m \sum_{k=0}^{K-1}(r_k - s_{m,k})^2 \ .$$

But, $\sum_{k=0}^{K-1}(r_k - s_{m,k})^2 = \left\|\underline{r} - \underline{s}_m\right\|^2$, the square of the distance between $\underline{r}$ and the point in the signal space diagram corresponding to $s_m(t)$. Therefore the ML decision rule reduces to:

$$\widehat{m} = \arg\min_m \left\|\underline{r} - \underline{s}_m\right\| \ .$$

In other words, the optimal decision is that symbol that is "closest" to $\underline{r}$ in the signal space.

<u>Example:</u> Suppose the $M = 4$ two dimensional signal constellation shown below is used for transmission, and that $\underline{r} = [0.3,\ 0.8]\sqrt{\mathcal{E}_s}$ is observed. The observation is marked in the signal space diagram shown below:

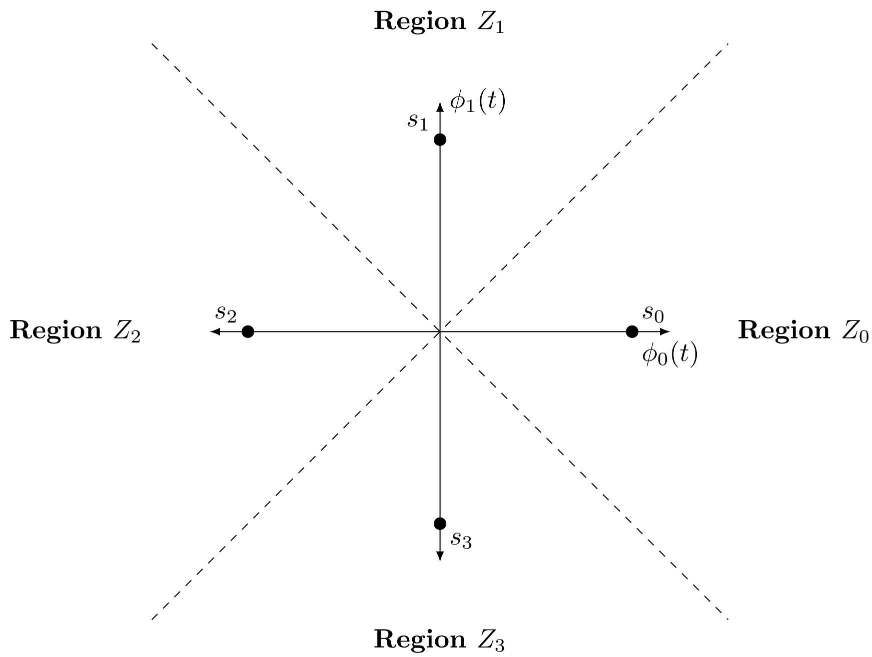


$$\left\|\underline{r} - \underline{s}_0\right\| = \sqrt{1.13\mathcal{E}_s}$$
$$\left\|\underline{r} - \underline{s}_1\right\| = \sqrt{0.13\mathcal{E}_s}$$
$$\left\|\underline{r} - \underline{s}_2\right\| = \sqrt{2.33\mathcal{E}_s}$$
$$\left\|\underline{r} - \underline{s}_3\right\| = \sqrt{3.33\mathcal{E}_s}$$

The decoder would choose $\widehat{m} = 1$.

Decision Regions:

Each possible received observation will be closest to one of the points in the signal constellation. For each signalling scheme, it is useful to draw the boundaries of the decision regions on the signal space diagram



Further Simplifications to the ML Decision Rule

The ML decision rule can also be expressed as:

$$\widehat{m} = \arg\min_m \sum_{k=0}^{K-1} (r_k^2 - 2r_k s_{m,k} + s_{m,k}^2)$$

or

$$\widehat{m} = \arg\min_m \left(\sum_{k=0}^{K-1} r_k^2 - 2 \sum_{k=0}^{K-1} r_k s_{m,k} + \sum_{k=0}^{K-1} s_{m,k}^2 \right)$$

or

$$\widehat{m} = \arg\min_m \left(-2 \sum_{k=0}^{K-1} r_k s_{m,k} + E_m \right)$$

or

$$\widehat{m} = \arg\max_m \left(\sum_{k=0}^{K-1} r_k s_{m,k} - E_m/2 \right) .$$

If all signals have equal energy (i.e., $E_m = E_l \ \forall \ m, l$), then the ML decision rule can be expressed as:

$$\widehat{m} = \arg\max_m \sum_{k=0}^{K-1} r_k s_{m,k} .$$