Lecture-11: Detection of Deterministic Signals in Gaussian Noise

The basic problem that we are dealing is the hypothesis testing problem:

$$H_0 : Y_k = N_k + S_{0k}, \quad k = 1, 2, 3, ...n$$

versus

$$H_1 : Y_k = N_k + S_{1k}, \quad k = 1, 2, 3, ...n$$

(1)

In the previous lectures, we have seen the cases where noise samples N_k were i.i.d. If the noise samples N_k in (1) are not independent of each other, then the likelihood ratio

$$L\left(\underline{y}\right) = \frac{E\{p_N(\underline{y} - \underline{S}_1)\}}{E\{p_N(\underline{y} - \underline{S}_0)\}} \quad \underline{y} \in \mathbb{R}^n$$

(2)

do not exhibit any particular structure, even for the case of deterministic signals. An important exception for this lack of structure is the situation in which the signals are deterministic $(\underline{S}_j = \underline{s}_j)$ and the noise vector $\underline{N}$ has multivariate Gaussian distribution. In this case, the optimum detectors have simple, easily implemented structures and performance of the optimum systems can be analyzed thoroughly. Thus it is of interest to consider this case in detail.

Let $\underline{S}_j$ for $j = \{0, 1\}$, take known values say, $\underline{s}_0, \underline{s}_1 \in \mathbb{R}^n$ and that noise vector $\underline{N}$ is a Gaussian random vector with mean vector $\underline{0}$ and covariance matrix $\sum_N$. That is,

$$\underline{N} \sim \mathcal{N}\left(\underline{0}, \Sigma_N\right)$$

(3)

The assumption of zero mean noise does not reduce the generality of the results since we can always subtract a non zero noise mean from $\underline{y}$ to produce a new observation with zero mean noise.

A Gaussian random vector in $\mathbb{R}^n$ with mean vector $\underline{\mu} \triangleq \mathbb{E}\{\underline{X}\} \in \mathbb{R}^n$ and $n \times n$ covariance matrix $\Sigma \triangleq \mathbb{E}\{(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})^T\}$ has probability density function

$$p_{\underline{X}}(\underline{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\{-\frac{1}{2}(\underline{x} - \underline{\mu})^T \Sigma^{-1}(\underline{x} - \underline{\mu})\} \quad \underline{x} \in \mathbb{R}^n$$

(4)

where $|\Sigma|$ denotes the determinant of Σ and Σ^{-1} denotes the inverse of Σ.
We will assume that Σ_N is positive definite. That is, $\underline{x}^T\Sigma_N\underline{x} > 0 \ \forall \ \underline{x} \in \mathbb{R}^n - \{\underline{0}\}$.
Positive definiteness of $\sum_N$ implies that $|\Sigma_N| > 0$ and Σ^{-1} exists.

The received vector $\underline{Y}$ has a Gaussian distribution with $\underline{s}_j$ as the mean and Σ_N as the covariance matrix, that is, $\underline{Y} \sim \mathcal{N}(\underline{s}_j, \Sigma_N), \ j \in \{0, 1\}$. The likelihood ratio for the received $\underline{y}$ can be written as

$$L(\underline{y}) = \frac{p_1(\underline{y})}{p_0(\underline{y})} = \frac{\frac{1}{(2\pi)^{n/2}|\Sigma_N|^{1/2}} \exp\{-\frac{1}{2}(\underline{y} - \underline{s}_1)^T\Sigma_N^{-1}(\underline{y} - \underline{s}_1)\}}{\frac{1}{(2\pi)^{n/2}|\Sigma_N|^{1/2}} \exp\{-\frac{1}{2}(\underline{y} - \underline{s}_0)^T\Sigma_N^{-1}(\underline{y} - \underline{s}_0)\}},$$

$$= \exp\left\{\underline{s}_1^T\Sigma_N^{-1}\underline{y} - \underline{s}_0^T\Sigma_N^{-1}\underline{y} - \frac{1}{2}\underline{s}_1^T\Sigma_N^{-1}\underline{s}_1 + \frac{1}{2}\underline{s}_0^T\Sigma_N^{-1}\underline{s}_0\right\}.$$

Since Σ_N and hence Σ_N^{-1} are symmetric matrices $\underline{s}_j^T\Sigma_N^{-1}\underline{y} = \underline{y}^T\Sigma_N^{-1}\underline{s}_j$. Hence the expression for $L(\underline{y})$ is

$$L(\underline{y}) = \exp\left\{(\underline{s}_1 - \underline{s}_0)^T\Sigma_N^{-1}\left(\underline{y} - \frac{\underline{s}_0 + \underline{s}_1}{2}\right)\right\} \quad \underline{y} \in \mathbb{R}^n. \tag{5}$$

We observe that the problem under consideration is the vector version of the simple location testing problem studied earlier, where the locations μ_0 and μ_1 are replaced with location vectors $\underline{s}_0$ and $\underline{s}_1$ and the noise variance σ^2 is replaced with noise covariance matrix Σ_N.

Taking logarithm on both sides of (5) we get,

$$\log L(\underline{y}) = (\underline{s}_1 - \underline{s}_0)^T\Sigma_N^{-1}\left(\underline{y} - \frac{\underline{s}_0 + \underline{s}_1}{2}\right).$$

Here, $\frac{1}{2}(\underline{s}_1 - \underline{s}_0)^T\Sigma_N^{-1}(\underline{s}_0 + \underline{s}_1)$ does not depend on $\underline{y}$. Hence it can be incorporated into the decision threshold. The optimum test becomes

$$\tilde{\delta}_0(\underline{y}) = \begin{cases} 1 & > \\ \gamma \quad (\underline{s}_1 - \underline{s}_0)^T \sum_N^{-1} \underline{y} \quad = \tau' \\ 0 & < \end{cases} \tag{6}$$

where $\tau' = \log\tau + \frac{1}{2}(\underline{s}_1 - \underline{s}_0)^T\Sigma_N^{-1}(\underline{s}_0 + \underline{s}_1)$. We can write,

$$(\underline{s}_1 - \underline{s}_0)^T\Sigma_N^{-1}\underline{y} = \underline{\tilde{s}}^T\underline{y} = \sum_{k=1}^{n} \tilde{s}_k y_k \tag{7}$$

where $\underline{\tilde{s}} = \Sigma_N^{-1}(\underline{s}_1 - \underline{s}_0)$ is called *derived signal* or *impulse response* of matched filter.

Now, the detector structure is identical to the optimum detector for coherent signals in i.i.d Gaussian noise with the actual signal $\underline{s}$ replaced by the "pseudo signal" $\underline{\tilde{s}}$. Hence the implementation of detector is no more difficult for dependent noise than for independent noise.

1 Performance of the Detector:

Observe that the quantity $T(\underline{Y}) \triangleq (\underline{s}_1 - \underline{s}_0)^T \Sigma_N^{-1} \underline{Y}$ is a linear transformation of the Gaussian random vector $\underline{Y}$. A basic property of the multivariate Gaussian distribution is that, linear transformation of Gaussian vectors are also Gaussian. In this case, since the transformation is to $\mathbb{R}$, $T(\underline{Y})$ is a Gaussian random variable. The distribution of $T(\underline{Y})$ under H_0 and H_1 can be completely characterized by finding its means and variances under the two hypothesis.

Under H_j, the mean of $T(\underline{Y})$ is given by

$$
\begin{aligned}
\mathbb{E}\{T(\underline{Y}|H_j)\} = \mathbb{E}\{\underline{\tilde{s}}^T \underline{Y}|H_j\} &= \underline{\tilde{s}}^T \mathbb{E}\{\underline{Y}|H_j\}, \\
&= \underline{\tilde{s}}^T \mathbb{E}\{\underline{s}_j + \underline{N}|H_j\}, \\
&= \underline{\tilde{s}}^T \mathbb{E}\{\underline{N}|H_j\} + \underline{\tilde{s}}^T \underline{s}_j, \\
&= \underline{\tilde{s}}^T \underline{s}_j \triangleq \tilde{\mu}_j, \quad j \in \{0,1\}.
\end{aligned}
\tag{8}
$$

Similarly, the variance of $T(\underline{Y})$ under H_j is

$$
\begin{aligned}
Var\{T(\underline{Y})|H_j\} &= \mathbb{E}\{(\underline{\tilde{s}}^T \underline{Y} - \underline{\tilde{s}}^T \underline{s}_j)^2|H_j\}, \\
&= \mathbb{E}\{(\underline{\tilde{s}}^T \underline{N})^2\}, \\
&= \mathbb{E}\{\underline{\tilde{s}}^T \underline{N} \underline{N}^T \underline{\tilde{s}}\}, \\
&= \underline{\tilde{s}}^T \mathbb{E}\{\underline{N} \underline{N}^T\} \underline{\tilde{s}}, \\
&= \underline{\tilde{s}}^T \Sigma_N \underline{\tilde{s}}, \\
&= (\underline{s}_1 - \underline{s}_0)^T \Sigma_N^{-1} (\underline{s}_1 - \underline{s}_0) \triangleq d^2.
\end{aligned}
\tag{9}
$$

Note that the variance of $T(\underline{Y})$ is independent of the hypothesis. Also, the positive definiteness of Σ_N implies that Σ_N^{-1} is also positive definite and thus $d^2 > 0$ unless $\underline{s}_1 = \underline{s}_0$. Thus, $T(\underline{Y}) \sim \mathcal{N}(\tilde{\mu}_j, d^2)$ under H_j for $j = \{0,1\}$. Since $T(\underline{Y})$ is continuous, there is no need for randomization. Hence γ in (6) is irrelevant.

$$
\begin{aligned}
P_j(\Gamma_1) &= \frac{1}{\sqrt{2\pi d^2}} \int_{\tau'}^{\infty} e^{-(x-\tilde{\mu}_j)/2d^2} dx, \\
&= 1 - \Phi\Big(\frac{\tau' - \tilde{\mu}_j}{d}\Big),
\end{aligned}
\tag{10}
$$

where d is the positive square root of d^2 and $\Phi(\cdot)$ is the cdf of standard Gaussian distribution. $P_j(\Gamma_1)$ can be written in terms of the original threshold τ as

$$
P_j(\Gamma_1) = \begin{cases} 1 - \Phi\Big(\dfrac{\log \tau}{d} + \dfrac{d}{2}\Big) & \text{for } j = 0 \\ 1 - \Phi\Big(\dfrac{\log \tau}{d} - \dfrac{d}{2}\Big) & \text{for } j = 1. \end{cases}
\tag{11}
$$

It can be observed that the Bayesian performance in the problem under consideration here is identical to the location testing problem with $n = 1$. Similarly, the performance for the Minimax and Neyman-Pearson problems under consideration here is identical to the location testing problem studied earlier.

Example 1.1. For Neyman-Pearson testing with false alarm probability of α, $P_{FA}(\tilde{\delta}) = \alpha = P_0(\Gamma_1)$. From (10) we see that

$$\alpha = 1 - \Phi\left(\frac{\tau' - \tilde{\mu}_0}{d}\right). \tag{12}$$

The corresponding detection probability is

$$
\begin{aligned}
P_D(\tilde{\delta}_{NP}) = P_1(\Gamma_1) &= 1 - \Phi\left(\frac{\tau' - \tilde{\mu}_1}{d}\right), \\
&= 1 - \Phi\left(\frac{d\Phi^{-1}(1-\alpha) + \tilde{\mu}_0 - \tilde{\mu}_1}{d}\right), \\
&= 1 - \Phi\left(\Phi^{-1}(1-\alpha) + \frac{\tilde{\mu}_0 - \tilde{\mu}_1}{d}\right).
\end{aligned}
$$

Consider,

$$
\begin{aligned}
\frac{\tilde{\mu}_0 - \tilde{\mu}_1}{d} &= \frac{\tilde{s}^T \underline{s}_0 - \tilde{s}^T \underline{s}_1}{d} \\
&= \frac{\left[\Sigma_N^{-1}(\underline{s}_1 - \underline{s}_0)\right]^T (\underline{s}_0 - \underline{s}_1)}{d} \\
&= \frac{(\underline{s}_1 - \underline{s}_0)^T \Sigma_N^{-1}(\underline{s}_0 - \underline{s}_1)}{d} = -d
\end{aligned}
$$

Therefore, the expression for detection probability is

$$P_D(\tilde{\delta}_{NP}) = 1 - \Phi(\Phi^{-1}(1-\alpha) - d) \tag{13}$$

Note 1. $P_D(\tilde{\delta}_{NP})$ is a monotonically increasing function in d as shown in Fig. 1. Higher the value of d, better is the detection probability.

Note 2. **Interpretation of d^2:** d^2 can be interpreted as a measure of signal to noise ratio. Consider the case when $\underline{s}_0 = \underline{0}$ and $\underline{s}_1 = \underline{s}$.

Case 1: When noise is i.i.d $\mathcal{N}(0, \sigma^2)$ which corresponds to the multivariate Gaussian case with $\Sigma_N = \sigma^2 I$, where I denotes a $n \times n$ identity matrix. In this case,

$$
\begin{aligned}
\Sigma_N^{-1} &= \sigma^{-2} I, \\
d^2 &= (\underline{s}_1 - \underline{s}_0)^T \Sigma_N^{-1}(\underline{s}_1 - \underline{s}_0), \\
&= \frac{\underline{s}^T I \underline{s}}{\sigma^2}, \\
&= \frac{1}{\sigma^2} \sum_{k=1}^{n} s_k^2 = \frac{n\bar{s}^2}{\sigma^2}.
\end{aligned} \tag{14}
$$

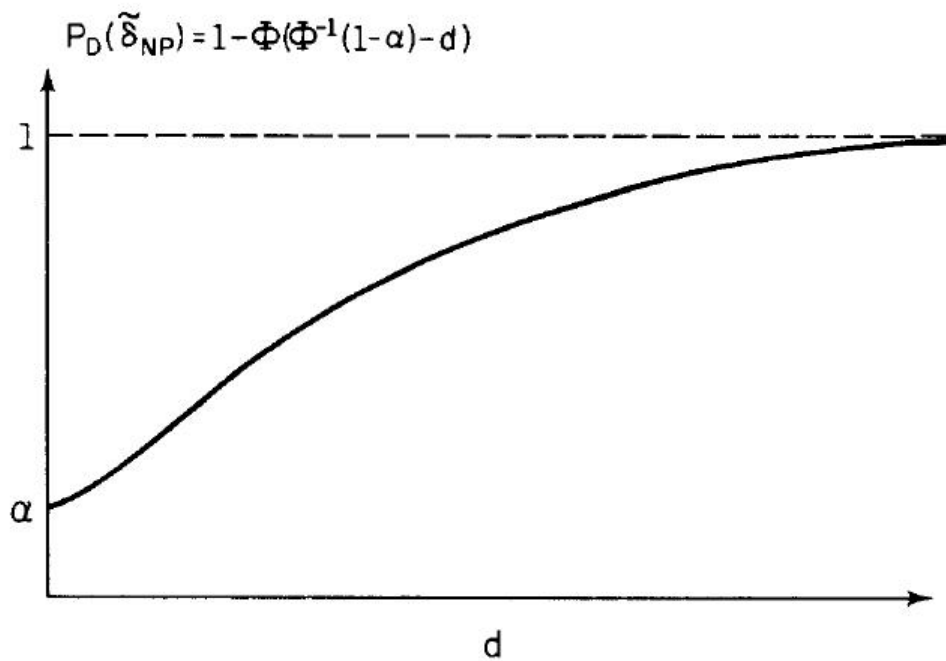$$P_D(\widetilde{\delta}_{NP}) = 1 - \Phi(\Phi^{-1}(1-\alpha) - d)$$

Figure 1: Power function for Neyman pearson test with Gaussian error
(Source: H. Vincent Poor, An Introduction to Signal Detection and Estimation(Second Edition), Figure II.D.3)

where $\bar{s}^2 \triangleq \frac{1}{n} \sum_{k=1}^{n} s_k^2$ is the average signal power, $\sigma^2 = \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\{N_k^2\}$ is the average noise power, so that d^2 is the product of average signal to noise power and the number of samples.

Case 2: A similar interpretation can be given to d^2 in non i.i.d case. Matched filter output for pure signal($\underline{s}$) at time n is

$$\underline{\tilde{s}}^T \underline{s} = \underline{s}^T \Sigma_N^{-1} \underline{s} = d^2.$$

Therefore, signal power at the output of matched filter at time n is d^4. Matched filter output power for input equal to pure noise($\underline{N}$) at time n is

$$\mathbb{E}\{(\underline{\tilde{s}}^T \ \underline{N})^2\} = \underline{s}^T \Sigma_N^{-1} \underline{s} = d^2.$$

Hence,

$$\frac{\text{signal power at the output of matched filter}}{\text{noise power at the output of matched filter}} = \frac{d^4}{d^2} = d^2.$$

Note 3. Matched filter output at time n is nothing but multiplication by $\tilde{s}^T$. We can write the quantity $\sum_{k=1}^{n} \tilde{s}_k y_k$ as the input at time n of an LTI filter with impulse response

$$\tilde{h}_k = \begin{cases} \tilde{s}_{n-k} & 0 \le k \le n-1 \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

The filter $\tilde{h}$ of (15) has maximum output SNR at time n among all linear filters with impulse response of length n.

Note 4. The quantity d^2 also has another interpretation for the i.i.d case with general signals $\underline{s}_0, \underline{s}_1$ and $\underline{N} \sim \mathcal{N}(\underline{0}, \sigma^2 I)$. In this case, we can write

$$d^2 = \frac{1}{\sigma^2} \|\underline{s}_1 - \underline{s}_0\|_2^2$$

where $\|\underline{s}_1 - \underline{s}_0\|$ denotes Euclidean distance between the signal vectors. Thus, the farther apart the signal vectors are, the better performance can be achieved.

2 Reduction to the i.i.d noise case

The observation vector $\underline{Y}$ can be transformed to give an equivalent observation with i.i.d noise. In particular, because Σ_N is positive definite, it can be written as

$$Sigma_N = CC^T \tag{16}$$

where C is an $n \times n$ invertible lower triangular matrix. Eq. 16 is called *Cholesky decomposition* of Σ_N.

$$\Sigma_N = CC^T,$$
$$(\Sigma_N)^{-1} = (CC^T)^{-1} = (C^T)^{-1}C^{-1}.$$

Let $D \triangleq C^{-1}$, then, $\Sigma_N^{-1} = D^T D$, D is also lower triangular. Let

$$\bar{\underline{Y}} \triangleq D\underline{Y} = D\underline{s}_j + D\underline{N},$$

$\bar{D} \triangleq D\underline{s}_j$ is the transformed signal and $\bar{\underline{N}} \triangleq D\underline{N}$ is the transformed noise. Clearly,

$$\mathbb{E}\{D\underline{N}\} = \underline{0},$$
$$Var\{D\underline{N}\} = \mathbb{E}\{D\underline{N}\,\underline{N}^T D^T\},$$
$$= D\Sigma_N D^T = C^{-1}\Sigma_N (C^{-1})^T,$$
$$= \Sigma_N \Sigma_N^{-1} = I.$$

Hence, $\bar{\underline{N}} \sim \mathcal{N}(\underline{0}, I)$. So, we now have an i.i.d noise situation and the optimum detection statistic is

$$T(\underline{Y}) = (\underline{s}_1 - \underline{s}_0)^T \Sigma_N^{-1} \underline{Y},$$
$$= (\underline{s}_1 - \underline{s}_0)^T D^T D\underline{Y},$$
$$= (\bar{\underline{s}}_1 - \bar{\underline{s}}_0)^T \bar{\underline{Y}}.$$

Note that the lower triangularity of C implies that C^{-1} is also lower triangular. Cholesky decomposition is preferred because, $\bar{\underline{Y}} = C^{-1}\underline{Y}$ can be implemented as the output at times $(0, 1, ..., n-1)$ of a causal (time varying) filter. Since the noise in the output of this filter is white (i.i.d), this filter is sometimes known as *whitening filter*. So, the optimum decoder can be represented as shown in Fig. 2.

The performance of coherent detection in dependent noise depends on how far apart the signals are when transformed to a coordinate system in which the noise components are i.i.d.

3   Optimal Signal Design

The performance of optimum coherent detection in Gaussian noise is improved by increasing the quantity $d^2 \triangleq (\underline{s}_1 - \underline{s}_0)^T \Sigma_N^{-1} (\underline{s}_1 - \underline{s}_0)$. We consider designing signal $\underline{s}_1 = \underline{s}$ such that these signals maximize d^2 given $noise \sim \mathcal{N}(\underline{0}, \Sigma_N)$.

Consider $\alpha-$ level N-P detection

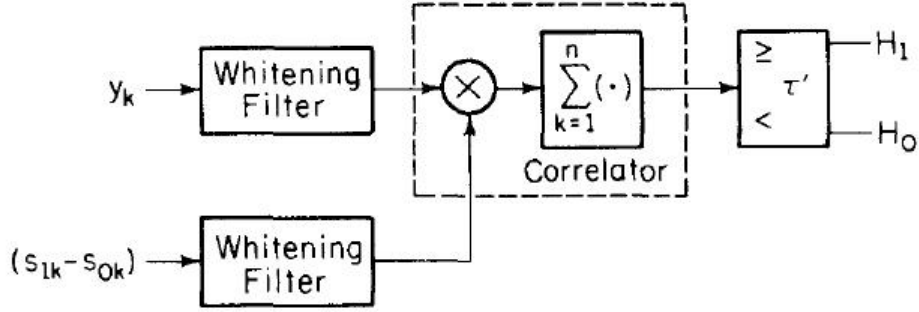$$P_D(\tilde{\delta}_{NP}) = 1 - \Phi(\Phi^{-1}(1-\alpha) - d)$$

Figure 2: Optimum detector for coherent signals in dependent Gaussian noise (Source: H. Vincent Poor, An Introduction to Signal Detection and Estimation(Second Edition), Figure III.B.8)

We have the following problem,

$$\max_{\underline{s}} d^2 = \underline{s}^T \Sigma_N^{-1} \underline{s} \quad \text{s.t} \ \|s\|_2^2 \leq E, \tag{17}$$

where E is the energy/power constraint. If $\lambda_1, \lambda_2, \ldots, \lambda_n$ are eigen values of Σ_N, since Σ_N is symmetric, we can find orthonormal eigen vectors $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_n$ corresponding to eigen values $\lambda_1, \lambda_2, \ldots, \lambda_n$ respectively. Σ_N can be expanded as

$$\Sigma_N = \sum_{i=1}^n \lambda_i \underline{v}_i \underline{v}_i^T,$$

$$\underline{x}^T \Sigma_N^{-1} \underline{x} = \underline{x}^T \left(\sum_{i=1}^n \lambda_i^{-1} \underline{v}_i \underline{v}_i^T \right) \underline{x},$$

$$= \sum_{i=1}^n \lambda_i^{-1} \underline{x}^T \underline{v}_i \underline{v}_i^T \underline{x},$$

$$\leq \sum_{i=1}^n (\lambda_{min}^{-1}) \underline{x}^T \underline{v}_i \underline{v}_i^T \underline{x} = \lambda_{min}^{-1} \|\underline{x}\|_2^2,$$

$$\leq \lambda_{min}^{-1} E.$$

The solution is by setting

$$\underline{s} = c\underline{v}_k \quad \text{s.t} \ \lambda_k = \lambda_{min} \quad \text{with} \quad \|\underline{s}\|_2^2 = c^2 = E.$$

8

Therefore,

$$
\begin{aligned}
d_{opt}^2 &= \underline{s}^T \Sigma_N^{-1} \underline{s}, \\
&= \sqrt{E} \underline{v}_k^T \Sigma_N^{-1} \underline{v}_k \sqrt{E}, \\
&= \frac{E}{\lambda_{min}}.
\end{aligned}
$$

The optimum detection probability for $\alpha-$ level N-P test with power constraint E is

$$
1 P_{D,opt} = 1 - \Phi \left(\Phi^{-1}(1 - \alpha) - \sqrt{\frac{E}{\lambda_{min}}} \right). \tag{18}
$$