

EE6617: 08 Statistical Decision Theory II

Major references:

- *Fundamentals of Statistical Processing, Volume II: Detection Theory*, 1/E, by Steven M. Kay, Prentice Hall, 1993
- *Statistical Signal Processing*, 1/E, by Louis L. Scharf, Prentice Hall, 1991

Last class we consider detecting a **signal in i.i.d $\mathcal{N}(0, \sigma^2)$ noise**.

What happens when we **remove** the constraint that $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$?

1 Generalized Matched Filter

Consider $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ where $\mathbf{C}$ is the covariance matrix and $\mathbf{C}$ is positive definite. Thus, $\mathbf{C}$ is symmetry and invertible.

Detect if a predetermined signal $s[0], \dots, s[N-1]$ is present in a noisy received signal $x[0], \dots, x[N-1]$.

$$\begin{aligned}\mathcal{H}_0 : \quad & \mathbf{x} = \mathbf{w} \\ \mathcal{H}_1 : \quad & \mathbf{x} = \mathbf{s} + \mathbf{w}.\end{aligned}$$

and $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$.

We already know the **NP detector is optimal**. To design NP detector, **formulate the likelihood test**. The PDFs under each hypothesis are

$$\begin{aligned}p(\mathbf{x}; \mathcal{H}_0) &= \frac{1}{(2\pi)^{N/2} \sqrt{\det(\mathbf{C})}} \exp\left(-\frac{1}{2} \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}\right) \\ p(\mathbf{x}; \mathcal{H}_1) &= \frac{1}{(2\pi)^{N/2} \sqrt{\det(\mathbf{C})}} \exp\left(-\frac{1}{2} (\mathbf{x} - \mathbf{s})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{s})\right).\end{aligned}$$

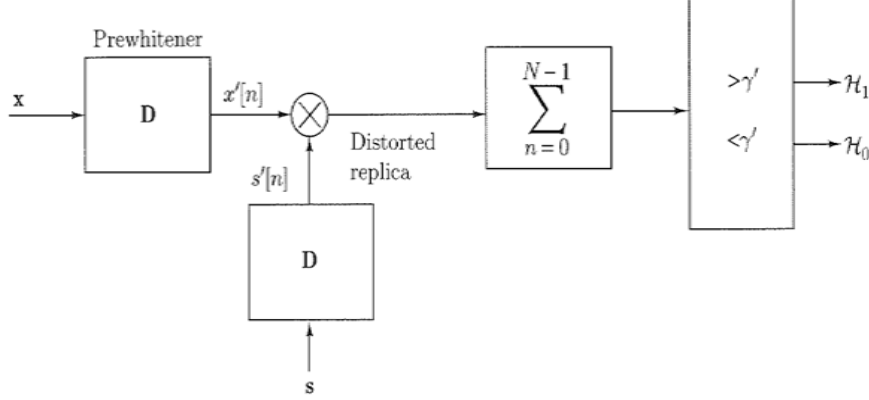
with

$$\begin{aligned}L(\mathbf{x}) &= \frac{p(\mathbf{x}; \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} \\ &= \exp\left(-\frac{1}{2} \left((\mathbf{x} - \mathbf{s})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{s}) - \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}\right)\right) > \gamma\end{aligned}$$

Since $\left((\mathbf{x} - \mathbf{s})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{s}) - \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}\right) = \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x} - 2\mathbf{x}^T \mathbf{C}^{-1} \mathbf{s} + \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s} - \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}$, taking log and moving terms return

$$T(\mathbf{x}) = \mathbf{x}^T \mathbf{C}^{-1} \mathbf{s} \underset{\gamma'}{\geqslant_0^1} \ln(\gamma) + \underbrace{\frac{1}{2} \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}}_{\gamma'}.$$

The test statistic can be rewritten by $T(\mathbf{x}) = \mathbf{x}^T \mathbf{C}^{-1} \mathbf{s} = \mathbf{x}'^T \mathbf{s}'$ where $\mathbf{x}' = \mathbf{D}\mathbf{x}$, $\mathbf{s}' = \mathbf{D}\mathbf{s}$, and $\mathbf{C}^{-1} = \mathbf{D}^T \mathbf{D}$. The filter $\mathbf{D}$ is the whitening filter.



1.1 Performance of Generalized Matched Filter

Since $\mathbf{x}$ is Gaussian under both hypotheses, we know $T(\mathbf{x})$ is Gaussian.

Thus, under $\mathcal{H}_0$, $\mathbf{x} = \mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ and

$$\begin{aligned} E[T(\mathbf{x})] &= E[\mathbf{x}^T \mathbf{C}^{-1} \mathbf{s}] = E[\mathbf{w}^T] \mathbf{C}^{-1} \mathbf{s} = 0 \\ \text{var}(T(\mathbf{x})) &= E[T(\mathbf{x})^2] = E[\mathbf{s}^T \mathbf{C}^{-1} \mathbf{w} \mathbf{w}^T \mathbf{C}^{-1} \mathbf{s}] = \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s} \\ T(\mathbf{x}) &\sim \mathcal{N}(0, \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}). \end{aligned}$$

Under $\mathcal{H}_1$, $\mathbf{x} = \mathbf{s} + \mathbf{w} \sim \mathcal{N}(\mathbf{s}, \mathbf{C})$ and

$$\begin{aligned} E[T(\mathbf{x})] &= E[\mathbf{x}^T] \mathbf{C}^{-1} \mathbf{s} = \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s} \\ \text{var}(T(\mathbf{x})) &= E[(\mathbf{x}^T \mathbf{C}^{-1} \mathbf{s})^2] - (\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s})^2 \\ &= \mathbf{s}^T \mathbf{C}^{-1} E[\mathbf{x} \mathbf{x}^T] \mathbf{C}^{-1} \mathbf{s} - (\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s})^2 \\ &\quad \text{where } E[\mathbf{x} \mathbf{x}^T] = \mathbf{s} \mathbf{s}^T + \mathbf{C} \\ &= \mathbf{s}^T \mathbf{C}^{-1} (\mathbf{s} \mathbf{s}^T + \mathbf{C}) \mathbf{C}^{-1} \mathbf{s} - (\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s})^2 \\ &= (\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s})^2 + \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s} - (\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s})^2 \\ &= \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s} \\ T(\mathbf{x}) &\sim \mathcal{N}(\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}, \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}). \end{aligned}$$

Thus, the variance is the same as $\mathcal{H}_0$.

We have

$$\begin{aligned} \mathcal{H}_0 : T(\mathbf{x}) &\sim \mathcal{N}(0, \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}) \\ \mathcal{H}_1 : T(\mathbf{x}) &\sim \mathcal{N}(\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}, \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}). \end{aligned}$$

For the target false alarm rate $P_{FA} = \alpha$, the threshold γ' can be determined such that

$$\begin{aligned} P_{FA} &= \int_{\gamma'}^{\infty} \frac{1}{\sqrt{2\pi \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}}} \exp\left(-\frac{1}{2\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}} t^2\right) dt \\ &= \int_{\frac{\gamma'}{\sqrt{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}}}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du \\ &= Q\left(\frac{\gamma'}{\sqrt{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}}}\right) = \alpha \end{aligned}$$

and $\gamma' = \sqrt{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}} Q^{-1}(\alpha)$.

We can characterize the detection performance

$$\begin{aligned} P_D &= \int_{\gamma'}^{\infty} \frac{1}{\sqrt{2\pi \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}}} \exp\left(-\frac{1}{2\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}} (t - \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s})^2\right) dt \\ &= Q\left(\frac{\gamma' - \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}}{\sqrt{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}}}\right) = Q\left(Q^{-1}(\alpha) - \sqrt{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}}\right). \end{aligned}$$

This P_D is the maximum detection probability given $P_{FA} = \alpha$ by the Neyman-Pearson theorem.

Q: What if we want to choose $\mathbf{s}$ to maximize the detector performance subject to a fixed $\|\mathbf{s}\|^2 = \xi$?

We will see this soon.

2 Multiple Hypothesis in IID Noise

Suppose we have N samples with M -ary hypotheses

$$\begin{aligned} \mathcal{H}_0 : \quad & \mathbf{x} = \mathbf{s}_0 + \mathbf{w} \\ & \vdots \\ \mathcal{H}_{M-1} : \quad & \mathbf{x} = \mathbf{s}_{M-1} + \mathbf{w}. \end{aligned}$$

and $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ and

$$p(\mathbf{x}|\mathcal{H}_i) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{s}_i\|^2\right)$$

Assume a **Bayesian framework**

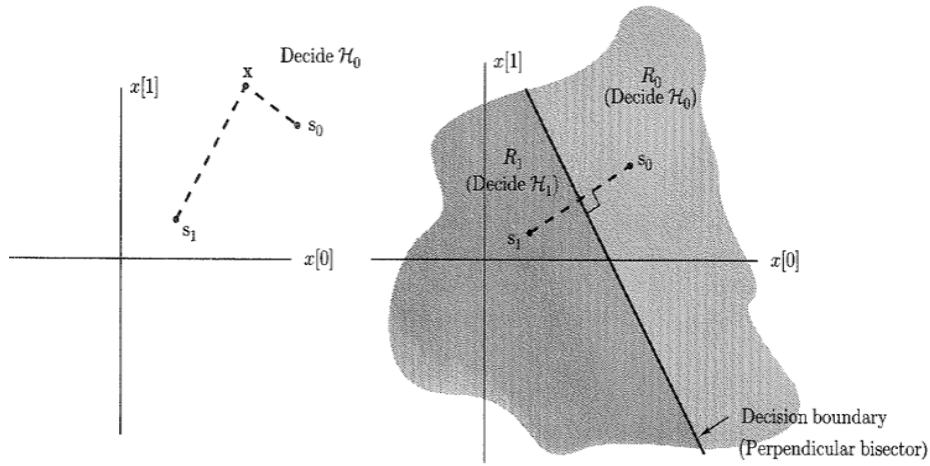
- (i) **Probability of error risk (MAP)**
- (ii) **Equally likely $\mathbf{s}_i$ (ML)**

In this case, the optimal detector is the ML detector

$$\begin{aligned}
i &= \operatorname{argmax}_{0 \leq \ell \leq M-1} p(\mathbf{x} | \mathcal{H}_\ell) \\
&= \operatorname{argmax}_{0 \leq \ell \leq M-1} \exp \left(-\frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{s}_\ell\|^2 \right) \\
&= \operatorname{argmin}_{0 \leq \ell \leq M-1} \|\mathbf{x} - \mathbf{s}_\ell\|^2.
\end{aligned}$$

This is often called a **minimum distance receiver**.

The detector decides the hypothesis whose signal minimizes Euclidean distance from the received vector.



Note that if we set $\|\mathbf{s}_\ell\|^2 = \xi_\ell$, the detector is

$$\begin{aligned}
i &= \operatorname{argmin}_{0 \leq \ell \leq M-1} \|\mathbf{x} - \mathbf{s}_\ell\|^2 \\
&= \operatorname{argmin}_{0 \leq \ell \leq M-1} (\|\mathbf{x}\|^2 - 2\mathbf{x}^T \mathbf{s}_\ell + \xi_\ell) \\
&= \operatorname{argmax}_{0 \leq \ell \leq M-1} \mathbf{x}^T \mathbf{s}_\ell - \frac{\xi_\ell}{2}.
\end{aligned}$$

If $\xi_0 = \xi_1 = \dots = \xi_{M-1}$, this is a correlated receiver.

2.1 Binary Error Probability

Consider the **binary scenario**. Assume $p(\mathcal{H}_0) = p(\mathcal{H}_1) = 1/2$.

$$P_e = \frac{1}{2}p(\mathcal{H}_1 | \mathcal{H}_0) + \frac{1}{2}p(\mathcal{H}_0 | \mathcal{H}_1).$$

Detector that minimize P_e is

$$T_i(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]s_i[n] - \frac{\xi_i}{2}$$

and

$$T_1(\mathbf{x}) \geq_0^1 T_0(\mathbf{x}) \Rightarrow T(\mathbf{x}) = T_1(\mathbf{x}) - T_0(\mathbf{x}) \geq_0^1 0.$$

$$\begin{aligned}\mathcal{H}_0 : \quad \mathbf{x} &= \mathbf{s}_0 + \mathbf{w} \\ \mathcal{H}_1 : \quad \mathbf{x} &= \mathbf{s}_1 + \mathbf{w}\end{aligned}$$

and $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$.

Under the hypothesis $\mathcal{H}_0$

$$\begin{aligned}E[T(\mathbf{x})] &= E[\mathbf{x}^T(\mathbf{s}_1 - \mathbf{s}_0)] - \frac{\xi_1 - \xi_0}{2} \\ &= E[\mathbf{x}^T](\mathbf{s}_1 - \mathbf{s}_0) - \frac{\xi_1 - \xi_0}{2} \\ &= \mathbf{s}_0^T \mathbf{s}_1 - \|\mathbf{s}_0\|^2 - \frac{\xi_1 - \xi_0}{2} \\ &= -\frac{1}{2}(\xi_1 - 2\mathbf{s}_0^T \mathbf{s}_1 + \xi_0) \\ &= -\frac{1}{2}\|\mathbf{s}_0 - \mathbf{s}_1\|^2\end{aligned}$$

$$\begin{aligned}\text{var}(T(\mathbf{x})) &= E\left[\left(\mathbf{x}^T(\mathbf{s}_1 - \mathbf{s}_0) - \frac{\xi_1 - \xi_0}{2}\right)^2\right] - \left(E[\mathbf{x}^T(\mathbf{s}_1 - \mathbf{s}_0)] - \frac{\xi_1 - \xi_0}{2}\right)^2 \\ &= (\mathbf{s}_1 - \mathbf{s}_0)^T E[\mathbf{x}\mathbf{x}^T](\mathbf{s}_1 - \mathbf{s}_0) - (\mathbf{s}_1 - \mathbf{s}_0)^T E[\mathbf{x}] E[\mathbf{x}]^T (\mathbf{s}_1 - \mathbf{s}_0) \\ &= \sigma^2 \|\mathbf{s}_1 - \mathbf{s}_0\|^2\end{aligned}$$

Under the hypothesis $\mathcal{H}_1$

$$\begin{aligned}E[T(\mathbf{x})] &= E[\mathbf{x}^T(\mathbf{s}_1 - \mathbf{s}_0)] - \frac{\xi_1 - \xi_0}{2} \\ &= \mathbf{s}_1^T \mathbf{s}_1 - \|\mathbf{s}_0\|^2 - \frac{\xi_1 - \xi_0}{2} \\ &= \frac{1}{2}(\xi_1 - 2\mathbf{s}_0^T \mathbf{s}_1 + \xi_0) \\ &= \frac{1}{2}\|\mathbf{s}_0 - \mathbf{s}_1\|^2\end{aligned}$$

$$\text{var}(T(\mathbf{x})) = \sigma^2 \|\mathbf{s}_1 - \mathbf{s}_0\|^2$$

To calculate the detection error probability,

$$\begin{aligned}
p(\mathcal{H}_1|\mathcal{H}_0) &= \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2 \|\mathbf{s}_1 - \mathbf{s}_0\|^2}} \exp\left(-\frac{1}{2\sigma^2 \|\mathbf{s}_1 - \mathbf{s}_0\|^2} \left(t + \frac{1}{2} \|\mathbf{s}_1 - \mathbf{s}_0\|^2\right)^2\right) dt \\
&= \int_{\frac{\frac{1}{2} \|\mathbf{s}_1 - \mathbf{s}_0\|^2}{\sqrt{\sigma^2 \|\mathbf{s}_1 - \mathbf{s}_0\|^2}}}^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du \\
&= Q\left(\sqrt{\frac{\|\mathbf{s}_1 - \mathbf{s}_0\|^2}{4\sigma^2}}\right)
\end{aligned}$$

and

$$\begin{aligned}
p(\mathcal{H}_0|\mathcal{H}_1) &= \int_{-\infty}^0 \frac{1}{\sqrt{2\pi\sigma^2 \|\mathbf{s}_1 - \mathbf{s}_0\|^2}} \exp\left(-\frac{1}{2\sigma^2 \|\mathbf{s}_1 - \mathbf{s}_0\|^2} \left(t - \frac{1}{2} \|\mathbf{s}_1 - \mathbf{s}_0\|^2\right)^2\right) dt \\
&= Q\left(\sqrt{\frac{\|\mathbf{s}_1 - \mathbf{s}_0\|^2}{4\sigma^2}}\right).
\end{aligned}$$

By $P_e = \frac{1}{2}p(\mathcal{H}_1|\mathcal{H}_0) + \frac{1}{2}p(\mathcal{H}_0|\mathcal{H}_1)$,

$$P_e = Q\left(\sqrt{\frac{\|\mathbf{s}_1 - \mathbf{s}_0\|^2}{4\sigma^2}}\right)$$

Q: Given $\xi_0 = \xi_1 = \xi$, what are the optimal $\mathbf{s}_1$ and $\mathbf{s}_0$ that minimize P_e ?

3 Implication of Channel Capacity

Consider M equally likely hypothesis

$$\mathcal{H}_i : \mathbf{x} = \mathbf{s}_i + \mathbf{w}, \quad \mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}), \quad i = 0, \dots, M-1$$

and $\mathbf{x}, \mathbf{s}_i, \mathbf{w} \in \mathbb{R}^N$

$$\begin{aligned}
\|\mathbf{s}_i\|^2 &= \xi = N\xi_d \\
|\mathbf{s}_i^T \mathbf{s}_j| &< \Delta \leq \xi, \quad \forall i \neq j.
\end{aligned}$$

We know the **optimal P_e detector is ML**. Thus, we decide $\mathcal{H}_i$ such that

$$i = \underset{0 \leq \ell \leq M-1}{\operatorname{argmax}} \underbrace{\frac{1}{N} \mathbf{x}^T \mathbf{s}_\ell}_{T_\ell(\mathbf{x})}.$$

Under $\mathcal{H}_i$,

$$\begin{aligned} T_i(\mathbf{x}) &\sim \mathcal{N}\left(\xi_d, \frac{\sigma^2}{N}\xi_d\right) \\ T_j(\mathbf{x}) &\sim \mathcal{N}\left(\frac{1}{N}\mathbf{s}_i^T \mathbf{s}_j, \frac{\sigma^2}{N}\xi_d\right), \quad i \neq j. \end{aligned}$$

Notice that $T_i(\mathbf{x}) \xrightarrow{N \rightarrow \infty} \xi_d$ and $T_j(\mathbf{x}) \xrightarrow{N \rightarrow \infty} 0$, $i \neq j$.

$$\Rightarrow p(\mathcal{H}_j|\mathcal{H}_i) \xrightarrow{N \rightarrow \infty} \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Thus,

$$P_e = \frac{1}{M} \sum_{i=0}^{M-1} P\left(\bigcup_{j \neq i} \{\mathcal{H}_j\} \middle| \mathcal{H}_i\right) \xrightarrow{N \rightarrow \infty} 0.$$

Q: Can M increase as N increases, while still having $P_e \rightarrow 0$ as $N \rightarrow \infty$?

Answer: Yes, $P_e \rightarrow 0$ as $N \rightarrow \infty$ iff

$$M \leq \left(1 + \frac{\xi_d}{\sigma^2}\right)^{N/2} = M_{max}.$$

$\Rightarrow \frac{\log_2(M_{max})}{N}$ is the **maximum number of bits to send during N channel uses (in N symbols) without any detection error**. This is called the **channel capacity per sample**.

$$\begin{aligned} \frac{\log_2\left(\left(1 + \frac{\xi_d}{\sigma^2}\right)^{N/2}\right)}{N} &= \frac{\frac{N}{2} \log_2\left(1 + \frac{\xi_d}{\sigma^2}\right)}{N} \\ &= \frac{1}{2} \log_2\left(1 + \frac{\xi_d}{\sigma^2}\right) = \underbrace{\frac{1}{2} \log_2(1 + SNR)}_{\text{channel capacity}}. \end{aligned}$$

In many ways detection/estimation & information theory are linked.