通訊系統 (II)

國立清華大學電機系暨通訊工程研究所
蔡育仁
台達館 821 室
Tel: 62210
E-mail: yrtsai@ee.nthu.edu.tw

Chapter 1
Signal-Space Analysis

Introduction

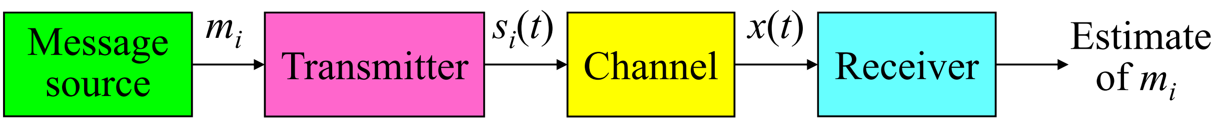
- We discuss some basic issues that relate to signal transmission over an **additive white Gaussian noise (AWGN)** channel
  - **Geometric representation** of signals with finite energy
  - **Maximum likelihood (ML)** procedure for signal detection in an AWGN channel
  - Derivation of the **correlation receiver** that is equivalent to the **matched filter receiver**
  - Probability of symbol error and the **union bound** approximation

Digital Communication Systems

Message

- Consider the most basic form of a **digital** communication system
  - A message source emits one symbol every T seconds, with the symbols belonging to an alphabet of M symbols denoted by $m_1, m_2, \ldots, m_M$
  - The ***a priori* probabilities** $p_1, p_2, \ldots, p_M$ specify the **message source** output
    - It is customary to assume that the M symbols of the alphabet are **equally likely**

$$p_i = P(m_i) = \frac{1}{M} \quad \text{for} \;\; i = 1, 2, \cdots, M$$



Transmitter

- The transmitter takes the message source output m_i and codes it into a distinct signal $s_i(t)$ **suitable for transmission** over the **analog channel**.
- The signal $s_i(t)$ is a **real-valued energy signal**

$$E_i = \int_0^T s_i^2(t)\, dt, \quad i = 1, 2, \cdots, M$$

  - A signal with **finite energy**
- The design of the signal $s_i(t)$ is a key issue in communication systems

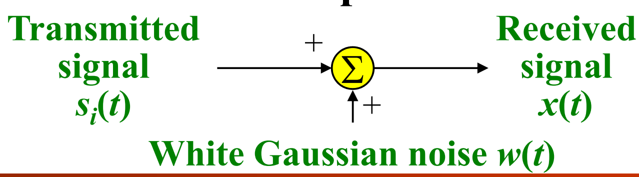
Channel

- Passing through the channel, the received signal $x(t)$ is

$$x(t) = \alpha s_i(t) + w(t), \quad 0 \le t \le T \text{ and } i = 1, 2, \cdots, M$$

  – α is the **complex-valued** channel gain
- The channel is assumed to have two characteristics:
  – The channel is **linear**, with a bandwidth that is **wide enough** to accommodate the transmission of signal $s_i(t)$
    - with **negligible or no distortion**
    - α includes the **attenuation** and **phase rotation**

  – The channel noise $w(t)$ is the sample function of a zero-mean **white Gaussian noise process**



Receiver

- The receiver has the task of
  – **Observing** the received signal $x(t)$ for a duration of T
  – Making a best **estimate** of the transmitted signal $s_i(t)$ (or m_i)
- However, owing to the presence of channel noise, the receiver will make occasional **errors**
  – To design the receiver so as to **minimize** the average **probability of symbol error**

$$P_e = \sum_{i=1}^{M} p_i P(\hat{m} \ne m_i | m_i)$$

  – where m_i is the transmitted symbol; $\hat{m}$ is the estimate produced by the receiver; $P(\hat{m} \ne m_i | m_i)$ is the **conditional** error probability given that the i-th symbol was sent

$\Rightarrow$ **Optimum in the minimum probability of error sense**

Geometric Representation of Signals

Linear Vector Space

- In signal analysis, we can represent signals as vectors
  - To remove some **redundancy** in the signals
  - To provide a more compact form for the signals
- The signal space could be constructed by **amplitude**, **phase**, **frequency** and/or **time**
- A vector space is called a **linear vector space** if it satisfies the following conditions:
  - 1: $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$
  - 2: $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$
  - 3: $\alpha (\mathbf{x} + \mathbf{y}) = \alpha \mathbf{x} + \alpha \mathbf{y}$
  - 4: $(\alpha + \beta) \mathbf{x} = \alpha \mathbf{x} + \beta \mathbf{x}$
  - where $\mathbf{x}$ and $\mathbf{y}$ are arbitrary vectors and α and β are scalars

Linear Vector Space (Cont.)

- In an N-dimensional linear vector space, we define a **inner product** as $\mathbf{x} \cdot \mathbf{y} \triangleq \sum_{i=1}^{N} x_i y_i$
  - where x_i and y_i are the elements of $\mathbf{x}$ and $\mathbf{y}$, respectively
- Two vectors $\mathbf{x}$ and $\mathbf{y}$ are said to be orthogonal if $\mathbf{x} \cdot \mathbf{y} = 0$.
- The **norm** (or the length) of a vector $\mathbf{x}$ is denoted by $||\mathbf{x}||$

$$\|\mathbf{x}\| \triangleq \sqrt{\mathbf{x} \cdot \mathbf{x}} = \sqrt{\sum_{i=1}^{N} x_i^2}$$

- This norm has the following properties:
  - 5: $||\mathbf{x}|| \geq 0$
  - 6: $||\mathbf{x}|| = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$
  - 7: $||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$
  - 8: $||\alpha\, \mathbf{x}|| = |\alpha| \cdot ||\mathbf{x}||$
- The Schwarz inequality: $|\mathbf{x} \cdot \mathbf{y}| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|$

Orthonormal Basis Functions

- The signal space is assumed to be an N-dimensional space
  - Constructed by N **orthonormal** basis functions
- The goal of Geometric Representation of Signals is to represent any set of M energy signals $\{s_i(t), i = 1, 2, \ldots, M\}$ as **linear combinations** of N **orthonormal** basis functions, $N \leq M$

$\Rightarrow$ Given a set of real-valued energy signals $s_1(t), s_2(t), \ldots, s_M(t)$

$$s_i(t) = \sum_{j=1}^{N} s_{ij}\phi_j(t), \quad 0 \leq t \leq T \text{ and } i = 1, 2, \cdots, M$$

$$s_{ij} = \int_0^T s_i(t)\phi_j(t)\, dt$$

Orthonormal Basis Functions (Cont.)

- The real-valued basis functions $\phi_1(t), \ldots, \phi_N(t)$ are **orthonormal**

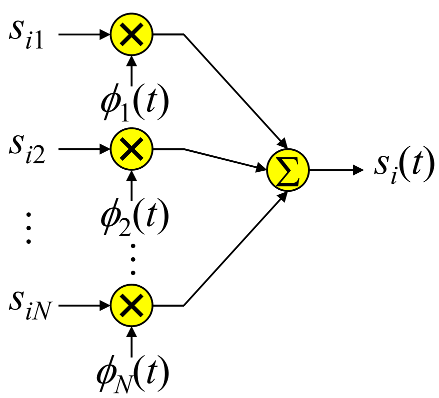
$$\int_0^T \phi_i(t)\phi_j(t)\,dt = \underline{\delta_{ij}} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Kronecker delta function

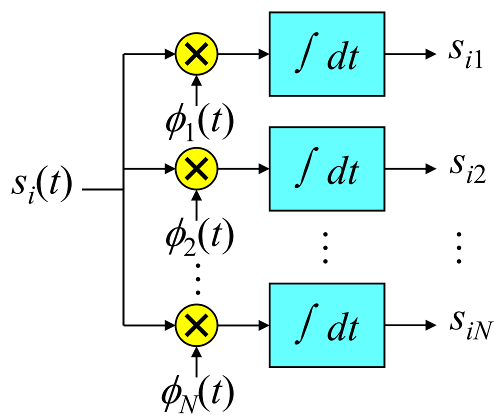
  - Each basis function is normalized to have **unit energy**
  - The basis functions $\phi_1(t), \phi_2(t), \ldots, \phi_N(t)$ are **orthogonal** with respect to each other over the interval $0 \leq t \leq T$
- The set of $\{s_{ij}\}_{j=1}^{N}$ may be viewed as an N-dimensional vector $\mathbf{s}_i$
  - A **one-to-one** relationship with the transmitted signal $s_i(t)$

Signal Synthesizer and Analyzer

- **Synthesizer:** given the N elements of the vectors $\mathbf{s}_i$ (i.e., s_{i1}, $s_{i2}, \ldots, s_{iN}$) as input to generate the signal $s_i(t)$
- **Analyzer:** given the signals $s_i(t)$, $i = 1, 2, \ldots, M$, as input to calculate the coefficients $s_{i1}, s_{i2}, \ldots, s_{iN}$



Synthesizer



Analyzer

Signal Vector

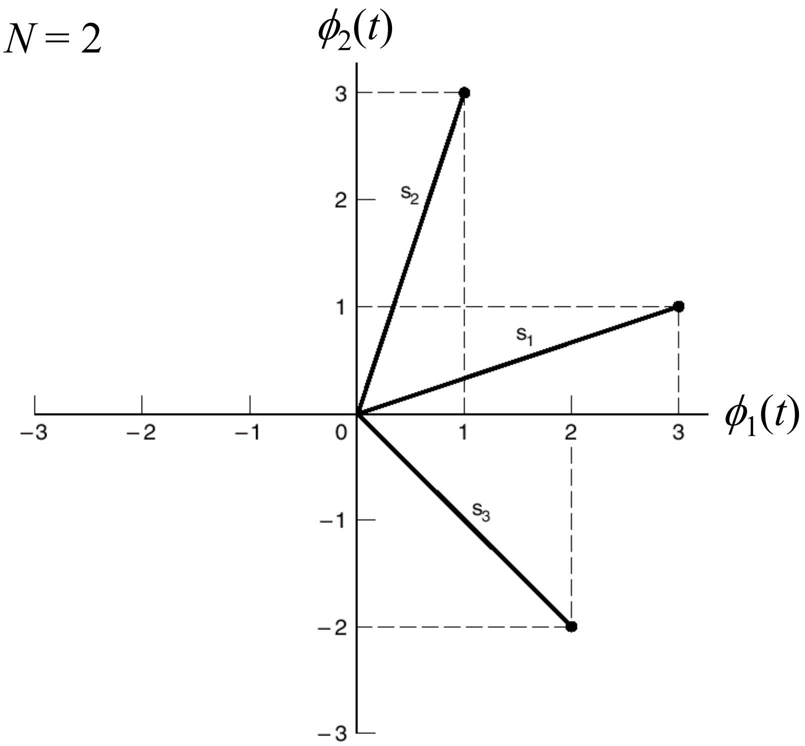
- Each signal in the set $\{s_i(t)\}$ is **completely determined** by the vector of its coefficients

$$\mathbf{s}_i = \begin{bmatrix} s_{i1} \\ s_{i2} \\ \vdots \\ s_{iN} \end{bmatrix}, \quad i = 1, 2, \cdots, M$$

- The vector $\mathbf{s}_i$ is called a **signal vector**
- Consider an ***N*-dimensional** Euclidean space
  - There are N mutually perpendicular axes labeled $\phi_1(t)$, $\phi_2(t)$, …, $\phi_N(t)$
  - The set of signal vectors $\{\mathbf{s}_i \mid i = 1, 2, ..., M\}$ defines a corresponding set of ***M* points** in the Euclidean space
  - This Euclidean space is called the **signal space**

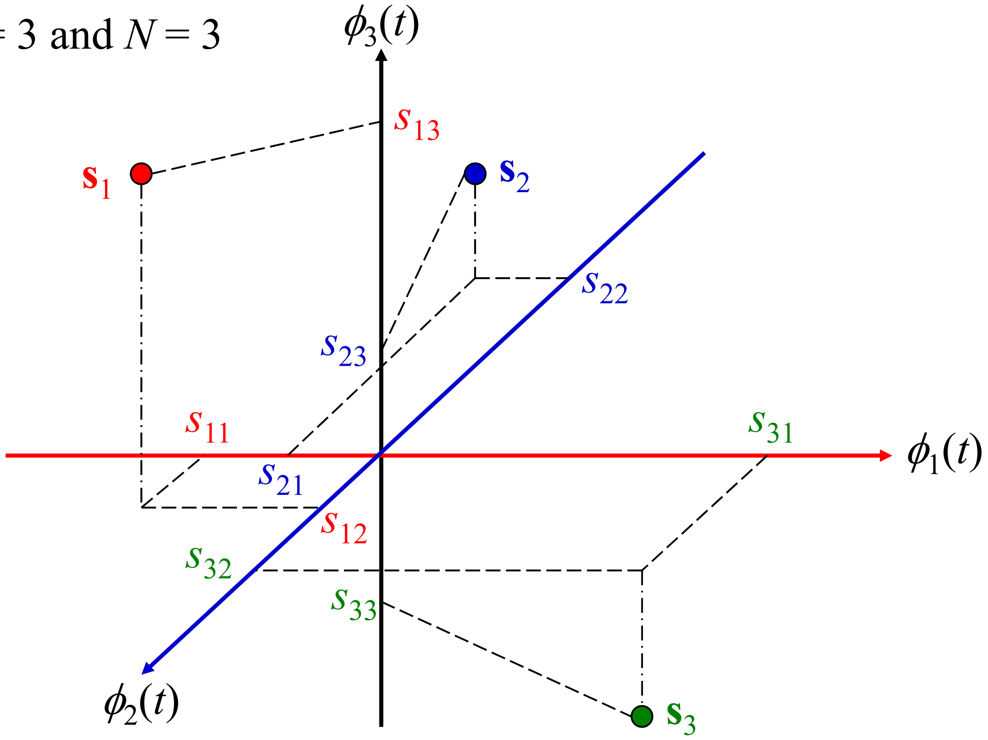
Signal Space

- $M = 3$ and $N = 2$



Signal Space

- $M = 3$ and $N = 3$



Signal Energy

- The squared-length of any signal vector $\mathbf{s}_i$ is defined as the **inner product** or **dot product** of $\mathbf{s}_i$ with itself

$$\|\mathbf{s}_i\|^2 = \mathbf{s}_i^T \mathbf{s}_i = \sum_{j=1}^{N} s_{ij}^2, \quad i = 1, 2, \cdots, M$$

- The **energy** of a signal $s_i(t)$ of duration T seconds is defined as

$$E_i = \int_0^T s_i^2(t)\, dt = \int_0^T \left[\sum_{j=1}^{N} s_{ij}\phi_j(t)\right]\left[\sum_{k=1}^{N} s_{ik}\phi_k(t)\right] dt$$

$$= \sum_{j=1}^{N}\sum_{k=1}^{N} s_{ij}s_{ik}\int_0^T \phi_j(t)\phi_k(t)\, dt$$

$$\boxed{s_i(t) = \sum_{j=1}^{N} s_{ij}\phi_j(t)}$$

- Since the $\phi_j(t)$, j = 1, 2, ..., N, form an **orthonormal set**

$$E_i = \sum_{j=1}^{N} s_{ij}^2 = \|\mathbf{s}_i\|^2 \qquad \boxed{\int_0^T \phi_j(t)\phi_k(t)\, dt = \delta_{jk}}$$

Distance and Angle

- For a pair of signals $s_i(t)$ and $s_k(t)$, the inner product of the signals over the interval [0, T] is

$$\int_0^T s_i(t)s_k(t)\,dt = \mathbf{s}_i^T\mathbf{s}_k$$

- For a specific pair of signals $s_i(t)$ and $s_k(t)$, the inner product is **invariant** to the choice of basis functions $\{\phi_j(t)\}_{j=1}^N$
  - The **rotation** of the coordinate system does not change the locations of signal points
- The **Euclidean distance** of two vectors $\mathbf{s}_i$ and $\mathbf{s}_k$ is

$$d_{ik} = \left\|\mathbf{s}_i - \mathbf{s}_k\right\|^2 = \sum_{j=1}^N (s_{ij} - s_{kj})^2 = \int_0^T \left[s_i(t) - s_k(t)\right]^2\,dt$$

- The **angle** θ_{ik} between two signal vectors $\mathbf{s}_i$ and $\mathbf{s}_k$ follows

$$\cos\theta_{ik} = \frac{\mathbf{s}_i^T\mathbf{s}_k}{\left\|\mathbf{s}_i\right\|\,\left\|\mathbf{s}_k\right\|}$$

Example 1: Schwarz Inequality

- Consider any pair of energy signals $s_1(t)$ and $s_2(t)$. The **Schwarz inequality** state that

$$\left(\int_{-\infty}^{\infty} s_1(t)s_2(t)\,dt\right)^2 \le \left(\int_{-\infty}^{\infty} s_1^2(t)\,dt\right)\left(\int_{-\infty}^{\infty} s_2^2(t)\,dt\right)$$

  - The equality holds if and only if $s_2(t) = cs_1(t)$, where c is a constant
- ***Proof***: Let $s_1(t)$ and $s_2(t)$ be expressed in terms of the pair of orthonormal basis functions $\phi_1(t)$ and $\phi_2(t)$ as follows:

$$s_1(t) = s_{11}\phi_1(t) + s_{12}\phi_2(t); \quad s_2(t) = s_{21}\phi_1(t) + s_{22}\phi_2(t)$$

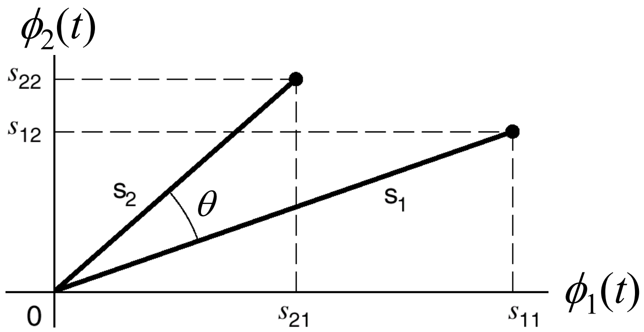
- The angle θ between two signal vectors $\mathbf{s}_1$ and $\mathbf{s}_2$ is

$$\cos\theta = \frac{\mathbf{s}_1^T\mathbf{s}_2}{\left\|\mathbf{s}_1\right\|\,\left\|\mathbf{s}_2\right\|} = \frac{\int_{-\infty}^{\infty} s_1(t)s_2(t)\,dt}{\left(\int_{-\infty}^{\infty} s_1^2(t)\,dt\right)^{1/2}\left(\int_{-\infty}^{\infty} s_2^2(t)\,dt\right)^{1/2}}$$

Example 1: Schwarz Inequality (Cont.)

- Since $|\cos\theta| \le 1 \Rightarrow$ the Schwarz inequality holds
- $|\cos\theta| = 1$ if and only if $\theta = 0 \Rightarrow \mathbf{s}_2 = c\mathbf{s}_1$
  - In other words, $s_2(t) = cs_1(t)$
- If $s_1(t)$ and $s_2(t)$ are **complex-valued** signals, the Schwarz inequality becomes

$$\left|\int_{-\infty}^{\infty} s_1(t)s_2^*(t)\,dt\right|^2 \le \left(\int_{-\infty}^{\infty} |s_1(t)|^2\,dt\right)\left(\int_{-\infty}^{\infty} |s_2(t)|^2\,dt\right)$$



Gram-Schmidt Orthogonalization Procedure

- Suppose we have a set of M energy signals: $s_1(t)$, $s_2(t)$, …, $s_M(t)$ with signal energy E_1, E_2, …, E_M
- We need to generate a complete orthonormal set of basis functions $\Rightarrow$ **Gram-Schmidt Orthogonalization Procedure**
- Starting with $s_1(t)$ chosen from this set **arbitrarily**,
  - The first basis function is defined by

$$\phi_1(t) = s_1(t)\big/\sqrt{E_1}\,, \quad s_1(t) = \sqrt{E_1}\phi_1(t) = s_{11}\phi_1(t)$$

- Next, using the signal $s_2(t)$, we define the coefficient s_{21} as

$$s_{21} = \int_0^T s_2(t)\phi_1(t)\,dt$$

- We obtain a new function $g_2(t)$ orthogonal to $\phi_1(t)$ over the interval $0 \le t \le T$

$$g_2(t) = s_2(t) - \underline{s_{21}\phi_1(t)}$$

G-S Orthogonalization Procedure (Cont.)

- Define the second basis function as

$$\phi_2(t) = \frac{g_2(t)}{\sqrt{\int_0^T g_2^2(t)\, dt}} = \frac{s_2(t) - s_{21}\phi_1(t)}{\sqrt{E_2 - s_{21}^2}}$$

$$s_2(t) = s_{21}\phi_1(t) + \sqrt{E_2 - s_{21}^2}\,\phi_2(t) = s_{21}\phi_1(t) + s_{22}\phi_2(t)$$

- Continuing in this fashion, we have

$$g_i(t) = s_i(t) - \sum_{j=1}^{i-1} s_{ij}\phi_j(t)$$

$$s_{ij} = \int_0^T s_i(t)\phi_j(t)\, dt, \quad j = 1, 2, \cdots, i-1$$

- Define the set of basis functions (an orthonormal set)

$$\phi_i(t) = \frac{g_i(t)}{\sqrt{\int_0^T g_i^2(t)\, dt}}, \quad i = 1, 2, \cdots, N$$

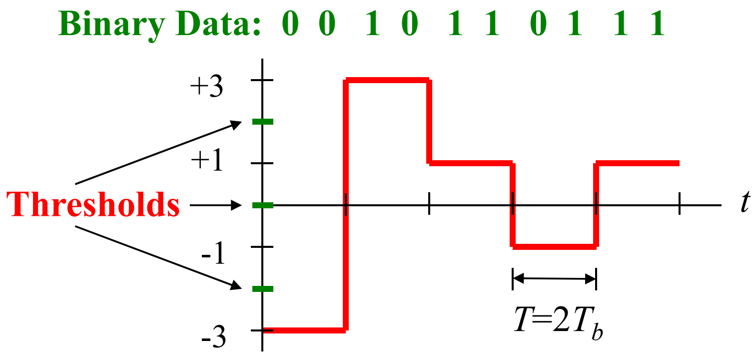
G-S Orthogonalization Procedure (Cont.)

- The dimension N is less than or equal to the number M
  - If the signals $s_1(t), s_2(t), \ldots, s_M(t)$ form a **linearly independent** set, we have $N = M$
  - If the signals $s_1(t), s_2(t), \ldots, s_M(t)$ are **not linearly independent**, we have $N < M$ and the function $g_i(t)$ is zero for $i > N$ ($s_i(t)$, $i > N$, is fully expanded by $\phi_i(t)$, $i = 1, \ldots, M$)
- The form of the basis functions $\phi_1(t), \phi_2(t), \ldots, \phi_N(t)$ has **not been specified**
  - $\phi_i(t)$ is **not restricted to** be either sinusoidal or sinc functions of time
- The expansion of the signal $s_i(t)$ is an **exact expression** where N and only N terms are significant

Example 2: 2B1Q Code

- For **baseband transmission**, the **Gray-encoded** 2B1Q code is used as the quaternary PAM signal
  - **Gray encode**: any symbol differs from an adjacent symbol in a **single bit position**

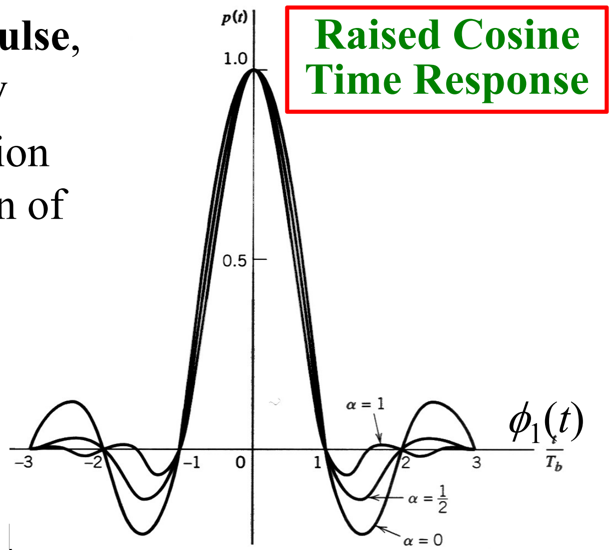
Symbol	Amplitude
10	+3
11	+1
01	–1
00	–3

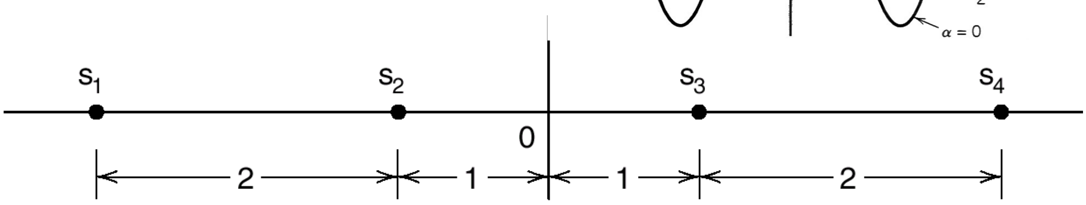


- The four possible signals, $s_1(t)$, $s_2(t)$, $s_3(t)$ and $s_4(t)$, are amplitude-scaled versions of a **Nyquist pulse**

Example 2: 2B1Q Code (Cont.)

- Let $\phi_1(t)$ denote the **Nyquist pulse**, normalized to have unit energy
  - $\phi_1(t)$ is the only basis function for the vector representation of the 2B1Q code
  - $M = 4$ and $N = 1$





Conversion of the Continuous AWGN Channel into a Vector Channel

Correlator Outputs of the Received Signal

- Let the input to the bank of N product correlators (**analyzer**) be the received signal $x(t) = s_i(t) + w(t)$
  - where $w(t)$ is a sample function of a white Gaussian noise process $W(t)$ of **zero mean** and **power spectral density N_0/2**
- The output of correlator j is the sample value of a random variable X_j

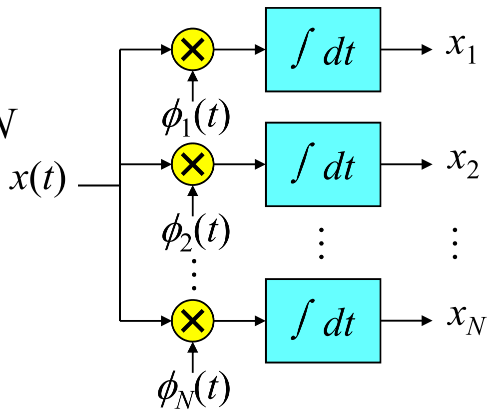
$$x_j = \int_0^T x(t)\phi_j(t)\,dt$$
$$= s_{ij} + w_j, \quad j = 1, 2, \cdots, N$$

  - where

$$s_{ij} = \int_0^T s_i(t)\phi_j(t)\,dt$$

  - The sample value of noise W_j

$$w_j = \int_0^T w(t)\phi_j(t)\,dt$$



Correlator Outputs of the Received Signal(Cont.)

- Consider a new random process $\tilde{X}(t)$ whose sample function is

$$\tilde{x}(t) = x(t) - \sum_{j=1}^{N} x_j \phi_j(t)$$
$$= s_i(t) + w(t) - \sum_{j=1}^{N} \left(s_{ij} + w_j \right) \phi_j(t)$$
$$= w(t) - \sum_{j=1}^{N} w_j \phi_j(t) \triangleq w'(t)$$

  - which depends solely on the channel noise $w(t)$
- The received signal can be expressed as

$$x(t) = \sum_{j=1}^{N} x_j \phi_j(t) + \tilde{x}(t) \triangleq \sum_{j=1}^{N} x_j \phi_j(t) + w'(t)$$

  - $w'(t)$ is the **remainder term** that **cannot be expanded** by the selected basis functions

Statistical Characterization

- The random process $X(t)$ is a **Gaussian process**
  - X_j is a Gaussian random variable for all j

$$\boxed{x_j = s_{ij} + w_j}$$

- The **mean** of X_j depends only on s_{ij},

$$\mu_{X_j} = E\left[X_j \right] = E\left[s_{ij} + W_j \right] = s_{ij} + E\left[W_j \right] = s_{ij}$$

- The **variance** of X_j is

$$\sigma_{X_j}^2 = \mathrm{var}\left[X_j \right] = E\left[(X_j - s_{ij})^2 \right] = E\left[W_j^2 \right]$$

- Note that the random variable W_j is defined as

$$W_j = \int_0^T W(t)\phi_j(t)\, dt$$

- Therefore, we have

$$\sigma_{X_j}^2 = E\left[\int_0^T W(t)\phi_j(t)\, dt \int_0^T W(u)\phi_j(u)\, du \right]$$
$$= E\left[\int_0^T \int_0^T \phi_j(t)\phi_j(u) W(t) W(u)\, dt\, du \right]$$

Statistical Characterization (Cont.)

- Interchanging the order of integration and expectation:

$$\sigma_{X_j}^2 = \int_0^T \int_0^T \phi_j(t)\phi_j(u)E\left[W(t)W(u)\right] dt\, du$$
$$= \int_0^T \int_0^T \phi_j(t)\phi_j(u)R_W(t,u)\, dt\, du$$

  - where $R_W(t, u)$ is the **autocorrelation function** of $W(t)$

$$R_W(t,u) = \frac{N_0}{2}\delta(t-u)$$

- Thus, we obtain

$$\sigma_{X_j}^2 = \frac{N_0}{2}\int_0^T \phi_j^2(t)\, dt$$

- Since the basis functions $\phi_j(t)$ have **unit energy**, we get

$$\sigma_{X_j}^2 = \frac{N_0}{2}, \quad \forall\ j$$

Statistical Characterization (Cont.)

- Since the basis functions $\phi_j(t)$ form an orthogonal set, X_j are **mutually uncorrelated**

$$\begin{aligned}\mathrm{cov}\left[X_j X_k\right] &= E\left[(X_j - \mu_{X_j})(X_k - \mu_{X_k})\right] \\ &= E\left[(X_j - s_{ij})(X_k - s_{ik})\right] = E\left[W_j W_k\right] \\ &= \int_0^T \int_0^T \phi_j(t)\phi_k(u)R_W(t,u)\, dt\, du \\ &= \frac{N_0}{2}\int_0^T \int_0^T \phi_j(t)\phi_k(u)\delta(t-u)\, dt\, du \\ &= \frac{N_0}{2}\int_0^T \phi_j(t)\phi_k(t)\, dt \\ &= 0, \quad \text{for } j \neq k\end{aligned}$$

- Since the X_j are Gaussian random variables, they are also **statistically independent** (Property of a Gaussian Process)

Statistical Characterization (Cont.)

- Define the **observation vector** of N random variables as

$$\mathbf{X} = \left[X_1 \quad X_2 \quad \cdots \quad X_N \right]^T$$

- The conditional probability density function (pdf) of $\mathbf{X}$, given that $s_i(t)$ or correspondingly the symbol m_i was transmitted, is

$$f_{\mathbf{X}}(\mathbf{x}|m_i) = \prod_{j=1}^{N} f_{X_j}(x_j|m_i), \quad i = 1, 2, \cdots, M$$

Statistically independent

- Since each X_j is a Gaussian random variable

$$f_{X_j}(x_j|m_i) = \frac{1}{\sqrt{\pi N_0}} \exp\left[-\frac{1}{N_0}\left(x_j - s_{ij}\right)^2 \right]$$

- The conditional pdf of $\mathbf{X}$ is

$$f_{\mathbf{X}}(\mathbf{x}|m_i) = \left(\pi N_0\right)^{-N/2} \exp\left[-\frac{1}{N_0}\sum_{j=1}^{N}\left(x_j - s_{ij}\right)^2 \right]$$

Statistical Characterization (Cont.)

- Note that the observation vector $\mathbf{X}$ completely characterizes the received signal $x(t)$, except for the **remaining noise term** $w'(t)$

$$x(t) = \sum_{j=1}^{N} x_j \phi_j(t) + w'(t)$$

- Since the noise process $W(t)$ is Gaussian with zero mean, the noise process $W'(t)$, with the sample function $w'(t)$, is also a **zero-mean Gaussian** process
- Any random variable $W'(t_k)$, derived from $W'(t)$, is **statistically independent** of the set of random variables $\{X_j\}$, i.e.,

$$E\left[X_j W'(t_k) \right] = 0, \quad j = 1, 2, \cdots, N$$

- $W'(t_k)$ is **irrelevant to** the message decision
  – $W'(t_k)$ is outside the N-dimensional signal space
  – The N correlator outputs are used for **decision-making**

Theorem of Irrelevance

- **Theorem of irrelevance**: For signal detection in **additive white Gaussian noise**, only the projections of the noise onto the basis functions of the signal set $\{s_i(t)\}$ affects the **sufficient statistics** (i.e., the statistics of **X**) of the detection problem; **the remainder of the noise is irrelevant**.

- The **AWGN channel** is equivalent to an ***N*-dimensional vector channel** described by the observation vector

$$\mathbf{x} = \mathbf{s}_i + \mathbf{w}, \quad i = 1, 2, \cdots, M$$

$$\mathbf{w} = \begin{bmatrix} w_1 & w_2 & \cdots & w_N \end{bmatrix}$$

  – where the dimension N is **the number of basis functions** involved in formulating the signal vector $\mathbf{s}_i$

Likelihood Functions

Likelihood Functions

- **At the receiver**, we are given the observation vector **x**
  - The requirement is **to estimate the message symbol** m_i that is responsible for generating **x**
- The definition of the **likelihood function**

$$L(m_i) = f_{\mathbf{X}}(\mathbf{x}|m_i), \quad i = 1, 2, \cdots, M$$

  - The **possibility** that the message symbol m_i was transmitted when the observation vector is **x**
- In practice, we generally use the **log-likelihood function**

$$l(m_i) = \log L(m_i), \quad i = 1, 2, \cdots, M$$

- The log-likelihood function bears a **one-to-one mapping** to the likelihood function
  - A probability density function is always **nonnegative**
  - The logarithmic function is **monotonically increasing**

Likelihood Functions

- For an **AWGN channel**, the log-likelihood function is

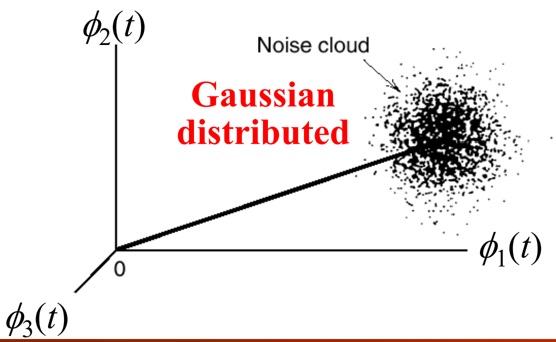
$$l(m_i) = -\frac{1}{N_0}\sum_{j=1}^{N}\left(x_j - s_{ij}\right)^2, \quad i = 1, 2, \cdots, M$$

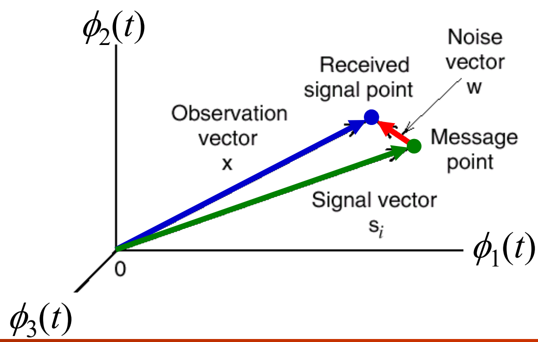
  - where the constant term $-(N/2)\log(\pi N_0)$ is ignored
  - The constant term is the same for different values of m_i
  - It bears no relation whatsoever to the message symbol m_i

Coherent Detection of Signals in Noise: Maximum Likelihood Decoding

Signal Points

- The transmitted signal $s_i(t)$ can be represented as a point in a **Euclidean space** of dimension $N \leq M$
  - The **transmitted signal point** or **message point** of $s_i(t)$
- The set of message points corresponding to the set of transmitted signals is called as a **signal constellation**
- The observation vector **x** (**received signal point**) differs from the transmitted signal vector $\mathbf{s}_i$ by a **random noise vector w**





Signal Detection – MAP Decision Rule

- Given the observation vector $\mathbf{x}$, perform a mapping from $\mathbf{x}$ to an **estimate** $\hat{m}$ of the transmitted symbol m_i
  - In a way that would **minimize the probability of error** in the decision-making process

$$P_e(m_i|\mathbf{x}) = P(m_i \text{ not sent}|\mathbf{x}) = 1 - P(m_i \text{ sent}|\mathbf{x})$$

- We can state the **optimum decision rule** as:

$$\boxed{\begin{array}{c}\text{Set } \hat{m} = m_i \text{ if} \\ P(m_i \text{ sent}|\mathbf{x}) \geq P(m_k \text{ sent}|\mathbf{x}) \quad \text{for all } k \neq i\end{array}}$$

- This decision rule is referred to as the **maximum *a posteriori* probability (MAP)** rule

Signal Detection – MAP Decision Rule (Cont.)

- Using **Bayes' rule**, we may restate the MAP rule as follows:

$$\text{Set } \hat{m} = m_i \text{ if } \frac{p_k f_{\mathbf{X}}(\mathbf{x}|m_k)}{f_{\mathbf{X}}(\mathbf{x})} \text{ is maximum for } k = i$$

The *a priori* probabilities are required

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(A \cap B)}{P(A)P(B)} = \frac{P(A)P(B|A)}{P(B)}}$$

  - p_k is the ***a priori* probability** of transmitting symbol m_k
  - $f_{\mathbf{X}}(\mathbf{x}|m_k)$ is the **conditional pdf** of $\mathbf{X}$ given the transmission of m_k, and $f_{\mathbf{X}}(\mathbf{x})$ is the unconditional pdf of $\mathbf{X}$
- The denominator $f_{\mathbf{X}}(\mathbf{x})$ is independent of the transmitted symbol
- If all the symbols are **equally likely**, $p_k = p_i$ for all i and k
  - The conditional pdf $f_{\mathbf{X}}(\mathbf{x}|m_k)$ bears a **one-to-one mapping** to the **log-likelihood function** $l(m_k)$

Maximum Likelihood Decision Rule

- Accordingly, we can restate the decision rule as follows:

$$\boxed{\begin{array}{c}\text{Set } \hat{m} = m_i \text{ if} \\ l(m_k) \text{ is maximum for } k = i\end{array}}$$

  - This decision rule is referred to as the **maximum likelihood (ML) rule** and the device for its implementation is the **maximum likelihood decoder**
- Based on the observation vector **x**, the decoder **computes** the log-likelihood functions as metrics for all the M possible message symbols, **compares** them, and then **decides** in favor of the maximum
- The ML decoder differs from the MAP decoder in that **it assumes equally likely message symbols**
  - The ML decoder **does not require** the *a priori* probabilities

Comparison: MAP and ML Decision Rules

- If the transmitting message symbols are **equally likely**, the MAP and ML decision rules have **the same** performance.
- If the transmitting message symbols are **not equally likely**, the MAP decision rule is **superior to** the ML decision rule
  - Since the information of *a priori* probabilities is available
- For example, there are two message symbols with the *a priori* probabilities $p_0 = 0.9999$ and $p_1 = 0.0001$
  - Considering the *a priori* probabilities is very important
- However, in general, the transmitting message symbols are **equally likely** for practical systems
  - The ML decision rule is commonly used

Maximum Likelihood Decision Rule (Cont.)

- Let Z denote the N-dimensional space (**observation space**) of all possible observation vectors **x**
- The observation space Z is partitioned into ***M*-decision regions**, $Z_1, Z_2, \ldots, Z_M$ (which are **non-overlapping**)
  - Accordingly, we can restate the ML decision rule as follows:

$$\text{Observation vector } \mathbf{x} \text{ lies in region } Z_i \text{ if}$$
$$l(m_k) \text{ is maximum for } k = i$$

- For an AWGN channel, the LLF $l(m_k)$ attains its maximum value when $\sum_{j=1}^{N}(x_j - s_{kj})^2$ is **minimized by the choice** $k = i$
  - Accordingly, we can restate the ML decision rule as follows:

$$\text{Observation vector } \mathbf{x} \text{ lies in region } Z_i \text{ if}$$
$$\sum_{j=1}^{N}(x_j - s_{kj})^2 \text{ is minimum for } k = i$$

Maximum Likelihood Decision Rule (Cont.)

- $\sum_{j=1}^{N}(x_j - s_{kj})^2$ is the Euclidean distance square between **x** and $\mathbf{s}_k$
  - Accordingly, we can restate the ML decision rule as follows:

$$\text{Observation vector } \mathbf{x} \text{ lies in region } Z_i \text{ if}$$
$$\text{the Euclidean distance } \left\| \mathbf{x} - \mathbf{s}_k \right\| \text{ is minimum for } k = i$$

- The ML decision rule is simply **to choose the message point closest to the received signal point**
- Note that $$\sum_{j=1}^{N}(x_j - s_{kj})^2 = \underbrace{\sum_{j=1}^{N} x_j^2}_{\text{Irrelevant to } k} - 2\sum_{j=1}^{N} x_j s_{kj} + \sum_{j=1}^{N} s_{kj}^2$$
  - Accordingly, we can restate the ML decision rule as follows:

$$\text{Observation vector } \mathbf{x} \text{ lies in region } Z_i \text{ if}$$
$$\sum_{j=1}^{N} x_j s_{kj} - E_k/2 \text{ is maximum for } k = i$$

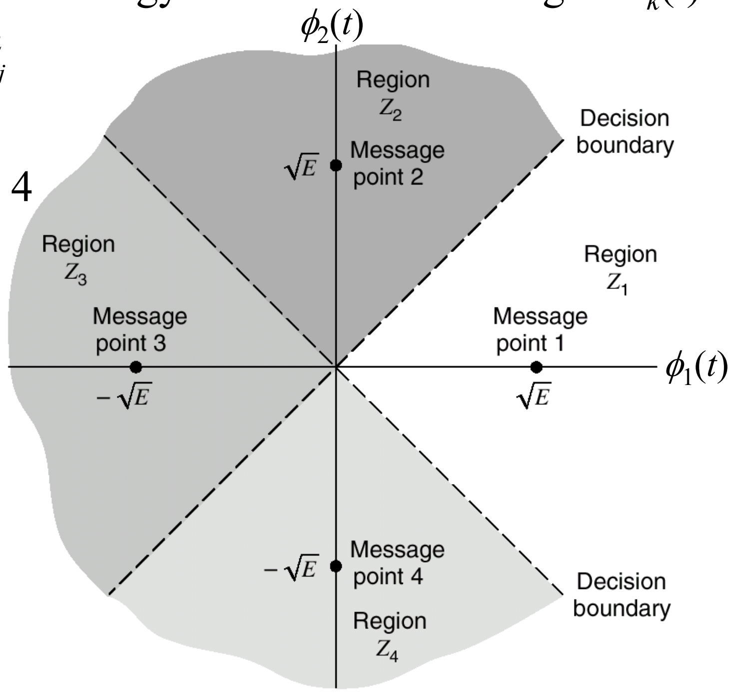
Maximum Likelihood Decision Rule (Cont.)

– where E_k is the energy of the transmitted signal $s_k(t)$

$$E_k = \sum_{j=1}^{N} s_{kj}^2$$

- Example:

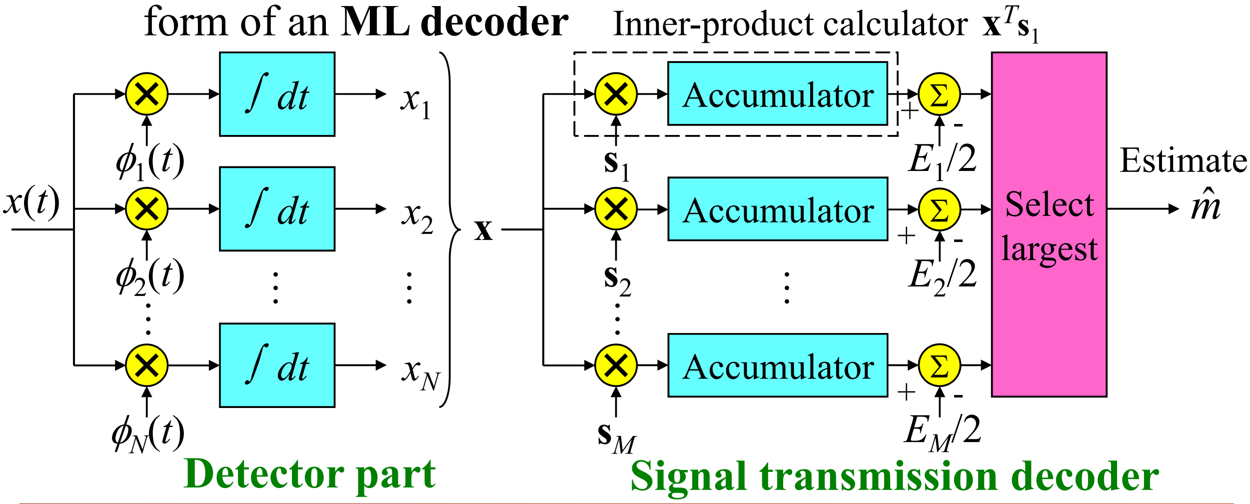
  $N = 2$ and $M = 4$



Correlation Receiver

Optimum Receiver

- The optimum receiver consists of two subsystems:
  - The **detector part**: It consists of a bank of M product-integrators or **correlators**
  - The **signal transmission decoder**: It is implemented in the form of an **ML decoder**



Equivalence: Correlation – Matched Filter

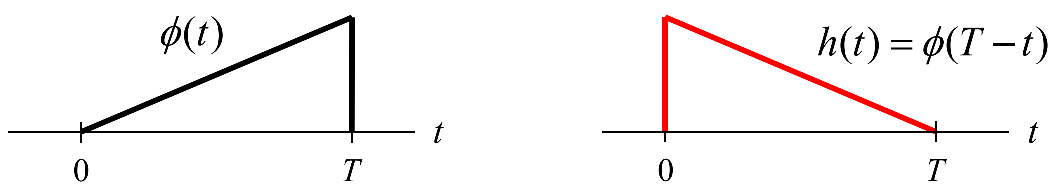
- We can use a corresponding set of **matched filters** to build the detector
- Consider a linear time-invariant filter with impulse response $h_j(t)$
- When the received signal $x(t)$ is used as the filter input, the resulting filter output $y_j(t)$ is

$$y_j(t) = \int_{-\infty}^{\infty} x(\tau) h_j(t-\tau)\, d\tau$$

$$\boxed{\begin{aligned} x(t) &= s_i(t) + w(t) \\ s_i(t) &= \sum_{j=1}^{N} s_{ij}\phi_j(t) \end{aligned}}$$

- The impulse response $h_j(t)$ **matched to** an input signal $\phi_j(t)$ is

$$h_j(t) = \phi_j(T-t)$$



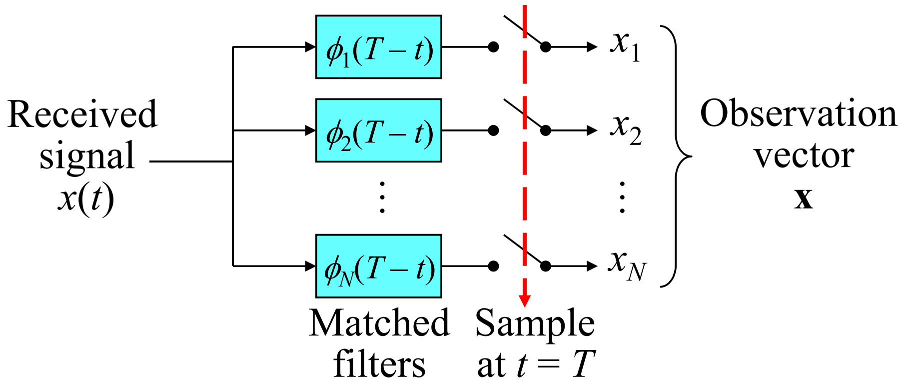
Equivalence: Correlation – Matched Filter(Cont.)

- Then the filter output is $y_j(t) = \int_{-\infty}^{\infty} x(\tau)\phi_j(T - t + \tau)\, d\tau$
- Sampling the output at time $t = T$, we get

$$y_j(T) = \int_{-\infty}^{\infty} x(\tau)\phi_j(\tau)\, d\tau$$

- Since $\phi_j(t)$ is zero outside the interval $0 \le t \le T$

$$y_j(T) = \int_{0}^{T} x(\tau)\phi_j(\tau)\, d\tau$$



Probability of Error

Noise Performance

- Suppose that symbol m_i is transmitted and an observation vector **x** is received
  - An **error** occurs whenever the received signal point does not fall inside region Z_i
- The average probability of symbol error is

$$P_e = \sum_{i=1}^{M} p_i P(\mathbf{x} \text{ does not lie in } Z_i | m_i \text{ sent})$$

$$= 1 - \frac{1}{M} \sum_{i=1}^{M} P(\mathbf{x} \text{ lies in } Z_i | m_i \text{ sent}) \quad \boxed{p_i = \frac{1}{M}, \quad \forall\, i}$$

- Since **x** is the sample value of random vector **X**, P_e can be expressed in terms of the **likelihood function** as follows:

$$P_e = 1 - \frac{1}{M} \sum_{i=1}^{M} \int_{Z_i} f_{\mathbf{X}}(\mathbf{x}|m_i)\, d\mathbf{x}$$

Invariance to Rotation and Translation

- For the ML detection of a signal in AWGN, **changes in the orientation of the signal constellation** with respect to both the **coordinate axes** and **origin** of the signal space **do not affect** the probability of symbol error P_e
  - In ML detection, P_e depends solely on the **relative Euclidean distances** between the message points
  - The AWGN is **spherically symmetric** in all directions
- The effect of a **rotation** applied to all the message points is equivalent to multiplying the signal vector $\mathbf{s}_i$ by an ***N*-by-*N* orthonormal matrix Q** for all i
  - where the matrix **Q** satisfies

$$\mathbf{Q}\mathbf{Q}^T = \mathbf{I} \longleftarrow \textbf{Identity matrix}$$

Invariance to Rotation and Translation (Cont.)

- The signal (noise) vector $\mathbf{s}_i$ ($\mathbf{w}$) is replaced by the rotated version

$$\mathbf{s}_{i,\text{rotate}} = \mathbf{Q}\mathbf{s}_i, \quad i = 1, 2, \cdots, M$$
$$\mathbf{w}_{\text{rotate}} = \mathbf{Q}\mathbf{w}$$

- The statistical characteristics of the noise vector are **unaffected**

$$E\left[\mathbf{w}_{\text{rotate}}\right] = \mathbf{0}; \quad E\left[\mathbf{w}_{\text{rotate}}\mathbf{w}_{\text{rotate}}^T\right] = \frac{N_0}{2}\mathbf{I}$$

- The observation vector for the **rotated signal constellation** is

$$\mathbf{x}_{\text{rotate}} = \mathbf{Q}\mathbf{s}_i + \mathbf{w}_{\text{rotate}} = \mathbf{Q}\mathbf{s}_i + \mathbf{w}, \quad i = 1, 2, \cdots, M$$

- The Euclidean distance between $\mathbf{x}_{\text{rotate}}$ and $\mathbf{s}_{i,\text{rotate}}$ is

$$\left\|\mathbf{x}_{\text{rotate}} - \mathbf{s}_{i,\text{rotate}}\right\| = \left\|\mathbf{x} - \mathbf{s}_i\right\|, \quad \text{for all } i$$

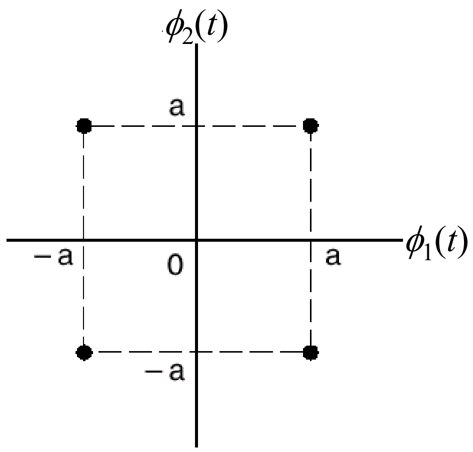
- Also note that

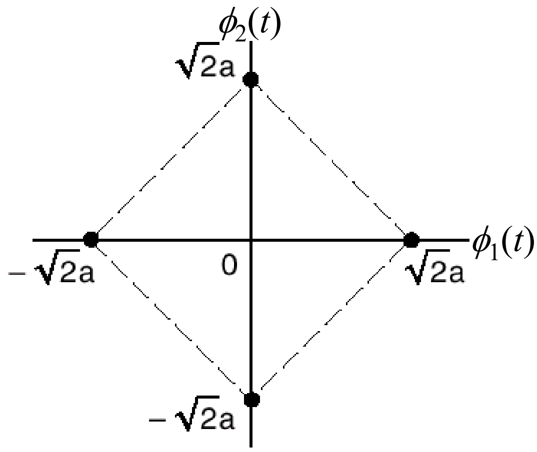
$$\mathbf{x} = \mathbf{Q}^T\mathbf{x}_{\text{rotate}} = \mathbf{Q}^T\mathbf{Q}\mathbf{s}_i + \mathbf{Q}^T\mathbf{w} = \mathbf{s}_i + \mathbf{w}'_{\text{rotate}} = \mathbf{s}_i + \mathbf{w}$$

Invariance to Rotation and Translation (Cont.)

- For translation, we have

$$\mathbf{s}_{i,\text{translate}} = \mathbf{s}_i - \mathbf{a}, \quad i = 1, 2, \cdots, M$$
$$\mathbf{x}_{\text{translate}} = \mathbf{x} - \mathbf{a},$$
$$\left\|\mathbf{x}_{\text{translate}} - \mathbf{s}_{i,\text{translate}}\right\| = \left\|\mathbf{x} - \mathbf{s}_i\right\|, \quad \text{for all } i$$





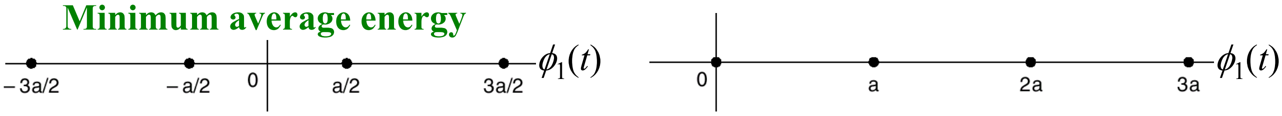
Minimum Energy Signals

- Based on the principle of **translational invariance**, we can translate the signal constellation to **minimize** the average energy
- The average energy of the signal constellation translated by a vector **a** is $\mathscr{E}_{\text{translate}} = \sum_{i=1}^{M} \|\mathbf{s}_i - \mathbf{a}\|^2 p_i$ $\quad \boxed{\|\mathbf{s}_i - \mathbf{a}\|^2 = \|\mathbf{s}_i\|^2 - 2\mathbf{a}^T\mathbf{s}_i + \|\mathbf{a}\|^2}$

$$\mathscr{E}_{\text{translate}} = \sum_{i=1}^{M} \|\mathbf{s}_i\|^2 p_i - 2\sum_{i=1}^{M} \mathbf{a}^T\mathbf{s}_i p_i + \sum_{i=1}^{M} \|\mathbf{a}\|^2 p_i = \mathscr{E} - 2\mathbf{a}^T E[\mathbf{s}] + \|\mathbf{a}\|^2$$

  - where $\mathscr{E}$ is the average energy of the **original** signal constellation and $E[\mathbf{s}] = \sum_{i=1}^{M} \mathbf{s}_i p_i$
- Differentiating $\mathscr{E}_{\text{translate}}$ with respect to **a** and setting it to zero
  - The minimizing translate is $\mathbf{a}_{\min} = E[\mathbf{s}]$
  - The minimum average energy $\mathscr{E}_{\text{translate,min}} = \mathscr{E} - \|\mathbf{a}_{\min}\|^2 = E[\mathbf{s}]$

Minimum average energy



Union Bound on the Probability of Error

- The average probability of symbol error P_e is

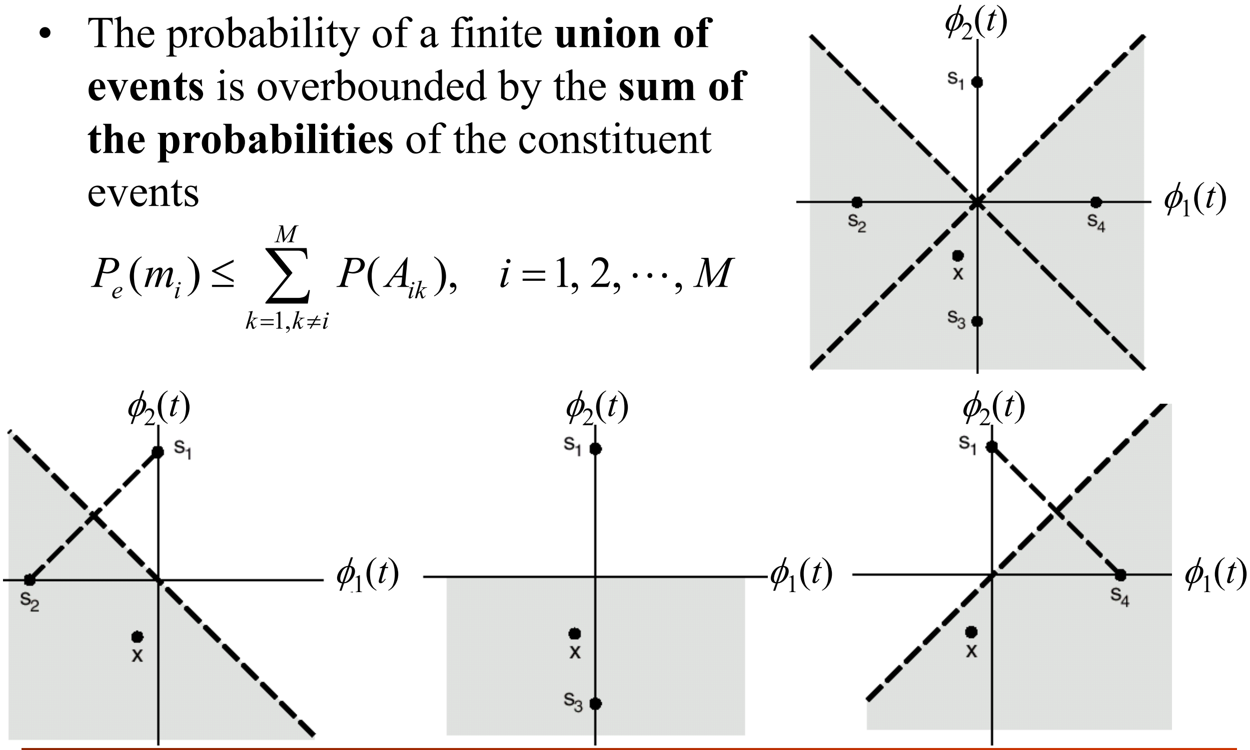
$$P_e = 1 - \frac{1}{M}\sum_{i=1}^{M} \int_{Z_i} f_{\mathbf{X}}(\mathbf{x}|m_i)\, d\mathbf{x}$$

- Numerical computation of the integral may be **impractical**
  - We can approximate P_e by **simplifying the integral** or **simplifying the region of integration**
- **Union bound:** a simple **upper bound** that bases on simplifying the region of integration $\quad \boxed{A_{ik} \text{ and } A_{ij} \text{ may be overlapped}}$
- Let A_{ik}, i, $k \in \{1, 2, \ldots, M\}$, denote the event that the observation vector **x** is **closer to $\mathbf{s}_k$ than to $\mathbf{s}_i$** when the symbol m_i is sent
- The conditional probability of symbol error $P_e(m_i)$ is equal to the probability of the **union events** A_{i1}, A_{i2}, …, $\underline{A_{i,i-1}, A_{i,i+1}}$, …, A_{iM}

Union Bound on the Probability of Error (Cont.)

- The probability of a finite **union of events** is overbounded by the **sum of the probabilities** of the constituent events

$$P_e(m_i) \le \sum_{k=1, k\ne i}^{M} P(A_{ik}), \quad i = 1, 2, \cdots, M$$



Union Bound on the Probability of Error (Cont.)

- Note that the probability $P(A_{ik})$ is **different from** the probability $P(\hat{m} = m_k | m_i)$
- $P(A_{ik}) \triangleq P_2(\mathbf{s}_i, \mathbf{s}_k)$ is the **pairwise error probability** in that the system uses only a pair of signals $\mathbf{s}_i$ and $\mathbf{s}_k$

$$P_2(\mathbf{s}_i, \mathbf{s}_k) = P(\mathbf{x} \text{ is closer to } \mathbf{s}_k \text{ than } \mathbf{s}_i, \text{ when } \mathbf{s}_i \text{ is sent})$$
$$= \int_{d_{ik}/2}^{\infty} \frac{1}{\sqrt{\pi N_0}} \exp\left(-v^2 / N_0\right) dv$$

  - $d_{ik} = \| \mathbf{s}_i - \mathbf{s}_k \|$ is the Euclidean distance between $\mathbf{s}_i$ and $\mathbf{s}_k$
- Setting $z = v/\sqrt{N_0}$,

$$\boxed{\mathrm{erfc}(u) = \frac{2}{\sqrt{\pi}} \int_u^{\infty} \exp(-z^2)\, dz}$$

$$P_2(\mathbf{s}_i, \mathbf{s}_k) = \frac{1}{2} \mathrm{erfc}\left(d_{ik} / 2\sqrt{N_0}\right)$$
$$\Rightarrow P_e(m_i) \le \frac{1}{2} \sum_{k=1, k\ne i}^{M} \mathrm{erfc}\left(d_{ik} / 2\sqrt{N_0}\right), \quad i = 1, 2, \cdots, M$$

Union Bound on the Probability of Error (Cont.)

- The probability of symbol error is overbounded as follows:

$$P_e = \sum_{i=1}^{M} p_i P_e(m_i) \le \frac{1}{2}\sum_{i=1}^{M}\sum_{k=1,k\ne i}^{M} p_i \,\mathrm{erfc}\left(d_{ik}\big/2\sqrt{N_0}\right)$$

- Suppose that the signal constellation is **circularly symmetric** about the origin $\Rightarrow P_e(m_i)$ is the same for all i

$$P_e \le \frac{1}{2}\sum_{k=1,k\ne i}^{M} \mathrm{erfc}\left(d_{ik}\big/2\sqrt{N_0}\right)$$

- Define $d_{\min}$ as the **minimum distance** between any two signals

$$P_e \le \frac{M-1}{2}\mathrm{erfc}\left(d_{\min}\big/2\sqrt{N_0}\right) \qquad \boxed{d_{\min} = \min_{k\ne i} d_{ik}, \text{ for all } i \text{ and } k}$$

- We can also further simplify the union bound on P_e as

$$P_e \le \frac{M-1}{2\sqrt{\pi}}\exp\left(-\underline{d_{\min}^2}\big/4N_0\right) \qquad \boxed{\mathrm{erfc}\left(\frac{d_{\min}}{2\sqrt{N_0}}\right) \le \frac{1}{\sqrt{\pi}}\exp\left(-\frac{d_{\min}^2}{4N_0}\right)}$$

Bit versus Symbol Error Probabilities

- For binary data transmission, it is more meaningful to consider the **bit error rate (BER)**: there are $K = \log_2 M$ bits per symbol
- **Case 1: Gray encode** is applied
  - Any two adjacent symbols differ in only **one bit** position
  - Given a symbol error, the most probable number of bit errors is **1** $\Rightarrow$ The BER is bounded as follows:

Only 1 bit is in error $\longrightarrow$ $\underline{P_e/\log_2 M} \le \mathrm{BER} \le P_e$ $\longleftarrow$ **All bits are in error**

- **Case 2:** All symbol errors occurs **equally likely**
  - The occurrence probability is $P_e/(M-1) = P_e/\left(2^k - 1\right)$
  - There are 2^{K-1} error symbols that the i-th bit is in error
  - The bit error rate is $\mathrm{BER} = \dfrac{2^{K-1}}{2^K - 1}P_e = \dfrac{M/2}{M-1}P_e \approx \underline{P_e/2}$