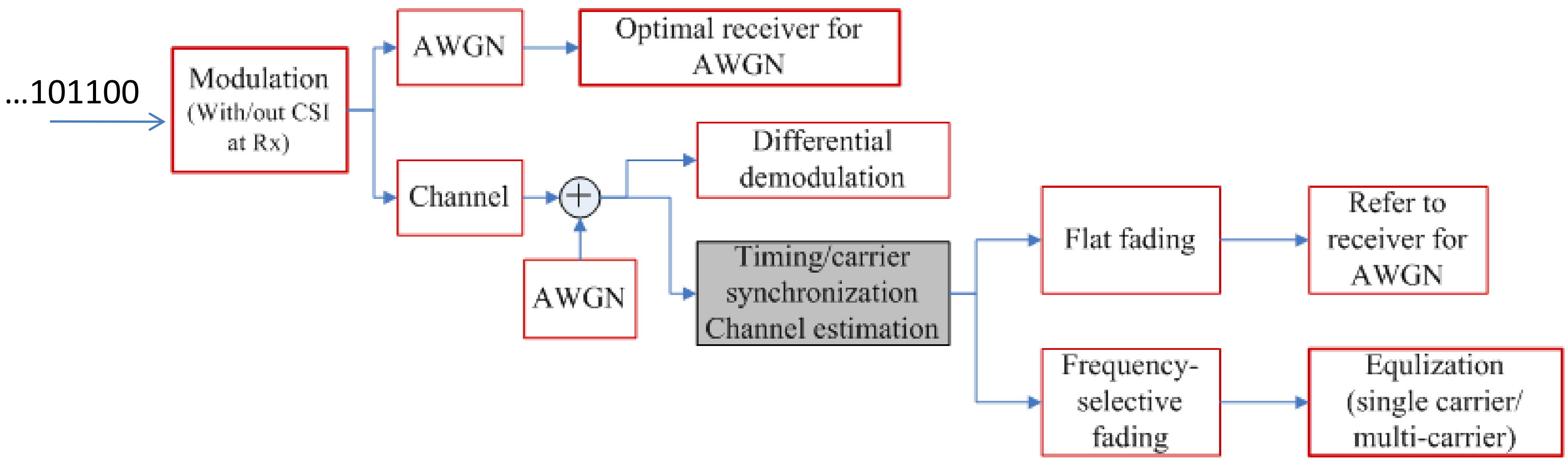
Principle of Digital Communication

---Lecture 3

Receiver design for AWGN channel

Part I

3.1 Introduction



A fundamental problem in receiver design is to decide, based on the received signal, which of the set of possible signals was actually sent. The task of the link designer is to make the probability of error in this decision as small as possible, given the system constraints.

When considering the practical channel, which at least attenuated and delayed the signal waveforms, and (if it is a passband signal) causes a change of carrier phase. **Thus, the model considered here applies to a receiver that can estimate the effects of the channel**. Such a receiver is termed a ***coherent receiver.*** Implementation of a coherent receiver involves synchronization in time, carrier frequency, and phase.

In this chapter, we assume that such synchronization functions have already been taken care of.

You can refer to Chapter 7 of Prof. Gallager's book or other related books, to acquire common terminology and important concepts in probability, random variables, and random processes.

3.2 Gaussian basics

The key reason why Gaussian random variables appear so often in both natural and manmade systems is the central limit theorem (CLT). In its elementary form, the CLT states that the sum of a number of independent and identically distributed random variables is well approximated as a Gaussian random variable.
The Gaussianity of receiver thermal noise is due to its arising from the movement of a large number of electrons.

However, because the CLT also works with a relatively small number of random variables, we shall see the CLT works in a number of other contexts, including performance analysis of equalizers in the presence of ISI as well as AWGN, and the modeling of multipath wireless channels.

Gaussian random variable

The random variable X is said to follow a *Gaussian,* or *normal distribution* if its density is of the form:

$$p(x) = \frac{1}{\sqrt{2\pi v^2}} \exp\left(-\frac{(x-m)^2}{2v^2}\right), \qquad -\infty < x < \infty, \tag{3.1}$$

where $m = \mathbb{E}[X]$ is the mean of X, and $v^2 = \mathrm{var}(X)$ is the variance of X.

The Gaussian density is therefore completely characterized by its mean and variance.

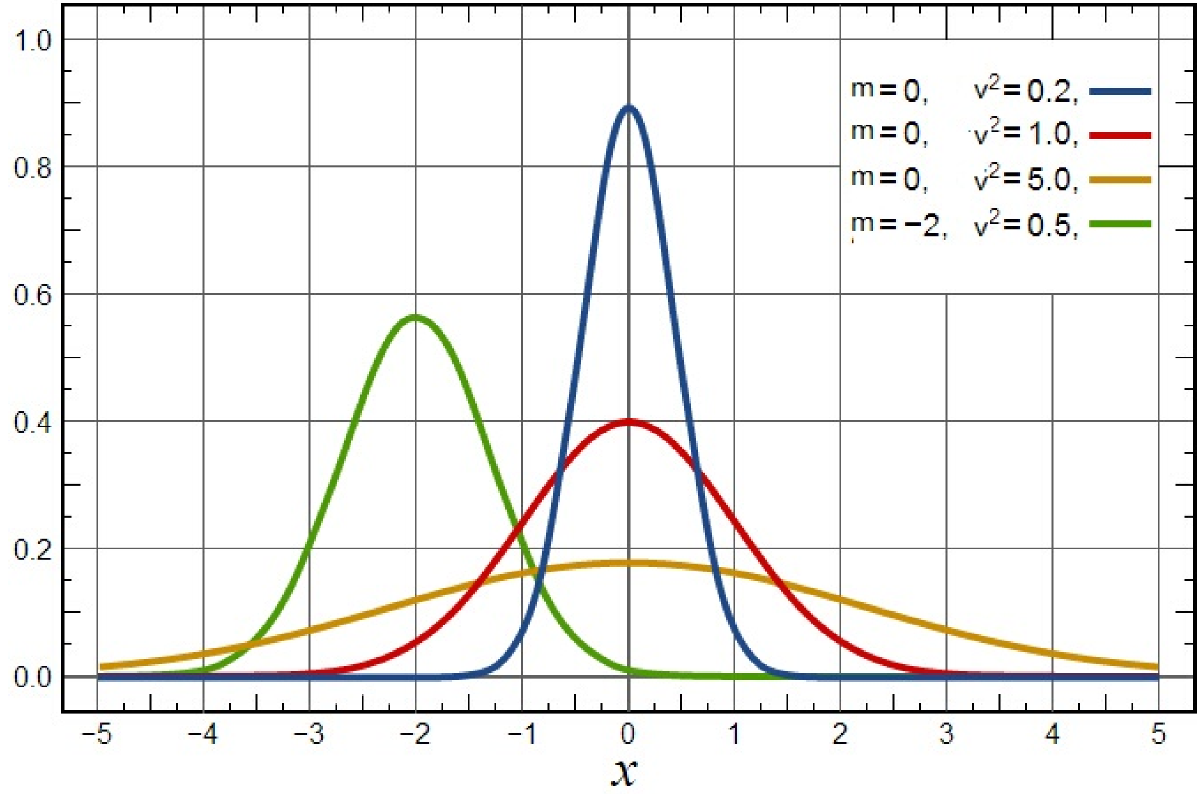


Fig. 3.1 The shape of Gaussian density.

the cumulative distribution function $\Phi(x)$ of a standard Gaussian random variable.

the complementary cumulative distribution function (CCDF) Q(x)

$$\Phi(x) = P[N(0,1) \le x] = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t, \qquad (3.2)$$

$$Q(x) = P[N(0,1) > x] = \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t. \qquad (3.3)$$

$\Phi(x) + Q(x) = 1$

$Q(-x) = \Phi(x)$

Q(−x) = 1−Q(x)

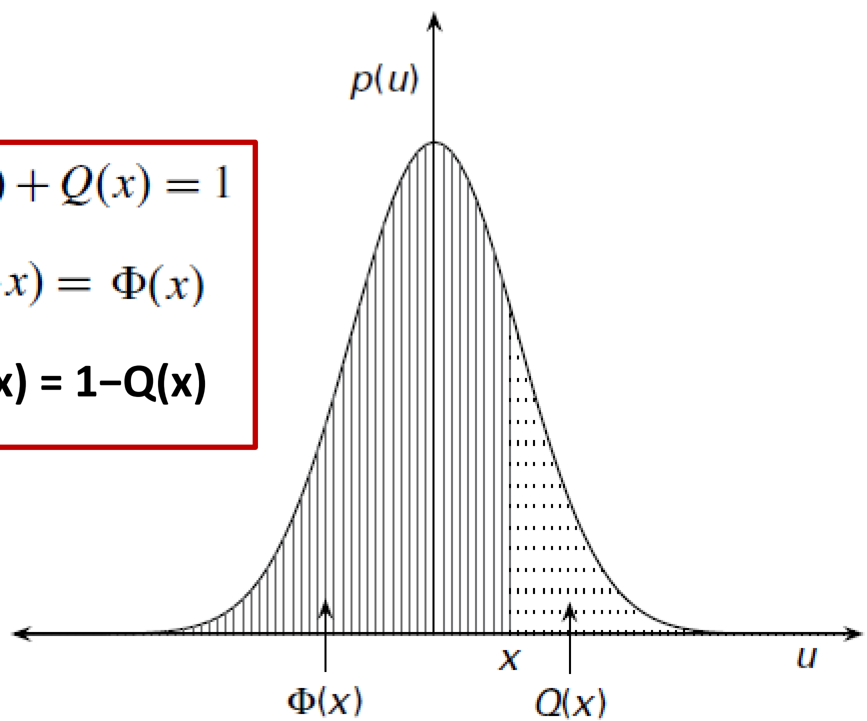


Fig. 3.2 The Φ and Q functions are obtained by integrating the N(0, 1) density over appropriate intervals.

Bounds on Q(x) for large arguments

Lower bound

Upper bound

$$\left(1-\frac{1}{x^2}\right)\frac{e^{-x^2/2}}{x\sqrt{2\pi}} \le Q(x) \le \frac{e^{-x^2/2}}{x\sqrt{2\pi}}, \qquad x \ge 0. \qquad (3.4)$$

These bounds are tight (the upper and lower bounds converge) for large values of x.

Upper bound on *Q(x)* useful for small arguments and for analysis

$$Q(x) \le \frac{1}{2}e^{-x^2/2}, \qquad x \ge 0. \qquad (3.5)$$

Upper bound 2

This bound is tight for small x, and gives the correct exponent of decay for large x. It is also useful for simplifying expressions involving a large number of Q functions.

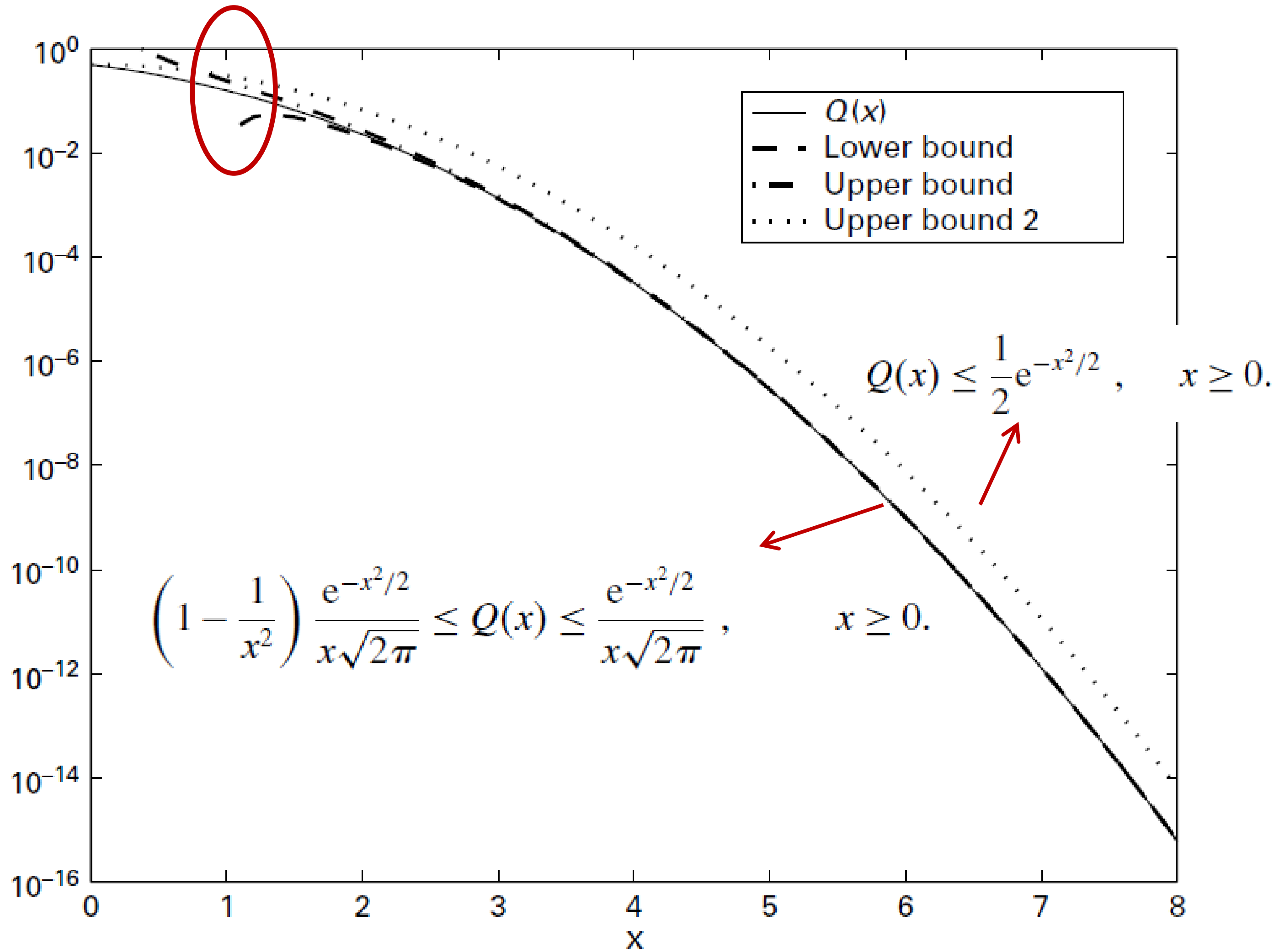


Fig. 3.4 The Q function and bounds.

3.3 Hypothesis testing basics

Hypothesis testing is a framework for deciding which of M possible hypotheses $H_1, \ldots, H_M$ "best" explains an observation Y .

$$\pi(i) = P[H_i] \quad i = 1, \ldots, M \qquad \sum_{i=1}^{M} \pi(i) = 1$$

We assume that the observation Y takes values in a finite-dimensional observation space ; that is, Y is a scalar or vector.

Example 3.3 (Basic Gaussian example) Consider binary hypothesis testing, in which H_0 corresponds to 0 being sent, H_1 corresponds to 1 being sent, and Y is a scalar decision statistic (e.g., generated by sampling the output of a receive filter or an equalizer). The conditional distributions for the observation given the hypotheses are $H_0 : Y \sim N(0, v^2)$ and $H_1 : Y \sim N(m, v^2)$, so that

$$p(y|0) = \frac{\exp\left(-\frac{y^2}{2v^2}\right)}{\sqrt{2\pi v^2}}; \qquad p(y|1) = \frac{\exp\left(-\frac{(y-m)^2}{2v^2}\right)}{\sqrt{2\pi v^2}}.$$

Decision rule

A decision rule $\delta : \Gamma \to \{1, \ldots, M\}$ is a mapping from the observation space to the set of hypotheses. Alternatively, a decision rule can be described in terms of a partition of the observation space into disjoint decision regions $\{\Gamma_i, i = 1, \ldots, M\}$, where

$$\Gamma_i = \{y \in \Gamma : \delta(y) = i\}.$$

That is, when $y \in \Gamma_i$ the decision rule says that H_i is true.

Example 3.4 A "sensible" decision rule for the basic Gaussian example (assuming that $m > 0$) is

$$\delta(y) = \begin{cases} 1, & y > \frac{m}{2}, \\ 0, & y \leq \frac{m}{2}. \end{cases}$$

This corresponds to the decision regions $\Gamma_1 = (\frac{m}{2}, \infty)$, and $\Gamma_0 = (-\infty, \frac{m}{2})$.

Is this “sensible” decision rule “good”?
What is “good”?

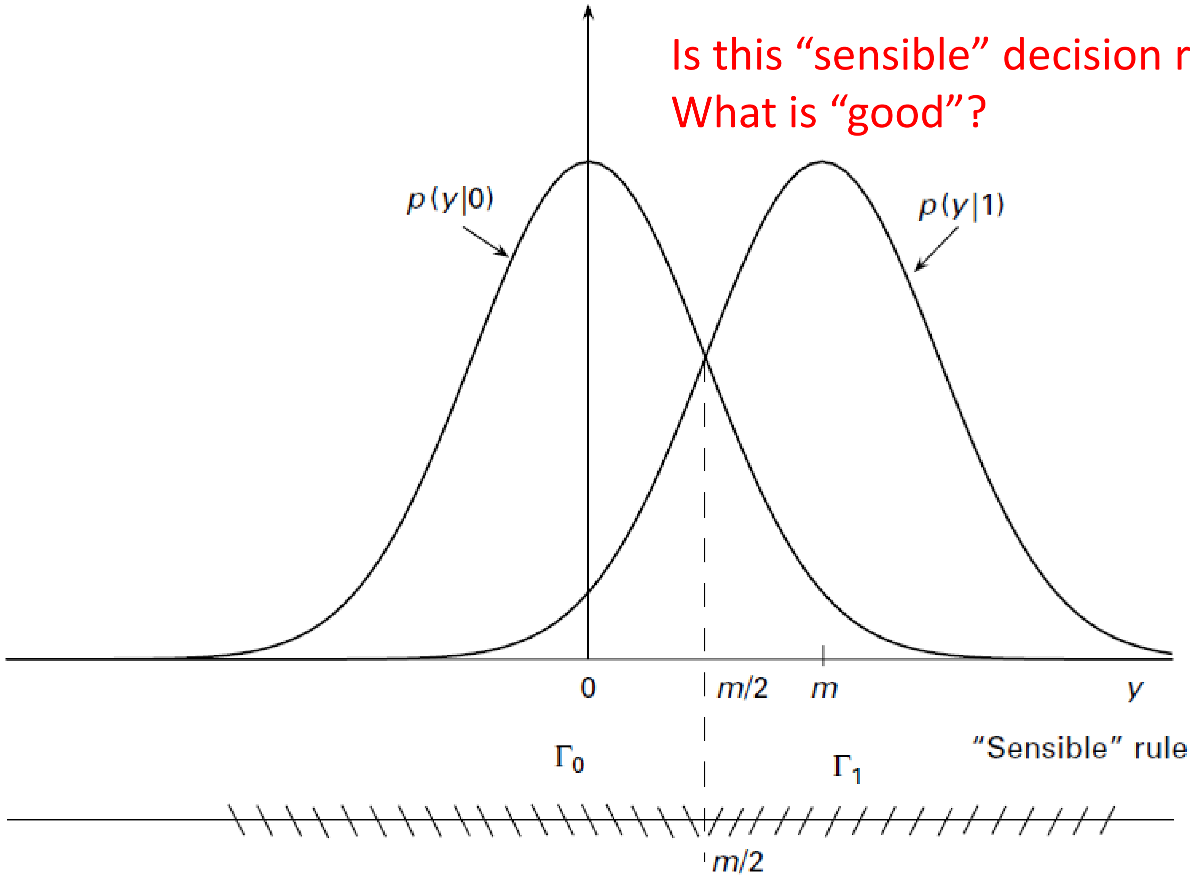


Fig. 3.6 The conditional densities and “sensible” decision rule for the basic Gaussian example.

Conditional error probability

For an M-ary hypothesis testing problem, the conditional error probability, conditioned on H_i, for a decision rule δ is defined as

$$P_{e|i} = P[\text{say } H_j \text{ for some } j \neq i | H_i \text{ is true}] = \sum_{j \neq i} P[Y \in \Gamma_j | H_i] = 1 - P[Y \in \Gamma_i | H_i], \quad (3.12)$$

The conditional probability of correct decision, given H_i

$$P_{c|i} = P[Y \in \Gamma_i | H_i]$$

If the prior probabilities are known, then we can define the (average) error probability as

$$P_e = \sum_{i=1}^{M} \pi(i) P_{e|i}. \quad (3.13)$$

Similarly, the average probability of a correct decision is given by

$$P_c = \sum_{i=1}^{M} \pi(i) P_{c|i} = 1 - P_e. \quad (3.14)$$

Example 3.5 The conditional error probabilities for the "sensible" decision rule $\delta(y)$ for the basic Gaussian example (Example **3.4**) are

$$P_{e|0} = P\left[Y > \frac{m}{2} | H_0\right] = Q\left(\frac{m}{2v}\right),$$

since $Y \sim N(0, v^2)$ under H_0, and

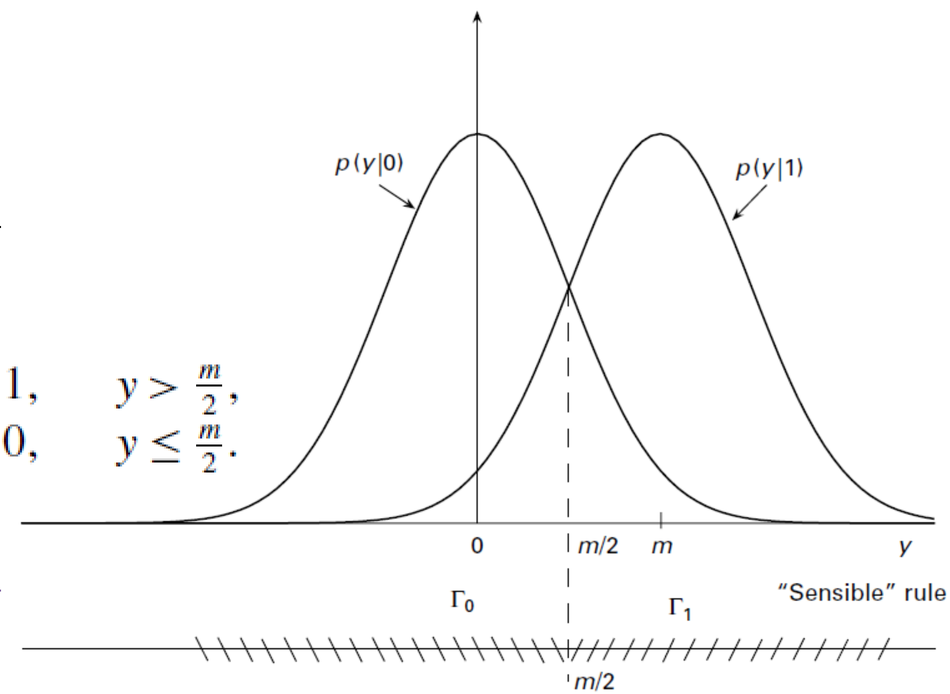
$$P_{e|1} = P\left[Y \le \frac{m}{2} | H_1\right] = \Phi\left(\frac{\frac{m}{2} - m}{v}\right) = Q\left(\frac{m}{2v}\right),$$

since $Y \sim N(m, v^2)$ under H_1. Furthermore, since $P_{e|1} = P_{e|0}$, the average error probability is also given by

$$P_e = Q\left(\frac{m}{2v}\right),$$

regardless of the prior probabilities.

$$\delta(y) = \begin{cases} 1, & y > \frac{m}{2}, \\ 0, & y \le \frac{m}{2}. \end{cases}$$



Minimum Probability of Error (MPE)

we want to minimize the error probability of a communication system, that is

$$\boxed{\left\{\Gamma_i^*\right\} = \underset{\{\Gamma_1,\Gamma_2,\ldots,\Gamma_M\}}{\arg\min}\ P_e}$$

Derivation of MPE rule:

When H_i is true， the correct detection probability is:

$$P_{c|i} = \Pr\left(y \in \Gamma_i \mid H_i\right) = \int_{\Gamma_i} p(y \mid i)dy$$

Minimum Probability of Error (MPE)

$$P_c = \sum_{i=1}^{M} \pi(i) P_{c|i} = \sum_{i=1}^{M} \pi(i) \int_{\Gamma_i} p(y\,|\,i)dy$$

$$= \sum_{i=1}^{M} \int_{\Gamma_i} \pi(i) p(y\,|\,i)dy$$

$$\Gamma = \bigcup_{i=1}^{M} \Gamma_i$$

$$= \sum_{i=1}^{M} \int_{\Gamma} 1_{(y\in\Gamma_i)} \pi(i) p(y\,|\,i)dy$$

$$= \int_{\Gamma} \sum_{i=1}^{M} 1_{(y\in\Gamma_i)} \pi(i) p(y\,|\,i)dy$$

$$\Rightarrow \quad \delta(y) = \arg\max_{1\le i\le M} \pi(i) p(y\,|\,i)$$

MPE rule

➢ Maximum Likelihood (ML)

We have the MPE rule:

$$\delta_{MPE}(y) = \arg\max_{1\le i\le M} \pi(i)\,p(y\,|\,i)$$

When equal prior probablity:

$$\pi(i) = \frac{1}{M}, \quad i = 1, 2, \ldots, M$$

ML rule

$$\boxed{\delta_{ML}(y) = \arg\max_{1\le i\le M} p(y\,|\,i)}$$

➢ Maximum A Posteriori Probability (MAP)

We have the MPE rule:

$$\delta_{MPE}(y) = \arg\max_{1 \le i \le M} \pi(i)\, p(y \mid i)$$

According to Bayes' rule

$$P(H_i \mid y) = \frac{\pi(i)\, p(y \mid i)}{p(y)} \qquad p(y) = \sum_j \pi(j)\, p(y \mid j)$$

MAP rule

$$\delta_{MAP}(y) = \arg\max_{1 \le i \le M} P(H_i \mid y)$$

Likelihood ratio test for binary hypothesis testing

For binary hypothesis testing, the MPE rule specializes to

$$\delta_{\mathrm{MPE}}(y) = \begin{cases} 1, & \pi(1)p(y|1) > \pi(0)p(y|0), \\ 0, & \pi(1)p(y|1) < \pi(0)p(y|0), \\ \text{don't care}, & \pi(1)p(y|1) = \pi(0)p(y|0), \end{cases} \tag{3.18}$$

which can be rewritten as

$$L(y) = \frac{p(y|1)}{p(y|0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\pi(0)}{\pi(1)}, \tag{3.19}$$

where $L(y)$ is called the **likelihood ratio (LR**). A test that compares the likelihood ratio with a threshold is called a **likelihood ratio test (LRT).**

We have just shown that the MPE rule is an LRT with threshold $\pi(0)/\pi(1)$
Similarly, the ML rule is an LRT with threshold **one**.

Example 3.3 (Basic Gaussian example) Consider binary hypothesis testing, in which H_0 corresponds to 0 being sent, H_1 corresponds to 1 being sent, and Y is a scalar decision statistic (e.g., generated by sampling the output of a receive filter or an equalizer). The conditional distributions for the observation given the hypotheses are $H_0 : Y \sim N(0, v^2)$ and $H_1 : Y \sim N(m, v^2)$, so that

$$p(y|0) = \frac{\exp\left(-\frac{y^2}{2v^2}\right)}{\sqrt{2\pi v^2}}; \quad p(y|1) = \frac{\exp\left(-\frac{(y-m)^2}{2v^2}\right)}{\sqrt{2\pi v^2}}.$$

Example 3.3 (Basic Gaussian example) Consider binary hypothesis testing, in which H_0 corresponds to 0 being sent, H_1 corresponds to 1 being sent, and Y is a scalar decision statistic (e.g., generated by sampling the output of a receive filter or an equalizer). The conditional distributions for the observation given the hypotheses are $H_0 : Y \sim N(0, v^2)$ and $H_1 : Y \sim N(m, v^2)$, so that

$$p(y|0) = \frac{\exp\left(-\frac{y^2}{2v^2}\right)}{\sqrt{2\pi v^2}}; \quad p(y|1) = \frac{\exp\left(-\frac{(y-m)^2}{2v^2}\right)}{\sqrt{2\pi v^2}}.$$

Let's see the likelihood ratio for the above basic Gaussian example

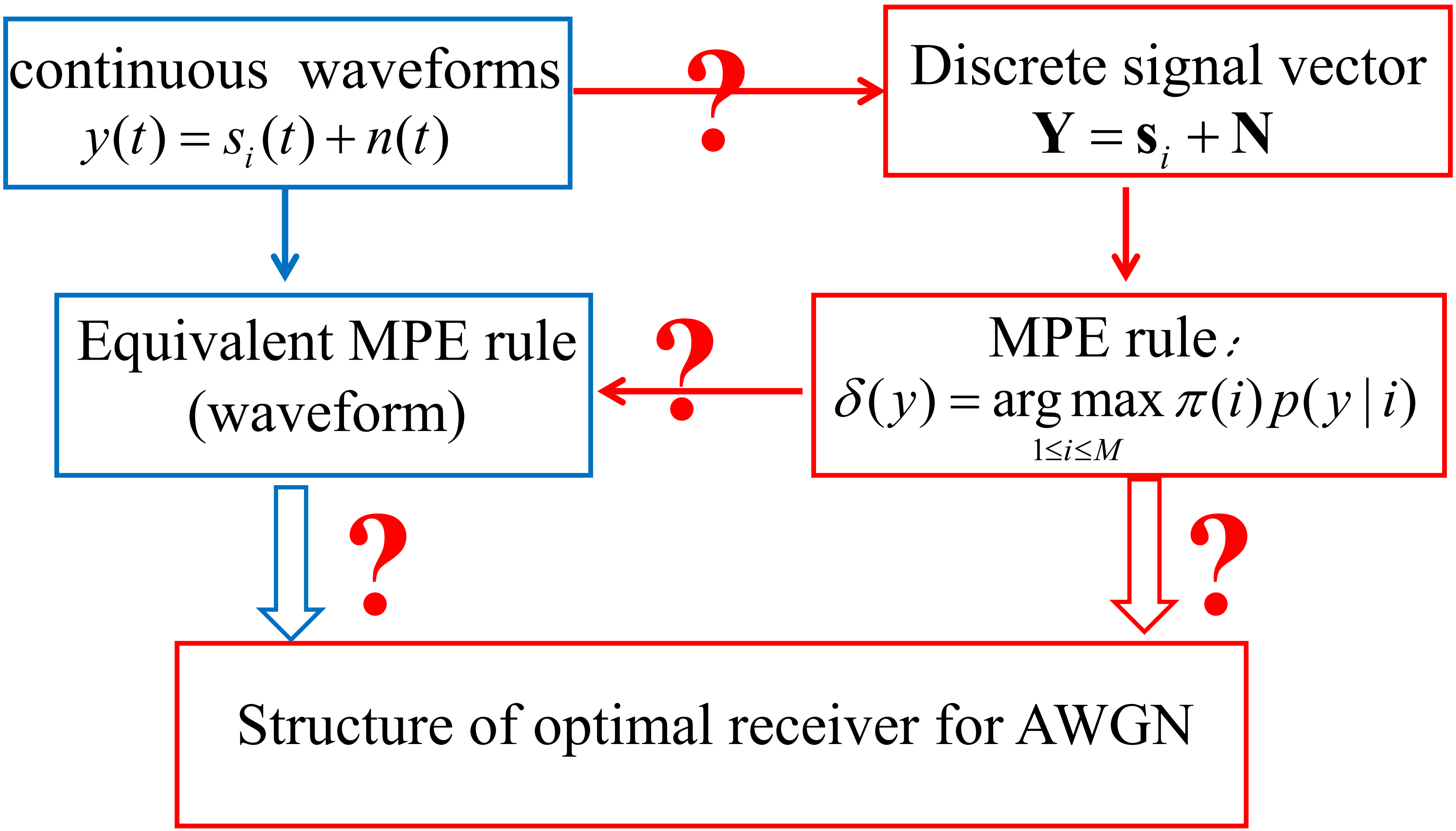
Substituting

$$p(y|0) = \frac{\exp\left(-\frac{y^2}{2v^2}\right)}{\sqrt{2\pi v^2}}; \quad p(y|1) = \frac{\exp\left(-\frac{(y-m)^2}{2v^2}\right)}{\sqrt{2\pi v^2}}.$$

into (3.19) $L(y) = \dfrac{p(y|1)}{p(y|0)}$

we obtain the likelihood ratio for the basic Gaussian example as

$$L(y) = \exp\left(\frac{1}{v^2}\left(my - \frac{m^2}{2}\right)\right).$$


continuous waveforms
$y(t) = s_i(t) + n(t)$
?
Discrete signal vector
$\mathbf{Y} = \mathbf{s}_i + \mathbf{N}$
Equivalent MPE rule
(waveform)
?
MPE rule:
$\delta(y) = \arg\max_{1 \le i \le M} \pi(i) p(y \mid i)$
?
?
Structure of optimal receiver for AWGN

3.4 Signal space concepts

Consider a communication system in which one of M continuous-time signals, $s_1(t), \ldots, s_M(t)$ is sent. The received signal equals the transmitted signal corrupted by AWGN. So we have **M hypotheses for** explaining **the received signals that are continuous-time signals**, with

$$H_i : y(t) = s_i(t) + n(t), \quad i = 1, \ldots, M, \tag{3.20}$$

where n(t) is WGN with PSD $\sigma^2 = N_0/2$

What is signal space?

The signal space is the finite-dimensional subspace (of dimension n ≤ M) spanned by $s_1(t), \ldots, s_M(t)$, which consists of all signals of the form $a_1 s_1(t) + \cdots + a_M s_M(t)$, where $a_1, \ldots, a_M$ are arbitrary scalars.

Let $\psi_1(t), \ldots, \psi_n(t)$ denote an orthonormal basis for the signal space. Such a basis can be constructed systematically by Gramm–Schmidt orthogonalization of the set of signals $s_1(t), \ldots, s_M(t)$

In general, for any signal set with M signals $\{s_i(t), i = 1, \ldots, M\}$ we can find an orthonormal basis $\{\psi_k, k = 1, \ldots, n\}$, where the dimension of the signal space, n, is at most equal to the number of signals, M.

The **vector representation of signal with respect to the basis** is

$$s_i[k] = \langle s_i, \psi_k \rangle, \quad i = 1, \ldots, M, \quad k = 1, \ldots, n.$$

$$\mathbf{s}_i = (s_i[1], \ldots, s_i[n])^T$$

1. The signal space spanned by the M possible received signals is finite dimensional, of dimension at most M. There is no signal energy outside this signal space, regardless of which signal is transmitted.

2. The component of WGN orthogonal to the signal space is independent of the component in the signal space, and its distribution does not depend on which signal was sent. It is therefore irrelevant to our hypothesis testing problem.

3. We can therefore **restrict attention to the signal and noise components lying in the signal space.** These can be represented by finite-dimensional vectors, thus simplifying the problem compared to the original problem of detection in continuous time.

Gramm–Schmidt orthogonalization $s_1(t), \ldots, s_M(t)$

Step 1 (Initialization) Let $\phi_1 = s_1$. If $\phi_1 \neq 0$, then set $\psi_1 = \phi_1/||\phi_1||$. Note that ψ_1 provides a basis function for $\mathcal{S}_1$.

Step $k+1$ Suppose that we have constructed an orthonormal basis $\mathcal{B}_k = \{\psi_1, \ldots \psi_m\}$ for the subspace $\mathcal{S}_k$ spanned by the first k signals (note that $m \leq k$). Define

$$\phi_{k+1}(t) = s_{k+1}(t) - \sum_{i=1}^{m} \langle s_{k+1}, \psi_i \rangle \psi_i(t).$$

The signal $\phi_{k+1}(t)$ is the component of $s_{k+1}(t)$ orthogonal to the subspace $\mathcal{S}_k$. If $\phi_{k+1} \neq 0$, define a new basis function $\psi_{m+1}(t) = \phi_{k+1}(t)/||\phi_{k+1}||$, and update the basis as $\mathcal{B}_{k+1} = \{\psi_1, \ldots, \psi_m, \psi_{m+1}\}$. If $\phi_{k+1} = 0$, then $s_{k+1} \in \mathcal{S}_k$, and it is not necessary to update the basis; in this case, we set $\mathcal{B}_{k+1} = \mathcal{B}_k = \{\psi_1, \ldots, \psi_m\}$.

The procedure terminates at step M, which yields a basis $\mathcal{B} = \{\psi_1, \ldots, \psi_n\}$ for the signal space $\mathcal{S} = \mathcal{S}_M$. The basis is not unique, and may depend (and typically does depend) on the order in which we go through the signals in the set.

We now project the received signal y(t) onto the signal space to obtain an n-dimensional vector **Y**. Specifically, set

$$\mathbf{Y} = (\langle y, \psi_1 \rangle, \ldots, \langle y, \psi_n \rangle)^T$$

Under hypothesis $H_i \ (i = 1, \ldots, M)$, projecting the signals and noise onto the signal space. We have

$$\mathbf{Y} = \mathbf{s}_i + \mathbf{N}, \quad i = 1, \ldots, M,$$

$$\mathbf{s}_i = (\langle s_i, \psi_1 \rangle, \ldots, \langle s_i, \psi_n \rangle)^T$$

$$\mathbf{N} = \underline{(\langle n, \psi_1 \rangle, \ldots, \langle n, \psi_n \rangle)^T}$$

what is the distribution of the noise components?

the vector $\mathbf{Y} = (y[1], \ldots, y[n])^T$ completely describes the component of the received signal y(t) in the signal space, which can be given by

$$y_S(t) = \sum_{j=1}^{n} \langle y, \psi_j \rangle \psi_j(t) = \sum_{j=1}^{n} y[j] \psi_j(t).$$

AWGN goes through a correlator

Any number obtained by linear processing of WGN can be expressed as the output of a correlation operation of the form

$$Z = \int_{-\infty}^{\infty} n(t)u(t)\mathrm{d}t = \langle n, u \rangle,$$

where u(t) is a deterministic, finite-energy, signal.

Since WGN is a Gaussian random process, we know that Z is a Gaussian random variable. To characterize its distribution, we need only compute its mean and variance.

Since n has zero mean, the mean of Z is seen to be zero by the following simple computation:

$$\mathbb{E}[Z] = \int_{-\infty}^{\infty} \mathbb{E}[n(t)]u(t)\mathrm{d}t = 0,$$

Proposition 3.1 (WGN through correlators)

Let u1(t) and u2(t) denote L2 waveforms (*finite-energy signals),* and let n(t) denote WGN with PSD $\sigma^2 = N_0/2$. Then $\langle n, u_1 \rangle$ and $\langle n, u_2 \rangle$ are jointly Gaussian with covariance

$$\text{cov}(\langle n, u_1 \rangle, \langle n, u_2 \rangle) = \sigma^2 \langle u_1, u_2 \rangle.$$

In particular, setting u1 = u2 = u, we obtain that

$$\text{var}(\langle n, u \rangle) = \text{cov}(\langle n, u \rangle, \langle n, u \rangle) = \sigma^2 ||u||^2.$$

Proof The random variables $\langle n, u_1 \rangle$ and $\langle n, u_2 \rangle$ are zero mean and jointly Gaussian, since n is zero mean and Gaussian. Their covariance is computed as

$$\begin{aligned}
\text{cov}(\langle n, u_1 \rangle, \langle n, u_2 \rangle) &= \mathbb{E}[\langle n, u_1 \rangle \langle n, u_2 \rangle] = \mathbb{E}[\int n(t) u_1(t) \mathrm{d}t \int n(s) u_2(s) \mathrm{d}s] \\
&= \int \int u_1(t) u_2(s) \mathbb{E}[n(t) n(s)] \mathrm{d}t \, \mathrm{d}s \\
&= \int \int u_1(t) u_2(s) \sigma^2 \delta(t - s) \mathrm{d}t \, \mathrm{d}s \\
&= \sigma^2 \int u_1(t) u_2(t) \mathrm{d}t = \sigma^2 \langle u_1, u_2 \rangle.
\end{aligned}$$

The component of y(t) orthogonal to the signal space is given by

$$y^{\perp}(t) = y(t) - y_S(t) = y(t) - \sum_{j=1}^{n} y_j \psi_j(t).$$

Now that the component of the received signal *orthogonal to the signal space,* $y^{\perp}(t)$ *, is irrelevant* for detection in AWGN.

Thus, it suffices to restrict attention to the finite-dimensional vector **Y** in the signal space for the purpose of optimal reception in AWGN.

$$y_S(t) = \sum_{j=1}^{n} \langle y, \psi_j \rangle \psi_j(t) = \sum_{j=1}^{n} y[j] \psi_j(t).$$

The original M hypotheses

$$H_i : y(t) = s_i(t) + n(t), \quad i = 1, \ldots, M,$$

The component of y(t) orthogonal to the signal space is given by

$$y^{\perp}(t) = y(t) - y_S(t) = y(t) - \sum_{j=1}^{n} y_j \psi_j(t).$$

$$\boxed{y_S(t) = \sum_{j=1}^{n} \langle y, \psi_j \rangle \psi_j(t) = \sum_{j=1}^{n} y[j] \psi_j(t).}$$

Theorem 3.1 (Restriction to signal space is optimal)

For the original hypotheses problem, there is no loss in detection performance in ignoring the component $y^{\perp}(t)$ of the received signal orthogonal to the signal space. Thus, it suffices to consider the equivalent hypothesis testing model given by

$$H_i : \mathbf{Y} = \mathbf{s}_i + \mathbf{N} \quad i = 1, \ldots, M.$$

Proof of Theorem 3.1 Conditioning on hypothesis H_i, we first note that $y^\perp$ does not have any signal contribution, since all of the M possible transmitted signals are in the signal space. That is, for $y(t) = s_i(t) + n(t)$, we have

$$\begin{aligned} y^\perp(t) &= y(t) - \sum_{j=1}^{n} \langle y, \psi_j \rangle \psi_j(t) = s_i(t) + n(t) - \sum_{j=1}^{n} \langle s_i + n, \psi_j \rangle \psi_j(t) \\ &= n(t) - \sum_{j=1}^{n} \langle n, \psi_j \rangle \psi_j(t) = n^\perp(t), \end{aligned}$$

where $n^\perp$ is the noise contribution orthogonal to the signal space. Next, we show that $n^\perp$ is independent of $\mathbf{N}$, the noise contribution in the signal space. Since $n^\perp$ and $\mathbf{N}$ are jointly Gaussian, it suffices to demonstrate that they are uncorrelated. Specifically, for any t and k, we have

$$\begin{aligned} \operatorname{cov}(n^\perp(t), N[k]) &= \mathbb{E}[n^\perp(t) N[k]] = \mathbb{E}[\{n(t) - \textstyle\sum_{j=1}^{n} N[j]\psi_j(t)\} N[k]] \\ &= \underbrace{\boxed{\mathbb{E}[n(t)N[k]]}}_{1} - \textstyle\sum_{j=1}^{n} \underbrace{\boxed{\mathbb{E}[N[j]N[k]]}}_{2} \psi_j(t). \end{aligned}$$

The first term on the extreme right-hand side can be simplified as

$$\mathbb{E}[n(t)\langle n, \psi_k\rangle] = \mathbb{E}[n(t)\int n(s)\psi_k(s)\mathrm{d}s]$$

1

$$= \int \mathbb{E}[n(t)n(s)]\psi_k(s)\mathrm{d}s = \int \sigma^2\delta(s-t)\psi_k(s)\mathrm{d}s = \sigma^2\psi_k(t).$$

noting that $\mathbb{E}[N[j]N[k]] = \sigma^2\delta_{jk}$, we obtain

2

$$1 - 2 = \mathrm{cov}(n^{\perp}(t), N[j]) = \sigma^2\psi_k(t) - \sigma^2\psi_k(t) = 0.$$

Thus, conditioned on H_i, $y^{\perp} = n^{\perp}$ does not contain any signal contribution, and is independent of the noise vector **N** in the signal space. It is therefore irrelevant to the detection problem

end

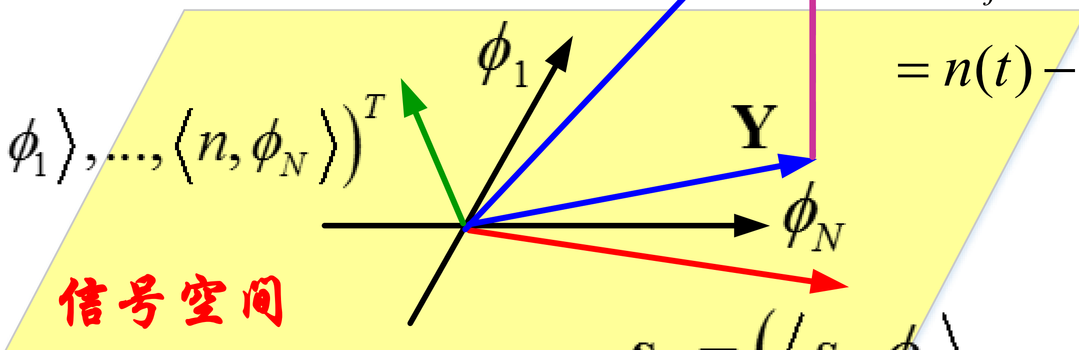
Signal space

$$y(t) = s_i(t) + n(t)$$

$$y(t) - \sum_{j=1}^{N} y_j \phi_j(t) = n(t) - \sum_{j=1}^{N} \langle n,\ \phi_j \rangle \phi_j(t) = n^{\perp}(t)$$

Irrelevant to the hypothesis problem

$$\mathbf{N} = \left(\langle n, \phi_1 \rangle, \ldots, \langle n, \phi_N \rangle\right)^T$$



$$\mathbf{s}_i = \left(\langle s_i, \phi_1 \rangle, \ldots, \langle s_i, \phi_N \rangle\right)^T$$

$$H_i : y(t) = s_i(t) + n(t), \qquad i = 1, \ldots, M$$

$$H_i : \ \mathbf{Y} = \mathbf{s}_i + \mathbf{N} \quad i = 1, \ldots, M$$

Example 3.6 (Application to two-dimensional linear modulation) Consider linear modulation in passband, for which the transmitted signal corresponding to a given symbol is of the form

$$s_{b_c,b_s}(t) = Ab_c p(t)(\sqrt{2}\cos 2\pi f_c t) - Ab_s p(t)(\sqrt{2}\sin 2\pi f_c t),$$

where the information is encoded in the pair of real numbers (b_c, b_s), and where $p(t)$ is a baseband pulse whose bandwidth is smaller than the carrier frequency f_c. We assume that there is no intersymbol interference, hence it suffices to consider each symbol separately. In this case, the signal space is two-dimensional, and a natural choice of basis functions for the signal space is $\psi_c(t) = \alpha p(t)\cos 2\pi f_c t$ and $\psi_s(t) = \alpha p(t)\sin 2\pi f_c t$, where α is a normalization constant.

$$\mathbf{y} = \begin{pmatrix} y_c \\ y_s \end{pmatrix} = \begin{pmatrix} b_c \\ b_s \end{pmatrix} + \begin{pmatrix} N_c \\ N_s \end{pmatrix},$$

where N_c, N_s are i.i.d. $N(0, \sigma^2)$ random variables.

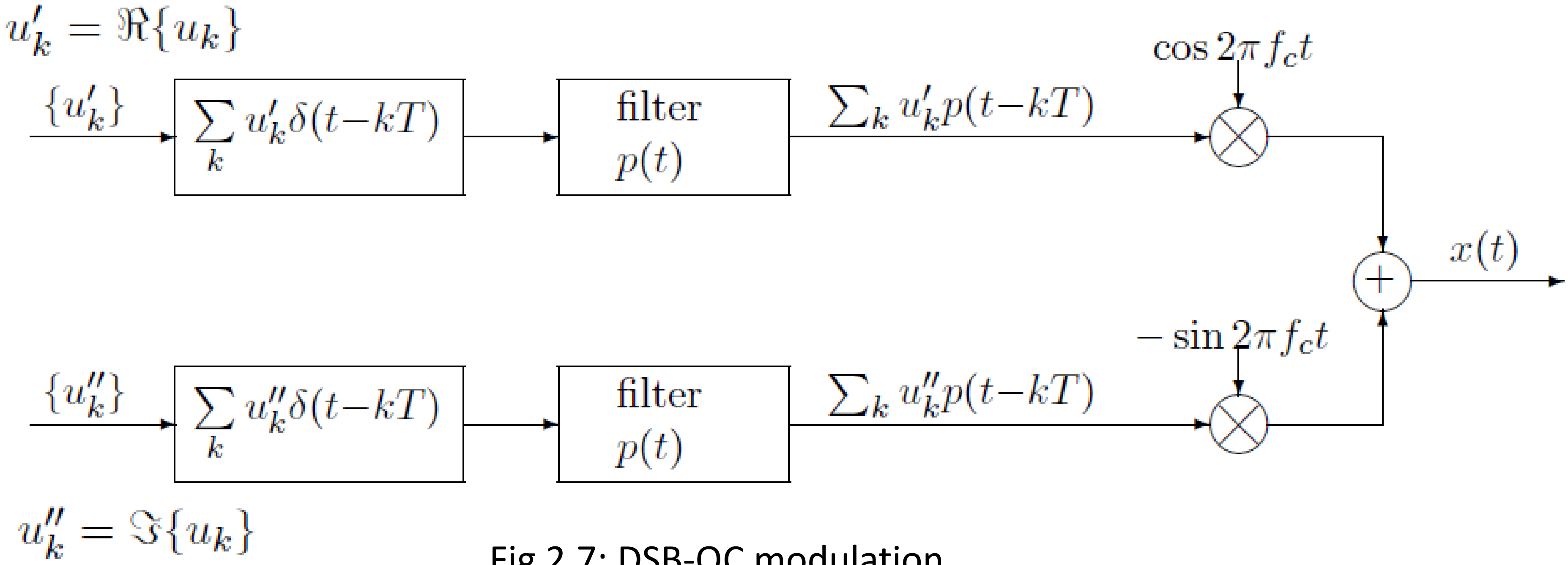
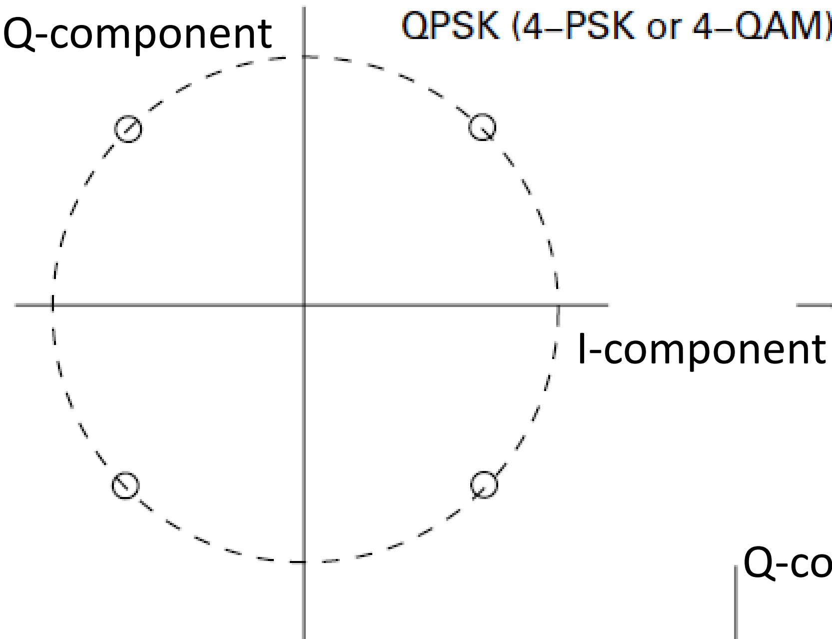
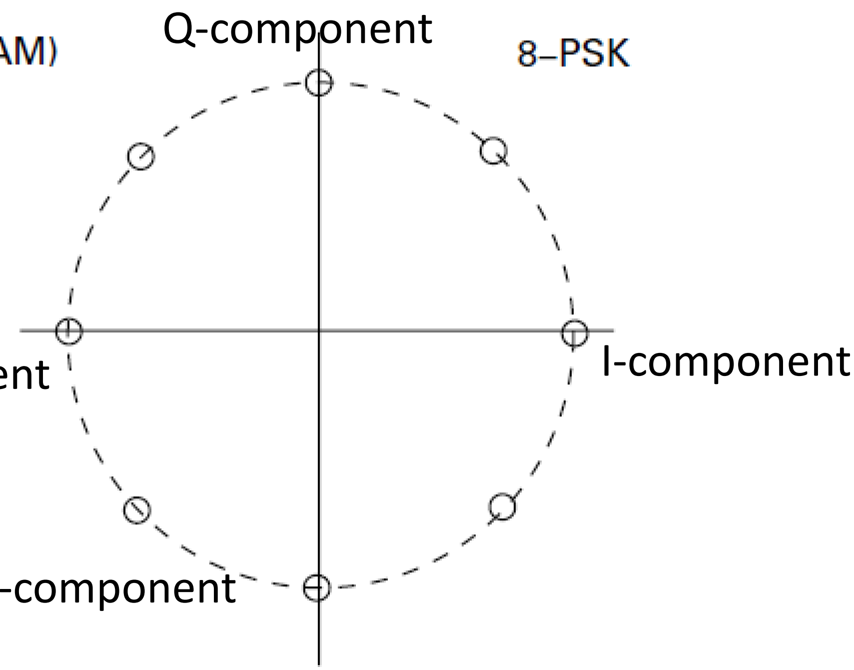


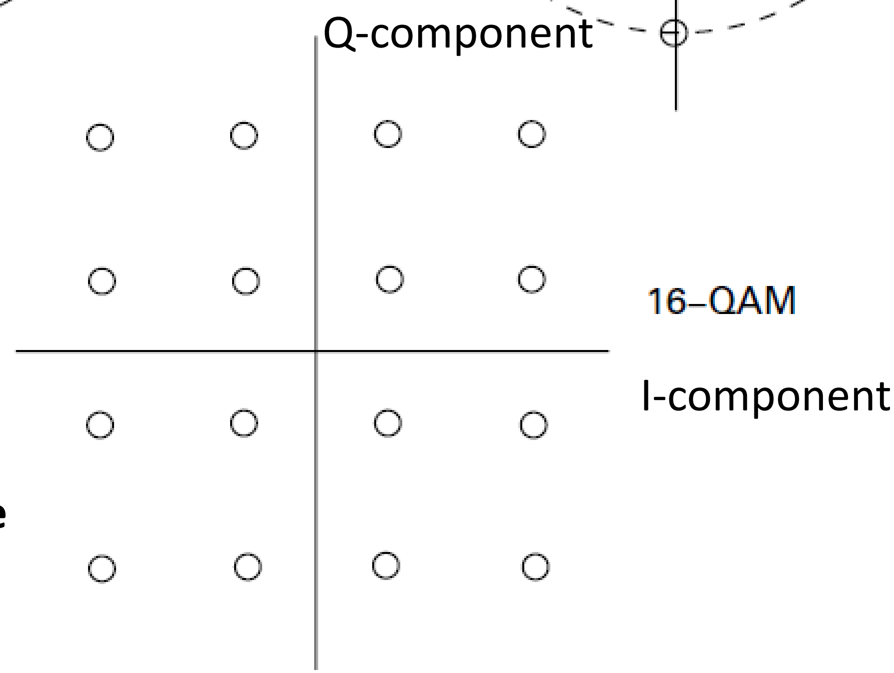
Fig 2.7: DSB-QC modulation

$$x(t) = 2\cos(2\pi f_c t)\left(\sum_k u'_k p(t-kT)\right) - 2\sin(2\pi f_c t)\left(\sum_k u''_k p(t-kT)\right).$$

This realization of QAM is called double-sideband quadrature-carrier (DSB-QC) modulation .







two-dimensional signal space

continuous waveforms
$y(t) = s_i(t) + n(t)$

Signal space

Discrete signal vector
$\mathbf{Y} = \mathbf{s}_i + \mathbf{N}$

Equivalent MPE rule
(waveform)

?

MPE rule：
$$\delta(y) = \arg\max_{1 \le i \le M} \pi(i)\, p(y \mid i)$$

?

?

Structure of optimal receiver for AWGN

3.5 Optimal reception in AWGN

When the received signal is a finite-dimensional vector.

Theorem 3.2 (Optimal detection in discrete-time AWGN)

Consider the finite-dimensional M-ary hypothesis testing problem where the observation is a random vector **Y** modeled as

$$H_i: \ \mathbf{Y} = \mathbf{s}_i + \mathbf{N} \quad i = 1, \ldots, M,$$

where $\mathbf{N} \sim N(0, \sigma^2 \mathbf{I})$ is discrete-time WGN.

(a) If hypothesis Hi has prior probability $\pi(i), i = 1, \ldots, M \ (\sum_{i=1}^{M} \pi(i) = 1)$, then the MPE decision rule is given by

$$\begin{aligned} \delta_{\mathrm{MPE}}(\mathbf{y}) &= \arg \min_{1 \le i \le M} ||\mathbf{y} - \mathbf{s}_i||^2 - 2\sigma^2 \log \pi(i) \\ &= \arg \max_{1 \le i \le M} \langle \mathbf{y}, \mathbf{s}_i \rangle - \frac{||\mathbf{s}_i||^2}{2} + \sigma^2 \log \pi(i). \end{aligned}$$

(b) When we observe **Y = y**, the ML decision rule is a "minimum distance rule," given by

$$\delta_{\mathrm{ML}}(\mathbf{y}) = \arg \min_{1 \le i \le M} ||\mathbf{y} - \mathbf{s}_i||^2 = \arg \max_{1 \le i \le M} \langle \mathbf{y}, \mathbf{s}_i \rangle - \frac{||\mathbf{s}_i||^2}{2}.$$

Proof

Under hypothesis Hi, Y is a Gaussian random vector with mean $\mathbf{s}_i$ and covariance matrix $\sigma^2\mathbf{I}$, so that

$$p_{\mathbf{Y}|i}(\mathbf{y}|H_i) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{||\mathbf{y}-\mathbf{s}_i||^2}{2\sigma^2}\right).$$

The MPE decision rule is

$$\delta_{\mathrm{MPE}}(y) = \arg\max_{1\leq i\leq M} \pi(i)p(y|i) = \arg\max_{1\leq i\leq M} \log\pi(i) + \log p(y|i).$$

so we get $\delta_{\mathrm{MPE}}(\mathbf{y}) = \arg\min_{1\leq i\leq M} ||\mathbf{y}-\mathbf{s}_i||^2 - 2\sigma^2\log\pi(i)$

$$= \arg\max_{1\leq i\leq M} \langle \mathbf{y}, \mathbf{s}_i\rangle - \frac{||\mathbf{s}_i||^2}{2} + \sigma^2\log\pi(i).$$

The MPE rule is $\delta_{ML}(y) = \arg\max_{1\leq i\leq M} p(y|i) = \arg\max_{1\leq i\leq M} \log p(y|i).$

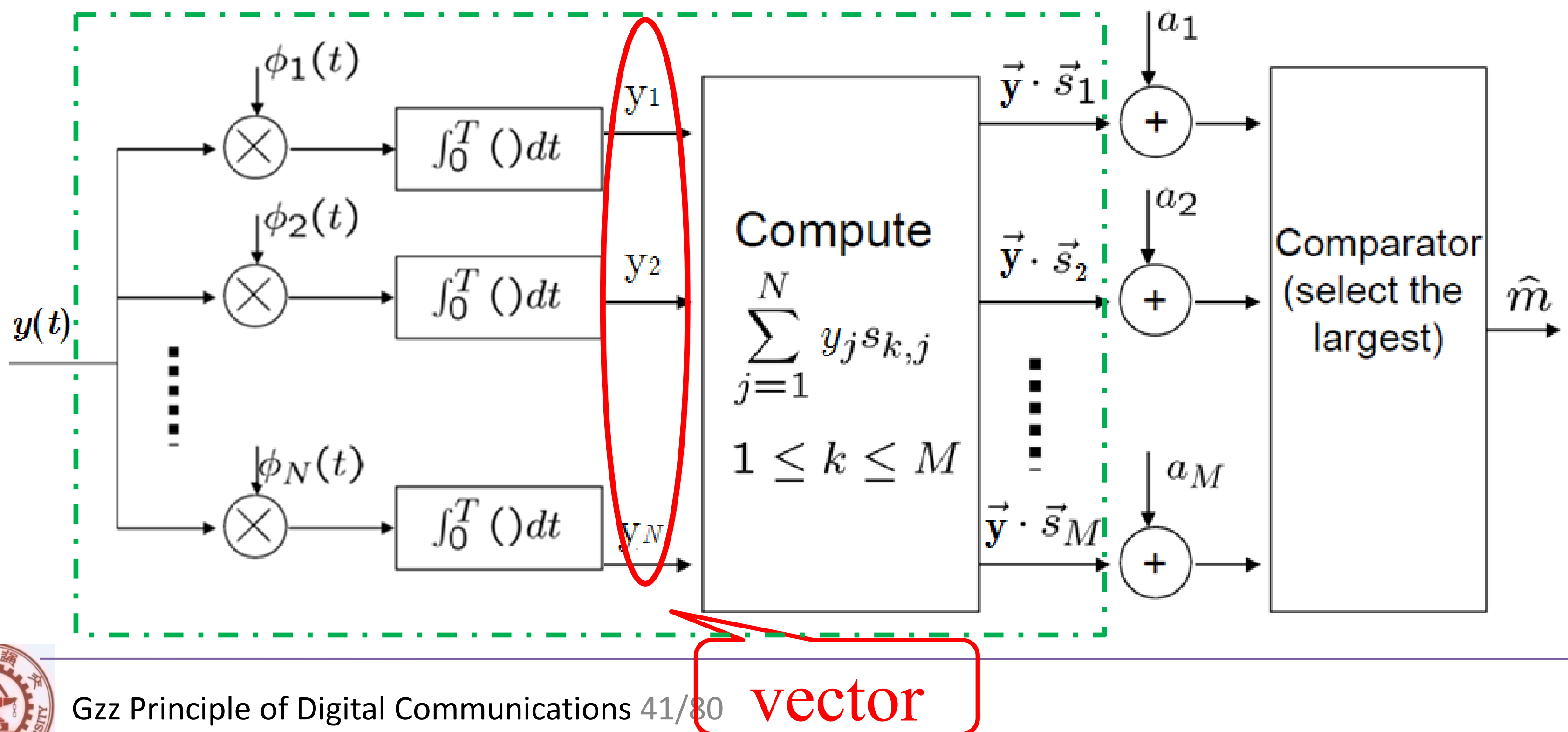
So we have $\delta_{\mathrm{ML}}(\mathbf{y}) = \arg\min_{1\leq i\leq M} ||\mathbf{y}-\mathbf{s}_i||^2 = \arg\max_{1\leq i\leq M} \langle \mathbf{y}, \mathbf{s}_i\rangle - \frac{||\mathbf{s}_i||^2}{2}.$

end

Optimal receiver structure 1– based on discrete vectors

$$\delta_{\mathrm{MPE}}(\mathbf{y}) = \arg\max_{1\le i\le M}\left(<\mathbf{y},\mathbf{s}_i> - \frac{\|\mathbf{s}_i\|^2}{2} + \frac{N_0}{2}\ln\pi(i)\right)$$

$$a_i = \frac{N_0}{2}\ln\pi(i) - \frac{E_i}{2}$$



Theorem 3.3 (Optimal coherent demodulation with real-valued signals)
For the continuous-time model $H_i : y(t) = s_i(t) + n(t), \quad i = 1, \ldots, M,$
the optimal detectors are given as follows:
(a) The ML decision rule is

$$\delta_{\mathrm{ML}}(y) = \arg \max_{1 \le i \le M} \langle y, s_i \rangle - \frac{||s_i||^2}{2}.$$

(b) If hypothesis Hi has prior probability $\pi(i),\ i = 1, \ldots, M\ (\sum_{i=1}^{M} \pi(i) = 1)$
then the MPE decision rule is given by

$$\delta_{\mathrm{MPE}}(\mathbf{y}) = \arg \max_{1 \le i \le M} \langle y, s_i \rangle - \frac{||s_i||^2}{2} + \sigma^2 \log \pi(i).$$

Proof:
Theorem 3.1 $H_i : y(t) = s_i(t) + n(t), \quad i = 1, \ldots, M,$

$$\Updownarrow$$

$$H_i : \ \mathbf{Y} = \mathbf{s}_i + \mathbf{N} \quad i = 1, \ldots, M.$$

$\boxed{\text{signal inner products are preserved}}$

Discrete time $$\delta_{\mathrm{ML}}(\mathbf{y}) = \arg \min_{1\le i\le M} ||\mathbf{y}-\mathbf{s}_i||^2 = \arg \max_{1\le i\le M} \langle \mathbf{y}, \mathbf{s}_i\rangle - \frac{||\mathbf{s}_i||^2}{2}.$$

Real waveform $$\delta_{\mathrm{ML}}(y) = \arg \max_{1\le i\le M} \langle y, s_i\rangle - \frac{||s_i||^2}{2}.$$

$||\mathbf{s}_i||^2 = ||s_i||^2.$

the inner product between the signals y_S and s_i (which both lie in the signal space) is the same as the inner product between their vector representations.

$$\langle y, s_i\rangle = \langle y_S + y^{\perp}, s_i\rangle = \langle y_S, s_i\rangle + \langle y^{\perp}, s_i\rangle,$$
$$= \langle y_S, s_i\rangle \qquad = \langle \mathbf{y}, \mathbf{s}_i\rangle,$$

$$\delta_{\mathrm{ML}}(\mathbf{y}) = \arg \min_{1\le i\le M} ||\mathbf{y}-\mathbf{s}_i||^2 = \arg \max_{1\le i\le M} \langle \mathbf{y}, \mathbf{s}_i\rangle - \frac{||\mathbf{s}_i||^2}{2}.$$

$$\Downarrow$$

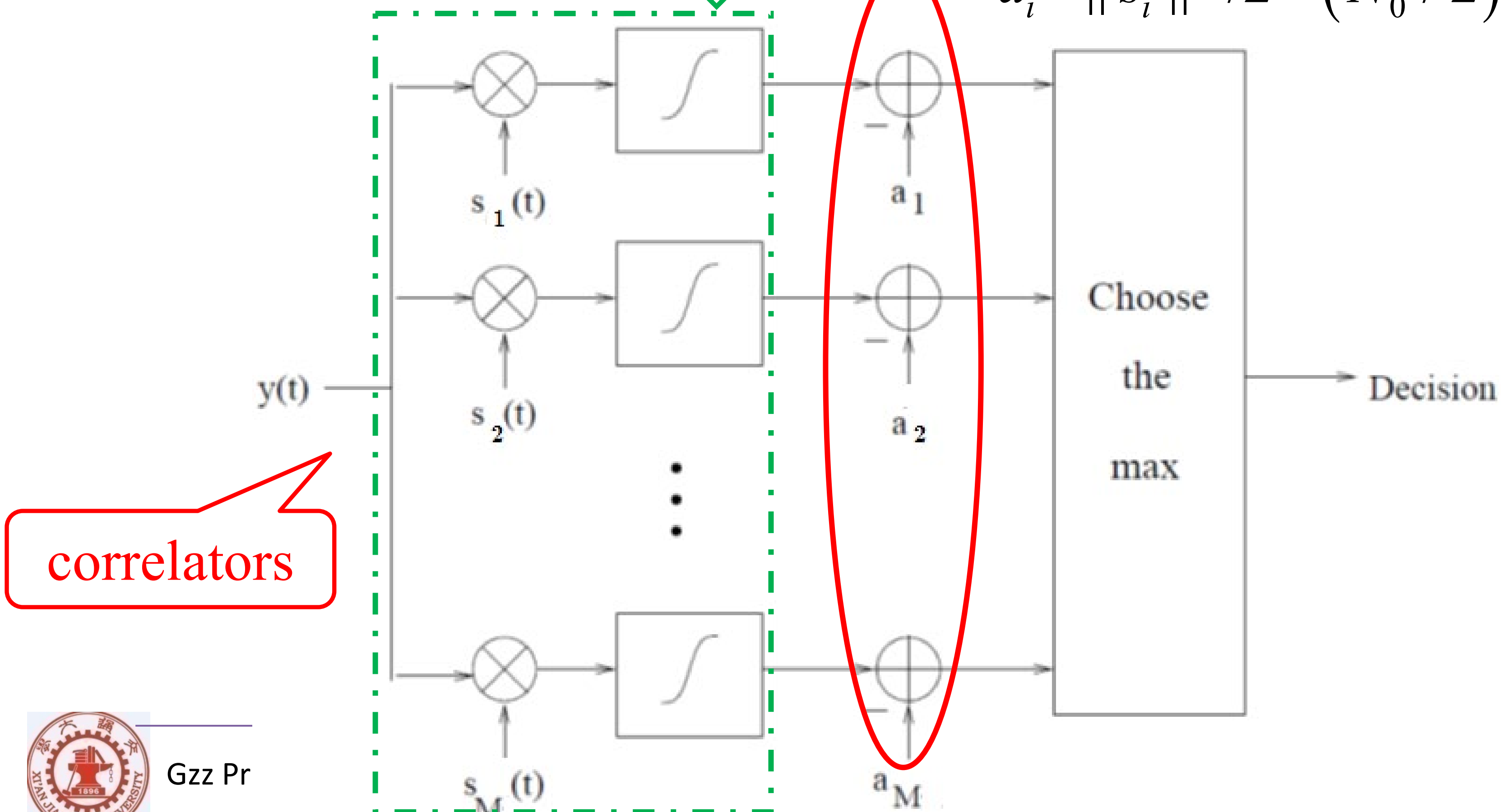
$$\delta_{\mathrm{ML}}(y) = \arg \max_{1\le i\le M} \langle y, s_i\rangle - \frac{||s_i||^2}{2}.$$

end

Optimal receiver structure 2– based on waveform

$$\hat{i} = \arg\max_{1\le i\le M}\left(< y, s_i > - \frac{\| s_i \|^2}{2} + \frac{N_0}{2}\ln \pi(i) \right)$$

$$a_i = \| s_i \|^2 /2 - \left(N_0/2\right)\ln \pi(i)$$



Example of MPE rule

Consider a set of binary signals

$$s_1(t) = -s_2(t) = \sqrt{E}\phi(t)$$

Assume that

$$\pi(m_1) \neq \pi(m_2)$$

Please give the decision region after the signals go through AWGN channels

Solution: the MPE rule

$$\delta_{\mathrm{MPE}}(\mathbf{y}) = \arg\min_{1\le i\le M}\left(\|\mathbf{y}-\mathbf{s}_i\|^2 - N_0 \ln \pi(i)\right)$$

$$\|\mathbf{y}-\mathbf{s}_1\|^2 - N_0 \ln \pi(m_1) \underset{m1}{\overset{m2}{\gtrless}} \|\mathbf{y}-\mathbf{s}_2\|^2 - N_0 \ln \pi(m_2)$$

Example of MPE rule

$$\| \mathbf{y} - \mathbf{s}_1 \|^2 - N_0 \ln \pi(m_1) \underset{m1}{\overset{m2}{\gtrless}} \| \mathbf{y} - \mathbf{s}_2 \|^2 - N_0 \ln \pi(m_2)$$

Let $d_1 = \| \mathbf{y} - \mathbf{s}_1 \|, \; d_2 = \| \mathbf{y} - \mathbf{s}_2 \|$

We have

$$d_1^2 - d_2^2 \underset{m1}{\overset{m2}{\gtrless}} N_0 \ln \frac{\pi(m_1)}{\pi(m_2)}$$

Constant C

Therefore, the decision regions are

$$R1: \; d_1^2 - d_2^2 < c$$

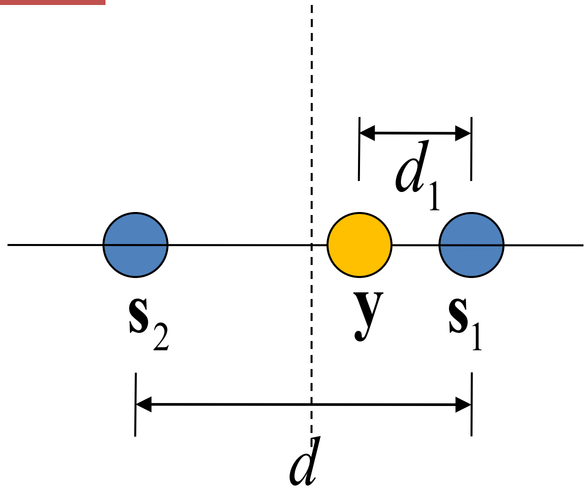
$$R2: \; d_1^2 - d_2^2 > c$$

Example of MPE rule

$$R1:\ d_1^2 - d_2^2 < c$$

$$R2:\ d_1^2 - d_2^2 > c$$

$$d = d_1 + d_2$$

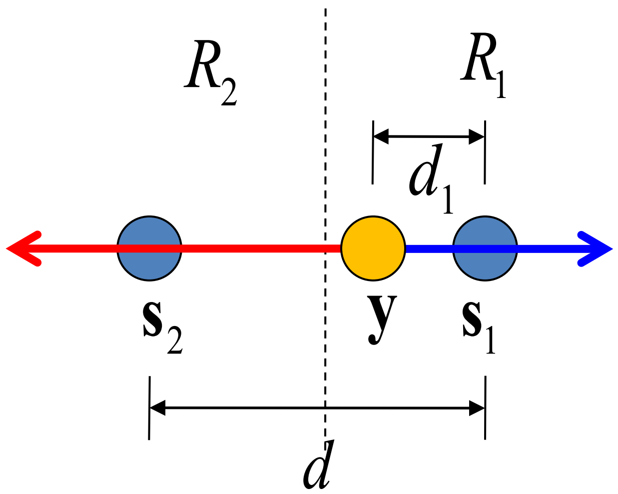


At the boundary of the decision regions:

$$d_1^2 - d_2^2 = d_1^2 - (d - d_1)^2 \equiv c$$

$$d_1^* = \frac{c + d^2}{2d} = \frac{d}{2} + \frac{N_0}{2d} \ln \frac{\pi(m_1)}{\pi(m_2)}$$

Example of MPE rule



$$d_1^* = \frac{c+d^2}{2d} = \frac{d}{2} + \frac{N_0}{2d}\ln\frac{\pi(m_1)}{\pi(m_2)}$$

$$d_1^* \begin{cases} = d/2, & if\ \ \pi(m_1) = \pi(m_2) \\ > d/2, & if\ \ \pi(m_1) > \pi(m_2) \\ < d/2, & if\ \ \pi(m_1) < \pi(m_2) \end{cases}$$

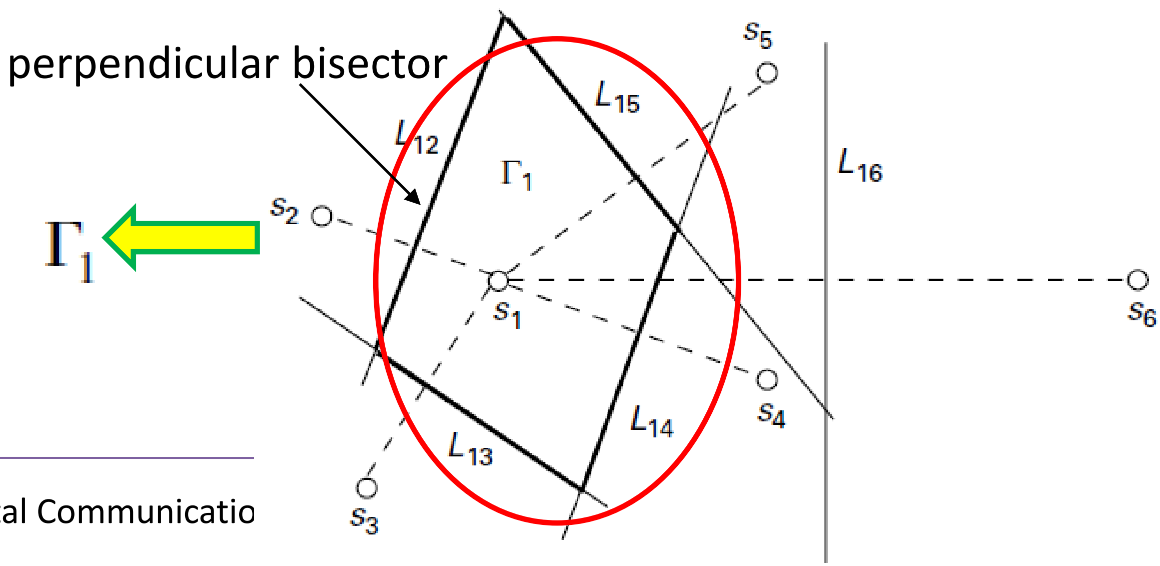
- With equal prior probability, the decision boundary lies in the middle of the connection of two signal points.
- With unequal prior probability, the decision region of the signal with smaller prior probability shrink.

3.5.1 Geometry of the ML decision rule

The ML decision rule is the minimum distance rule.

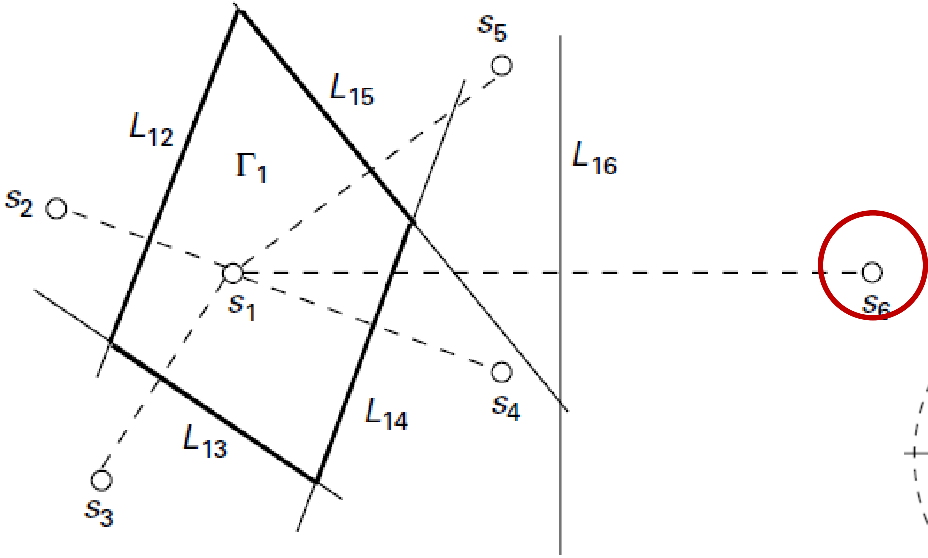
1. The signal vectors si, and the received vector y, are points in n-dimensional signal space. Take the two dimensional (n = 2) space as an example.

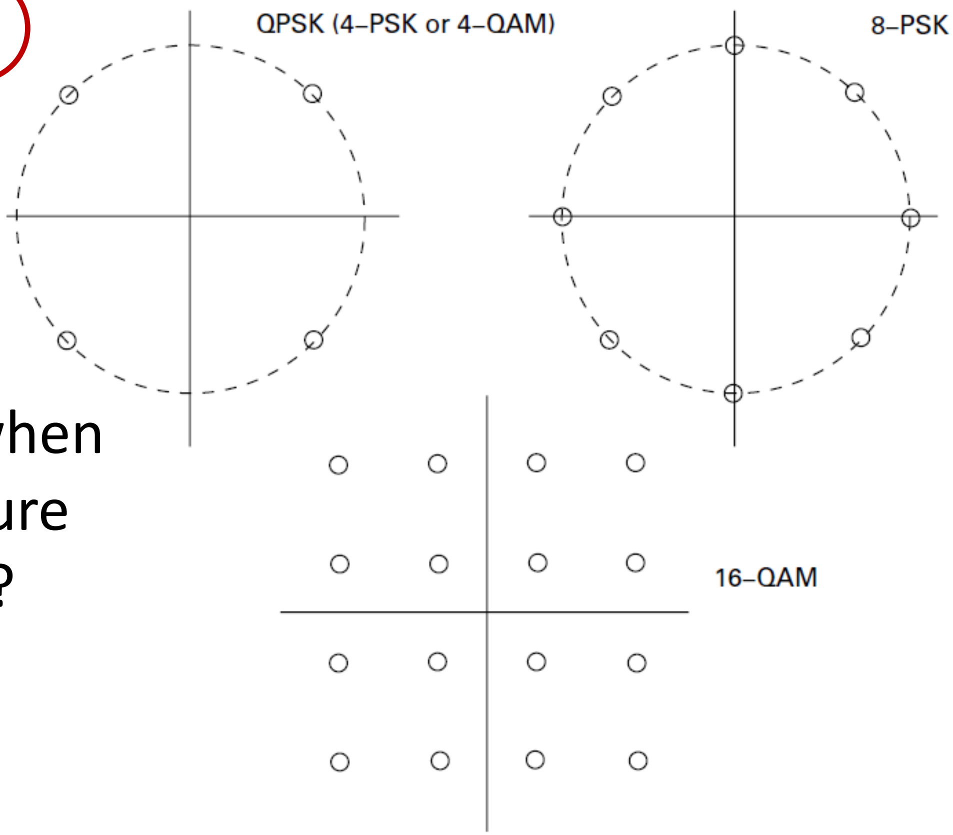
2. For any given i, draw a line between **si** and **sj** for all $j \neq i$. The intersection of these lines defines the ith decision region.



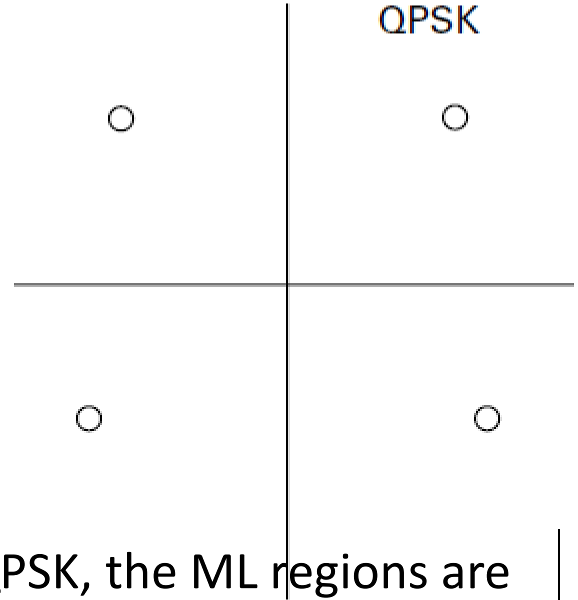
Note that L16 plays no role in determining Γ_1 , since signal **s**6 is "too far" from **s**1.

We can do the similar procedure to get other decision regions.

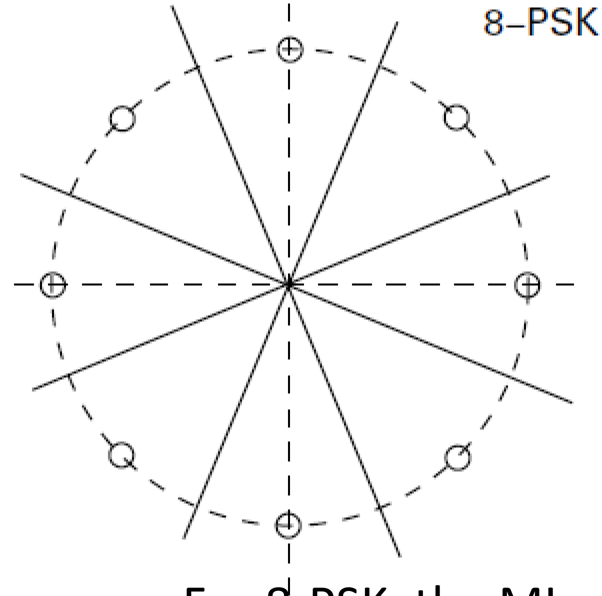




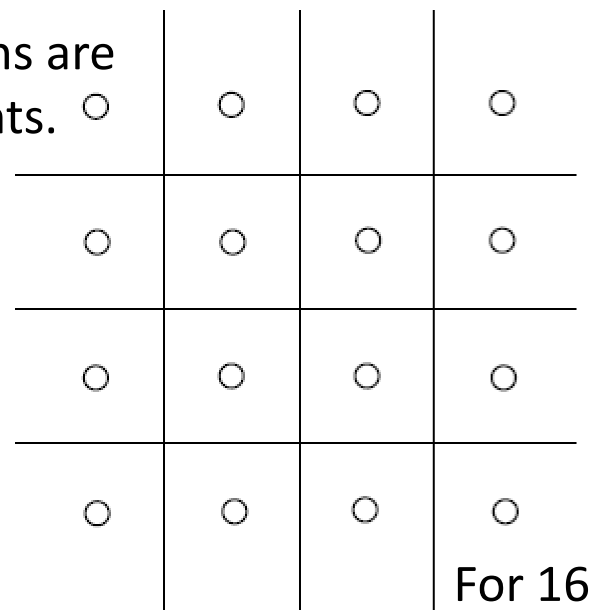
What will be decision regions when applying this preceding procedure to the standard constellations ?



For QPSK, the ML regions are simply the four quadrants.



For 8-PSK, the ML regions are sectors of a circle.



16–QAM

For 16-QAM, the ML regions take a rectangular form.

3.5.2 Soft decisions

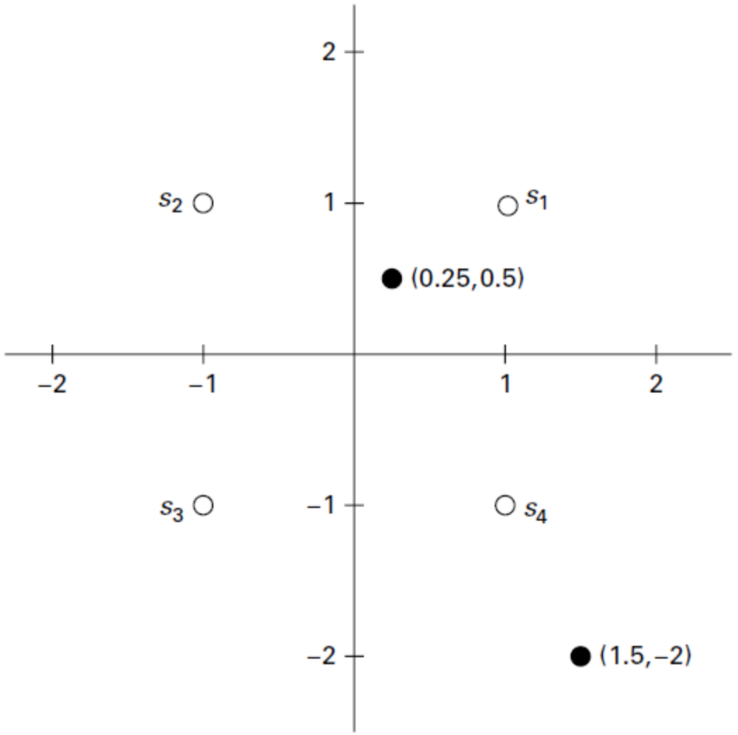
Maximum likelihood and MPE demodulation correspond to “hard” decisions regarding which of M signals have been sent.

$$H_i:\ \mathbf{Y} = \mathbf{s}_i + \mathbf{N} \quad i = 1, \ldots, M,$$

$$\delta_{\mathrm{ML}}(\mathbf{y}) = \arg \min_{1 \le i \le M} ||\mathbf{y} - \mathbf{s}_i||^2 = \arg \max_{1 \le i \le M} \langle \mathbf{y}, \mathbf{s}_i \rangle - \frac{||\mathbf{s}_i||^2}{2}.$$

What is the result when applying ML?

We will find more when using soft decisions.



Why soft decision?

“hard” decisions consider which of M signals have been sent. Each such M-ary “symbol” corresponds to $\log_2 M$ bits.

Often, however, we may employ an error-correcting code over the entire sequence of transmitted symbols or bits. In such a situation, the decisions from the demodulator, which performs M-ary hypothesis testing for each symbol, must be fed to a decoder which accounts for the structure of the error-correcting code to produce more reliable decisions. It becomes advantageous in such a situation to feed the decoder more information than that provided by hard decisions.

Soft decisions are means of quantifying our estimate of the reliability of our decisions. While there are many mechanisms that could be devised for conveying more information than hard decisions, the maximal amount of information that the demodulator can provide is the posterior probabilities

$$\pi(i|\mathbf{y}) = P[\mathbf{s}_i \text{ sent}|\mathbf{y}] = P[H_i|\mathbf{y}],$$

where **y** is the value taken by the observation **Y**. These posterior probabilities can be computed using Bayes' rule, as follows:

$$\pi(i|\mathbf{y}) = P[H_i|\mathbf{y}] = \frac{p(\mathbf{y}|i)P[H_i]}{p(\mathbf{y})} = \frac{p(\mathbf{y}|i)P[H_i]}{\sum_{j=1}^{M} p(\mathbf{y}|j)P[H_j]}.$$

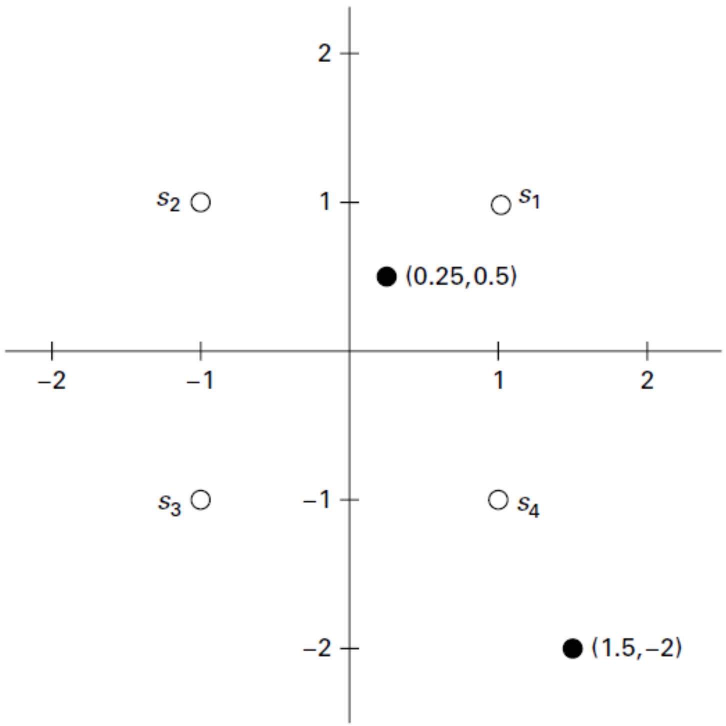
Calculating the conditional densities $p(\mathbf{y}|j)$ and setting $\pi(i) = P[H_i]$, we obtain

$$\pi(i|\mathbf{y}) = \frac{\pi(i)\exp\left(-\frac{||\mathbf{y}-\mathbf{s}_i||^2}{2\sigma^2}\right)}{\sum_{j=1}^{M} \pi(j)\exp\left(-\frac{||\mathbf{y}-\mathbf{s}_j||^2}{2\sigma^2}\right)}.$$

Let's consider the problem again.

For the considered problem, suppose that $\sigma^2 = 1$ and $\pi(i) \equiv 1/4$

Then we compute the posterior probabilities as follows.



$$\pi(i|\mathbf{y}) = \frac{\pi(i)\exp\left(-\frac{||\mathbf{y}-\mathbf{s}_i||^2}{2\sigma^2}\right)}{\sum_{j=1}^{M}\pi(j)\exp\left(-\frac{||\mathbf{y}-\mathbf{s}_j||^2}{2\sigma^2}\right)}.$$

Table 3.1 Posterior probabilities for the QPSK constellation in Figure 3.13, assuming equal priors and $\sigma^2 = 1$.

$\pi(i\|\mathbf{y})$	$\mathbf{y} = (0.25, 0.5)$	$\mathbf{y} = (1.5, -2)$
1	0.455	0.017
2	0.276	0
3	0.102	0.047
4	0.167	0.935

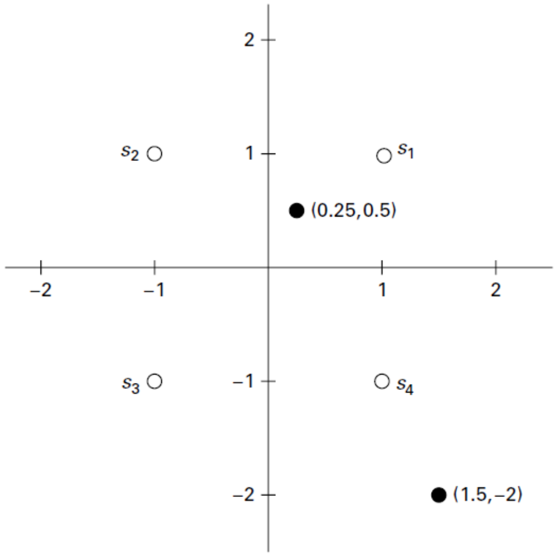


Table 3.1 Posterior probabilities for the QPSK constellation in Figure 3.13, assuming equal priors and $\sigma^2 = 1$.

$\pi(i\|\mathbf{y})$	$\mathbf{y} = (0.25, 0.5)$	$\mathbf{y} = (1.5, -2)$
1	0.455	0.017
2	0.276	0
3	0.102	0.047
4	0.167	0.935

•The observation $\mathbf{y} = (0.25, 0.5)$ falls in the decision region for **s**1, but is close to the decision boundary. The posterior probabilities in Table 3.1 reflect the uncertainty.

•The observation $\mathbf{y} = (1.5, -2)$ in the decision region for **s**4, is far away from the decision boundaries, hence we would expect it to provide a reliable decision. The posterior probabilities reflect this: the posterior probability for **s**4 is significantly larger than that of the other possible symbol values.

Table 3.2 illustrates what happens when the noise variance is increased to $\sigma^2 = 4$ for the scenario.

Table 3.2 Posterior probabilities for the QPSK constellation assuming equal priors and $\sigma^2 = 4$.

$\pi(i\|\mathbf{y})$	$\mathbf{y} = (0.25, 0.5)$	$\mathbf{y} = (1.5, -2)$
1	0.299	0.183
2	0.264	0.086
3	0.205	0.235
4	0.233	0.497

The posteriors for $\mathbf{y} = (0.25, 0.5)$ which is close to the decision boundaries, are close to uniform, which indicates that the observation is highly unreliable.

Even for $\mathbf{y} = (1.5, -2)$, the posterior probabilities for symbols other than **s**2 are significant.

Table 3.2 illustrates what happens when the noise variance is increased to $\sigma^2 = 4$ for the scenario.

Table 3.2 Posterior probabilities for the QPSK constellation assuming equal priors and $\sigma^2 = 4$.

$\pi(i\|\mathbf{y})$	$\mathbf{y} = (0.25, 0.5)$	$\mathbf{y} = (1.5, -2)$
1	0.299	0.183
2	0.264	0.086
3	0.205	0.235
4	0.233	0.497

Unlike ML hard decisions, which depend only on the distances between the observation and the signal points, the posterior probabilities also depend on the noise variance. If the noise variance is higher, then the decision becomes more unreliable.

Example The receiver in a binary communication system employs a decision statistic Z which behaves as follows:

$Z = N$ if 0 is sent,

$Z = 4 + N$ if 1 is sent,

where N is modeled as Laplacian with density

$$p_N(x) = \frac{1}{2}\mathrm{e}^{-|x|}, \quad -\infty < x < \infty.$$

Note Parts (a) and (b) can be done independently.

(a) Give a sensible decision rule, and the corresponding average error probability.

(b) Find $P_{\mathrm{e}|1}$, the conditional error probability given that 1 is sent, for the decision rule

$$\delta(z) = \begin{cases} 0, & z < 1, \\ 1, & z \geq 1. \end{cases}$$

Example The receiver in a binary communication system employs a decision statistic Z which behaves as follows:

$Z = N$ if 0 is sent,

$Z = 4 + N$ if 1 is sent,

where N is modeled as Laplacian with density

$$p_N(x) = \frac{1}{2}e^{-|x|}, \quad -\infty < x < \infty.$$

Note Parts (a) and (b) can be done independently.

(c) Find and sketch, as a function of z, the log likelihood ratio

$$K(z) = \log L(z) = \log \frac{p(z|1)}{p(z|0)},$$

where $p(z|i)$ denotes the conditional density of Z given that i is sent ($i = 0, 1$).

(d) Is the rule in (c) the MPE rule for any choice of prior probabilities? If so, specify the prior probability $\pi_0 = P[0 \text{ sent}]$ for which it is the MPE rule. If not, say why not.

$$f(x \mid \mu, b) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b} \right)$$