

ECS 452 (EC3)

Asst. Prof. Dr. Prapun Suksompong

8 Optimal Detection for Additive Noise Channels: 1-D Case

We now derive the optimal demodulator for the waveform channel. From the previous chapter, we have seen that instead of analyzing the waveform channel, we can convert it to an equivalent vector channel. The length of the vector is the same as the size K of the orthonormal basis for the waveforms $s_1(t), s_2(t), \dots, s_M(t)$. In this chapter, we will assume $K = 1$. This is the case, for example, when we use PAM.

Definition 8.1. Detection Problem: When $K = 1$, our problem under consideration is simply that of **detecting** the scalar message S in the presence of **additive noise N** . The received signal R is given by

$$R = S + N.$$

- S is selected from an alphabet \mathcal{S} containing M possible values $s^{(1)}, s^{(2)}, \dots, s^{(M)}$.
- $p_S(s^{(j)}) = P[S = s^{(j)}] \equiv p_j$.
- S and N are independent.

A detector's job is to guess the value of the channel input S from the value of the received channel output R . We denote this guessed value by \hat{S} . An optimal detector is the one that minimizes the (symbol) error probability $P(\mathcal{E}) = P[\hat{S} \neq S]$.

8.2. The analysis here is very similar to what we have done in Chapter 3. Here, for clarity, we note some important differences:

- In Chapter 3, The channel input and output are denoted by X and Y , respectively. Here, they are denoted by S and R .
- In Chapter 3, the transition probabilities are arbitrary and summarized by the matrix \mathbf{Q} . Here, the transition probabilities is basically controlled by the additive noise.
- In Chapter 3, both X and Y are discrete. Here, S is discrete. However, because noise is continuous, R will be a continuous random variable.

Ch 3



Ch 8



Even with these differences, several techniques that we used in Chapter 3 will be applicable here.

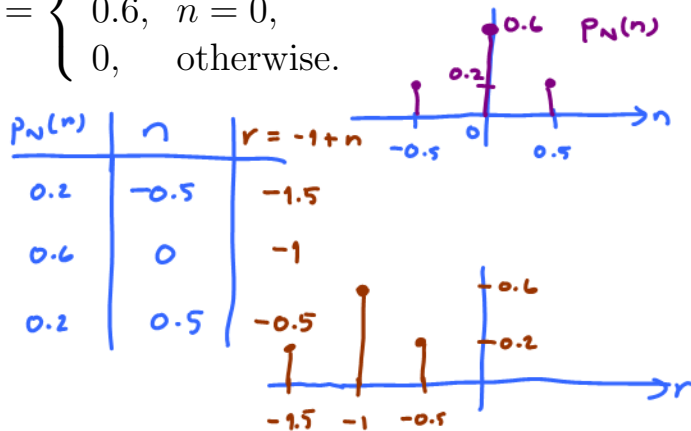
Example 8.3. Review: To re-connect with what we studied in Chapter 3, let's try to find the \mathbf{Q} matrix when the additive noise is discrete. Suppose

$$p_S(s) = \begin{cases} 0.3, & s = -1, \\ 0.7, & s = 1, \\ 0, & \text{otherwise,} \end{cases} \quad \text{and } p_N(n) = \begin{cases} 0.2, & n \in \{-0.5, +0.5\}, \\ 0.6, & n = 0, \\ 0, & \text{otherwise.} \end{cases}$$

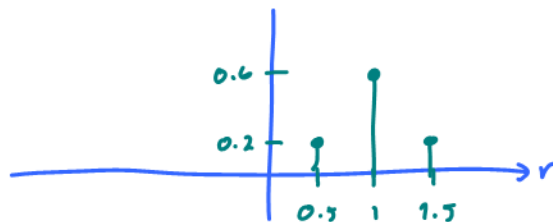
Because $R = S + N$, we know that

(a) given $S = -1$, we have $R = -1 + N$:

$$P[R=r | S=-1] = \begin{cases} 0.2, & r \in \{-1.5, -0.5\}, \\ 0.6, & r = -1, \\ 0, & \text{otherwise.} \end{cases}$$



(b) given $S = 1$, we have $R = 1 + N$:



The \mathbf{Q} matrix is given by

$$\mathbf{Q} = \begin{matrix} & \begin{matrix} r \\ \downarrow \end{matrix} & \begin{matrix} -1.5 & -1 & -0.5 & 0.5 & 1 & 1.5 \end{matrix} \\ \begin{matrix} s \\ \downarrow \end{matrix} & \begin{matrix} -1 \\ 1 \end{matrix} & \begin{bmatrix} 0.2 & 0.6 & 0.2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.2 & 0.6 & 0.2 \end{bmatrix} \end{matrix}$$

Note that each row of the \mathbf{Q} matrix is simply a shifted copy of the noise pmf. The amount of shift is the corresponding value of s for that row.

8.4. Formula-wise, when the additive noise is discrete, each row of the \mathbf{Q} matrix (as in Example 8.3) is given by

$$p_{R|S}(r|s) = p_N(r - s). \tag{45}$$

8.5. When the additive noise is continuous, there are uncountably many possible values for the channel output R . Hence, representing conditional probabilities in the form of a matrix \mathbf{Q} does not make sense here.

When R is continuous, the conditional pmf $p_{R|S}(r|s)$ is replaced by the conditional pdf $f_{R|S}(r|s)$. For additive noise N with pdf $f_N(n)$, we have

$$f_{R|S}(r|s) = f_N(r - s). \quad (46)$$

Example 8.6. Suppose the discrete additive noise in Example 8.3 is replaced by a continuous additive noise:

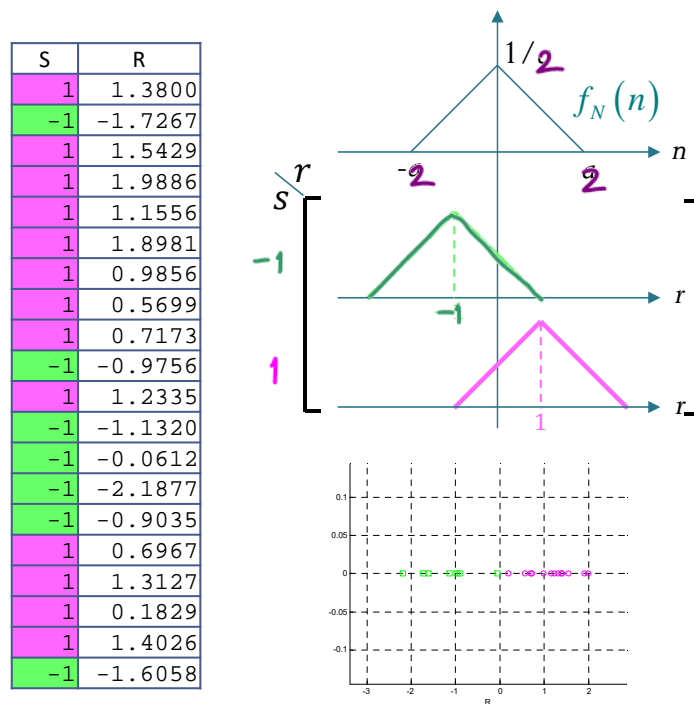


Figure 40: Binary PAM under “Triangular” Noise

8.7. The optimal detector, which minimizes the error probability, is the **MAP detector**:

$$\hat{s}_{\text{MAP}}(r) = \arg \max_{s \in \mathcal{S}} p_S(s) f_{R|S}(r|s) = \arg \max_{s \in \mathcal{S}} p_S(s) f_N(r - s). \quad (47)$$

Because event $[W = j]$ is the same as event $[S = s^{(j)}]$, we also have

$$\hat{w}_{\text{MAP}}(r) = \arg \max_{j \in \{1, 2, \dots, M\}} p_j f_N(r - s^{(j)}). \quad (48)$$

When the prior probabilities are ignored, we have the (sub-optimal) **ML detector**:

$$\hat{s}_{\text{ML}}(r) = \arg \max_{s \in \mathcal{S}} f_{R|S}(r|s) = \arg \max_{s \in \mathcal{S}} f_N(r-s). \quad (49)$$

and

$$\hat{w}_{\text{ML}}(r) = \arg \max_{j \in \{1,2,\dots,M\}} f_N(r-s^{(j)}). \quad (50)$$

8.8. Graphically, here are the steps to find the MAP detector:

(a) Plot $p_1 f_N(r-s^{(1)})$, $p_2 f_N(r-s^{(2)})$, \dots , $p_M f_N(r-s^{(M)})$.

- Note that they are functions of r .
- This is similar to scaling the rows of the \mathbf{Q} matrix by the corresponding prior probabilities in Chapter 3 to get the \mathbf{P} matrix.

(b) Select the maximum plot for each (observed) r value.

- If there are multiple max values, select any.
- The corresponding $s^{(j)}$ is the value of \hat{s}_{MAP} at r .

Example 8.9. Back to Example 8.6.

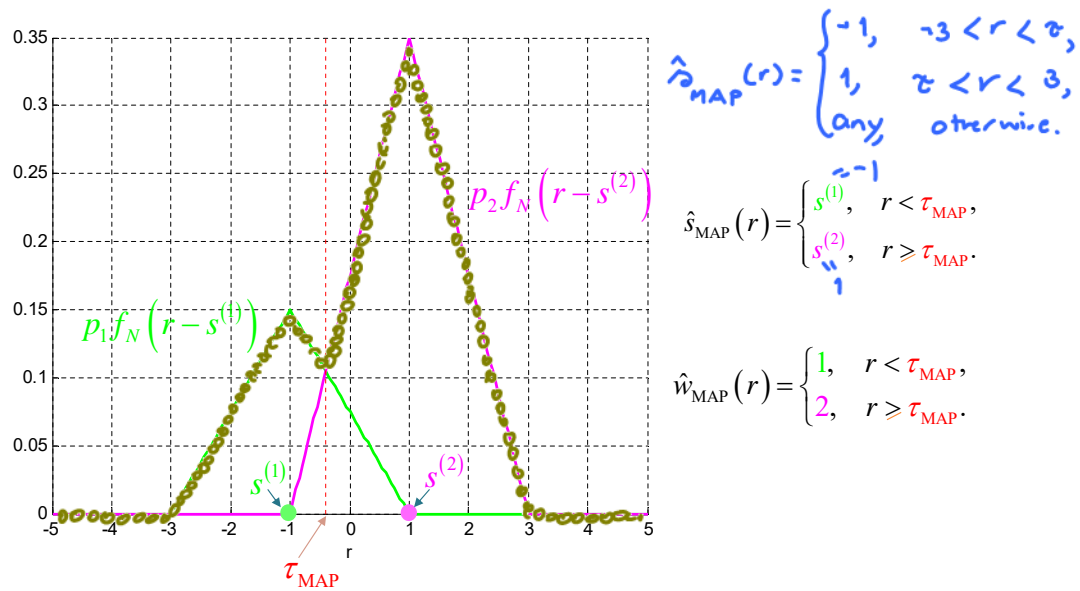


Figure 41: MAPD for Binary PAM under “Triangular” Noise

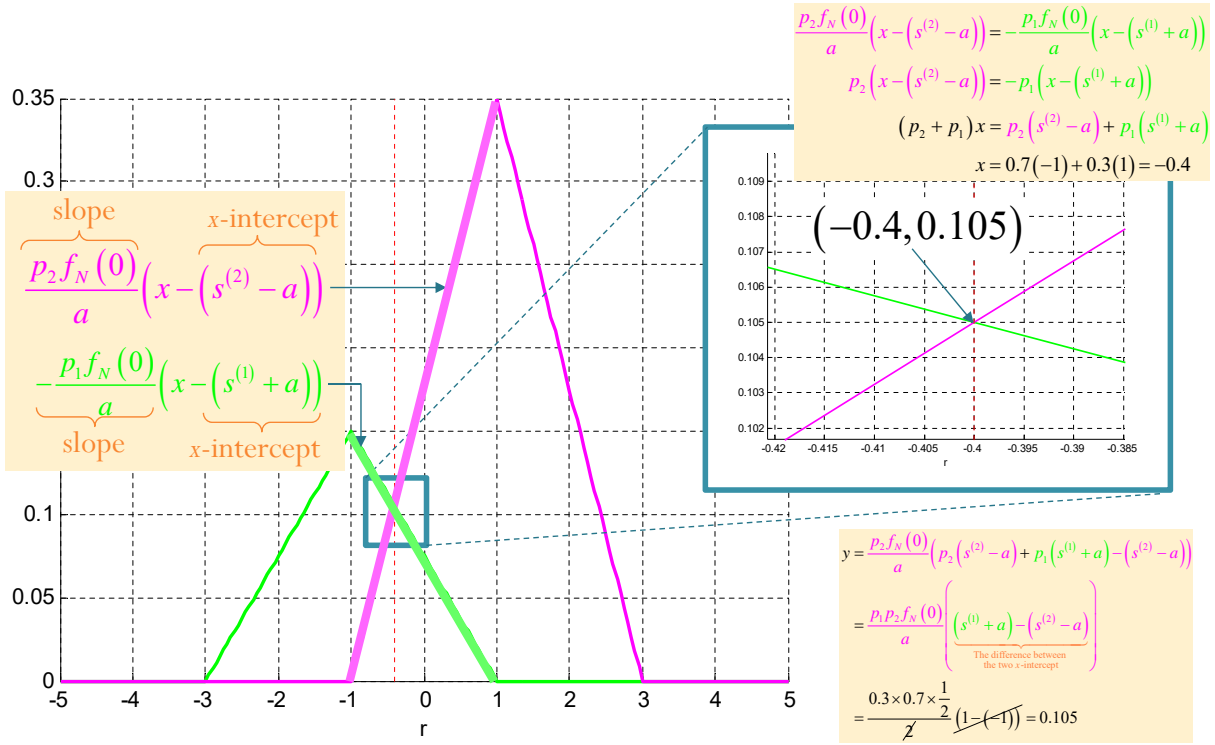


Figure 42: Solving for τ_{MAP} in MAPD for Binary PAM under “Triangular” Noise

Definition 8.10. The i th decision “region”, denoted by \mathcal{D}_i for a decoder $\hat{s}(r)$ is defined as the collection of all the r values at which r is decoded as $s^{(i)}$.

- The collection $\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_M$ should partition the whole observable values (support) of R .

Example 8.11. Back to Example 8.6.

$$\hat{s}_{\text{MAP}}(r) = \begin{cases} s^{(1)}, & r \in \mathcal{D}_1, \\ s^{(2)}, & r \in \mathcal{D}_2, \end{cases} \quad \text{where} \quad \begin{aligned} \mathcal{D}_1 &= (-\infty, \tau_{\text{MAP}}), \\ \mathcal{D}_2 &= [\tau_{\text{MAP}}, \infty). \end{aligned}$$

Figure 43: Decision Regions in MAPD for Binary PAM under “Triangular” Noise

8.12. The error probability of a detector can be found via its success probability

$$\begin{aligned}
P(\mathcal{C}) &= \sum_{i=1}^M P(\mathcal{C} | S = s^{(i)}) P[S = s^{(i)}] = \sum_{i=1}^M P[R \in D_i | S = s^{(i)}] p_i \\
&= \sum_{i=1}^M p_i P[S + N \in D_i | S = s^{(i)}] = \sum_{i=1}^M p_i P[N + s^{(i)} \in D_i] \\
&= \sum_{i=1}^M p_i \int_{D_i} f_N(r - s^{(i)}) dr = \sum_{i=1}^M \int_{D_i} p_i f_N(r - s^{(i)}) dr.
\end{aligned}$$

This gives

$$\begin{aligned}
P(\mathcal{E}) &= 1 - P(\mathcal{C}) \\
&= \sum_{i=1}^M p_i \int_{D_i^c} f_N(r - s^{(i)}) dr = \sum_{i=1}^M \int_{D_i^c} p_i f_N(r - s^{(i)}) dr.
\end{aligned}$$

Although, at first, the above expressions may look complicated, it is similar to what we did in Chapter 3: graphically, the area under the max (selected) plot is $P(\mathcal{C})$.

Example 8.13. Back to Example 8.6.

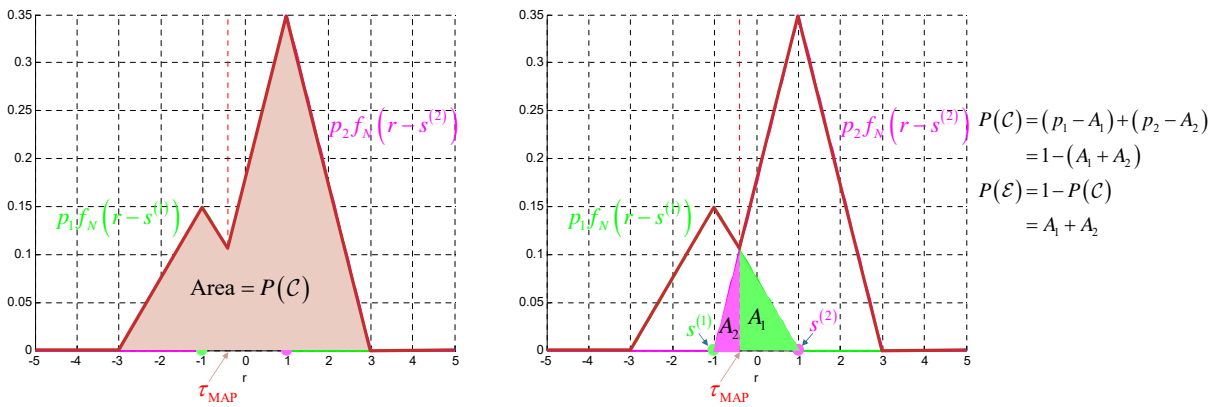


Figure 44: Probability of Successful Detection for Binary PAM under “Triangular” Noise

8.14. Gaussian Noise: When the noise N is Gaussian with mean 0 and standard deviation σ_N ,

$$f_N(n) = \frac{1}{\sqrt{2\pi}\sigma_N} e^{-\frac{1}{2}\left(\frac{n}{\sigma_N}\right)^2}.$$

Definition 8.15. In general, a **Gaussian (normal) random variable** X with mean m and standard deviation σ is characterized by its probability density function (PDF):

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}.$$

To talk about such X , we usually write $X \sim \mathcal{N}(m, \sigma^2)$. Probability involving X can be evaluated by

$$P[X \in A] = \int_A f_X(x) dx.$$

In particular,

$$P[X \in [a, b]] = \int_a^b f_X(x) dx = F_X(b) - F_X(a)$$

where $F_X(x) = \int_{-\infty}^x f_X(t) dt$ is called the cumulative distribution function (CDF) of X .

We usually express probability involving Gaussian random variable via the Q function which is defined by

$$Q(z) = \int_z^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Note that $Q(z)$ is the same as $P[S > z]$ where $S \sim \mathcal{N}(0, 1)$; that is $Q(z)$ is the probability of the “tail” of $\mathcal{N}(0, 1)$.

It can be shown that

- Q is a decreasing function
- $Q(0) = \frac{1}{2}$
- $Q(-z) = 1 - Q(z)$

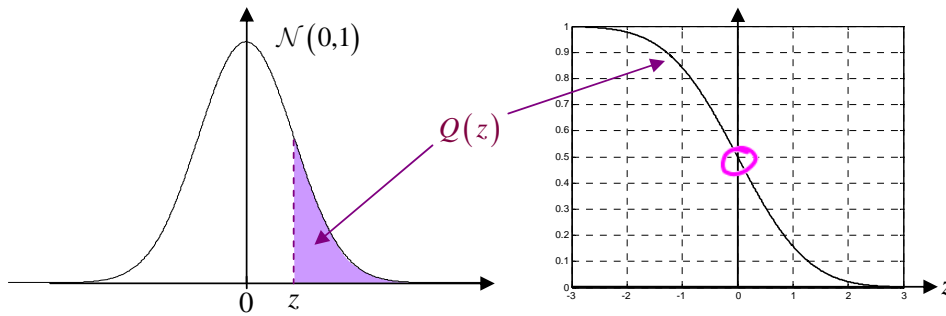
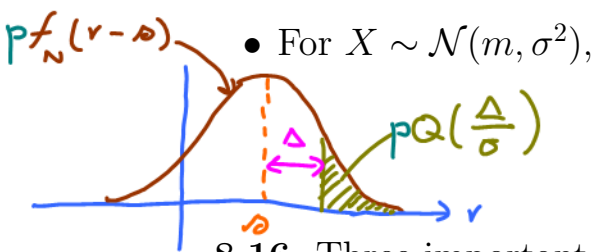


Figure 45: Q -function

- o This is useful for converting the argument of the Q function to positive value.



• For $X \sim \mathcal{N}(m, \sigma^2)$,

$$P[X > c] = Q\left(\frac{c - m}{\sigma}\right).$$

$$P[N > a] - P[N > b]$$

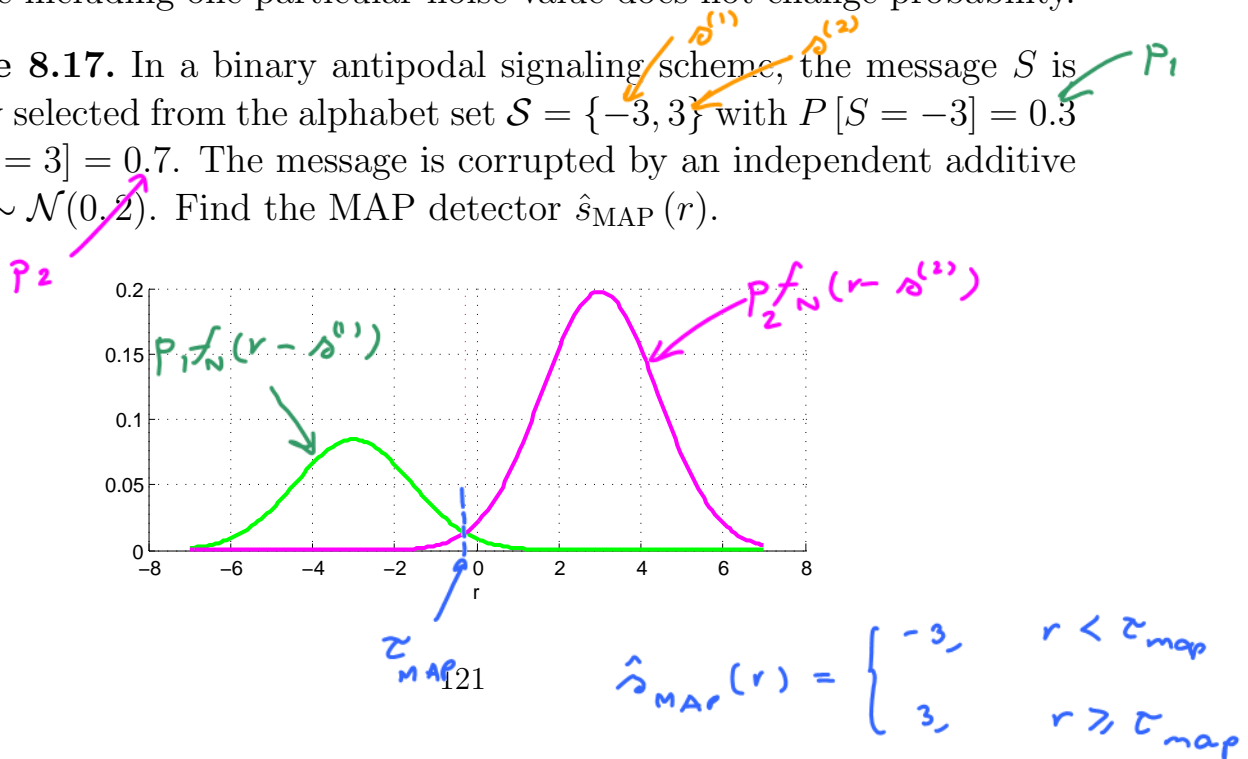
8.16. Three important noise probabilities for $N \sim \mathcal{N}(0, \sigma_N^2)$:



$$P[N > c] = Q\left(\frac{c}{\sigma_N}\right), P[N < c] = 1 - Q\left(\frac{c}{\sigma_N}\right), P[a < N < b] = Q\left(\frac{a}{\sigma_N}\right) - Q\left(\frac{b}{\sigma_N}\right)$$

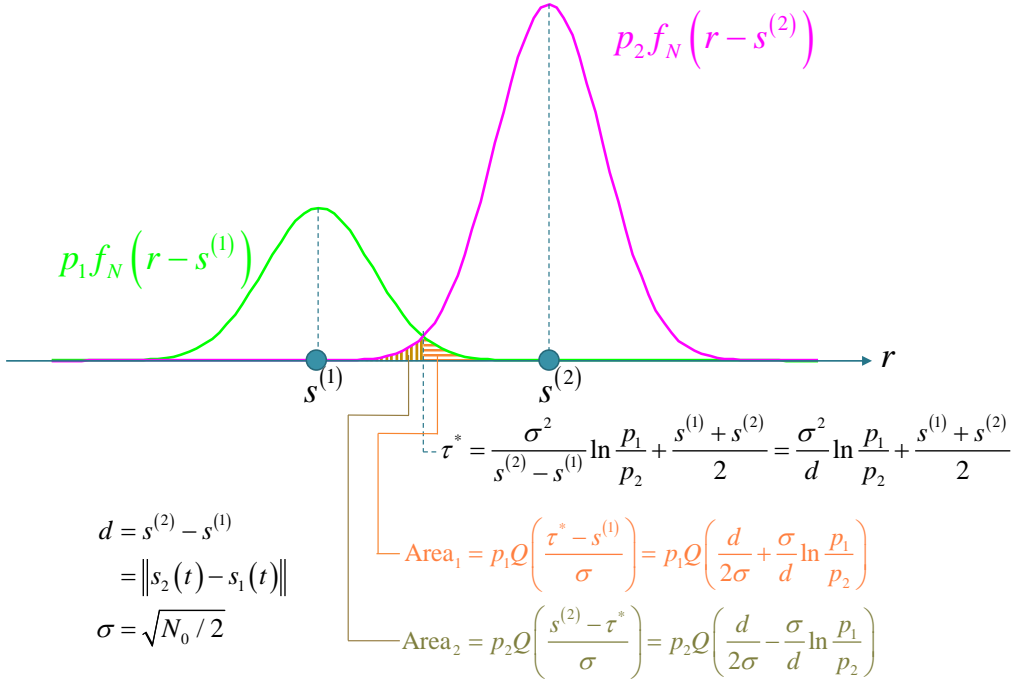
Note that all strict inequalities above can also be replaced by the ones that also include equalities because the noise is a continuous random variable and hence including one particular noise value does not change probability.

Example 8.17. In a binary antipodal signaling scheme, the message S is randomly selected from the alphabet set $\mathcal{S} = \{-3, 3\}$ with $P[S = -3] = 0.3$ and $P[S = 3] = 0.7$. The message is corrupted by an independent additive noise $N \sim \mathcal{N}(0, 2)$. Find the MAP detector $\hat{s}_{\text{MAP}}(r)$.



For a given r , we select $\hat{s}(r) = s^{(2)}$ if and only if

$$\begin{aligned}
 p_2 f_N(r - s^{(2)}) &> p_1 f_N(r - s^{(1)}) \\
 p_2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{r-s^{(2)}}{\sigma}\right)^2} &> p_1 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{r-s^{(1)}}{\sigma}\right)^2} \\
 r > \frac{\sigma^2}{s^{(2)} - s^{(1)}} \ln \frac{p_1}{p_2} + \frac{s^{(1)} + s^{(2)}}{2} &= \frac{\sigma^2}{d} \ln \frac{p_1}{p_2} + \frac{s^{(1)} + s^{(2)}}{2}
 \end{aligned}$$



$$\begin{aligned}
 P(\mathcal{E}) &= p_1 Q\left(\frac{\tau^* - s^{(1)}}{\sigma}\right) + p_2 Q\left(\frac{s^{(2)} - \tau^*}{\sigma}\right) \\
 &= p_1 Q\left(\frac{d}{2\sigma} + \frac{\sigma}{d} \ln \frac{p_1}{p_2}\right) + p_2 Q\left(\frac{d}{2\sigma} - \frac{\sigma}{d} \ln \frac{p_1}{p_2}\right)
 \end{aligned}$$

We can see from the last expression that the error probability of the optimal (MAP) detector depends on $s^{(1)}$ and $s^{(2)}$ only through their distance d .

8.18. For M -ary PAM under additive noise channel where $R = S + N$, $S \perp\!\!\!\perp N$, and $N \sim \mathcal{N}(0, \sigma_N^2)$

$$\hat{s}_{\text{MAP}}(r) = \arg \max_{s \in \mathcal{S}} (p_S(s) f_N(r - s)) \quad (51)$$

$$\stackrel{\text{ln}}{\times 2 \sigma_N^2} \left(= \arg \max_{s \in \mathcal{S}} \left(p_S(s) \frac{1}{\sqrt{2\pi\sigma_N^2}} e^{-\frac{1}{2} \left(\frac{r-s}{\sigma_N} \right)^2} \right) \right) \quad (52)$$

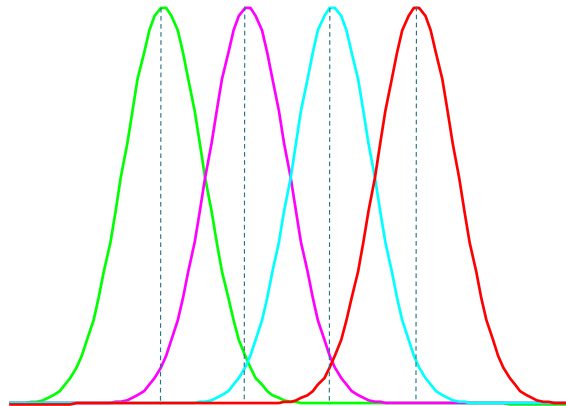
$$= \arg \max_{s \in \mathcal{S}} \left(2\sigma_N^2 \ln p_S(s) - (r - s)^2 \right) \quad (53)$$

$$= \arg \max_{s \in \mathcal{S}} \left(\sigma_N^2 \ln p_S(s) - \frac{E_s}{2} + s \cdot r \right), \quad (54)$$

and from (53), when we set all $p_S(s)$ to be the same (or ignore the $p_S(s)$ in our calculation)

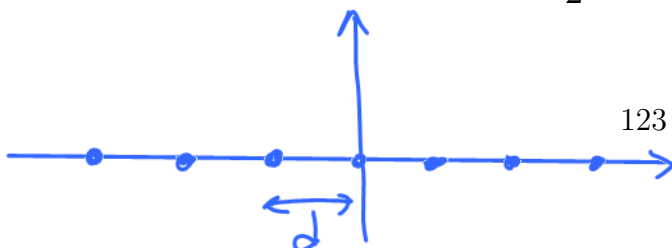
$$\hat{s}_{\text{ML}}(r) = \arg \min_{s \in \mathcal{S}} (r - s)^2 = \arg \min_{s \in \mathcal{S}} d(r, s).$$

Definition 8.19. In “standard” multi-level PAM, we required that the spacing between all adjacent signals to be the same. Furthermore, all M signals are equally likely. To minimize the average energy, we also require that the constellation is “centered” around zero.

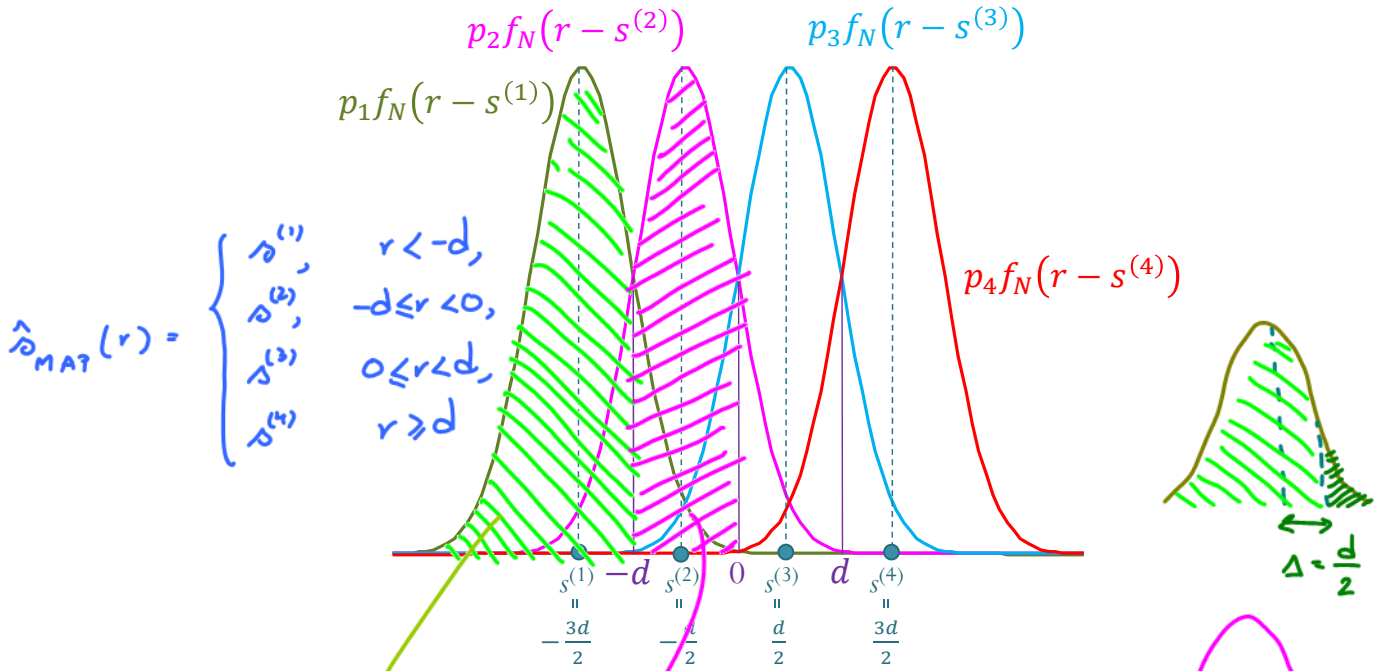


Suppose the distance between adjacent signals is d , then the M signals are represented in the constellation by

$$s^{(j)} = \frac{d}{2} (2j - 1 - M).$$



Example 8.20. Probability of Detection Error for Standard Quaternary PAM



If we define $q = Q\left(\frac{d}{2\sigma}\right)$, then

$$P(\mathcal{C}) = \frac{1}{4}(1 - q) + \frac{1}{4}(1 - 2q) + \frac{1}{4}(1 - 2q) + \frac{1}{4}(1 - q) = 1 - \frac{6}{4}q = 1 - \frac{3}{2}q.$$

Therefore, the detection error probability is

$$P(\mathcal{E}) = 1 - P(\mathcal{C}) = \frac{3}{2}q = \frac{3}{2}Q\left(\frac{d}{2\sigma}\right). \quad (55)$$

For easier comparison with other modulation schemes, we express $P(\mathcal{E})$ in terms of SNR $= \frac{E_b}{\sigma^2}$. To find the average energy per bit E_b , we first find the average energy per symbol:

$$E_s = \sum_{j=1}^M p_j E_j = \frac{1}{4} \left(\left(-\frac{3d}{2}\right)^2 + \left(-\frac{d}{2}\right)^2 + \left(\frac{d}{2}\right)^2 + \left(\frac{3d}{2}\right)^2 \right) = \frac{5}{4}d^2.$$

Scaling E_s by $\log_2 M$, which is the number of bits per symbol, we get the average energy per bit:

$$E_b = \frac{E_s}{\log_2 M} = \frac{E_s}{\log_2 4} = \frac{\frac{5}{4}d^2}{\log_2 4} = \frac{5}{8}d^2.$$

Hence, we can replace the distance d in (55) by

$$d = \sqrt{\frac{8}{5}E_b} = 2\sqrt{\frac{2}{5}E_b}$$

which gives

$$P(\mathcal{E}) = \frac{3}{2}Q\left(\frac{2\sqrt{\frac{2}{5}E_b}}{2\sigma}\right) = \frac{3}{2}Q\left(\sqrt{\frac{2E_b}{5\sigma^2}}\right). \quad (56)$$

Suppose the noise process in the waveform channel is AWGN with PSD $\frac{N_0}{2}$. Then, $\sigma^2 = \frac{N_0}{2}$ and

$$P(\mathcal{E}) = \frac{3}{2}Q\left(\sqrt{\frac{4E_b}{5N_0}}\right). \quad (57)$$

8.21. At this point, we know that we are able to find the optimal detector (the MAP detector) and the corresponding detection error probability by first converting the waveform channel into the vector channel. Note that to do this, we will first need to find the orthonormal basis functions (possibly by GSOP) can projects all the signals into the signal space. However, it turns out that we don't have to do any of these at all. Once we have gone through all the derivation of the optimal detector and the corresponding detection error probability and got answers in the vector channel, we can convert our answers back to the waveform calculations. This is what we will do next.

8.22. Correlation detector: Recall, from (54), that for additive noise channel where $R = S + N$, $S \perp\!\!\!\perp N$, and $N \sim \mathcal{N}(0, \sigma_N^2)$,

$$\hat{s}_{\text{MAP}}(r) = \arg \max_{s \in S} \left(\sigma_N^2 \ln p_S(s) - \frac{E_s}{2} + r \cdot s \right). \quad (58)$$

Now,

$$r \cdot s^{(j)} = \langle r(t), s_j(t) \rangle = \int_{-\infty}^{\infty} r(t) s_j(t) dt.$$

Therefore, the calculation involved in finding the optimal detector can be performed directly on the original waveform functions of the signals.

This mathematical equivalence means the optimal detector can be implemented with correlators as shown in Figure 46. Note that

$$\eta_j = \sigma_N^2 \ln p_j - \frac{E_j}{2}.$$

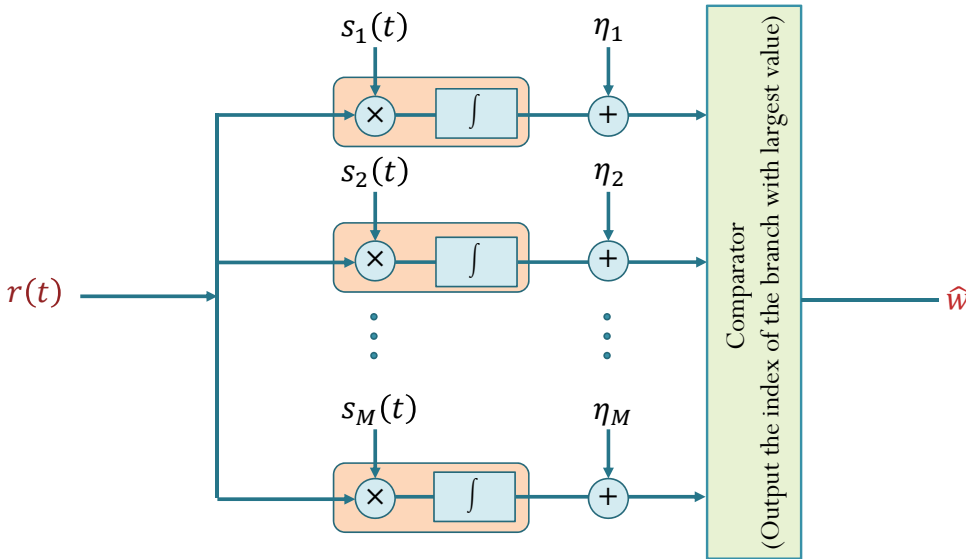


Figure 46:
Correlation
detector.

8.23. Matched filter implementation of the optimal detector: In practice, to calculate the correlation (inner-product) $\langle r(t), s(t) \rangle$ above, we use filtering. Such filter is called the matched filter.

Recall that when a signal $r(t)$ passes through a filter whose impulse response is $h(t)$, the output of the filter is given by

$$\{r * h\}(t) = \int_{-\infty}^{\infty} r(\tau)h(t - \tau) d\tau$$

Let's try $h(t) = s(T - t)$ for some constant T . This filter is called the matched filter. Note that

$$h(t - \tau) = s(T - (t - \tau)) = s(T - t + \tau).$$

Therefore,

$$\{r * h\}(t) = \int_{-\infty}^{\infty} r(\tau)s(T - t + \tau) d\tau.$$

In particular,

$$\{r * h\}(T) = \int_{-\infty}^{\infty} r(\tau) s(\tau) d\tau = \langle r(t), s(t) \rangle.$$

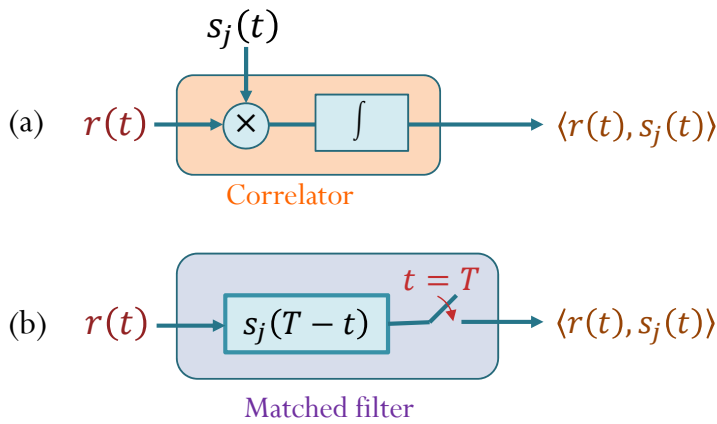


Figure 47: Two equivalent implementations of the inner-product between the received signal and the signal waveform: (a) via correlator and (b) via matched filter.

So, now we have an alternative way to calculate the inner-product term in 58. This new calculation via matched filter is shown in Figure 47b where it is also compared with the correlator implementation. When all the correlators in Figure 46 are replaced by the corresponding matched filters, we obtain Figure 48 which is an implementation of the optimal (MAP) detector via matched filters.

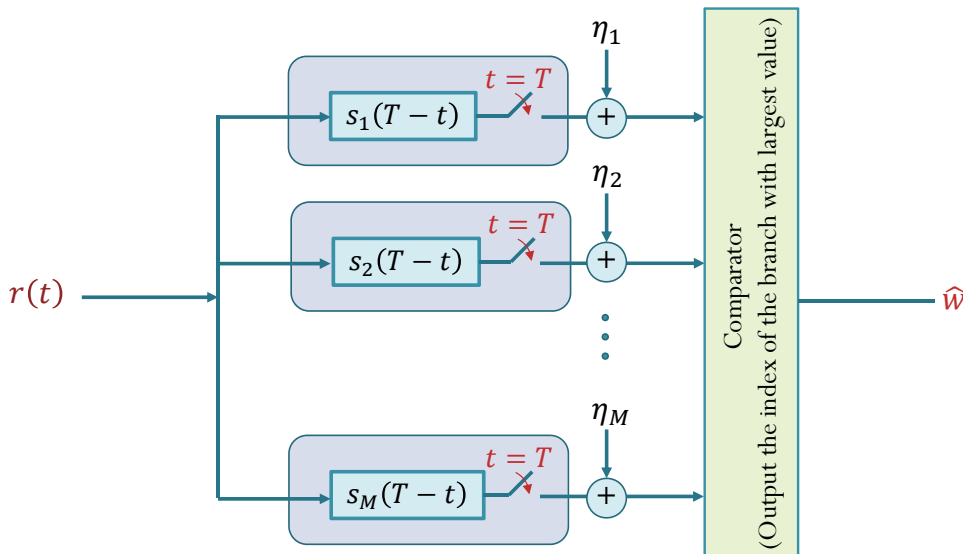


Figure 48: Matched filter implementation of the optimal detector

8.24. Capacity of the AWGN channel:

Back when we finished the study of optimal decoder for the discrete communication channel in Chapter 3, we then proceeded to find the limit of the communication rate via the concept of channel capacity in Chapter 4.

For the waveform channel under additive white Gaussian noise (AWGN) introduced in Chapter 7, similar information-theoretic quantities and analysis can be defined and performed to find the channel capacity. Its meaning is still the same, being the maximum rate that information can be *reliably* sent over the channel. Additionally, we will assume that the mean square value of the channel input can not exceed P . It turns out that, for the AWGN channel with power constraint P and the mean square value of the noise is σ_N^2 , the channel capacity is

$$C = \frac{1}{2} \log_2 \left(1 + \frac{P}{\sigma_N^2} \right).$$

For a **band-limited AWGN channel** for which signals are bandlimited to B [Hz]. If the power spectral density of the noise is $\frac{N_0}{2}$ as discussed in Chapter 7, then

$$\sigma_N^2 = \mathbb{E} [N^2(t)] = R_N(0) = \int_{-\infty}^{\infty} S_N(f) df = \int_{-B}^B \frac{N_0}{2} df = \frac{N_0}{2} \times 2B = N_0 B.$$

Recall that, from the sampling theorem, a maximum of $2B$ independent samples of information per second can be transmitted, errorfree, over such channel. Therefore, the corresponding channel capacity is

$$C = 2B \times \frac{1}{2} \log_2 \left(1 + \frac{P}{N_0 B} \right) = B \log_2 \left(1 + \frac{P}{N_0 B} \right)$$

bits per second.

Remarks:

- (a) Capacity can be made infinite by increasing the signal power P to infinity.
- (b) Suppose we increase the bandwidth. In the limit as $B \rightarrow \infty$, the channel capacity approaches a limit of $(\log_2 e) \frac{P}{N_0}$.