Coherent Detection

Reading:

- Ch. 4 in Kay-II.
- (Part of) Ch. III.B in Poor.

Coherent Detection (of A Known Deterministic Signal) in Independent, Identically Distributed (I.I.D.) Noise whose Pdf/Pmf is Exactly Known

$$\mathcal{H}_0 : \qquad x[n] = w[n], \quad n = 1, 2, \ldots, N \quad \text{versus}$$
$$\mathcal{H}_1 : \qquad x[n] = s[n] + w[n], \quad n = 1, 2, \ldots, N$$

where

- $s[n]$ is a known deterministic signal and
- $w[n]$ is i.i.d. noise with exactly known probability density or mass function (pdf/pmf).

The scenario where the signal $s[n]$, $n = 1, 2, \ldots, N$ is *exactly known* to the designer is sometimes referred to as the *coherent-detection scenario*.

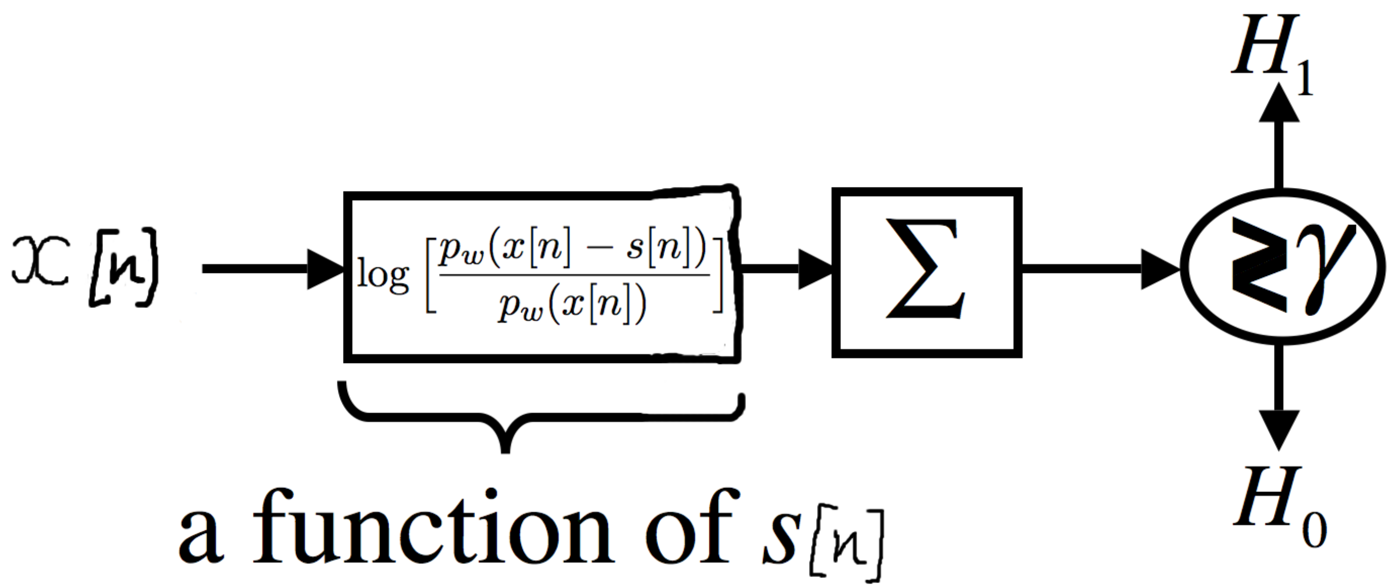
The likelihood ratio for this problem is

$$\Lambda(\boldsymbol{x}) = \frac{\prod_{n=1}^{N} p_w(x[n] - s[n])}{\prod_{n=1}^{N} p_w(x[n])}$$

and its logarithm is

$$\log \Lambda(\boldsymbol{x}) = \sum_{n=0}^{N-1} \log \left[\frac{p_w(x[n] - s[n])}{p_w(x[n])} \right]$$

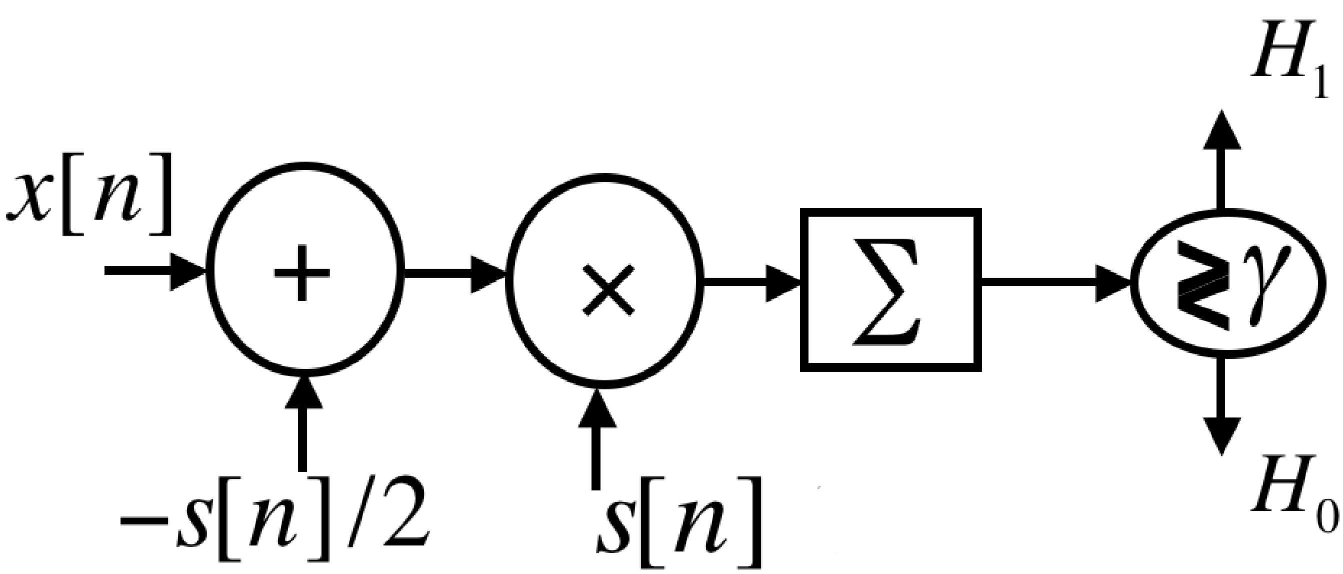
which needs to be compared with a threshold γ (say). Here is a schematic of our *coherent detector*:



Example: Coherent Detection in AWGN (Ch. 4 in Kay-II)

If the noise $w[n] \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ (i.e. additive white Gaussian noise, AWGN) and noise variance σ^2 is known, the likelihood-ratio test reduces to (upon taking log and scaling by σ^2):

$$\sigma^2 \log \Lambda(\boldsymbol{x}) = \sum_{n=1}^{N} (x[n]s[n] - s^2[n]/2) \overset{\mathcal{H}_1}{\gtrless} \gamma \text{ (a threshold)} \quad (1)$$



and is known as the *correlator detector* or simply *correlator*. This name is due to the fact that, in the Neyman-Pearson setting, we could absorb the $s^2[n]$ terms into the threshold, leading to the test statistic proportional to the sample correlation between $x[n]$ and $s[n]$:

$$T(\boldsymbol{x}) = \sum_{n=1}^{N} x[n]\, s[n]. \qquad (2)$$

We decide $\mathcal{H}_1$ if $T(\boldsymbol{x}) > \gamma$ (the "new" threshold, after absorbing the $s^2[n]$ terms).

This structure is also known as the *matched-filter* receiver, due to the fact that it operates by comparing the output of a linear, time-invariant (LTI) system (filter) to a threshold. Indeed, we can write (2) as

$$T(\boldsymbol{x}) = \sum_{n=1}^{N} x[n]\, s[n] = \sum_{n=-\infty}^{\infty} x[n]\, h[N-n] = \{x \star h\}[N]$$

where $\{x \star h\}[n]$ denotes convolution of sequences $\{x[n]\}$ and $\{h[n]\}$ and, in this case,

$$h[n] = \begin{cases} s[N-n], & 0 \le n \le N-1 \\ 0, & \text{otherwise} \end{cases}.$$

Therefore, this system

- inputs the observation sequence $x[1], x[2], \ldots, x[N]$ to a digital finite impulse response LTI filter
  and then

- samples the output

$$y[n] = \{x \star h\}[n]$$

at time $n = N$ for comparison with a threshold.

We can express the filter output $y[n]$ in the Fourier domain as follows:

$$\begin{aligned} y[n] &= \int_{-1/2}^{1/2} H(f)\, X(f)\, \exp(j\, 2\pi f n)\, df \\ &= \int_{-1/2}^{1/2} [S(f)]^*\, X(f)\, \exp[j\, 2\pi f (n - N)]\, df \end{aligned}$$

where

$$\begin{aligned} H(f) &= \text{Fourier transform}\{s[N-n]\} \\ &= \sum_{n=0}^{N-1} s[N-n] \exp(-j2\pi f n) \\ &\overset{k=N-n}{=} \sum_{k=1}^{N} s[k] \exp[-j2\pi f (N-k)] \\ &= \exp(-j2\pi f N) \cdot [S(f)]^*, \quad f \in [-\tfrac{1}{2}, \tfrac{1}{2}]. \end{aligned}$$

Finally, sampling the filter output at $n = N$ yields

$$T(\boldsymbol{x}) = y[N] = \int_{-1/2}^{1/2} [S(f)]^* \, X(f) \, df$$

which is not surprising (recall the Parseval's theorem).

Example: Coherent Detection in I.I.D. Laplacian Noise

The Laplacian-noise model is sometimes used to represent the behavior of *impulsive noise* in communication receivers.

If the noise $w[n] \overset{\text{i.i.d.}}{\sim}$ Laplacian:

$$p(w[n]) = \frac{1}{\sqrt{2\,\sigma^2}} \cdot \exp\Big(- \sqrt{\frac{2}{\sigma^2}} \cdot |w[n]|\Big), \quad n = 1, 2, \ldots, N$$

and σ^2 is known, then the log likelihood is

$$\begin{aligned} \log \Lambda(\boldsymbol{x}) &= \sum_{n=1}^{N} \log \Big[\frac{p_w(x[n] - s[n])}{p_w(x[n])}\Big] \\ &= \sqrt{\frac{2}{\sigma^2}} \cdot \sum_{n=1}^{N} (-|x[n] - s[n]| + |x[n]|). \end{aligned}$$

Now, our coherent detector can be written as

$$\sum_{n=1}^{N} (-|x[n] - s[n]| + |x[n]|) \overset{\mathcal{H}_1}{\gtrless} \gamma$$

Interestingly, applying the *maximum-likelihood test* (i.e. the Bayes' decision rule for a 0-1 loss and *a priori* equiprobable hypotheses) corresponds to setting $\gamma = 0$, implying that the maximum-likelihood detector *does not* require the knowledge of the noise parameter σ^2 to declare its decision. Again, the knowledge of σ^2 is *key* to assessing the detection performance.

An intuitive detector follows by defining

$$y[n] = x[n] - \tfrac{1}{2}\, s[n].$$

Then

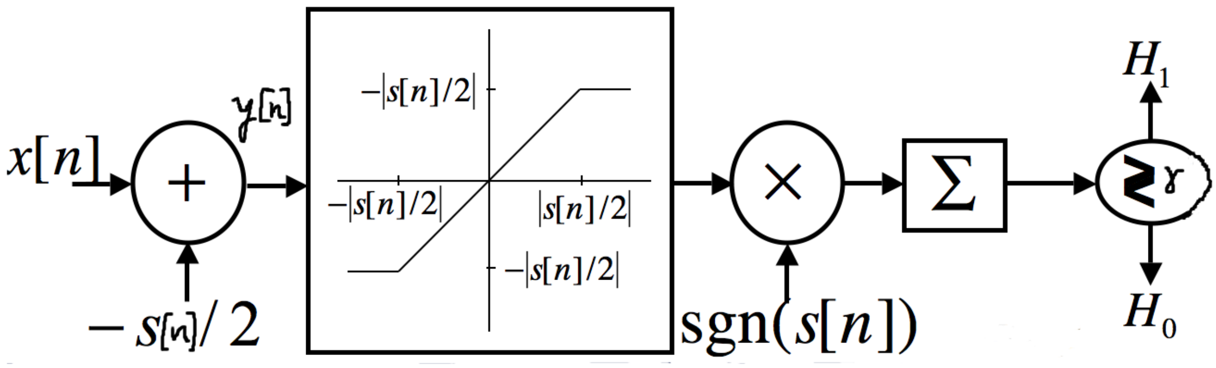
$$\left.\begin{array}{rcl} \mathrm{E}\,\{y[n]\,|\,\mathcal{H}_0\} & = & -\frac{1}{2}\, s[n] \\ \mathrm{E}\,\{y[n]\,|\,\mathcal{H}_1\} & = & \frac{1}{2}\, s[n] \end{array}\right\} \quad \text{(symmetrized version)}$$

and

$$\sum_{n=1}^{N} (-|x[n] - s[n]| + |x[n]|) \overset{\mathcal{H}_1}{\gtrless} \gamma$$

$$\iff \quad \sum_{n=1}^{N} \Big(- |y[n] - \tfrac{1}{2}\, s[n]| + |y[n] + \tfrac{1}{2}\, s[n]| \Big) \overset{\mathcal{H}_1}{\gtrless} \gamma$$

depicted as follows:



see also pp. 49–50 in Poor.

It is of interest to contrast the correlator detector on p. 4 (optimal for AWGN) with the optimal detector for i.i.d. Laplacian noise:

- Both systems *center* the observations by subtracting $s[n]/2$ from each $x[n]$.

- The correlator detector on p. 4 then
  - *correlates* the centered data with the known signal $s[n]$ and
  - *compares* the correlator output with a threshold γ.

- Alternatively, upon centering the observations, the optimal detector for Laplacian noise
  - *soft-limits* the centered data,
  - *correlates* these soft-limited centered observations with the sequence of signal signs $\mathrm{sgn}(s[n[)$, and
  - *compares* the correlator output with a threshold γ.

Soft-limiting in the optimal detector for Laplacian noise *reduces* the effect of large observations on the correlator sum and, therefore, makes the system robust to large noise values (which will occur frequently due to the heavy tails of the Laplacian pdf).

Coherent Detection in Gaussian Noise With Known Covariance (Generalized Matched Filters, Ch. 4 in Kay-II)

Assume that we have observed $\boldsymbol{x} = \begin{bmatrix} x[1] \\ x[2] \\ \vdots \\ x[N] \end{bmatrix}$, which, given the parameter vector $\boldsymbol{\mu}$, is distributed as $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$, where Σ is a *known* positive definite covariance matrix:

$$p(\boldsymbol{x} \mid \boldsymbol{\mu}) = \mathcal{N}(\boldsymbol{\mu}, \Sigma). \tag{3}$$

Consider the following simple hypothesis test:

$$\begin{aligned} \mathcal{H}_0 : \quad & \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \text{versus} \\ \mathcal{H}_1 : \quad & \boldsymbol{\mu} = \boldsymbol{\mu}_1 \end{aligned}$$

which can be viewed as *coherent signal detection in correlated noise with known covariance* (also referred to as the *generalized matched filter detection* in Kay-II).

Let us find the likelihood ratio for this problem:

$$\Lambda(\boldsymbol{x}) = \frac{p(\boldsymbol{x}\,|\,\boldsymbol{\mu}_1)}{p(\boldsymbol{x}\,|\,\boldsymbol{\mu}_0)} = \frac{\frac{1}{\sqrt{(2\pi)^N|\Sigma|}} \cdot \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu}_1)\right]}{\frac{1}{\sqrt{(2\pi)^N|\Sigma|}} \cdot \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_0)^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu}_0)\right]}$$

$$= \exp[\boldsymbol{\mu}_1^T \Sigma^{-1} \boldsymbol{x} - \boldsymbol{\mu}_0^T \Sigma^{-1} \boldsymbol{x} - \tfrac{1}{2}\boldsymbol{\mu}_1^T \Sigma^{-1} \boldsymbol{\mu}_1 + \tfrac{1}{2}\boldsymbol{\mu}_0^T \Sigma^{-1} \boldsymbol{\mu}_0]$$

$$= \exp\left\{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T \Sigma^{-1} \underbrace{[\boldsymbol{x} - \tfrac{1}{2}(\boldsymbol{\mu}_0 + \boldsymbol{\mu}_1)]}_{\text{centering the observations}}\right\} \overset{\mathcal{H}_1}{\gtrless} \tau. \qquad (4)$$

Maximum-likelihood test (i.e. the Bayes' decision rule for a 0-1 loss and *a priori* equiprobable hypotheses): $\Lambda(\boldsymbol{x}) \overset{\mathcal{H}_1}{\gtrless} 1$ or, equivalently,

$$(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T \Sigma^{-1} [\boldsymbol{x} - \tfrac{1}{2}(\boldsymbol{\mu}_0 + \boldsymbol{\mu}_1)] \overset{\mathcal{H}_1}{\gtrless} 0. \qquad (5)$$

Neyman-Pearson setting: In this case, we can absorb the $-\frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T \Sigma^{-1} (\boldsymbol{\mu}_0 + \boldsymbol{\mu}_1)$ term into the threshold, which leads to the test statistic:

$$T(\boldsymbol{x}) = (\underbrace{\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0}_{\boldsymbol{s}})^T \Sigma^{-1} \boldsymbol{x} = \boldsymbol{s}^T \Sigma^{-1} \boldsymbol{x} \qquad (6)$$

Note that we have defined

$$\boldsymbol{s} \stackrel{\triangle}{=} \boldsymbol{\mu}_1 - \boldsymbol{\mu}_0.$$

As usual, we decide $\mathcal{H}_1$ if $T(\boldsymbol{x}) > \gamma$, the "new" threshold, obtained upon absorbing the $-\frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T \Sigma^{-1} (\boldsymbol{\mu}_0 + \boldsymbol{\mu}_1)$ term.

We can use *prewhitening* to handle the noise correlation. Consider the Cholesky factorization $\Sigma^{-1} = D^T D$, whose application to prewhitening has already been discussed in handout # 3. Then, $T(\boldsymbol{x})$ can be written as

$$T(\boldsymbol{x}) = \overbrace{\boldsymbol{s}^T D^T}^{(\boldsymbol{s}^{\mathrm{transf}})^T} \; \overbrace{D\,\boldsymbol{x}}^{\boldsymbol{x}^{\mathrm{transf}}} = \underbrace{\sum_{n=1}^{N} x^{\mathrm{transf}}[n]\, s^{\mathrm{transf}}[n]}_{\text{correlator}}. \qquad (7)$$

An interesting and useful fact about the Cholesky factorization is that D is a lower-triangular matrix, implying that we can compute $x^{\mathrm{transf}}[n]$ by passing the raw data $x[n]$ through a bank of *causal* time-varying linear filters (having coefficients $d_{n,i}$, $i = 1, 2, \ldots, n$, $n = 1, 2, \ldots, N$):

$$x^{\mathrm{transf}}[n] = \sum_{i=1}^{n} \underbrace{d_{n,i}}_{(n,\,i) \text{ element of } D} x[i].$$

Special Case (Coherent Detection in AWGN, see pp. 4–7):

Let us choose $\boldsymbol{\mu}_0 = \mathbf{0}$, $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0 = \boldsymbol{\mu}_1 = \boldsymbol{s} \triangleq \begin{bmatrix} s[1] \\ s[2] \\ \vdots \\ s[N] \end{bmatrix}$, and

$\Sigma = \sigma^2 I$, where σ^2 is known noise variance and I denotes the identity matrix. Substituting these choices into (6) and scaling it by the positive constant σ^2 yields the familiar *correlator* test statistic:

$$\boldsymbol{s}^T \boldsymbol{x} = \underbrace{\sum_{n=1}^{N} x[n]\, s[n]}_{\text{correlator}}.$$

Here, the maximum-likelihood test is [after scaling by the positive constant σ^2, see (5)]

$$\boldsymbol{s}^T \left(\boldsymbol{x} - \tfrac{1}{2}\boldsymbol{s}\right) \overset{\mathcal{H}_1}{\gtrless} 0 \quad \Longleftrightarrow \quad \sum_{n=1}^{N} x[n]\, s[n] \overset{\mathcal{H}_1}{\gtrless} \tfrac{1}{2} \sum_{n=1}^{N} (s[n])^2$$

which, not surprisingly, can be implemented *without knowing* the noise variance σ^2.

Coherent Detection in Gaussian Noise with Known Covariance: Performance Analysis Under the Neyman-Pearson Setting

Consider the conditional pdf of $T(\boldsymbol{x})$ in (6)

$$T(\boldsymbol{x}) = \boldsymbol{s}^T \Sigma^{-1} \boldsymbol{x}$$

given $\boldsymbol{\mu}$. Note that we define

$$\boldsymbol{s} \stackrel{\triangle}{=} \boldsymbol{\mu}_1 - \boldsymbol{\mu}_0.$$

Given $\boldsymbol{\mu}$, $T(\boldsymbol{x})$ is a linear combination of Gaussian random variables, implying that it is also Gaussian, with mean and variance:

$$\begin{aligned} \mathrm{E}\,[T(\boldsymbol{X}) \,|\, \boldsymbol{\mu}] &= \boldsymbol{s}^T \Sigma^{-1} \boldsymbol{\mu} \\ \mathrm{var}[T(\boldsymbol{X}) \,|\, \boldsymbol{\mu}] &= \boldsymbol{s}^T \Sigma^{-1} \boldsymbol{s} \quad \text{(not a function of } \boldsymbol{\mu}\text{)}. \end{aligned}$$

Now

$$
\begin{aligned}
P_{\mathrm{FA}} &= P[T(\boldsymbol{X}) > \gamma \,|\, \boldsymbol{\mu} = \boldsymbol{\mu}_0] \\
&= P\Big[\overbrace{\frac{T(\boldsymbol{X}) - \boldsymbol{s}^T\,\Sigma^{-1}\boldsymbol{\mu}_0}{\sqrt{\boldsymbol{s}^T\,\Sigma^{-1}\,\boldsymbol{s}}}}^{\text{standard normal random variable}} > \frac{\gamma - \boldsymbol{s}^T\,\Sigma^{-1}\boldsymbol{\mu}_0}{\sqrt{\boldsymbol{s}^T\,\Sigma^{-1}\boldsymbol{s}}} \Big] \\
&= Q\left(\frac{\gamma - \boldsymbol{s}^T\,\Sigma^{-1}\boldsymbol{\mu}_0}{\sqrt{\boldsymbol{s}^T\,\Sigma^{-1}\boldsymbol{s}}} \right) \qquad (8)
\end{aligned}
$$

and

$$
\begin{aligned}
P_{\mathrm{D}} &= P[T(\boldsymbol{X}) > \gamma \,|\, \boldsymbol{\mu} = \boldsymbol{\mu}_1] \\
&= P\Big[\overbrace{\frac{T(\boldsymbol{X}) - \boldsymbol{s}^T\,\Sigma^{-1}\boldsymbol{\mu}_1}{\sqrt{\boldsymbol{s}^T\,\Sigma^{-1}\,\boldsymbol{s}}}}^{\text{standard normal random variable}} > \frac{\gamma - \boldsymbol{s}^T\,\Sigma^{-1}\boldsymbol{\mu}_1}{\sqrt{\boldsymbol{s}^T\,\Sigma^{-1}\,\boldsymbol{s}}} \Big] \\
&= Q\left(\frac{\gamma - \boldsymbol{s}^T\,\Sigma^{-1}\boldsymbol{\mu}_1}{\sqrt{\boldsymbol{s}^T\,\Sigma^{-1}\,\boldsymbol{s}}} \right).
\end{aligned}
$$

We use (8) to obtain a γ that satisfies the specified P_{FA}:

$$
\frac{\gamma}{\sqrt{\boldsymbol{s}^T\,\Sigma^{-1}\,\boldsymbol{s}}} = Q^{-1}(P_{\mathrm{FA}}) + \frac{\boldsymbol{s}^T\,\Sigma^{-1}\,\boldsymbol{\mu}_0}{\sqrt{\boldsymbol{s}^T\,\Sigma^{-1}\,\boldsymbol{s}}}
$$

implying

$$
P_{\mathrm{D}} = Q\Big(Q^{-1}(P_{\mathrm{FA}}) - \sqrt{\boldsymbol{s}^T\,\Sigma^{-1}\,\boldsymbol{s}} \Big). \qquad (9)
$$

Here,

$$\boldsymbol{s}^T \Sigma^{-1} \boldsymbol{s} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$$

is the deflection coefficient or (a reasonable definition for) the signal-to-noise ratio (SNR).

Special Case (Coherent Detection in AWGN, see p. 4):

Again, let us choose $\boldsymbol{\mu}_0 = \boldsymbol{0}$, $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0 = \boldsymbol{\mu}_1 = \boldsymbol{s} \triangleq \begin{bmatrix} s[1] \\ s[2] \\ \vdots \\ s[N] \end{bmatrix}$,

and $\Sigma = \sigma^2 I$, where σ^2 is a known noise variance. Substituting these choices into (9) yields

$$P_{\text{D}} = Q\left(Q^{-1}(P_{\text{FA}}) - \sqrt{\frac{\boldsymbol{s}^T \boldsymbol{s}}{\sigma^2}}\right) = Q\left(Q^{-1}(P_{\text{FA}}) - \sqrt{\frac{\sum_{n=1}^{N} (s[n])^2}{\sigma^2}}\right).$$

Note that the detection performance depends on the signal $s[n]$ *only through its energy*

$$\mathcal{E} = \boldsymbol{s}^T \boldsymbol{s} = \sum_{n=1}^{N} (s[n])^2$$

i.e. the shape of $s[n]$ is irrelevant! This is *not true* for correlated Gaussian noise, see the discussion on pp. 22–25.

Coherent Detection in Gaussian Noise with Known Covariance: Minimum Average Error Probability (for Bayesian Decision-theoretic Approach with 0-1 Loss)

Consider minimizing the average error probability for the practically most interesting case of equiprobable hypotheses:

$$\pi(\boldsymbol{\mu}_0) = \pi(\boldsymbol{\mu}_1) = \tfrac{1}{2} \tag{10}$$

which leads to the *maximum-likelihood test*:

$$\underbrace{\frac{p(\boldsymbol{x} \mid \boldsymbol{\mu}_1)}{p(\boldsymbol{x} \mid \boldsymbol{\mu}_0)}}_{\text{likelihood ratio } \Lambda(\boldsymbol{x})} \quad \overset{\mathcal{H}_1}{\gtrless} \frac{\pi(\boldsymbol{\mu}_0)}{\pi(\boldsymbol{\mu}_1)} = 1$$

and reduces to (upon taking the log)

$$\begin{aligned}
\log \Lambda(\boldsymbol{x}) &= -\tfrac{1}{2}\,(\boldsymbol{x} - \boldsymbol{\mu}_1)^T\, \Sigma^{-1}\, (\boldsymbol{x} - \boldsymbol{\mu}_1)^T \\
&\quad + \tfrac{1}{2}\,(\boldsymbol{x} - \boldsymbol{\mu}_0)^T\, \Sigma^{-1}\, (\boldsymbol{x} - \boldsymbol{\mu}_0)^T & (11) \\
&= \underbrace{(\overbrace{\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0}^{\boldsymbol{s}})^T\, \Sigma^{-1}\, [\boldsymbol{x} - \tfrac{1}{2}\,(\boldsymbol{\mu}_0 + \boldsymbol{\mu}_1)]}_{\triangleq\, \Upsilon(\boldsymbol{x})} \overset{\mathcal{H}_1}{\gtrless} 0. & (12)
\end{aligned}$$

Note that we have defined

$$\begin{aligned}\Upsilon(\boldsymbol{x}) &= \boldsymbol{s}^T \Sigma^{-1} [\boldsymbol{x} - \tfrac{1}{2}(\boldsymbol{\mu}_0 + \boldsymbol{\mu}_1)] \quad \text{(log-likelihood ratio)}\\ \boldsymbol{s} &= \boldsymbol{\mu}_1 - \boldsymbol{\mu}_0.\end{aligned}$$

To determine the minimum average error probability:

$$\begin{aligned}\text{min av. error prob.} &= \tfrac{1}{2} P[\Upsilon(\boldsymbol{X}) > 0 \mid \boldsymbol{\mu} = \boldsymbol{\mu}_0]\\ &\quad + \tfrac{1}{2} P[\Upsilon(\boldsymbol{X}) < 0 \mid \boldsymbol{\mu} = \boldsymbol{\mu}_1]\end{aligned}$$

we note that, assuming (5), $\Upsilon(\boldsymbol{x})$ is conditionally Gaussian given $\boldsymbol{\mu}$, with mean and variance:

$$\begin{aligned}\mathrm{E}[\Upsilon(\boldsymbol{X}) \mid \boldsymbol{\mu}] &= \boldsymbol{s}^T \Sigma^{-1} [\boldsymbol{\mu} - \tfrac{1}{2}(\boldsymbol{\mu}_0 + \boldsymbol{\mu}_1)]\\ \mathrm{var}[\Upsilon(\boldsymbol{X}) \mid \boldsymbol{\mu}] &= \boldsymbol{s}^T \Sigma^{-1} \boldsymbol{s} \quad \text{(not a function of } \boldsymbol{\mu}\text{)}\end{aligned}$$

implying

$$\mathrm{E}[\Upsilon(\boldsymbol{X}) \mid \boldsymbol{\mu} = \boldsymbol{\mu}_1] = -\mathrm{E}[\Upsilon(\boldsymbol{X}) \mid \boldsymbol{\mu} = \boldsymbol{\mu}_0] = \tfrac{1}{2} \boldsymbol{s}^T \Sigma^{-1} \boldsymbol{s}.$$

Therefore,

$$
\begin{aligned}
& P[\Upsilon(\boldsymbol{X}) > 0 \,|\, \boldsymbol{\mu} = \boldsymbol{\mu}_0] \\
&= P\Big[\underbrace{\frac{\Upsilon(\boldsymbol{X}) + \frac{1}{2}\,\boldsymbol{s}^T\,\Sigma^{-1}\,\boldsymbol{s}}{\sqrt{\boldsymbol{s}^T\,\Sigma^{-1}\,\boldsymbol{s}}}}_{\text{standard normal random variable}} > \frac{\frac{1}{2}\,\boldsymbol{s}^T\,\Sigma^{-1}\,\boldsymbol{s}}{\sqrt{\boldsymbol{s}^T\,\Sigma^{-1}\,\boldsymbol{s}}} \,\Big|\, \boldsymbol{\mu} = \boldsymbol{\mu}_0 \Big] \\
&= Q\Big(\tfrac{1}{2}\sqrt{\boldsymbol{s}^T\,\Sigma^{-1}\,\boldsymbol{s}}\Big)
\end{aligned}
$$

and, by symmetry,

$$
P[\Upsilon(\boldsymbol{X}) < 0 \,|\, \boldsymbol{\mu} = \boldsymbol{\mu}_1] = Q\Big(\tfrac{1}{2}\sqrt{\boldsymbol{s}^T\,\Sigma^{-1}\,\boldsymbol{s}}\Big)
$$

implying, finally,

$$
\text{min av. error prob.} = Q\Big(\tfrac{1}{2}\sqrt{\boldsymbol{s}^T\,\Sigma^{-1}\,\boldsymbol{s}}\Big). \tag{13}
$$

Optimal System Design Based on Detection Performance: General Comments (Useful for Many Applications)

- The detection performance is a function of the experiment parameters. Hence, we can optimize it by adjusting these parameters.

- There are two common performance criteria:
  - maximizing $P_{\rm D}$ for a given $P_{\rm FA}$ (in the Neyman-Pearson spirit), or
  - minimizing the preposterior (Bayes) risk (in general) and average error probability in particular (which is equal to the preposterior risk for a 0-1 loss), popular in communications.

- Examples of adjustable parameters:
  - Exciting signal: direction of incidence, intensity, waveform, polarity, etc.
  - System: sensor placement etc.

Example: Optimal Signal-waveform Design for Coherent Detection in Correlated Gaussian Noise with Known Covariance

Consider a coherent radar/sonar/NDE scenario with $\boldsymbol{\mu}_0 = \mathbf{0}$, $\boldsymbol{\mu}_1 = \boldsymbol{s}$, and correlated Gaussian noise with known covariance. In the Neyman-Pearson spirit, let us maximize P_{D} for a given P_{FA}. First, recall the expression (9) for the detection probability:

$$P_{\mathrm{D}} = Q\Big(Q^{-1}(P_{\mathrm{FA}}) - \sqrt{\boldsymbol{s}^T\,\Sigma^{-1}\,\boldsymbol{s}}\Big).$$

Note that P_{D} *increases monotonically* as the deflection coefficient $\boldsymbol{s}^T\,\Sigma^{-1}\,\boldsymbol{s}$ grows. Since here larger signal energy clearly improves the detection performance and our focus is on optimizing the signal shape, we now impose the energy constraint:

$$\boldsymbol{s}^T\boldsymbol{s} = \mathcal{E} \quad (\mathcal{E} \text{ is specified}).$$

Hence, we have the following optimization problem:

$$\max_{\boldsymbol{s}} \boldsymbol{s}^T\,\Sigma^{-1}\,\boldsymbol{s} \quad \text{subject to } \boldsymbol{s}^T\,\boldsymbol{s} = \mathcal{E}$$

or, (almost) equivalently,

$$\max_{\boldsymbol{s}} \frac{\boldsymbol{s}^T\,\Sigma^{-1}\,\boldsymbol{s}}{\boldsymbol{s}^T\,\boldsymbol{s}} = \frac{1}{\min_i \lambda_i(\Sigma)}$$

where $\lambda_i(\Sigma)$ are the eigenvalues of Σ. The optimal $\boldsymbol{s}$ is proportional to an eigenvector of Σ that corresponds to its smallest eigenvalue. Therefore, the optimized detection performance is (for a specified signal energy $\mathcal{E}$):

$$P_{\mathrm{D,opt}} = Q\left(Q^{-1}(P_{\mathrm{FA}}) - \sqrt{\frac{\mathcal{E}}{\min_i \lambda_i(\Sigma)}} \right).$$

(Similar to) Examples 4.5 and III.B.4 in Kay-II and Poor (respectively): Consider a special case with $N = 2$ observations and

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

where σ^2 is known variance and $-1 < \rho < 1$ is known correlation coefficient. It is easy to show that the eigenvalues of this matrix are

$$\lambda_1 = \sigma^2 (1 - \rho), \quad \lambda_2 = \sigma^2 (1 + \rho) \quad \text{(eigenvalues)}$$

and the corresponding eigenvectors are:

$$\boldsymbol{u}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \boldsymbol{u}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{(eigenvectors)}.$$

Thus, if $\rho > 0$, $\min_i \lambda_i(\Sigma) = \lambda_1$ and an optimal signal $\boldsymbol{s}$ is

$$\boldsymbol{s} \propto \boldsymbol{u}_1 \quad \Longrightarrow \quad \boldsymbol{s} = \sqrt{\frac{\mathcal{E}}{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

On the other hand, if $\rho < 0$, $\min_i \lambda_i(\Sigma) = \lambda_2$ and an optimal signal $\boldsymbol{s}$ is

$$\boldsymbol{s} \propto \boldsymbol{u}_1 \quad \Longrightarrow \quad \boldsymbol{s} = \sqrt{\frac{\mathcal{E}}{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

and

$$P_{\mathrm{D,opt}} = Q\left(Q^{-1}(P_{\mathrm{FA}}) - \sqrt{\frac{\mathcal{E}}{\sigma^2 \, (1 - |\rho|)}}\right).$$

Two Known Signals in AWGN: Minimum Average-Error Probability Approach to Coherent Detection (Bayesian Decision Theory for 0-1 Loss)

We first consider the *binary case*:

$$\begin{aligned}
\mathcal{H}_0 : \quad & x[n] = s_0[n] + w[n], \quad n = 1, 2, \ldots, N \quad \text{versus} \\
\mathcal{H}_1 : \quad & x[n] = s_1[n] + w[n], \quad n = 1, 2, \ldots, N
\end{aligned}$$

and the standard AWGN model for $w[n]$:

$$w[n] \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$$

with known variance σ^2. This problem fits into our coherent detection formulation for Gaussian noise:

$$\begin{aligned}
\mathcal{H}_0 : \quad & \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \text{versus} \\
\mathcal{H}_1 : \quad & \boldsymbol{\mu} = \boldsymbol{\mu}_1
\end{aligned}$$

with

$$\Sigma = \sigma^2 I, \;\; \boldsymbol{\mu}_0 = \begin{bmatrix} s_0[1] \\ s_0[2] \\ \vdots \\ s_0[N] \end{bmatrix}, \;\; \boldsymbol{\mu}_1 = \begin{bmatrix} s_1[1] \\ s_1[2] \\ \vdots \\ s_1[N] \end{bmatrix}. \qquad (14)$$

Example. Binary Phase-shift Keying (BPSK):

$$s_0[n] = \cos(2\pi f_0\,(n-1)),\;\; s_1[n] = \cos(2\pi f_0\,(n-1)+\pi).$$

Consider minimizing the average error probability for the practically most interesting case of equiprobable hypotheses:

$$\pi(\boldsymbol{\mu}_0) = \pi(\boldsymbol{\mu}_1) = \tfrac{1}{2} \tag{15}$$

which leads to the *maximum-likelihood test*. Substituting (14) into the likelihood ratio (11)–(12) yields the maximum-likelihood test:[1]

$$\begin{aligned}
\Lambda(\boldsymbol{x}) &= \exp\Big(-\frac{1}{2\,\sigma^2}\,\|\boldsymbol{x}-\boldsymbol{\mu}_1\|^2 + \frac{1}{2\,\sigma^2}\,\|\boldsymbol{x}-\boldsymbol{\mu}_0\|^2\Big) && (16)\\
&= \exp\Big\{\frac{1}{\sigma^2}\cdot(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)^T\,[\boldsymbol{x}-\tfrac{1}{2}\,(\boldsymbol{\mu}_0+\boldsymbol{\mu}_1)]\Big\} \overset{\mathcal{H}_1}{\gtrless} 1 && (17)
\end{aligned}$$

Equivalently, choose the following optimal decision regions (using the notation from handout # 5):

$$\mathcal{X}_m^\star = \Big\{\boldsymbol{x} : m = \arg\max_{i\in\{0,1\}} \Big[\underbrace{(\sum_{n=0}^{N-1} x[n]s_i[n])}_{\text{correlator}} - \underbrace{\tfrac{1}{2}\,\mathcal{E}_i}_{\text{bias term}}\Big]\Big\}$$

[1] Here, $\|\boldsymbol{y}\| = \sqrt{\boldsymbol{y}^T\boldsymbol{y}}$ denotes the Euclidean norm of a vector $\boldsymbol{y}$.

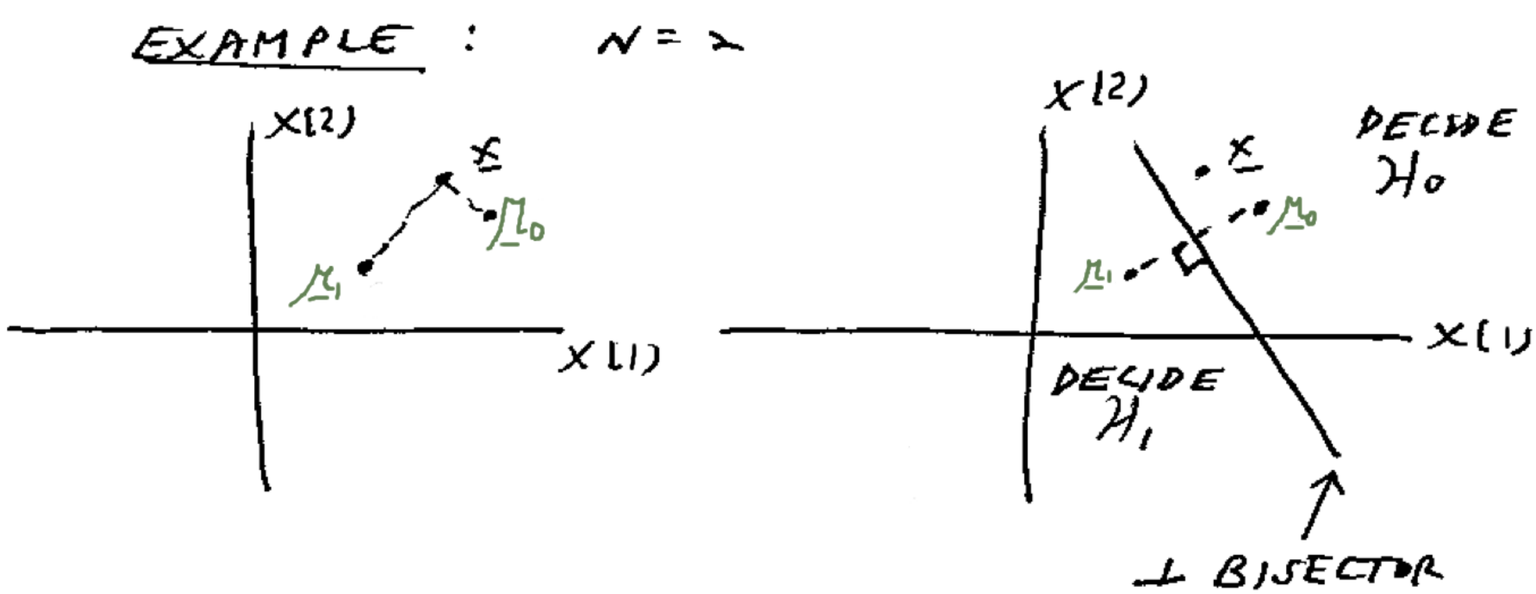
for $m = 0, 1$, where

$$\mathcal{E}_i = \sum_{n=1}^{N} (s_i[n])^2 \quad \text{(energy of } s_i[n]\text{)}, \; i = 0, 1.$$

Equation (16) provides an alternative interpretation of the optimal decision region:

$$\mathcal{X}_m^\star = \left\{ \boldsymbol{x} : m = \arg \min_{i \in \{0,1\}} \|\boldsymbol{x} - \boldsymbol{\mu}_l\| \right\}$$

which is the *minimum-distance receiver*: decide $\mathcal{H}_0$ if $\boldsymbol{\mu}_0$ is closer in Euclidean distance to $\boldsymbol{x}$; otherwise decide $\mathcal{H}_1$.



Note: This maximum-likelihood/minimum-distance receiver *does not* require the knowledge of the noise variance σ^2 to make the decision. However, the knowledge of σ^2 is *key* to assessing the error-probability performance of this (and any other) receiver in AWGN.

If $\mathcal{E}_0 = \mathcal{E}_1$, simply select the hypothesis yielding the larger correlation:

$$\mathcal{X}_m^\star \quad = \quad \Big\{\boldsymbol{x} : m = \arg \max_{i \in \{0,1\}} \underbrace{\Big(\sum_{n=0}^{N-1} x[n] s_i[n] \Big)}_{\text{correlator}} \Big\}.$$

Performance Analysis. Substituting (14) into (13) yields

$$\text{min av. error prob.} = Q\Big(\tfrac{1}{2}\,\frac{\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\|}{\sigma}\Big) \tag{18}$$

which has been derived for equiprobable hypotheses (10), see also the derivation on pp. 19–21. As expected, this minimum average error probability *decreases* as the separation between $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_0$ (quantified by $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\|$) *increases*.

Recall that we are focusing on the communication scenario here. Due to the FCC regulations or the physics of the transmitting device, we must impose an energy constraint. Let us constrain the *average signal energy* [assuming equiprobable hypotheses (10)]:

$$\overline{\mathcal{E}} = \tfrac{1}{2}\,(\mathcal{E}_0 + \mathcal{E}_1) \quad (\overline{\mathcal{E}} \text{ is specified}).$$

Now

$$\begin{aligned} \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\|^2 \quad &= \quad \boldsymbol{\mu}_1^T \boldsymbol{\mu}_1 - 2\boldsymbol{\mu}_1^T \boldsymbol{\mu}_0 + \boldsymbol{\mu}_0^T \boldsymbol{\mu}_0 = 2\,\overline{\mathcal{E}} - 2\,\boldsymbol{\mu}_1^T \boldsymbol{\mu}_0 \\ &= \quad 2\,\overline{\mathcal{E}}\,\Big(1 - \frac{\boldsymbol{\mu}_1^T \boldsymbol{\mu}_0}{\overline{\mathcal{E}}}\Big) \end{aligned}$$

which suggests the following definition of the signal correlation:

$$\rho_s \triangleq \frac{\boldsymbol{\mu}_1^T \boldsymbol{\mu}_0}{\overline{\mathcal{E}}}.$$

Then

$$\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\|^2 = 2\,\overline{\mathcal{E}}\,(1 - \rho_s)$$

and, upon substituting into (18), we obtain

$$\text{min av. error prob.} = Q\Big(\sqrt{\frac{\overline{\mathcal{E}}\,(1-\rho_s)}{2\,\sigma^2}}\Big)$$

which is minimized (for a given $\overline{\mathcal{E}}$) when $\rho_s = -1$, i.e. $\boldsymbol{\mu}_1 = -\boldsymbol{\mu}_0$.

Example. BPSK:

$$s_0[n] = \cos(2\pi f_0 (n-1)), \quad s_1[n] = \cos(2\pi f_0 (n-1) + \pi)$$

for $n = 1, 2, \ldots, N$. In this case,

$$\boldsymbol{\mu}_1 = -\boldsymbol{\mu}_0 \quad \text{(antipodal signaling)}$$

yielding

$$\rho_s = -1$$

which is the optimal signaling choice that minimizes the average error probability, and

$$\text{min av. error probability} = Q\Big(\sqrt{\frac{\overline{\mathcal{E}}}{\sigma^2}}\Big).$$

where

$$\overline{\mathcal{E}} = \mathcal{E}_0 = \mathcal{E}_1 \approx \frac{NA^2}{2}.$$

Therefore, BPSK is the optimal signaling scheme for coherent binary detection in AWGN with equiprobable hypotheses.

Example. Frequency-shift keying (FSK):

$$\begin{aligned} s_0[n] &= A\cos\big(2\pi f_0(n-1)\big) \\ s_1[n] &= A\cos\big(2\pi f_1(n-1)\big), \quad n = 1, 2, \ldots, N. \end{aligned}$$

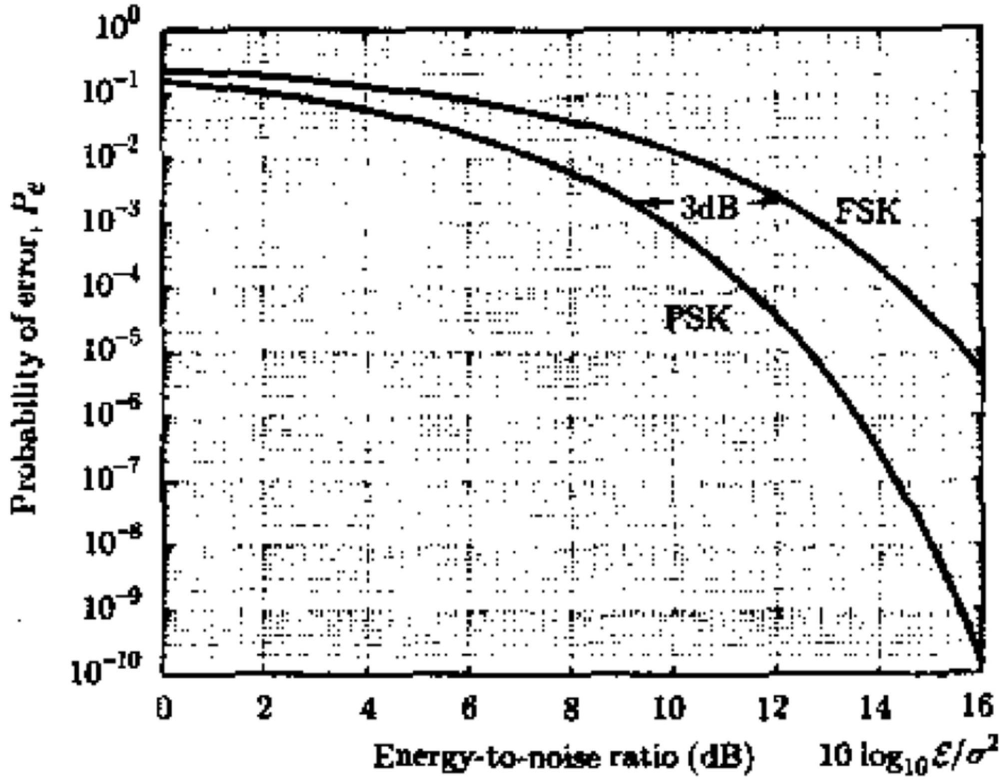
For $|f_1 - f_0| \gg 1/N$, we have

$$\sum_{n=0}^{N-1} s_0[n]s_1[n] \approx 0 \quad \text{i.e.} \quad \rho_s \approx 0$$

and

$$\text{min av. error probability} = Q\Big(\sqrt{\frac{\overline{\mathcal{E}}}{2\,\sigma^2}}\Big)$$

implying that FSK is 3 dB poorer than BPSK at min av. error probability $= 10^{-3}$. See Fig. 4.12 in Kay-II:



Multiple Known Signals in AWGN: Minimum Average-Error Probability Approach to Coherent Detection (Bayesian Decision Theory for 0-1 Loss) (Ch. 4.5.3 in Kay-II)

Now, consider *M-ary hypothesis testing*:

$$\begin{aligned}
\mathcal{H}_0 : \quad & x[n] = s_0[n] + w[n], \quad n = 1, 2, \ldots, N \quad \text{versus} \\
\mathcal{H}_1 : \quad & x[n] = s_1[n] + w[n], \quad n = 1, 2, \ldots, N \quad \text{versus} \\
\vdots \quad & \\
\mathcal{H}_{M-1} : \quad & x[n] = s_{M-1}[n] + w[n], \quad n = 1, 2, \ldots, N
\end{aligned}$$

and the standard AWGN model for $w[n]$:

$$w[n] \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$$

with known variance σ^2. Define

$$\boldsymbol{\mu}_i = \begin{bmatrix} s_i[1] \\ s_i[2] \\ \vdots \\ s_i[N] \end{bmatrix}, \quad i = 0, 1, \ldots, M-1$$

and consider minimizing the average error probability for the practically most interesting case of equiprobable hypotheses:

$$\pi(\boldsymbol{\mu}_0) = \pi(\boldsymbol{\mu}_1) = \cdots = \pi(\boldsymbol{\mu}_{M-1}) = \frac{1}{M} \tag{19}$$

which leads to the *maximum-likelihood test* [obtained by specializing eq. (23) from handout # 5 to the above model]:

$$\mathcal{X}_m^\star = \Big\{ \boldsymbol{x} : m = \arg \max_{i \in \{0,1,\ldots,M-1\}} \Big[\underbrace{\Big(\sum_{n=0}^{N-1} x[n] s_i[n]\Big)}_{\text{correlator}} - \underbrace{\tfrac{1}{2}\,\mathcal{E}_i}_{\text{bias term}} \Big] \Big\}$$

for $m = 0, 1, \ldots, M-1$, where

$$\mathcal{E}_i = \sum_{n=1}^{N} (s_i[n])^2 \quad \text{(energy of } s_i[n]\text{)}, \; i = 0, 1, \ldots, M-1.$$

Can we compute a closed-form expression for the minimum average error probability in the M-ary case? In general, no, but we can obtain an upper bound by applying standard tricks such as (combining) Chernoff and union bounds.

If the signals $s_i[n]$ and $s_m[n]$ are *orthogonal*:

$$\sum_{n=1}^{N} s_i[n] s_m[n] = 0, \quad \forall\, i \neq m \tag{20}$$

then it is fairly easy to obtain the exact expression for the minimum average error probability. To further simplify matters, we assume that all signals have equal energy:

$$\mathcal{E}_i = \mathcal{E}, \quad i = 0, 1, \ldots, M-1.$$

We also define

$$T_i(\boldsymbol{x}) = \Big(\sum_{n=0}^{N-1} x[n] s_i[n]\Big) - \tfrac{1}{2}\,\mathcal{E}.$$

Now, (20) implies

$$\begin{aligned}
&\mathrm{cov}[T_i(\boldsymbol{x}), T_m(\boldsymbol{x}) \,|\, \boldsymbol{\mu} = \boldsymbol{\mu}_l] && (21)\\
&\quad = \mathrm{E}\Big[\big(\sum_{n_1=1}^{N} w[n_1] s_i[n_1]\big)\,\big(\sum_{n_2=1}^{N} w[n_2] s_i[n_2]\big)\Big] && (22)\\
&\quad = \begin{cases} 0, & i \neq m \\ \sigma^2 \sum_{n=1}^{N} (s_i[n])^2 = \sigma^2\,\mathcal{E}, & i = m \end{cases} && (23)
\end{aligned}$$

i.e. the test statistics $T_i(\boldsymbol{x})$, $i = 1, 2, \ldots, M-1$ are mutually independent (since, assuming (5), they are conditionally Gaussian given $\boldsymbol{\mu}$) which allows the analytical computation of the minimum average error-probability. An error occurs if $\mathcal{H}_m$ is the true hypothesis but $T_m(\boldsymbol{x})$ *is not the largest* among the test statistics $T_i(\boldsymbol{x})$, $i = 0, 1, \ldots, M-1$. Therefore, using (19), we obtain the expression for the minimum average error probability:

min av. error probability

$$= \sum_{m=0}^{M-1} \frac{1}{M} P\Big[T_m(\boldsymbol{X}) < \max_{i \in \{0,1,\ldots,M-1\} \backslash \{m\}} [T_i(\boldsymbol{X})] \,\Big|\, \boldsymbol{\mu} = \boldsymbol{\mu}_m\Big]$$

which simplifies, by symmetry, to a single conditional probability:

$$\text{min av. error prob.} = P\Big[T_0(\boldsymbol{X}) < \max_{i\in\{1,2,\dots,M-1\}}[T_i(\boldsymbol{X})] \,\Big|\, \boldsymbol{\mu} = \boldsymbol{\mu}_0\Big].$$

Since $T_i(\boldsymbol{X})$ are affine functions of Gaussian random variables, they are also Gaussian under $\mathcal{H}_0$:

$$p(T_i(\boldsymbol{x}) \,|\, \boldsymbol{\mu} = \boldsymbol{\mu}_0) = \begin{cases} \mathcal{N}(\frac{1}{2}\,\mathcal{E}, \sigma^2\,\mathcal{E}), & i = 0 \\ \mathcal{N}(0, \sigma^2\,\mathcal{E}), & i \neq 0 \end{cases}$$

and thus [using independence in (23), which is critical for tractability]

min av. error probability

$$= 1 - P\Big[T_0(\boldsymbol{X}) > \max_{i\in\{1,2,\dots,M-1\}}[T_i(\boldsymbol{X})] \,\Big|\, \boldsymbol{\mu} = \boldsymbol{\mu}_0\Big]$$

$$= 1 - P\Big[T_0(\boldsymbol{X}) > T_1(\boldsymbol{X}), T_0(\boldsymbol{X}) > T_2(\boldsymbol{X}), \dots, T_0(\boldsymbol{X}) > T_M(\boldsymbol{X}) \,\Big|\, \boldsymbol{\mu} = \boldsymbol{\mu}_0\Big]$$

$$\stackrel{\text{total. prob.}}{=} 1 - \int P\Big[t_0 > T_1(\boldsymbol{X}), t_0 > T_2(\boldsymbol{X}), \dots, t_0 > T_M(\boldsymbol{X}) \,\Big|\, T_0(\boldsymbol{x}) = t_0, \boldsymbol{\mu} = \boldsymbol{\mu}_0\Big]\, p_{T_0}(t_0)\, dt_0$$

$$\stackrel{\text{independence}}{=} 1 - \int_{-\infty}^{\infty} \Big\{ \prod_{i=1}^{M-1} P[T_i(\boldsymbol{X}) < t_0 \,|\, \boldsymbol{\mu} = \boldsymbol{\mu}_0] \Big\}\, p_{T_0}(t_0)\, dt_0$$

which can finally be simplified to [see Ch. 4.5 in Kay-II]:

min av. error probability

$$= 1 - \int_{-\infty}^{\infty} \Phi^{M-1}(u) \, \frac{1}{\sqrt{2\pi}} \exp\Big[-\tfrac{1}{2} \Big(u - \sqrt{\frac{\mathcal{E}}{\sigma^2}} \Big)^2 \Big] \, du$$

where $\Phi(\cdot)$ denotes the standard normal cdf. We define the signal-to-noise ratio (SNR) as

$$\text{SNR} = \frac{\mathcal{E}}{\sigma^2}.$$