ECE 630: Statistical Communication Theory

Dr. B.-P. Paris
Dept. Electrical and Comp. Engineering
George Mason University

Last updated: February 26, 2019

© 2018, B.-P. Paris

Part I

Introduction

Elements of a Digital Communications System

Source: produces a sequence of information symbols b.

Transmitter: maps symbol sequence to analog signal $s(t)$.

Channel: models corruption of transmitted signal $s(t)$.

Receiver: produces reconstructed sequence of information symbols $\hat{b}$ from observed signal $R(t)$.

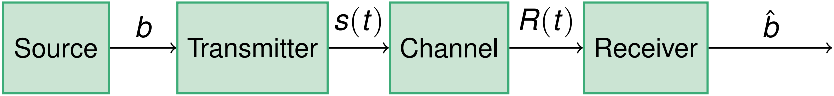


Figure: Block Diagram of a Generic Digital Communications System

The Source

- The source models the statistical properties of the digital information source.
- Three main parameters:
  - **Source Alphabet:** list of the possible information symbols the source produces; also called *Signal Constellation*
    - Example: $\mathcal{A} = \{0, 1\}$; symbols are called bits.
    - Alphabet for a source with M (typically, a power of 2) symbols: e.g., $\mathcal{A} = \{\pm 1, \pm 3, \ldots, \pm(M-1)\}$.
    - Alphabet with positive and negative symbols is often more convenient.
    - Symbols may be complex valued; e.g., $\mathcal{A} = \{\pm 1, \pm j\}$.

- **A priori Probability:** relative frequencies with which the source produces each of the symbols.
  - Example: a binary source that produces (on average) equal numbers of 0 and 1 bits has $\pi_0 = \pi_1 = \frac{1}{2}$.
  - Notation: π_n denotes the probability of observing the n-th symbol.
  - Typically, a-priori probabilities are all equal, i.e., $\pi_n = \frac{1}{M}$.
  - A source with M symbols is called an M-ary source.
    - binary ($M = 2$)
    - quaternary ($M = 4$)

Bit 1	Bit 2	Symbol
0	0	-3
0	1	-1
1	1	$+1$
1	0	$+3$

Table: Example: Representing two bits in one quaternary symbol.

- **Symbol Rate:** The number of information symbols the source produces per second. Also called the baud rate R.
  - Related: information rate R_b, indicates number of bits source produces per second.
  - Relationship: $R_b = R \cdot \log_2(M)$.
  - Also, $T = 1/R$ is the symbol period.
  - Note: for most communication systems, the bandwidth W occupied by the transmitted signal is approximately equal to the baud rate R,
$$W \approx R$$

Remarks

- This view of the source is simplified.
- We have omitted important functionality normally found in the source, including
  - error correction coding and interleaving, and
  - Usually, a block that maps bits to symbols is broken out separately.
- This simplified view is sufficient for our initial discussions.
- Missing functionality will be revisited when needed.

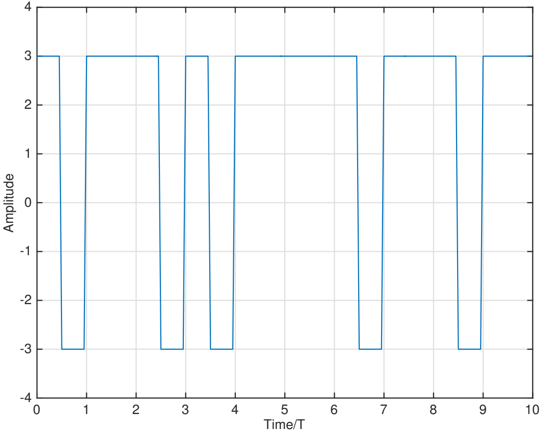
The Transmitter

- The transmitter translates the information symbols at its input into signals that are "appropriate" for the channel, e.g.,
  - meet bandwidth requirements due to regulatory or propagation considerations,
  - provide good receiver performance in the face of channel impairments:
    - noise,
    - distortion (i.e., undesired linear filtering),
    - interference.
- A digital communication system transmits only a discrete set of information symbols.
  - Correspondingly, only a discrete set of possible signals is employed by the transmitter.
  - The transmitted signal is an analog (continuous-time, continuous amplitude) signal.

Illustrative Example

- The sources produces symbols from the alphabet $\mathcal{A} = \{0, 1\}$.
- The transmitter uses the following rule to map symbols to signals:
  - If the n-th symbol is $b_n = 0$, then the transmitter sends the signal
$$s_0(t) = \begin{cases} A & \text{for } (n-1)T \leq t < nT \\ 0 & \text{else.} \end{cases}$$
  - If the n-th symbol is $b_n = 1$, then the transmitter sends the signal
$$s_1(t) = \begin{cases} A & \text{for } (n-1)T \leq t < (n-\frac{1}{2})T \\ -A & \text{for } (n-\frac{1}{2})T \leq t < nT \\ 0 & \text{else.} \end{cases}$$

Symbol Sequence $b = \{1, 0, 1, 1, 0, 0, 1, 0, 1, 0\}$



The Communications Channel

- The communications channel models the degradation the transmitted signal experiences on its way to the receiver.
- For wireless communications systems, we are concerned primarily with:
  - **Noise:** random signal added to received signal.
    - Mainly due to thermal noise from electronic components in the receiver.
    - Can also model interference from other emitters in the vicinity of the receiver.
    - Statistical model is used to describe noise.
  - **Distortion:** undesired filtering during propagation.
    - Mainly due to multi-path propagation.
    - Both deterministic and statistical models are appropriate depending on time-scale of interest.
    - Nature and dynamics of distortion is a key difference between wireless and wired systems.

Thermal Noise

- At temperatures above absolute zero, electrons move randomly in a conducting medium, including the electronic components in the front-end of a receiver.
- This leads to a random waveform.
  - The power of the random waveform equals $P_N = kT_0 W$.
    - k: Boltzmann's constant ($1.38 \times 10^{-23}\,\mathrm{W\,s/K}$).
    - T_0: temperature in degrees Kelvin (room temperature ≈ 290 K).
    - For bandwidth W equal to 1 Hz, $P_N \approx 4 \times 10^{-21}$ W (-174 dBm).
- Noise power is small, but power of received signal decreases rapidly with distance from transmitter.
  - Noise provides a fundamental limit to the range and/or rate at which communication is possible.

Exercise: Path Loss and Signal-to-Noise Ratio

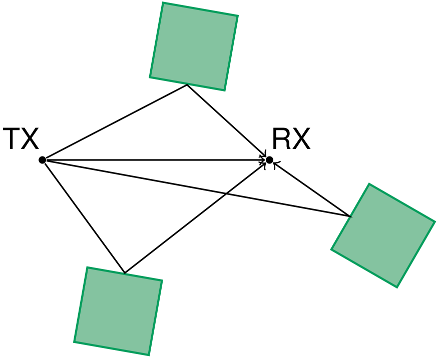
- A transmitter emits a signal with:
  - bandwidth $W = 1\,\mathrm{MHz}$
  - transmitted power $P_t = 1\,\mathrm{mW}$
  - carrier frequency $f_c = 1\,\mathrm{GHz}$
- During propagation from transmitter to receiver, the signal's power decreases; the received power follows Friis law:

$$P_r = P_t \cdot \left(\frac{c}{4\pi f_c d} \right)^2$$

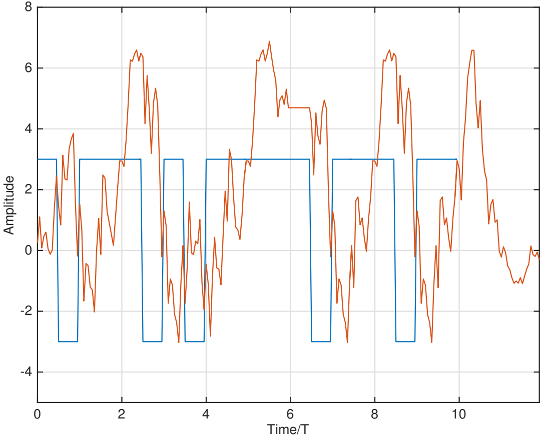
where $c = 3 \times 10^8\,\mathrm{m/s}$ is the speed of light and d is the distance between transmitter and receiver (in meters).
- Find:
  - the power of the received signal P_r for $d = 10\,\mathrm{km}$
  - the noise power P_N in the bandwidth W occupied by the transmitted signal
  - the ratio $\frac{P_r}{P_N}$; this is called the signal-to-noise ratio (SNR)

Multi-Path

- In a multi-path environment, the receiver sees the combination of multiple scaled and delayed versions of the transmitted signal.



Distortion from Multi-Path



- Received signal "looks" very different from transmitted signal.
- Inter-symbol interference (ISI).
- Multi-path is a very serious problem for wireless systems.

The Receiver

- The receiver is designed to reconstruct the original information sequence b.
- Towards this objective, the receiver uses
  - the received signal $R(t)$,
  - knowledge about how the transmitter works,
    - Specifically, the receiver knows how symbols are mapped to signals.
  - the a-priori probability and rate of the source.
- The transmitted signal typically contains information that allows the receiver to gain information about the channel, including
  - training sequences to estimate the impulse response of the channel,
  - synchronization preambles to determine symbol locations and adjust amplifier gains.

The Receiver

- The receiver input is an analog signal and it's output is a sequence of discrete information symbols.
  - Consequently, the receiver must perform analog-to-digital conversion (sampling).
- Correspondingly, the receiver can be divided into an analog front-end followed by digital processing.
  - Many receivers have (relatively) simple front-ends and sophisticated digital processing stages.
  - Digital processing is performed on standard digital hardware (from ASICs to general purpose processors).
  - Moore's law can be relied on to boost the performance of digital communications systems.

Measures of Performance

- The receiver is expected to perform its function optimally.
- **Question:** optimal in what sense?
  - Measure of performance must be statistical in nature.
    - observed signal is random, and
    - transmitted symbol sequence is random.
  - Metric must reflect the reliability with which information is reconstructed at the receiver.
- **Objective:** Design the receiver that minimizes the probability of a symbol error.
  - Also referred to as symbol error rate.
  - Closely related to bit error rate (BER).

Learning Objectives

1. Understand the mathematical foundations that lead to the design of optimal receivers in AWGN channels.
   - statistical hypothesis testing
   - signal spaces
2. Understand the principles of digital information transmission.
   - baseband and passband transmission
   - relationship between data rate and bandwidth
3. Apply receiver design principles to communication systems with additional channel impairments
   - random amplitude or phase
   - linear distortion (e.g., multi-path)

Course Outline

- Mathematical Prerequisites
  - Basics of Gaussian Random Variables and Random Processes
  - Signal space concepts
- Principles of Receiver Design
  - Optimal decision: statistical hypothesis testing
  - Receiver frontend: the matched filter
- Signal design and modulation
  - Baseband and passband
  - Linear modulation
  - Bandwidth considerations
- Advanced topics
  - Synchronization in time, frequency, phase
  - Introduction to equalization

Part II

Mathematical Prerequisites

Gaussian Random Variables — Why we Care

- Gaussian random variables play a critical role in modeling many random phenomena.
  - By central limit theorem, Gaussian random variables arise from the superposition (sum) of many random phenomena.
    - Pertinent example: random movement of very many electrons in conducting material.
    - Result: thermal noise is well modeled as Gaussian.
  - Gaussian random variables are mathematically tractable.
    - In particular: any linear (more precisely, affine) transformation of Gaussians produces a Gaussian random variable.
- Noise added by channel is modeled as being Gaussian.
  - Channel noise is the most fundamental impairment in a communication system.

Gaussian Random Variables

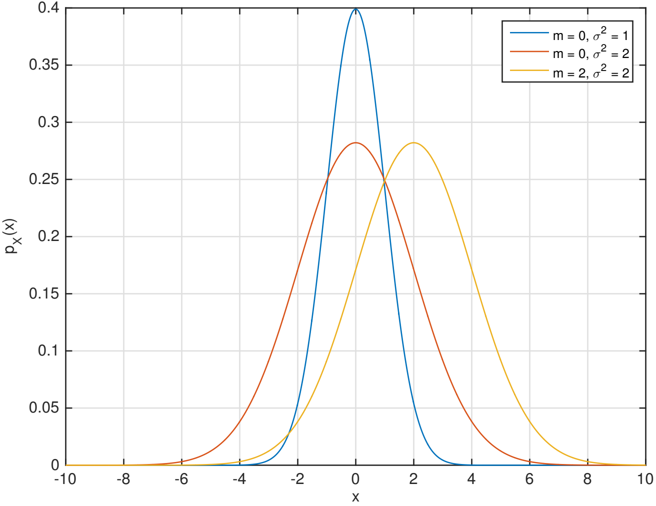
- A random variable X is said to be Gaussian (or Normal) if its pdf is of the form

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

- All properties of a Gaussian are determined by the two parameters m and σ^2.
- **Notation:** $X \sim \mathcal{N}(m, \sigma^2)$.
- **Moments:**

$$\begin{aligned} \mathbf{E}[X] &= \int_{-\infty}^{\infty} x \cdot p_X(x)\,dx = m \\ \mathbf{E}[X^2] &= \int_{-\infty}^{\infty} x^2 \cdot p_X(x)\,dx = m^2 + \sigma^2. \end{aligned}$$

Plot of Gaussian pdf's



The Gaussian Error Integral — Q(x)

- We are often interested in $\Pr\{X > x\}$ for Gaussian random variables X.
- These probabilities cannot be computed in closed form since the integral over the Gaussian pdf does not have a closed form expression.
- Instead, these probabilities are expressed in terms of the Gaussian error integral

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

The Gaussian Error Integral — Q(x)

- **Example:** Suppose $X \sim \mathcal{N}(1, 4)$, what is $\Pr\{X > 5\}$?

$$\begin{aligned} \Pr\{X > 5\} &= \int_5^\infty \frac{1}{\sqrt{2\pi \cdot 2^2}} e^{-\frac{(x-1)^2}{2 \cdot 2^2}}\, dx \qquad \text{substitute } z = \frac{x-1}{2} \\ &= \int_2^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz = Q(2) \end{aligned}$$

Exercises

- Let $X \sim \mathcal{N}(-3, 4)$, find expressions in terms of $Q(\cdot)$ for the following probabilities:
  1. $\Pr\{X > 5\}$?
  2. $\Pr\{X < -1\}$?
  3. $\Pr\{X^2 + X > 2\}$?

Bounds for the Q-function

- Since no closed form expression is available for $Q(x)$, bounds and approximations to the Q-function are of interest.
- The following bounds are tight for large values of x:

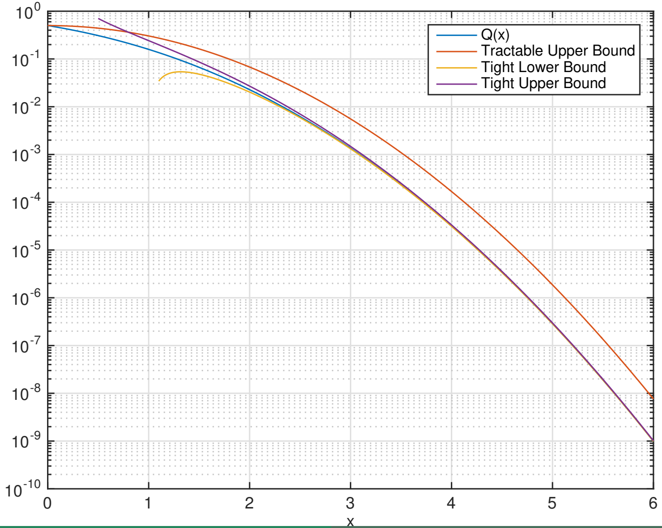
$$\left(1 - \frac{1}{x^2}\right) \frac{e^{-\frac{x^2}{2}}}{x\sqrt{2\pi}} \leq Q(x) \leq \frac{e^{-\frac{x^2}{2}}}{x\sqrt{2\pi}}.$$

- The following bound is not as quite as tight but very useful for analysis

$$Q(x) \leq \frac{1}{2} e^{-\frac{x^2}{2}}.$$

- Note that all three bounds are dominated by the term $e^{-\frac{x^2}{2}}$; this term determines the asymptotic behaviour of $Q(x)$.

Plot of $Q(x)$ and Bounds



Exercise: Chernoff Bound

- For a random variable X, the Chernoff Bound provides a tight upper bound on the probability $\Pr\{X > x\}$.
- The Chernoff bound is given by

$$\Pr\{X > x\} \leq \min_{t>0} \frac{\mathbf{E}[e^{tX}]}{e^{tx}}.$$

- Let $X \sim \mathcal{N}(0, 1)$; use the Chernoff bound to show that

$$\Pr\{X > x\} = Q(x) \leq e^{-x^2/2}$$

Gaussian Random Vectors

- A length N random vector $\vec{X}$ is said to be Gaussian if its pdf is given by

$$p_{\vec{X}}(\vec{x}) = \frac{1}{(2\pi)^{N/2}|K|^{1/2}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{m})^T K^{-1}(\vec{x}-\vec{m})\right).$$

- **Notation:** $\vec{X} \sim \mathcal{N}(\vec{m}, K)$.
- Mean vector

$$\vec{m} = \mathbf{E}[\vec{X}] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \vec{x} p_{\vec{X}}(\vec{x})\, d\vec{x}.$$

- Covariance matrix

$$K = \mathbf{E}[(\vec{X}-\vec{m})(\vec{X}-\vec{m})^T] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (\vec{x}-\vec{m})(\vec{x}-\vec{m})^T p_{\vec{X}}(\vec{x})\, d$$

  - $|K|$ denotes the determinant of K.
  - K must be positive definite, i.e., $\vec{z}^T K \vec{z} > 0$ for all $\vec{z}$.

Exercise: Important Special Case: N=2

- Consider a length-2 Gaussian random vector with

$$\vec{m} = \vec{0} \text{ and } K = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \text{ with } |\rho| \leq 1.$$

- Find the pdf of $\vec{X}$.
- Answer:

$$p_{\vec{X}}(\vec{x}) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left(\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2\sigma^2(1-\rho^2)}\right)$$

Important Properties of Gaussian Random Vectors

1. If the N Gaussian random variables X_n comprising the random vector $\vec{X}$ are uncorrelated ($\text{Cov}[X_i, X_j] = 0$, for $i \neq j$), then they are statistically independent.
2. Any affine transformation of a Gaussian random vector is also a Gaussian random vector.
   - Let $\vec{X} \sim \mathcal{N}(\vec{m}, K)$
   - Affine transformation: $\vec{Y} = A\vec{X} + \vec{b}$
   - Then, $\vec{Y} \sim \mathcal{N}(A\vec{m} + \vec{b}, AKA^T)$

Exercise: Generating Correlated Gaussian Random Variables

- Let $\vec{X} \sim \mathcal{N}(\vec{m}, K)$, with

$$\vec{m} = \vec{0} \text{ and } K = \sigma^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

  - The elements of $\vec{X}$ are uncorrelated.
- Transform $\vec{Y} = A\vec{X}$, with

$$A = \begin{pmatrix} \sqrt{1-\rho^2} & \rho \\ 0 & 1 \end{pmatrix}$$

- Find the pdf of $\vec{Y}$.

Random Processes — Why we Care

- Random processes describe signals that change randomly over time.
  - Compare: deterministic signals can be described by a mathematical expression that describes the signal exactly for all time.
  - Example: $x(t) = 3\cos(2\pi f_c t + \pi/4)$ with $f_c = 1$GHz.
- We will encounter three types of random processes in communication systems:
  1. (nearly) deterministic signal with a random parameter — Example: sinusoid with random phase.
  2. signals constructed from a sequence of random variables — Example: digitally modulated signals with random symbols
  3. noise-like signals
- **Objective:** Develop a framework to describe and analyze random signals encountered in the receiver of a

Random Process — Formal Definition

- Random processes can be defined completely analogous to random variables over a probability triple space $(\Omega, \mathcal{F}, P)$.
- **Definition:** A random process is a mapping from each element ω of the sample space Ω to a function of time (i.e., a signal).
- Notation: $X_t(\omega)$ — we will frequently omit ω to simplify notation.
- Observations:
  - We will be interested in both real and complex valued random processes.
  - Note, for a given random outcome ω_0, $X_t(\omega_0)$ is a *deterministic* signal.
  - Note, for a fixed time t_0, $X_{t_0}(\omega)$ is a *random variable*.

Sample Functions and Ensemble

- For a given random outcome ω_0, $X_t(\omega_0)$ is a deterministic signal.
  - Each signal that that can be produced by a our random process is called a sample function of the random process.
- The collection of all sample functions of a random process is called the ensemble of the process.
- **Example:** Let $\Theta(\omega)$ be a random variable with four equally likely, possible values $\Omega = \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$. Define the random process $X_t(\omega) = \cos(2\pi f_0 t + \Theta(\omega))$.
  The ensemble of this random process consists of the four sample functions:

$$X_t(\omega_1) = \cos(2\pi f_0 t) \qquad X_t(\omega_2) = -\sin(2\pi f_0 t)$$
$$X_t(\omega_3) = -\cos(2\pi f_0 t) \qquad X_t(\omega_4) = \sin(2\pi f_0 t)$$

Probability Distribution of a Random Process

- For a given time instant t, $X_t(\omega)$ is a random variable.
- Since it is a random variable, it has a pdf (or pmf in the discrete case).
  - We denote this pdf as $p_{X_t}(x)$.
- The statistical properties of a random process are specified completely if the joint pdf

$$p_{X_{t_1},\ldots,X_{t_n}}(x_1,\ldots,x_n)$$

  is available for all n and t_i, $i = 1,\ldots,n$.
  - This much information is often not available.
  - Joint pdfs with many sampling instances can be cumbersome.
  - We will shortly see a more concise summary of the statistics for a random process.

Random Process with Random Parameters

- A deterministic signal that depends on a random parameter is a random process.
  - Note, the sample functions of such random processes do not “look” random.
- Running Examples:
  - **Example (discrete phase):** Let $\Theta(\omega)$ be a random variable with four equally likely, possible values $\Omega = \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$. Define the random process $X_t(\omega) = \cos(2\pi f_0 t + \Theta(\omega))$.
  - **Example (continuous phase):** same as above but phase $\Theta(\omega)$ is uniformly distributed between 0 and 2π, $\Theta(\omega) \sim U[0, 2\pi)$.
- For both of these processes, the complete statistical description of the random process can be found.

Example: Discrete Phase Process

- **Discrete Phase Process:** Let $\Theta(\omega)$ be a random variable with four equally likely, possible values $\Omega = \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$. Define the random process $X_t(\omega) = \cos(2\pi f_0 t + \Theta(\omega))$.
- Find the first-order density $p_{X_t}(x)$ for this process.
- Find the second-order density $p_{X_{t_1} X_{t_2}}(x_1, x_2)$ for this process.
  - Note, since the phase values are discrete the above pdfs must be expressed with the help of δ-functions.
  - Alternatively, one can derive a probability mass function.

Solution: Discrete Phase Process

- First-order density function:

$$p_{X_t}(x) = \frac{1}{4}(\delta(x - \cos(2\pi f_0 t)) + \delta(x + \sin(2\pi f_0 t)) + \delta(x + \cos(2\pi f_0 t)) + \delta(x - \sin(2\pi f_0 t)))$$

- Second-order density function:

$$\begin{aligned} p_{X_{t_1} X_{t_2}}(x_1, x_2) = \frac{1}{4}(&\delta(x_1 - \cos(2\pi f_0 t_1)) \cdot \delta(x_2 - \cos(2\pi f_0 t_2)) + \\ &\delta(x_1 + \sin(2\pi f_0 t_1)) \cdot \delta(x_2 + \sin(2\pi f_0 t_2)) + \\ &\delta(x_1 + \cos(2\pi f_0 t_1)) \cdot \delta(x_2 + \cos(2\pi f_0 t_2)) + \\ &\delta(x_1 - \sin(2\pi f_0 t_1)) \cdot \delta(x_2 - \sin(2\pi f_0 t_2))) \end{aligned}$$

Example: Continuous Phase Process

- **Continuous Phase Process:** Let $\Theta(\omega)$ be a random variable that is uniformly distributed between 0 and 2π, $\Theta(\omega) \sim [0, 2\pi)$. Define the random process $X_t(\omega) = \cos(2\pi f_0 t + \Theta(\omega))$.
- Find the first-order density $p_{X_t}(x)$ for this process.
- Find the second-order density $p_{X_{t_1} X_{t_2}}(x_1, x_2)$ for this process.

Solution: Continuous Phase Process

- First-order density:

$$p_{X_t}(x) = \frac{1}{\pi\sqrt{1-x^2}} \qquad \text{for } |x| \leq 1.$$

Notice that $p_{X_t}(x)$ does **not** depend on t.

- Second-order density:

$$p_{X_{t_1} X_{t_2}}(x_1, x_2) = \frac{1}{\pi\sqrt{1-x_2^2}} \cdot [\frac{1}{2} \cdot \\ \delta(x_1 - \cos(2\pi f_0(t_1 - t_2) + \arccos(x_2))) + \\ \delta(x_1 - \cos(2\pi f_0(t_1 - t_2) - \arccos(x_2)))]$$

Random Processes Constructed from Sequence of Random Experiments

- Model for digitally modulated signals.
- Example:
  - Let $X_k(\omega)$ denote the outcome of the k-th toss of a coin:

$$X_k(\omega) = \begin{cases} 1 & \text{if heads on } k\text{-th toss} \\ -1 & \text{if tails on } k\text{-th toss.} \end{cases}$$

  - Let $p(t)$ denote a pulse of duration T, e.g.,

$$p(t) = \begin{cases} 1 & \text{for } 0 \leq t \leq T \\ 0 & \text{else.} \end{cases}$$

  - Define the random process X_t

$$X_t(\omega) = \sum_k X_k(\omega) p(t - nT)$$

Probability Distribution

- Assume that heads and tails are equally likely.
- Then the first-order density for the above random process is

$$p_{X_t}(x) = \frac{1}{2}(\delta(x-1) + \delta(x+1)).$$

- The second-order density is:

$$p_{X_{t_1} X_{t_2}}(x_1, x_2) = \begin{cases} \delta(x_1 - x_2) p_{X_{t_1}}(x_1) & \text{if } nT \leq t_1, t_2 \leq (n+1)T \\ p_{X_{t_1}}(x_1) p_{X_{t_2}}(x_2) & \text{else.} \end{cases}$$

  - These expression become more complicated when $p(t)$ is not a rectangular pulse.

Probability Density of Random Processs Defined Directly

- Sometimes the n-th order probability distribution of the random process is given.
  - Most important example: Gaussian Random Process
    - Statistical model for noise.
  - **Definition:** The random process X_t is Gaussian if the vector $\vec{X}$ of samples taken at times $t_1, \ldots, t_n$

$$\vec{X} = \begin{pmatrix} X_{t_1} \\ \vdots \\ X_{t_n} \end{pmatrix}$$

is a Gaussian random vector for all $t_1, \ldots, t_n$.

Second Order Description of Random Processes

- Characterization of random processes in terms of n-th order densities is
  - frequently not available
  - mathematically cumbersome
- A more tractable, practical alternative description is provided by the second order description for a random process.
- **Definition:** The second order description of a random process consists of the
  - mean function and the
  - autocorrelation function

  of the process.
- Note, the second order description can be computed from the (second-order) joint density.
  - The converse is not true — at a minimum the distribution must be specified (e.g., Gaussian).

Mean Function

- The second order description of a process relies on the mean and autocorrelation functions — these are defined as follows
- **Definition:** The mean of a random process is defined as:

$$\mathbf{E}[X_t] = m_X(t) = \int_{-\infty}^{\infty} x \cdot p_{X_t}(x)\, dx$$

  - Note, that the mean of a random process is a deterministic signal.
  - The mean is computed from the first oder density function.

Autocorrelation Function

- **Definition:** The autocorrelation function of a random process is defined as:

$$R_X(t,u) = \mathbf{E}[X_t X_u] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \cdot p_{X_t,X_u}(x,y)\, dx\, dy$$

  - Autocorrelation is computed from second order density

Autocovariance Function

- Closely related: autocovariance function:

$$
\begin{aligned}
C_X(t,u) &= \mathbf{E}[(X_t - m_X(t))(X_u - m_X(u))] \\
&= R_X(t,u) - m_X(t)m_X(u)
\end{aligned}
$$

Exercise: Discrete Phase Example

- Find the second-order description for the discrete phase random process.
  - **Discrete Phase Process:** Let $\Theta(\omega)$ be a random variable with four equally likely, possible values $\Omega = \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$. Define the random process $X_t(\omega) = \cos(2\pi f_0 t + \Theta(\omega))$.
- **Answer:**
  - Mean: $m_X(t) = 0$.
  - Autocorrelation function:

$$R_X(t, u) = \frac{1}{2} \cos(2\pi f_0 (t - u)).$$

Exercise: Continuous Phase Example

- Find the second-order description for the continuous phase random process.
  - **Continuous Phase Process:** Let $\Theta(\omega)$ be a random variable that is uniformly distributed between 0 and 2π, $\Theta(\omega) \sim [0, 2\pi)$. Define the random process $X_t(\omega) = \cos(2\pi f_0 t + \Theta(\omega))$.
- **Answer:**
  - Mean: $m_X(t) = 0$.
  - Autocorrelation function:

$$R_X(t, u) = \frac{1}{2} \cos(2\pi f_0 (t - u)).$$

Properties of the Autocorrelation Function

- The autocorrelation function of a (real-valued) random process satisfies the following properties:
  1. $R_X(t,t) \geq 0$
  2. $R_X(t,u) = R_X(u,t)$ (symmetry)
  3. $|R_X(t,u)| \leq \frac{1}{2}(R_X(t,t) + R_X(u,u))$
  4. $|R_X(t,u)|^2 \leq R_X(t,t) \cdot R_X(u,u)$

Stationarity

- The concept of stationarity is analogous to the idea of time-invariance in linear systems.
- **Interpretation:** For a stationary random process, the statistical properties of the process do not change with time.
- **Definition:** A random process X_t is strict-sense stationary (sss) to the n-th order if:

$$p_{X_{t_1},\ldots,X_{t_n}}(x_1,\ldots,x_n) = p_{X_{t_1+T},\ldots,X_{t_n+T}}(x_1,\ldots,x_n)$$

  for all T.
  - The statistics of X_t do not depend on *absolute* time but only on the time differences between the sample times.

Wide-Sense Stationarity

- A simpler and more tractable notion of stationarity is based on the second-order description of a process.
- **Definition:** A random process X_t is wide-sense stationary (wss) if
  1. the mean function $m_X(t)$ is constant **and**
  2. the autocorrelation function $R_X(t, u)$ depends on t and u only through $t - u$, i.e., $R_X(t, u) = R_X(t - u)$
- **Notation:** for a wss random process, we write the autocorrelation function in terms of the single time-parameter $\tau = t - u$:

$$R_X(t, u) = R_X(t - u) = R_X(\tau).$$

Exercise: Stationarity

- **True or False:** Every random process that is strict-sense stationarity to the second order is also wide-sense stationary.
  - **Answer:** True
- **True or False:** Every random process that is wide-sense stationary must be strict-sense stationarity to the second order.
  - **Answer:** False
- **True or False:** The discrete phase process is strict-sense stationary.
  - **Answer:** False; first order density depends on t, therefore, not even first-order sss.
- **True or False:** The discrete phase process is wide-sense stationary.
  - **Answer:** True

White Gaussian Noise

- **Definition:** A (real-valued) random process X_t is called white Gaussian Noise if
  - X_t is Gaussian for each time instance t
  - Mean: $m_X(t) = 0$ for all t
  - Autocorrelation function: $R_X(\tau) = \frac{N_0}{2}\delta(\tau)$
  - White Gaussian noise is a good model for noise in communication systems.
  - Note, that the variance of X_t is infinite:

$$\mathrm{Var}(X_t) = \mathbf{E}[X_t^2] = R_X(0) = \frac{N_0}{2}\delta(0) = \infty.$$

  - Also, for $t \neq u$: $\mathbf{E}[X_t X_u] = R_X(t, u) = R_X(t - u) = 0.$

Integrals of Random Processes

- We will see, that receivers always include a linear, time-invariant system, i.e., a filter.
- Linear, time-invariant systems *convolve* the input random process with the impulse response of the filter.
  - Convolution is fundamentally an integration.
- We will establish conditions that ensure that an expression like

$$Z(\omega) = \int_a^b X_t(\omega) h(t)\, dt$$

  is "well-behaved".
  - The result of the (definite) integral is a random variable.
- **Concern:** Does the above integral *converge*?

Mean Square Convergence

- There are different senses in which a sequence of random variables may converge: *almost surely*, *in probability*, *mean square*, and *in distribution*.
- We will focus exclusively on mean square convergence.
- For our integral, mean square convergence means that the Rieman sum and the random variable Z satisfy:
  - Given $\epsilon > 0$, there exists a $\delta > 0$ so that

$$\mathbf{E}[(\sum_{k=1}^{n} X_{\tau_k} h(\tau_k)(t_k - t_{k-1}) - Z)^2] \leq \epsilon.$$

with:

  - $a = t_0 < t_1 < \cdots < t_n = b$
  - $t_{k-1} \leq \tau_k \leq t_k$
  - $\delta = \max_k (t_k - t_{k-1})$

Mean Square Convergence — Why We Care

- It can be shown that the integral converges if

$$\int_a^b \int_a^b R_X(t,u) h(t) h(u) \, dt \, du < \infty$$

  - We will see shortly that this implies $\mathbf{E}[|Z|^2] < \infty$.
- **Important:** When the integral converges, then the order of integration and expectation can be interchanged, e.g.,

$$\mathbf{E}[Z] = \mathbf{E}[\int_a^b X_t h(t) \, dt] = \int_a^b \mathbf{E}[X_t] h(t) \, dt = \int_a^b m_X(t) h(t) \, dt$$

- Throughout this class, we will focus exclusively on cases where $R_X(t,u)$ and $h(t)$ are such that our integrals converge.

Exercise: Brownian Motion

- **Definition:** Let N_t be white Gaussian noise with $\frac{N_0}{2} = \sigma^2$. The random process

$$W_t = \int_0^t N_s \, ds \quad \text{for } t \geq 0$$

is called Brownian Motion or Wiener Process.

- Compute the mean and autocorrelation functions of W_t.
- **Answer:** $m_W(t) = 0$ and $R_W(t,u) = \sigma^2 \min(t,u)$

Integrals of Gaussian Random Processes

- Let X_t denote a Gaussian random process with second order description $m_X(t)$ and $R_X(t,s)$.
- Then, the integral

$$Z = \int_a^b X(t)h(t)\,dt$$

is a Gaussian random variable.
- Moreover mean and variance are given by

$$\mu = \mathbf{E}[Z] = \int_a^b m_X(t)h(t)\,dt$$

$$\begin{aligned}\mathrm{Var}[Z] = \mathbf{E}[(Z-\mathbf{E}[Z])^2] &= \mathbf{E}[(\int_a^b (X_t - m_x(t))h(t)\,dt)^2] \\ &= \int_a^b\int_a^b C_X(t,u)h(t)h(u)\,dt\,du\end{aligned}$$

Jointly Defined Random Processes

- Let X_t and Y_t be jointly defined random processes.
  - E.g., input and output of a filter.
- Then, joint densities of the form $p_{X_t Y_u}(x, y)$ can be defined.
- Additionally, second order descriptions that describe the correlation between samples of X_t and Y_t can be defined.

Crosscorrelation and Crosscovariance

- **Definition:** The crosscorrelation function $R_{XY}(t,u)$ is defined as:

$$R_{XY}(t,u) = \mathbf{E}[X_t Y_u] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy p_{X_t Y_u}(x,y)\,dx\,dy.$$

- **Definition:** The crosscovariance function $C_{XY}(t,u)$ is defined as:

$$C_{XY}(t,u) = R_{XY}(t,u) - m_X(t)m_Y(u).$$

- **Definition:** The processes X_t and Y_t are called jointly wide-sense stationary if:
  1. $R_{XY}(t,u) = R_{XY}(t-u)$ **and**
  2. $m_X(t)$ and $m_Y(t)$ are constants.

Filtering of Random Processes

Filtered Random Process

$$X_t \longrightarrow \boxed{h(t)} \longrightarrow Y_t$$

© 2018, B.-P. Paris

Filtering of Random Processes

- Clearly, X_t and Y_t are jointly defined random processes.
- Standard LTI system — convolution:

$$Y_t = \int h(t-\sigma) X_\sigma \, d\sigma = h(t) * X_t$$

- Recall: this convolution is "well-behaved" if

$$\int\int R_X(\sigma, \nu) h(t-\sigma) h(t-\nu) \, d\sigma \, d\nu < \infty$$

  - E.g.: $\int\int R_X(\sigma, \nu) \, d\sigma \, d\nu < \infty$ **and** $h(t)$ stable.

Second Order Description of Output: Mean

- The expected value of the filter's output Y_t is:

$$\begin{aligned}\mathbf{E}[Y_t] &= \mathbf{E}[\int h(t-\sigma)X_\sigma \, d\sigma] \\ &= \int h(t-\sigma)\mathbf{E}[X_\sigma] \, d\sigma \\ &= \int h(t-\sigma) m_X(\sigma) \, d\sigma\end{aligned}$$

- For a wss process X_t, $m_X(t)$ is constant. Therefore,

$$\mathbf{E}[Y_t] = m_Y(t) = m_X \int h(\sigma) \, d\sigma$$

is also constant.

Crosscorrelation of Input and Output

- The crosscorrelation between input and ouput signals is:

$$
\begin{aligned}
R_{XY}(t,u) = \mathbf{E}[X_t Y_u] &= \mathbf{E}[X_t \int h(u-\sigma) X_\sigma \, d\sigma \\
&= \int h(u-\sigma) \mathbf{E}[X_t X_\sigma] \, d\sigma \\
&= \int h(u-\sigma) R_X(t,\sigma) \, d\sigma
\end{aligned}
$$

- For a wss input process

$$
\begin{aligned}
R_{XY}(t,u) &= \int h(u-\sigma) R_X(t,\sigma) \, d\sigma = \int h(\nu) R_X(t, u-\nu) \, d\nu \\
&= \int h(\nu) R_X(t-u+\nu) \, d\nu = R_{XY}(t-u)
\end{aligned}
$$

  - Input and output are jointly stationary.

Autocorelation of Output

- The autocorrelation of Y_t is given by

$$R_Y(t,u) = \mathbf{E}[Y_t Y_u] = \mathbf{E}[\int h(t-\sigma) X_\sigma \, d\sigma \int h(u-\nu) X_\nu \, d\nu]$$
$$= \int\int h(t-\sigma) h(u-\nu) R_X(\sigma,\nu) \, d\sigma \, d\nu$$

- For a wss input process:

$$R_Y(t,u) = \int\int h(t-\sigma) h(u-\nu) R_X(\sigma,\nu) \, d\sigma \, d\nu$$
$$= \int\int h(\lambda) h(\lambda-\gamma) R_X(t-\lambda, u-\lambda+\gamma) \, d\lambda \, d\gamma$$
$$= \int\int h(\lambda) h(\lambda-\gamma) R_X(t-u-\gamma) \, d\lambda \, d\gamma = R_Y(t-u)$$

  - Define $R_h(\gamma) = \int h(\lambda) h(\lambda-\gamma) \, d\lambda = h(\lambda) * h(-\lambda)$.
  - Then, $R_Y(\tau) = \int R_h(\gamma) R_X(\tau-\gamma) \, d\gamma = R_h(\tau) * R_X(\tau)$

Exercise: Filtered White Noise Process

- Let the white Gaussian noise process X_t be input to a filter with impulse response

$$h(t) = e^{-at}u(t) = \begin{cases} e^{-at} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases}$$

- Compute the second order description of the output process Y_t.
- **Answers:**
  - Mean: $m_Y = 0$
  - Autocorrelation:

$$R_Y(\tau) = \frac{N_0}{2} \frac{e^{-a|\tau|}}{2a}$$

Power Spectral Density — Concept

- **Power Spectral Density** (PSD) measures how the power of a random process is distributed over frequency.
  - Notation: $S_X(f)$
  - Units: Watts per Hertz (W/Hz)
- Thought experiment:
  - Pass random process X_t through a narrow bandpass filter:
    - center frequency f
    - bandwidth Δf
    - denote filter output as Y_t
  - Measure the power P at the output of bandpass filter:

$$P = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} |Y_t|^2 \, dt$$

  - Relationship between power and (PSD)

$$P \approx S_X(f) \cdot \Delta f.$$

Relation to Autocorrelation Function

- For a wss random process, the power spectral density is closely related to the autocorrelation function $R_X(\tau)$.
- **Definition:** For a random process X_t with autocorrelation function $R_X(\tau)$, the power spectral density $S_X(f)$ is defined as the Fourier transform of the autocorrelation function,

$$S_X(f) = \int_{-\infty}^{\infty} R_X(\tau) e^{j2\pi f\tau}\, d\tau.$$

  - For non-stationary processes, it is possible to define a spectral represenattion of the process.
  - However, the spectral contents of a non-stationary process will be time-varying.
- **Example:** If N_t is white noise, i.e., $R_N(\tau) = \frac{N_0}{2}\delta(\tau)$, then

$$S_X(f) = \frac{N_0}{2} \quad \text{for all } f$$

Properties of the PSD

- Inverse Transform:
$$R_X(\tau) = \int_{-\infty}^{\infty} S_X(f) e^{-j2\pi f \tau} \, df.$$
- The total power of the process is
$$\mathbf{E}[|X_t|^2] = R_X(0) = \int_{-\infty}^{\infty} S_X(f) \, df.$$
- $S_X(f)$ is even and non-negative.
  - Evenness of $S_X(f)$ follows from evenness of $R_X(\tau)$.
  - Non-negativeness is a consequence of the autocorrelation function being positive definite
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(t) f^*(u) R_X(t,u) \, dt \, du \geq 0$$
for all choices of $f(\cdot)$, including $f(t) = e^{-j2\pi f t}$.

Filtering of Random Processes

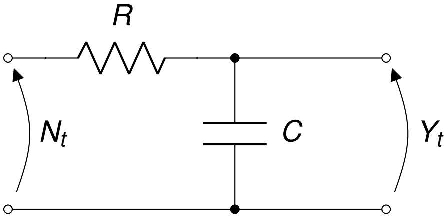
- Random process X_t with autocorrelation $R_X(\tau)$ and PSD $S_X(f)$ is input to LTI filter with impuse response $h(t)$ and frequency response $H(f)$.
- The PSD of the output process Y_t is

$$S_Y(f) = |H(f)|^2 S_X(f).$$

  - Recall that $R_Y(\tau) = R_X(\tau) * C_h(\tau)$,
  - where $C_h(\tau) = h(\tau) * h(-\tau)$.
  - In frequency domain: $S_Y(f) = S_X(f) \cdot \mathcal{F}\{C_h(\tau)\}$
  - With

$$\begin{aligned} \mathcal{F}\{C_h(\tau)\} &= \mathcal{F}\{h(\tau) * h(-\tau)\} \\ &= \mathcal{F}\{h(\tau)\} \cdot \mathcal{F}\{h(-\tau)\} \\ &= H(f) \cdot H^*(f) = |H(f)|^2. \end{aligned}$$

Exercise: Filtered White Noise



- Let N_t be a white noise process that is input to the above circuit. Find the power spectral density of the output process.
- **Answer:**

$$S_Y(f) = \left| \frac{1}{1 + j2\pi fRC} \right|^2 \frac{N_0}{2} = \frac{1}{1 + (2\pi fRC)^2} \frac{N_0}{2}.$$

Signal Space Concepts — Why we Care

- **Signal Space Concepts** are a powerful tool for the analysis of communication systems and for the design of optimum receivers.
- **Key Concepts:**
  - Orthonormal basis functions — tailored to signals of interest — span the signal space.
  - *Representation theorem:* allows any signal to be represented as a (usually finite dimensional) vector
    - Signals are interpreted as points in signal space.
  - For random processes, representation theorem leads to random signals being described by random vectors with uncorrelated components.
    - *Theorem of Irrelavance* allows us to disregrad nearly all components of noise in the receiver.
- We will briefly review key ideas that provide underpinning for signal spaces.

Linear Vector Spaces

- The basic structure needed by our signal spaces is the idea of linear vector space.
- **Definition:** A linear vector space $\mathcal{S}$ is a collection of elements ("vectors") with the following properties:
  - Addition of vectors is defined and satisfies the following conditions for any $x, y, z \in \mathcal{S}$:
    1. $x + y \in \mathcal{S}$ (closed under addition)
    2. $x + y = y + x$ (commutative)
    3. $(x + y) + z = x + (y + z)$ (associative)
    4. The zero vector $\vec{0}$ exists and $\vec{0} \in \mathcal{S}$. $x + \vec{0} = x$ for all $x \in \mathcal{S}$.
    5. For each $x \in \mathcal{S}$, a unique vector $(-x)$ is also in $\mathcal{S}$ and $x + (-x) = \vec{0}$.

Linear Vector Spaces — continued

- **Definition — continued:**
  - Associated with the set of vectors in $\mathcal{S}$ is a set of scalars. If a, b are scalars, then for any $x, y \in \mathcal{S}$ the following properties hold:
    1. $a \cdot x$ is defined and $a \cdot x \in \mathcal{S}$.
    2. $a \cdot (b \cdot x) = (a \cdot b) \cdot x$
    3. Let 1 and 0 denote the multiplicative and additive identies of the field of scalars, then $1 \cdot x = x$ and $0 \cdot x = \vec{0}$ for all $x \in \mathcal{S}$.
    4. Associative properties:

$$
\begin{aligned}
a \cdot (x + y) &= a \cdot x + a \cdot y \\
(a + b) \cdot x &= a \cdot x + b \cdot x
\end{aligned}
$$

Running Examples

- The space of length-N vectors $\mathbb{R}^N$

$$\begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_N + y_N \end{pmatrix} \text{ and } a \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} = \begin{pmatrix} a \cdot x_1 \\ \vdots \\ a \cdot x_N \end{pmatrix}$$

- The collection of all square-integrable signals over $[T_a, T_b]$, i.e., all signals $x(t)$ satisfying

$$\int_{T_a}^{T_b} |x(t)|^2 \, dt < \infty.$$

  - Verifying that this is a linear vector space is easy.
  - This space is called $L^2(T_a, T_b)$ (pronounced: ell-two).

Inner Product

- To be truly useful, we need linear vector spaces to provide
  - means to measure the length of vectors and
  - to measure the distance between vectors.
- Both of these can be achieved with the help of inner products.
- **Definition:** The inner product of two vectors $x, y, \in \mathcal{S}$ is denoted by $\langle x, y \rangle$. The inner product is a *scalar* assigned to x and y so that the following conditions are satisfied:
  1. $\langle x, y \rangle = \langle y, x \rangle$ (for complex vectors $\langle x, y \rangle = \langle y, x \rangle^*$)
  2. $\langle a \cdot x, y \rangle = a \cdot \langle x, y \rangle$, with scalar a
  3. $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$, with vector z
  4. $\langle x, x \rangle > 0$, except when $x = \vec{0}$; then, $\langle x, x \rangle = 0$.

Exercise: Valid Inner Products?

- $x, y \in \mathbb{R}^N$ with
$$\langle x, y \rangle = \sum_{n=1}^{N} x_n y_n$$
- **Answer:** Yes; this is the standard *dot product*.
- $x, y \in \mathbb{R}^N$ with
$$\langle x, y \rangle = \sum_{n=1}^{N} x_n \cdot \sum_{n=1}^{N} y_n$$
- **Answer:** No; last condition does not hold, which makes this inner product useless for measuring distances.

Exercise: Valid Inner Products?

- $x(t), y(t) \in L^2(a,b)$ with
$$\langle x(t), y(t) \rangle = \int_a^b x(t) y(t)\, dt$$
- **Answer:** Yes; continuous-time equivalent of the dot-product.
- $x, y \in \mathbb{C}^N$ with
$$\langle x, y \rangle = \sum_{n=1}^{N} x_n y_n^*$$
- **Answer:** Yes; the conjugate complex is critical to meet the last condition (e.g., $\langle j, j \rangle = -1 < 0$).
- $x, y \in \mathbb{R}^N$ with
$$\langle x, y \rangle = x^T K y = \sum_{n=1}^{N} \sum_{m=1}^{N} x_n K_{n,m} y_m$$

Exercise: Valid Inner Products?

- $x, y \in \mathbb{R}^N$ with

$$\langle x, y \rangle = x^T K y = \sum_{n=1}^{N} \sum_{m=1}^{N} x_n K_{n,m} y_m$$

  with K an $N \times N$-matrix
- **Answer:** Only if K is positive definite (i.e., $x^T K x > 0$ for all $x \neq \vec{0}$).

Norm of a Vector

- **Definition:** The norm of vector $x \in \mathcal{S}$ is denoted by $\|x\|$ and is defined via the inner product as

$$\|x\| = \sqrt{\langle x, x \rangle}.$$

  - Notice that $\|x\| > 0$ unless $x = \vec{0}$, then $\|x\| = 0$.
  - The norm of a vector measures the length of a vector.
  - For signals $\|x(t)\|^2$ measures the *energy* of the signal.
- **Example:** For $x \in \mathbb{R}^N$, Cartesian length of a vector

$$\|x\| = \sqrt{\sum_{n=1}^{N} |x_n|^2}$$

Norm of a Vector — continued

- **Illustration:**

$$\|a \cdot x\| = \sqrt{\langle a \cdot x, a \cdot x \rangle} = |a| \|x\|$$

  - Scaling the vector by a, scales its length by a.

Inner Product Space

- We call a linear vector space with an associated, valid inner product an inner product space.
  - **Definition:** An inner product space is a linear vector space in which a inner product is defined for all elements of the space and the norm is given by $\|x\| = \langle x, x \rangle$.
- **Standard Examples:**
  1. $\mathbb{R}^N$ with $\langle x, y \rangle = \sum_{n=1}^{N} x_n y_n$.
  2. $L^2(a, b)$ with $\langle x(t), y(t) \rangle = \int_a^b x(t) y(t)\, dt$.

Schwartz Inequality

- The following relationship between norms and inner products holds for all inner product spaces.
- **Schwartz Inequality:** For any $x, y \in \mathcal{S}$, where $\mathcal{S}$ is an inner product space,

$$|\langle x, y\rangle| \leq \|x\| \cdot \|y\|$$

  with equality if and only if $x = c \cdot y$ with scalar c
  - Proof follows from $\|x + a \cdot y\|^2 \geq 0$ with $a = -\frac{\langle x,y\rangle}{\|y\|^2}$.

Orthogonality

- **Definition:** Two vectors are orthogonal if the inner product of the vectors is zero, i.e.,

$$\langle x, y \rangle = 0.$$

- **Example:** The standard basis vectors e_m in $\mathbb{R}^N$ are orthogonal; recall

$$e_m = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \quad \text{the 1 occurs on the } m\text{-th row}$$

Orthogonality

- **Example:** The basis functions for the Fourier Series expansion $w_m(t) \in L^2(0, T)$ are orthogonal; recall

$$w_m(t) = \frac{1}{\sqrt{T}} e^{j2\pi mt/T}.$$

Distance between Vectors

- **Definition:** The distance d between two vectors is defined as the norm of their difference, i.e.,
$$d(x,y) = \|x - y\|$$
- **Example:** The Cartesian (or Euclidean) distance between vectors in $\mathbb{R}^N$:
$$d(x,y) = \|x - y\| = \sqrt{\sum_{n=1}^{N} |x_n - y_n|^2}.$$
- **Example:** The root-mean-squared error (RMSE) between two signals in $L^2(a,b)$ is
$$d(x(t),y(t)) = \|x(t) - y(t)\| = \sqrt{\int_a^b |x(t) - y(t)|^2 \, dt}$$

Properties of Distances

- Distance measures defined by the norm of the difference between vectors x, y have the following properties:
  1. $d(x, y) = d(y, x)$
  2. $d(x, y) = 0$ if and only if $x = y$
  3. $d(x, y) \leq d(x, z) + d(y, z)$ for all vectors z (Triangle inequality)

Exercise: Prove the Triangle Inequality

- Begin like this:

$$
\begin{aligned}
d^2(x,y) &= \|x-y\|^2 \\
&= \|(x-z)+(z-y)\|^2 \\
&= \langle (x-z)+(z-y), (x-z)+(z-y) \rangle
\end{aligned}
$$

-

$$
\begin{aligned}
d^2(x,y) &= \langle x-z, x-z \rangle + 2\langle x-z, z-y \rangle + \langle z-y, z-y \rangle \\
&\leq \langle x-z, x-z \rangle + 2|\langle x-z, z-y \rangle| + \langle z-y, z-y \rangle \\
(Schwartz): &\leq \langle x-z, x-z \rangle + 2\|x-z\| \cdot \|z-y\| + \langle z-y, z- \\
&= d(x,z)^2 + 2d(x,z) \cdot d(y,z) + d(y,z)^2 \\
&= (d(x,z) + d(y,z))^2
\end{aligned}
$$

Hilbert Spaces — Why we Care

- We would like our vector spaces to have one more property.
  - We say the sequence of vectors $\{x_n\}$ converges to vector x, if
    $$\lim_{n\to\infty} \|x_n - x\| = 0.$$
  - We would like the limit point x of any sequence $\{x_n\}$ to be in our vector space.
  - Integrals and derivatives are fundamentally limits; we want derivatives and integrals to stay in the vector space.
  - A vector space is said to be closed if it contains all of its limit points.
- **Definition:** A closed, inner product space is A Hilbert Space.

Hilbert Spaces — Examples

- **Examples:** Both $\mathbb{R}^N$ and $L^2(a, b)$ are Hilbert Spaces.
- **Counter Example:** The space of rational number $\mathbb{Q}$ is **not** closed (i.e., not a Hilbert space)
  - E.g.,

$$\sum_{n=0}^{\infty} \frac{1}{n!} = e \notin \mathbb{Q},$$

even though all $\frac{1}{n!} \in \mathbb{Q}$.

Subspaces

- **Definition:** Let $\mathcal{S}$ be a linear vector space. The space $\mathcal{L}$ is a subspace of $\mathcal{S}$ if
  1. $\mathcal{L}$ is a *subset* of $\mathcal{S}$ and
  2. $\mathcal{L}$ is *closed*.

  - If $x, y \in \mathcal{L}$ then also $x, y, \in \mathcal{S}$.
  - And, $a \cdot x + b \cdot y \in \mathcal{L}$ for all scalars a, b.
- **Example:** Let $\mathcal{S}$ be $L^2(T_a, T_b)$. Define $\mathcal{L}$ as the set of all sinusoids of frequency f_0, i.e., signals of the form $x(t) = A\cos(2\pi f_0 t + \phi)$, with $0 \leq A < \infty$ and $0 \leq \phi < 2\pi$
  1. All such sinusoids are square integrable.
  2. Linear combination of two sinusoids of frequency f_0 is a sinusoid of the same frequency.

Projection Theorem

- **Definition:** Let $\mathcal{L}$ be a subspace of the Hilbert Space $\mathcal{H}$. The vector $x \in \mathcal{H}$ (and $x \notin \mathcal{L}$) is orthogonal to the subspace $\mathcal{L}$ if $\langle x, y \rangle = 0$ for every $y \in \mathcal{L}$.
- **Projection Theorem:** Let $\mathcal{H}$ be a Hilbert Space and $\mathcal{L}$ is a subspace of $\mathcal{H}$.
  Every vector $x \in \mathcal{H}$ has a unique decomposition
  $$x = y + z$$
  with $y \in \mathcal{L}$ and z orthogonal to $\mathcal{L}$.
  Furthermore,
  $$\|z\| = \|x - y\| = \min_{\nu \in \mathcal{L}} \|x - \nu\|.$$
  - y is called the projection of x onto $\mathcal{L}$.
  - Distance from x to all elements of $\mathcal{L}$ is minimized by y.

Exercise: Fourier Series

- Let $x(t)$ be a signal in the Hilbert space $L^2(0, T)$.
- Define the subspace $\mathcal{L}$ of signals $\nu_n(t) = A_n \cos(2\pi nt/T)$ for a fixed n and T.
- Find the signal $y(t) \in \mathcal{L}$ that minimizes

$$\min_{y(t)\in\mathcal{L}} \|x(t) - y(t)\|^2.$$

- **Answer:** $y(t)$ is the sinusoid with amplitude

$$A_n = \frac{2}{T}\int_0^T x(t)\cos(2\pi nt/T)\,dt = \frac{2}{T}\langle x(t), \cos(2\pi nt/T)\rangle.$$

  - Note that this is (part of the trigonometric form of) the Fourier Series expansion.
  - Note that the inner product involves the projection of $x(t)$ onto an element of $\mathcal{L}$.

Projection Theorem

- The Projection Theorem is most useful when the subspace $\mathcal{L}$ has certain structural properties.
- In particular, we will be interested in the case when $\mathcal{L}$ is spanned by a set of orthonormal vectors.
  - Let's define what that means.

Separable Vector Spaces

- **Definition:** A Hilbert space $\mathcal{H}$ is said to be separable if there exists a set of vectors $\{\Phi_n\}$, $n = 1, 2, \ldots$ that are elements of $\mathcal{H}$ and such that every element $x \in \mathcal{H}$ can be expressed as

$$x = \sum_{n=1}^{\infty} X_n \Phi_n.$$

  - The coefficients X_n are scalars associated with vectors Φ_n.
  - *Equality* is taken to mean

$$\lim_{n\to\infty} \left\| x - \sum_{n=1}^{\infty} X_n \Phi_n \right\|^2 = 0.$$

Representation of a Vector

- The set of vectors $\{\Phi_n\}$ is said to be complete if the above is valid for every $x \in \mathcal{H}$.
- A complete set of vectors $\{\Phi_n\}$ is said to form a basis for $\mathcal{H}$.
- **Definition:** The representation of the vector x (with respect to the basis $\{\Phi_n\}$) is the sequence of coefficients $\{X_n\}$.
- **Definition:** The number of vectors Φ_n that is required to express every element x of a separable vector space is called the dimension of the space.

Example: Length-N column Vectors

- The space $\mathbb{R}^N$ is separable and has dimension N.
  - Basis vectors ($m = 1, \ldots, N$):

$$\Phi_m = e_m = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \text{ the 1 occurs on the } m\text{-th row}$$

  - There are N basis vectors; dimension is N.

Example: Length-N column Vectors — continued

- (con't)
  - For any vector $x \in \mathbb{R}^N$:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} = \sum_{m=1}^{N} x_m e_m$$

Examples: L^2

- **Fourier Bases:** The following is a complete basis for $L^2(0, T)$

$$\begin{aligned} \Phi_{2n}(t) &= \sqrt{\frac{2}{T}} \cos(2\pi nt/T) \\ \Phi_{2n+1}(t) &= \sqrt{\frac{2}{T}} \sin(2\pi nt/T) \end{aligned} \quad n = 0, 1, 2, \ldots$$

  - This implies that $L^2(0, T)$ is a separable vector space.
  - $L^2(0, T)$ is infinite-dimensional.

Examples: L^2

- **Piecewise Linear Signals:** The set of vectors (signals)

$$\Phi_n(t) = \begin{cases} \frac{1}{\sqrt{T}} & (n-1)T \leq t < nT \\ 0 & \text{else} \end{cases}$$

is **not** a basis for $L^2(0, \infty)$.

  - Only piecewise constant signals can be represented.
  - But, this is a basis for the subspace of L^2 consisting of piecewise constant signals.

Orthonormal Bases

- **Definition:** A basis for a separable vector space is an orthonormal basis if the elements of the vectors that constitute the basis satisfy
  1. $\langle \Phi_n, \Phi_m \rangle = 0$ for all $n \neq m$. (*ortho*gonal)
  2. $\|\Phi_n\| = 1$, for all $n = 1, 2, \ldots$ (*normal*ized)
- Note:
  - Not every basis is orthonormal.
    - We will see shortly, every basis can be turned into an orthonormal basis.
  - Not every set of orthonornal vectors constitutes a basis.
    - Example: Piecewise Linear Signals.

Representation with Orthonormal Basis

- An orthonormal basis is much prefered over an arbitrary basis because the representation of vector x is very easy to compute.
- The representation $\{X_n\}$ of a vector x

  $$x = \sum_{n=1}^{\infty} X_n \Phi_n$$

  with respect to an orthonormal basis $\{\Phi_n\}$ is computed using

  $$X_n = \langle x, \Phi_n \rangle.$$

  The representation X_n is obtained by projecting x onto the basis vector Φ_n!
  - In contrast, when bases are not orthonormal, finding the representation of x requires solving a system of linear equations.

Parsevals Relationship

- **Parsevals Theorem:** If vectors x and y are represented with respect to an orthonormal basis $\{\Phi_n\}$ by $\{X_n\}$ and $\{Y_n\}$, respectively, then

$$\langle x, y \rangle = \sum_{n=1}^{\infty} X_n \cdot Y_n$$

Parsevals Relationship

- Parsevals theorem implies

$$\|x\|^2 = \sum_{n=1}^{\infty} X_n^2$$

and

$$\|x-y\|^2 = \sum_{n=1}^{\infty} |X_n - Y_n|^2$$

- Inner products, norms, and distances can be computed using vectors or their representations; the results are the same.

Back to the Projection Theorem

- We claimed earlier that the projection theorem is particularly useful when the subspace $\mathcal{L}$ is structured.
- Specifically, let $\mathcal{L}$ be a subspace of $\mathcal{S}$ spanned by a (usually finite) orthonormal basis $\{\Phi_n\}_{n=0}^{N-1}$.
  - Note that $\{\Phi_n\}_{n=0}^{N-1}$ is **not** a complete basis for $\mathcal{S}$.
  - There are $x \in \mathcal{S}$ that cannot be represented by this basis.
- Then, the projection $y \in \mathcal{L}$ of a vector $x \in \mathcal{S}$ is simply

$$y = \sum_{n=0}^{N-1} Y_n \Phi_n \text{ with } Y_n = \langle x, \Phi_n \rangle.$$

- **Examples:**
  - Band-limited Fourier series expansion
  - Polynomial regression with Legendre polynomials

Exercise: Orthonormal Basis

- Show that for orthonormal basis $\{\Phi_n\}$, the representation X_n of x is obtained by projection

$$\langle x, \Phi_n \rangle = X_n$$

  - **Hint:** You need to find

$$\hat{X}_n = \arg\min_{X_n} \|x - X_n\Phi_n - \sum_{m \neq n} X_m\Phi_m\|^2$$

- Show that Parsevals theorem is true.

The Gram-Schmidt Procedure

- An arbitrary basis $\{\Phi_n\}$ can be converted into an orthonormal basis $\{\Psi_n\}$ using an algorithm known as the Gram-Schmidt procedure:

Step 1: $\Psi_1 = \frac{\Phi_1}{\|\Phi_1\|}$ (normalize Φ_1)

Step 2 (a): $\tilde{\Psi}_2 = \Phi_2 - \langle \Phi_2, \Psi_1 \rangle \cdot \Psi_1$ (make $\tilde{\Psi}_2 \perp \Psi_1$)

Step 2 (b): $\Psi_2 = \frac{\tilde{\Psi}_2}{\|\tilde{\Psi}_2\|}$

$$\vdots$$

Step k (a): $\tilde{\Psi}_k = \Phi_k - \sum_{n=1}^{k-1} \langle \Phi_k, \Psi_n \rangle \cdot \Psi_n$

Step k (b): $\Psi_k = \frac{\tilde{\Psi}_k}{\|\tilde{\Psi}_k\|}$

  - Whenever $\tilde{\Psi}_n = 0$, the basis vector is omitted.

Gram-Schmidt Procedure

- Note:
  - By construction, $\{\Psi\}$ is an orthonormal set of vectors.
  - If the orginal basis $\{\Phi\}$ is complete, then $\{\Psi\}$ is also complete.
    - The Gram-Schmidt construction implies that every Φ_n can be represented in terms of Ψ_m, with $m = 1, \ldots, n$.
- Because
  - any basis can be normalized (using the Gram-Schmidt procedure) and
  - the benefits of orthonormal bases when computing the representation of a vector

  a basis is usually assumed to be orthonormal.

Exercise: Gram-Schmidt Procedure

- The following three basis functions are given

$$\Phi_1(t) = I_{[0,\frac{T}{2}]}(t) \quad \Phi_2(t) = I_{[0,T]}(t) \quad \Phi_3(t) = I_{[\frac{T}{2},T]}(t)$$

where $I_{[a,b]}(t) = 1$ for $a \leq t \leq b$ and zero otherwise.

1. Compute an *orthonormal* basis from the above basis functions.
2. Compute the representation of $\Phi_n(t)$, $n = 1, 2, 3$ with respect to this orthonormal basis.
3. Compute $\|\Phi_1(t)\|$ and $\|\Phi_2(t) - \Phi_3(t)\|$

Answers for Exercise

1. Orthonormal bases:

$$\Psi_1(t) = \sqrt{\frac{2}{T}} I_{[0,\frac{T}{2}]}(t) \quad \Psi_2(t) = \sqrt{\frac{2}{T}} I_{[\frac{T}{2},T]}(t)$$

2. Representations:

$$\phi_1 = \begin{pmatrix} \sqrt{\frac{T}{2}} \\ 0 \end{pmatrix} \quad \begin{pmatrix} \sqrt{\frac{T}{2}} \\ \sqrt{\frac{T}{2}} \end{pmatrix} \quad \begin{pmatrix} 0 \\ \sqrt{\frac{T}{2}} \end{pmatrix}$$

3. Distances: $\|\Phi_1(t)\| = \sqrt{\frac{T}{2}}$ and $\|\Phi_2(t) - \Phi_3(t)\| = \sqrt{\frac{T}{2}}$.

A Hilbert Space for Random Processes

- A vector space for random processes X_t that is analogous to $L^2(a, b)$ is of greatest interest to us.
  - This vector space contains random processes that satisfy, i.e.,
  $$\int_a^b \mathbf{E}[X_t^2]\, dt < \infty.$$
  - **Inner Product:** cross-correlation
  $$\mathbf{E}[\langle X_t, Y_t \rangle] = \mathbf{E}[\int_a^b X_t Y_t\, dt].$$
  - Fact: This vector space is separable; therefore, an orthonormal basis $\{\Phi\}$ exists.

A Hilbert Space for Random Processes

- (con't)
  - **Representation:**

$$X_t = \sum_{n=1}^{\infty} X_n \Phi_n(t) \quad \text{for } a \leq t \leq b$$

with

$$X_n = \langle X_t, \Phi_n(t) \rangle = \int_a^b X_t \Phi_n(t)\, dt.$$

    - Note that X_n is a *random variable*.
  - For this to be a valid Hilbert space, we must interprete equality of processes X_t and Y_t in the mean squared sense, i.e.,

$$X_t = Y_t \text{ means } \mathbf{E}[|X_t - Y_t|^2] = 0.$$

Karhunen-Loeve Expansion

- **Goal:** Choose an orthonormal basis $\{\Phi\}$ such that the representation $\{X_n\}$ of the random process X_t consists of *uncorrelated random variables*.
  - The resulting representation is called Karhunen-Loeve expansion.
- Thus, we want

$$\mathbf{E}[X_n X_m] = \mathbf{E}[X_n]\mathbf{E}[X_m] \quad \text{for } n \neq m.$$

Karhunen-Loeve Expansion

- It can be shown, that for the representation $\{X_n\}$ to consist of uncorrelated random variables, the orthonormal basis vectors $\{\Phi\}$ must satisfy

$$\int_a^b K_X(t,u) \cdot \Phi_n(u)\, du = \lambda_n \Phi_n(t)$$

  - where $\lambda_n = \text{Var}[X_n]$.
  - $\{\Phi_n\}$ and $\{\lambda_n\}$ are the eigenfunctions and eigenvalues of the autocovariance function $K_X(t,u)$, respectively.

Example: Wiener Process

- For the Wiener Process, the autocovariance function is

$$K_X(t,u) = R_X(t,u) = \sigma^2 \min(t,u).$$

- It can be shown that

$$\begin{aligned} \Phi_n(t) &= \sqrt{\frac{2}{T}} \sin((n-\frac{1}{2})\pi\frac{t}{T}) \\ \lambda_n &= \left(\frac{\sigma T}{(n-\frac{1}{2})\pi}\right)^2 \end{aligned} \quad \text{for } n = 1, 2, \ldots.$$

Properties of the K-L Expansion

- The eigenfunctions of the autocovariance function form a complete basis.
- If X_t is Gaussian, then the representation $\{X_n\}$ is a vector of independent, Gaussian random variables.
- For white noise, $K_X(t,u) = \frac{N_0}{2}\delta(t-u)$. Then, the eigenfunctions must satisfy

$$\int \frac{N_0}{2}\delta(t-u)\Phi(u)\,du = \lambda\Phi(t).$$

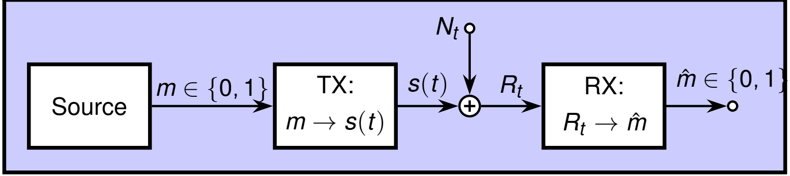
  - Any orthonormal set of bases $\{\Phi\}$ satisfies this condition!
  - Eigenvalues λ are all equal to $\frac{N_0}{2}$.
  - If X_t is white, Gaussian noise then the representation $\{X_n\}$ are independent, identically distributed random variables.
    - zero mean
    - variance $\frac{N_0}{2}$

Part III

Optimum Receivers in AWGN Channels

A Simple Communication System

Simple Communication System



- **Objectives:** For the above system
  - describe the optimum receiver and
  - find the probability of error for that receiver.

Assumptions

Noise: N_t is a white Gaussian noise process with spectral height $\frac{N_0}{2}$:

$$R_N(\tau) = \frac{N_0}{2}\delta(\tau).$$

- Additive White Gaussian Noise (AWGN).

Source: characterized by the *a priori* probabilities

$$\pi_0 = \Pr\{m = 0\} \quad \pi_1 = \Pr\{m = 1\}.$$

- For this example, will assume $\pi_0 = \pi_1 = \frac{1}{2}$.

Assumptions (cont'd)

Transmitter: maps information bits m to signals:

$$m \to s(t) : \begin{cases} s_0(t) = \sqrt{\frac{E_b}{T}} & \text{if } m = 0 \\ s_1(t) = -\sqrt{\frac{E_b}{T}} & \text{if } m = 1 \end{cases}$$

for $0 \leq t \leq T$.

- Note that we are considering the transmission of a single bit.
- In AWGN channels, each bit can be considered in isolation.

Objective

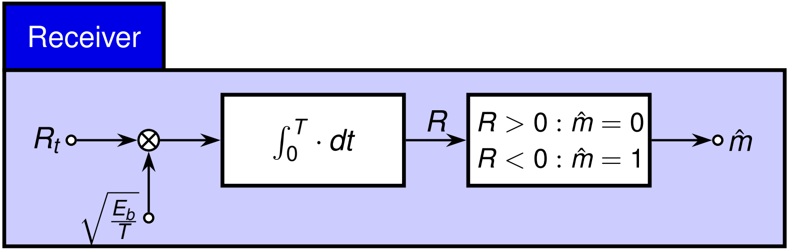
- In general, the objective is to find the receiver that minimizes the probability of error:

$$
\begin{aligned}
\Pr\{e\} &= \Pr\{\hat{m} \neq m\} \\
&= \pi_0 \Pr\{\hat{m} = 1 | m = 0\} + \pi_1 \Pr\{\hat{m} = 0 | m = 1\}.
\end{aligned}
$$

  - For this example, optimal receiver will be given (next slide).
- Also, compute the probability of error for the communication system.
  - That is the focus of this example.

Receiver

- We will see that the following receiver minimizes the probability of error for *this* communication system.



- **RX Frontend** computes $R = \int_0^T R_t \sqrt{\frac{E_b}{T}} \, dt = \langle R_t, s_0(t) \rangle$.
- **RX Backend** compares R to a threshold to arrive at decision $\hat{m}$.

Plan for Finding $\Pr\{e\}$

- Analysis of the receiver proceeds in the following steps:
  1. Find the *conditional* distribution of the output R from the receiver frontend.
     - Conditioning with respect to each of the possibly transmitted signals.
     - This boils down to finding conditional mean and variance of R.
  2. Find the conditional error probabilities $\Pr\{\hat{m} = 0 | m = 1\}$ and $\Pr\{\hat{m} = 1 | m = 0\}$.
     - Involves finding the probability that R exceeds a threshold.
  3. Total probability of error:

$$\Pr\{e\} = \pi_0 \Pr\{\hat{m} = 0 | m = 1\} + \pi_1 \Pr\{\hat{m} = 0 | m = 1\}.$$

Conditional Distribution of R

- There are two random effects that affect the received signal:
  - the additive white Gaussian noise N_t and
  - the random information bit m.
- By conditioning on m — thus, on $s(t)$ — randomness is caused by the noise only.
- Conditional on m, the output R of the receiver frontend is a Gaussian random variable:
  - N_t is a Gaussian random process; for given $s(t)$, $R_t = s(t) + N_t$ is a Gaussian random process.
  - The frontend performs a linear transformation of R_t: $R = \langle R_t, s_0(t) \rangle$.
- We need to find the conditional means and variances

Conditional Distribution of R

- The conditional means and variance of the frontend output R are

$$\mathbf{E}[R|m=0] = E_b \qquad \mathrm{Var}[R|m=0] = \frac{N_0}{2}E_b$$

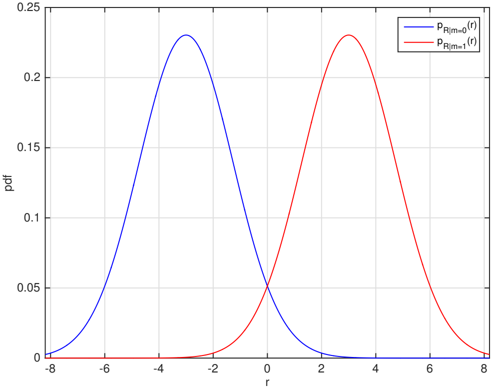
$$\mathbf{E}[R|m=1] = -E_b \qquad \mathrm{Var}[R|m=1] = \frac{N_0}{2}E_b$$

- Therefore, the conditional distributions of R are

$$p_{R|m=0}(r) \sim N(E_b, \frac{N_0}{2}E_b) \qquad p_{R|m=1}(r) \sim N(-E_b, \frac{N_0}{2}E_b)$$

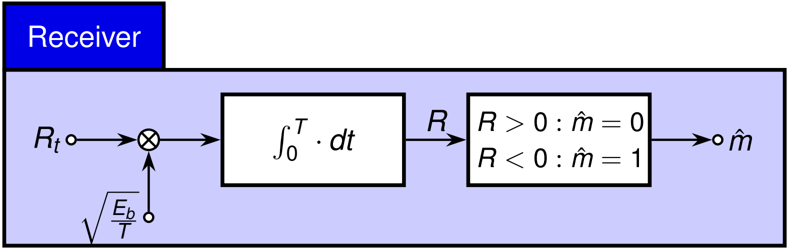
  - The two conditional distributions differ in the mean and have equal variances.

Conditional Distribution of R



- The two conditional pdfs are shown in the plot above, with
  - $E_b = 3$
  - $\frac{N_0}{2} = 1$

Conditional Probability of Error



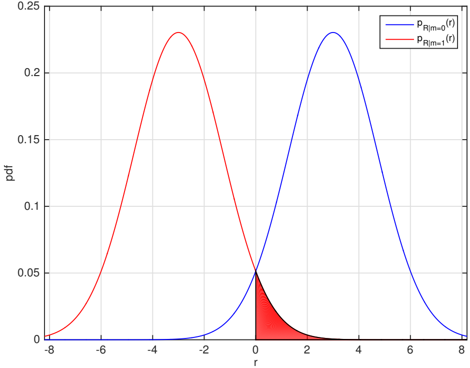
- The receiver backend decides:

$$\hat{m} = \begin{cases} 0 & \text{if } R > 0 \\ 1 & \text{if } R < 0 \end{cases}$$

- Two conditional error probabilities:

$$\Pr\{\hat{m} = 0 | m = 1\} \quad \text{and} \quad \Pr\{\hat{m} = 1 | m = 0\}$$

Error Probability $\Pr\{\hat{m} = 0 | m = 1\}$



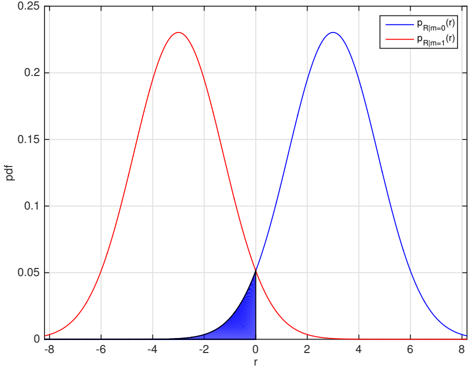
- Conditional error probability $\Pr\{\hat{m} = 0 | m = 1\}$ corresponds to shaded area.

$$\Pr\{\hat{m} = 0 | m = 1\} = \Pr\{R > 0 | m = 1\}$$

$$= \int_0^\infty p_{R|m=1}(r)\, dr = Q\left(\sqrt{\frac{2E_b}{N0}}\right)$$

© 2018, B.-P. Paris

Error Probability $\Pr\{\hat{m} = 1 | m = 0\}$



- Conditional error probability $\Pr\{\hat{m} = 1 | m = 0\}$ corresponds to shaded area.

$$\Pr\{\hat{m} = 1 | m = 0\} = \Pr\{R < 0 | m = 0\}$$
$$= \int_{-\infty}^{0} p_{R|m=0}(r)\, dr = Q\left(\sqrt{\frac{2E_b}{N0}}\right).$$

Average Probability of Error

- The (average) probability of error is the average of the two conditional probabilities of error.
  - The average is weighted by the a priori probabilities π_0 and π_1.
- Thus,

$$\Pr\{e\} = \pi_0 \Pr\{\hat{m} = 1 | m = 0\} + \pi_1 \Pr\{\hat{m} = 0 | m = 1\}.$$

- With the above conditional error probabilities and equal priors $\pi_0 = \pi_1 = \frac{1}{2}$

$$\Pr\{e\} = \frac{1}{2} \mathrm{Q}\left(\sqrt{\frac{2E_b}{N0}}\right) + \frac{1}{2} \mathrm{Q}\left(\sqrt{\frac{2E_b}{N0}}\right) = \mathrm{Q}\left(\sqrt{\frac{2E_b}{N0}}\right).$$

- Note that the error probability depends on the ratio $\frac{E_b}{N_0}$,
  - where E_b is the energy of signals $s_0(t)$ and $s_1(t)$.
  - This ratio is referred to as the signal-to-noise ratio.

Exercise - Compute Probability of Error

- Compute the probability of error for the example system if the only change in the system is that signals $s_0(t)$ and $s_1(t)$ are changed to triangular signals:

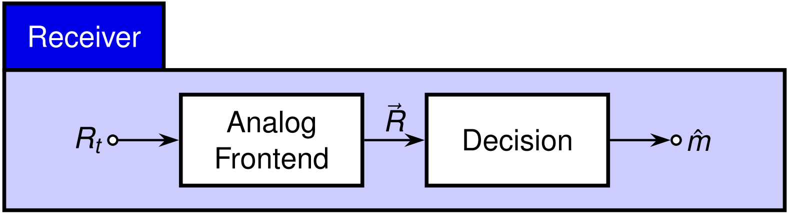
$$s_0(t) = \begin{cases} \frac{2A}{T} \cdot t & \text{for } 0 \leq t \leq \frac{T}{2} \\ 2A - \frac{2A}{T} \cdot t & \text{for } \frac{T}{2} \leq t \leq T \\ 0 & \text{else} \end{cases} \qquad s_1(t) = -s_0(t)$$

  with $A = \sqrt{\frac{3E_b}{T}}$.

- **Answer:**

$$\Pr\{e\} = Q\left(\sqrt{\frac{3E_b}{2N_0}}\right)$$

Structure of a Generic Receiver



- Receivers consist of:
  - an *analog frontend*: maps observed signal R_t to decision statistic $\vec{R}$.
  - *decision device*: determines which symbol $\hat{m}$ was sent based on observation of $\vec{R}$.
- Optimum design of decision device will be considered first.

Problem Setup

- **Given:**
  - a random vector $\vec{R} \in \mathbb{R}^n$ of observations and
  - hypotheses, H_0 and H_1, providing statistical models for $\vec{R}$:

$$H_0\colon \vec{R} \sim p_{\vec{R}|H_0}(\vec{r}|H_0)$$

$$H_1\colon \vec{R} \sim p_{\vec{R}|H_1}(\vec{r}|H_1)$$

  with known *a priori* probabilities $\pi_0 = \Pr\{H_0\}$ and $\pi_1 = \Pr\{H_1\}$ ($\pi_0 + \pi_1 = 1$).
- **Problem:** Decide which of the two hypotheses is best supported by the observation $\vec{R}$.
  - Specific objective: minimize the probability of error

$$\begin{aligned}\Pr\{e\} &= \Pr\{\text{decide } H_0 \text{ when } H_1 \text{ is true}\} \\ &\quad + \Pr\{\text{decide } H_1 \text{ when } H_0 \text{ is true}\} \\ &= \Pr\{\text{decide } H_0|H_1\}\Pr\{H_1\} + \Pr\{\text{decide } H_1|H_0\}\Pr\{H_0\}\end{aligned}$$

Generic Decision Rule

- The decision device performs a mapping that assigns a decision, H_0 or H_1, to each possible observation $\vec{R} \in \mathbb{R}^n$.
- A generic way to realize such a mapping is:
  - partition the space of all possible observations, $\mathbb{R}^n$, into two disjoint, complementary decision regions Γ_0 and Γ_1:

$$\Gamma_0 \cup \Gamma_1 = \mathbb{R}^n \text{ and } \Gamma_0 \cap \Gamma_1 = \emptyset.$$

  - **Decision Rule:**

$$\text{If } \vec{R} \in \Gamma_0\text{: decide } H_0$$

$$\text{If } \vec{R} \in \Gamma_1\text{: decide } H_1$$

Probability of Error

- The probability of error can now be expressed in terms of the decision regions Γ_0 and Γ_1:

$$\begin{aligned}\Pr\{e\} &= \Pr\{\text{decide } H_0|H_1\}\Pr\{H_1\} + \Pr\{\text{decide } H_1|H_0\}\Pr\{H_0\} \\ &= \pi_1 \int_{\Gamma_0} p_{\vec{R}|H_1}(\vec{r}|H_1)\, d\vec{r} + \pi_0 \int_{\Gamma_1} p_{\vec{R}|H_0}(\vec{r}|H_0)\, d\vec{r}\end{aligned}$$

- Our objective becomes to find the decision regions Γ_0 and Γ_1 that minimize the probability of error.

Probability of Error

- Since $\Gamma_0 \cup \Gamma_1 = \mathbb{R}^n$ it follows that $\Gamma_1 = \mathbb{R}^n \setminus \Gamma_0$

$$
\begin{aligned}
\Pr\{e\} &= \pi_1 \int_{\Gamma_0} p_{\vec{R}|H_1}(\vec{r}|H_1)\, d\vec{r} + \pi_0 \int_{\mathbb{R}^n \setminus \Gamma_0} p_{\vec{R}|H_0}(\vec{r}|H_0)\, d\vec{r} \\
&= \pi_0 \int_{\mathbb{R}^n} p_{\vec{R}|H_0}(\vec{r}|H_0)\, d\vec{r} \\
&+ \int_{\Gamma_0} (\pi_1 p_{\vec{R}|H_1}(\vec{r}|H_1) - \pi_0 p_{\vec{R}|H_0}(\vec{r}|H_0))\, d\vec{r} \\
&= \pi_0 - \int_{\Gamma_0} (\pi_0 p_{\vec{R}|H_0}(\vec{r}|H_0) - \pi_1 p_{\vec{R}|H_1}(\vec{r}|H_1))\, d\vec{r}.
\end{aligned}
$$

- $\Pr\{e\}$ is minimized by chosing Γ_0 to contain all $\vec{r}$ for which the integrand $(\pi_0 p_{\vec{R}|H_0}(\vec{r}|H_0) - \pi_1 p_{\vec{R}|H_1}(\vec{r}|H_1)) < 0$.

Minimum $\Pr\{e\}$ (MPE) Decision Rule

- Thus, the decision region Γ_0 that minimizes the probability of error is given by:

$$
\begin{aligned}
\Gamma_0 &= \left\{\vec{r} : (\pi_0 p_{\vec{R}|H_0}(\vec{r}|H_0) - \pi_1 p_{\vec{R}|H_1}(\vec{r}|H_1)) > 0\right\} \\
&= \left\{\vec{r} : \pi_0 p_{\vec{R}|H_0}(\vec{r}|H_0) > \pi_1 p_{\vec{R}|H_1}(\vec{r}|H_1))\right\} \\
&= \left\{\vec{r} : \frac{p_{\vec{R}|H_1}(\vec{r}|H_1)}{p_{\vec{R}|H_0}(\vec{r}|H_0)} < \frac{\pi_0}{\pi_1}\right\}
\end{aligned}
$$

- The decision region Γ_1 follows

$$
\Gamma_1 = \Gamma_0^C = \left\{\vec{r} : \frac{p_{\vec{R}|H_1}(\vec{r}|H_1)}{p_{\vec{R}|H_0}(\vec{r}|H_0)} > \frac{\pi_0}{\pi_1}\right\}
$$

Likelihood Ratio

- The MPE decision rule can be written as

$$\text{If } \frac{p_{\vec{R}|H_1}(\vec{R}|H_1)}{p_{\vec{R}|H_0}(\vec{R}|H_0)} \begin{cases} > \frac{\pi_0}{\pi_1} & \text{decide } H_1 \\ < \frac{\pi_0}{\pi_1} & \text{decide } H_0 \end{cases}$$

  - **Notation:**

$$\frac{p_{\vec{R}|H_1}(\vec{R}|H_1)}{p_{\vec{R}|H_0}(\vec{R}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\pi_0}{\pi_1}$$

- The ratio of conditional density functions

$$\Lambda(\vec{R}) = \frac{p_{\vec{R}|H_1}(\vec{R}|H_1)}{p_{\vec{R}|H_0}(\vec{R}|H_0)}$$

is called the likelihood ratio.

Log-Likelihood Ratio

- Many of the densities of interest are exponential functions (e.g., Gaussian).
- For these densities, it is advantageous to take the log of both sides of the decision rule.
  - **Important:** This does not change the decision rule because the logarithm is monotonically increasing!
- The MPE decision rule can be written as:

$$L(\vec{R}) = \ln\left(\frac{p_{\vec{R}|H_1}(\vec{R}|H_1)}{p_{\vec{R}|H_0}(\vec{R}|H_0)}\right) \underset{H_0}{\overset{H_1}{\gtrless}} \ln\left(\frac{\pi_0}{\pi_1}\right)$$

- $L(\vec{R}) = \ln(\Lambda(\vec{R}))$ is called the log-likelihood ratio.

Example: Gaussian Hypothesis Testing

- The most important hypothesis testing problem for communications over AWGN channels is

$$H_0 : \vec{R} \sim N(\vec{m}_0, \sigma^2 I)$$

$$H_1 : \vec{R} \sim N(\vec{m}_1, \sigma^2 I)$$

- This problem arises when
  - one of two known signals is transmitted over an AWGN channel, and
  - a linear analog frontend is used.
- Note that
  - the conditional means are different — reflecting different signals
  - covariance matrices are the same — since they depend on noise only.
  - components of $\vec{R}$ are independent — indicating that the frontend projects R_t onto orthogonal bases.

Resulting Log-Likelihood Ratio

- For this problem, the log-likelihood ratio simplifies to

$$
\begin{aligned}
L(\vec{R}) &= \frac{1}{2\sigma^2} \sum_{k=1}^{n} (R_k - m_{0k})^2 - (R_k - m_{1k})^2 \\
&= \frac{1}{2\sigma^2} (\|\vec{R} - \vec{m}_0\|^2 - \|\vec{R} - \vec{m}_1\|^2) \\
&= \frac{1}{2\sigma^2} \left(2\langle \vec{R}, \vec{m}_1 - \vec{m}_0 \rangle - (\|\vec{m}_1\|^2 - \|\vec{m}_0\|^2) \right)
\end{aligned}
$$

  - The second expressions shows that the *Euclidean distance* between observations $\vec{R}$ and means $\vec{m}_i$ plays a central role in Gaussian hypothesis testing.
  - The last expression highlights the projection of the observation $\vec{R}$ onto the difference between the means $\vec{m}_i$.

MPE Decision Rule

- With the above log-liklihood ratio, the MPE decision rule becomes equivalently
  - either

$$\langle \vec{R}, \vec{m}_1 - \vec{m}_0 \rangle \underset{H_0}{\overset{H_1}{\gtrless}} \sigma^2 \ln\left(\frac{\pi_0}{\pi_1}\right) + \frac{\|\vec{m}_1\|^2 - \|\vec{m}_0\|^2}{2}$$

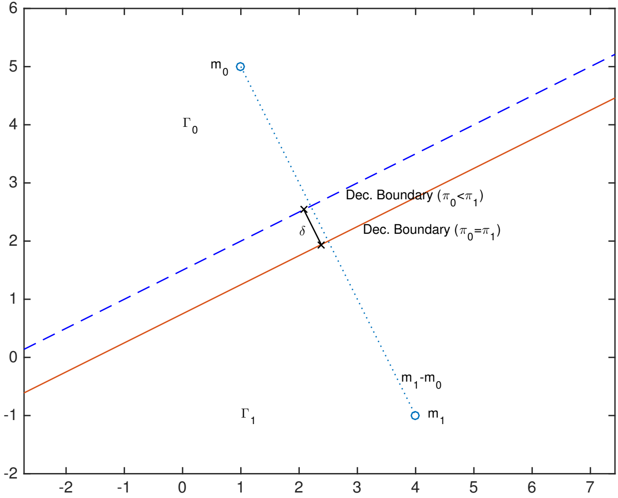
  - or

$$\|\vec{R} - \vec{m}_0\|^2 - 2\sigma^2 \ln(\pi_0) \underset{H_0}{\overset{H_1}{\gtrless}} \|\vec{R} - \vec{m}_1\|^2 - 2\sigma^2 \ln(\pi_1)$$

Decision Regions

- The MPE decision rule divides $\mathbb{R}^n$ into two half planes that are the decision regions Γ_0 and Γ_1.
- The dividing line (decision boundary) between the regions is *perpendicular to* $\vec{m}_1 - \vec{m}_0$.
  - This is a consequence of the inner product in the first form of the decision rule.
- If the priors π_0 and π_1 are equal, then the decision boundary passes through the midpoint $\frac{\vec{m}_0+\vec{m}_1}{2}$.
  - For unequal priors, the decision boundary is shifted towards the mean of the *less likely* hypothesis.
  - The distance of this shift equals $\delta = \frac{2\sigma^2 |\ln(\pi_0/\pi_1)|}{\|\vec{m}_1 - \vec{m}_0\|}$.
  - This follows from the (squared) distances in the second form of the decision rule.

Decision Regions



Probability of Error

- **Question:** What is the probability of error with the MPE decision rule?
  - Using MPE decision rule

$$\langle \vec{R}, \vec{m}_1 - \vec{m}_0 \rangle \underset{H_0}{\overset{H_1}{\gtrless}} \sigma^2 \ln\left(\frac{\pi_0}{\pi_1}\right) + \frac{\|\vec{m}_1\|^2 - \|\vec{m}_0\|^2}{2}$$

- **Plan:**
  - Find conditional densities of $\langle \vec{R}, \vec{m}_1 - \vec{m}_0 \rangle$ under H_0 and H_1.
  - Find conditional error probabilities

$$\int_{\Gamma_i} p_{\vec{R}|H_j}(\vec{r}|H_j)\, d\vec{r} \text{ for } i \neq j.$$

  - Find average probability of error.

Conditional Distributions

- Since $\langle \vec{R}, \vec{m}_1 - \vec{m}_0 \rangle$ is a linear transformation and $\vec{R}$ is Gaussian, the conditional distributions are Gaussian.

$$H_0\colon N(\underbrace{\langle \vec{m}_0, \vec{m}_1 \rangle - \|\vec{m}_0\|^2}_{\mu_0}, \underbrace{\sigma^2 \|\vec{m}_0 - \vec{m}_1\|^2}_{\sigma_m^2})$$

$$H_1\colon N(\underbrace{\|\vec{m}_1\|^2 - \langle \vec{m}_0, \vec{m}_1 \rangle}_{\mu_1}, \underbrace{\sigma^2 \|\vec{m}_0 - \vec{m}_1\|^2}_{\sigma_m^2})$$

Conditional Error Probabilities

- The MPE decision rule compares

$$\langle \vec{R}, \vec{m}_1 - \vec{m}_0 \rangle \underset{H_0}{\overset{H_1}{\gtrless}} \underbrace{\sigma^2 \ln\left(\frac{\pi_0}{\pi_1}\right) + \frac{\|\vec{m}_1\|^2 - \|\vec{m}_0\|^2}{2}}_{\gamma}$$

- Resulting conditional probabilities of error

$$\Pr\{e|H_0\} = \mathrm{Q}\left(\frac{\gamma - \mu_0}{\sigma_m}\right) = \mathrm{Q}\left(\frac{\|\vec{m}_0 - \vec{m}_1\|}{2\sigma} + \frac{\sigma \ln(\pi_0/\pi_1)}{\|\vec{m}_0 - \vec{m}_1\|}\right)$$

$$\Pr\{e|H_1\} = \mathrm{Q}\left(\frac{\mu_1 - \gamma}{\sigma_m}\right) = \mathrm{Q}\left(\frac{\|\vec{m}_0 - \vec{m}_1\|}{2\sigma} - \frac{\sigma \ln(\pi_0/\pi_1)}{\|\vec{m}_0 - \vec{m}_1\|}\right)$$

Average Probability of Error

- The average error probability equals

$$\begin{aligned}\Pr\{e\} =& \Pr\{\text{decide } H_0|H_1\}\Pr\{H_1\} + \Pr\{\text{decide } H_1|H_0\}\Pr\{H_0\}\\ =&\pi_0 \mathrm{Q}\left(\frac{\|\vec{m}_0-\vec{m}_1\|}{2\sigma} + \frac{\sigma\ln(\pi_0/\pi_1)}{\|\vec{m}_0-\vec{m}_1\|}\right) +\\ &\pi_1 \mathrm{Q}\left(\frac{\|\vec{m}_0-\vec{m}_1\|}{2\sigma} - \frac{\sigma\ln(\pi_0/\pi_1)}{\|\vec{m}_0-\vec{m}_1\|}\right)\end{aligned}$$

- Important special case: $\pi_0 = \pi_1 = \frac{1}{2}$

$$\Pr\{e\} = \mathrm{Q}\left(\frac{\|\vec{m}_0-\vec{m}_1\|}{2\sigma}\right)$$

  - The error probability depends on the ratio of
    - distance between means $\|\vec{m}_0-\vec{m}_1\|$
    - and noise standard deviation

Maximum-Likelihood (ML) Decision Rule

- The maximum-likelihood decision rule disregards priors and decides for the hypothesis with higher likelihood.
- **ML Decision rule:**

$$\Lambda(\vec{R}) = \frac{p_{\vec{R}|H_1}(\vec{R}|H_1)}{p_{\vec{R}|H_0}(\vec{R}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} 1$$

or equivalently, in terms of the log-likelihood,

$$L(\vec{R}) = \ln\left(\frac{p_{\vec{R}|H_1}(\vec{R}|H_1)}{p_{\vec{R}|H_0}(\vec{R}|H_0)}\right) \underset{H_0}{\overset{H_1}{\gtrless}} 0$$

- Obviously, the ML decision is equivalent to the MPE rule when the priors are equal.
- In the Gaussian case, the ML rule does not require knowledge of the noise variance.

A-Posteriori Probability

- By Bayes rule, the probability of hypothesis H_i *after* observing $\vec{R}$ is
$$\Pr\{H_i|\vec{R}=\vec{r}\} = \frac{\pi_i p_{\vec{R}|H_i}(\vec{r}|H_i)}{p_{\vec{R}}(\vec{r})},$$
where $p_{\vec{R}}(\vec{r})$ is the unconditional pdf of $\vec{R}$
$$p_{\vec{R}}(\vec{r}) = \sum_i \pi_i p_{\vec{R}|H_i}(\vec{r}|H_i).$$
- **Maximum A-Posteriori (MAP) decision rule:**
$$\Pr\{H_1|\vec{R}=\vec{r}\} \underset{H_0}{\overset{H_1}{\gtrless}} \Pr\{H_0|\vec{R}=\vec{r}\}$$
  - **Interpretation:** Decide in favor of the hypothesis that is more likely given the observed signal $\vec{R}$.

The MAP and MPE Rules are Equivalent

- The MAP and MPE rules are equivalent: the MAP decision rule achieves the minimum probability of error.
- The MAP rule can be written as

$$\frac{\Pr\{H_1|\vec{R}=\vec{r}\}}{\Pr\{H_0|\vec{R}=\vec{r}\}} \underset{H_0}{\overset{H_1}{\gtrless}} 1.$$

- Inserting $\Pr\{H_i|\vec{R}=\vec{r}\} = \frac{\pi_i p_{\vec{R}|H_i}(\vec{r}|H_i)}{p_{\vec{R}}(\vec{r})}$ yields

$$\frac{\pi_1 p_{\vec{R}|H_1}(\vec{r}|H_1)}{\pi_0 p_{\vec{R}|H_0}(\vec{r}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} 1$$

  - This is obviously equal to the MPE rule

$$\frac{p_{\vec{R}|H_1}(\vec{r}|H_1)}{p_{\vec{R}|H_0}(\vec{r}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\pi_0}{\pi_1}.$$

More than Two Hypotheses

- Frequently, more than two hypotheses must be considered:

$$
\begin{aligned}
H_0 &: \vec{R} \sim p_{\vec{R}|H_0}(\vec{r}|H_0) \\
H_1 &: \vec{R} \sim p_{\vec{R}|H_1}(\vec{r}|H_1) \\
&\vdots \\
H_M &: \vec{R} \sim p_{\vec{R}|H_M}(\vec{r}|H_M)
\end{aligned}
$$

- In these cases, it is no longer possible to reduce the decision rules to
  - the computation of the likelihood ratio
  - followed by comparison to a threshold

More than Two Hypotheses

- Instead the decision rules take the following forms
  - **MPE rule:**

$$\hat{m} = \arg \max_{i \in \{0,\dots,M-1\}} \pi_i p_{\vec{R}|H_i}(\vec{r}|H_i)$$

  - **ML rule:**

$$\hat{m} = \arg \max_{i \in \{0,\dots,M-1\}} p_{\vec{R}|H_i}(\vec{r}|H_i)$$

  - **MAP rule:**

$$\hat{m} = \arg \max_{i \in \{0,\dots,M-1\}} \Pr\{H_i|\vec{R} = \vec{r}\}$$

More than Two Hypotheses: The Gaussian Case

- When the hypotheses are of the form H_i: $\vec{R} \sim N(\vec{m}_i, \sigma^2 I)$, then the decision rules become:
  - **MPE and MAP decision rules:**

$$\begin{aligned}\hat{m} &= \arg \min_{i \in \{0,\ldots,M-1\}} \|\vec{r} - \vec{m}_i\|^2 - 2\sigma^2 \ln(\pi_i) \\ &= \arg \max_{i \in \{0,\ldots,M-1\}} \langle \vec{r}, \vec{m}_i \rangle + \sigma^2 \ln(\pi_i) - \frac{\|\vec{m}_i\|^2}{2}\end{aligned}$$

  - **ML decision rule:**

$$\begin{aligned}\hat{m} &= \arg \min_{i \in \{0,\ldots,M-1\}} \|\vec{r} - \vec{m}_i\|^2 \\ &= \arg \max_{i \in \{0,\ldots,M-1\}} \langle \vec{r}, \vec{m}_i \rangle - \frac{\|\vec{m}_i\|^2}{2}\end{aligned}$$

    - This is also the MPE rule when the priors are all equal.

Take-Aways

- The conditional densities $p_{\vec{R}|H_i}(\vec{r}|H_i)$ play a key role.
- **MPE decision rule:**
  - Binary hypotheses:

$$\Lambda(\vec{R}) = \frac{p_{\vec{R}|H_1}(\vec{R}|H_1)}{p_{\vec{R}|H_0}(\vec{R}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\pi_0}{\pi_1}$$

  - M hypotheses:

$$\hat{m} = \arg \max_{i \in \{0,\ldots,M-1\}} \pi_i p_{\vec{R}|H_i}(\vec{r}|H_i).$$

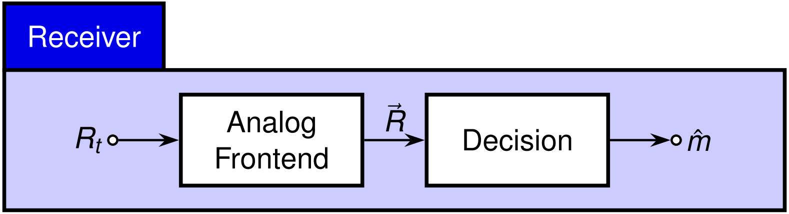
Take-Aways

- For the Gaussian case (different means, equal variance), decisions are based on the Euclidean distance between observations $\vec{R}$ and conditional means $\vec{m}_i$:

$$\hat{m} = \arg \min_{i \in \{0,\dots,M-1\}} \|\vec{r} - \vec{m}_i\|^2 - 2\sigma^2 \ln(\pi_i)$$

$$= \arg \max_{i \in \{0,\dots,M-1\}} \langle \vec{r}, \vec{m}_i \rangle + \sigma^2 \ln(\pi_i) - \frac{\|\vec{m}_i\|^2}{2}$$

Structure of a Generic Receiver



- Receivers consist of:
  - an *analog frontend*: maps observed signal R_t to decision statistic $\vec{R}$.
  - *decision device*: determines which symbol $\hat{m}$ was sent based on observation of $\vec{R}$.
- Focus on designing optimum frontend.

Problem Formulation and Assumptions

- In terms of the received signal R_t, we can formulate the following decision problem:

$$H_0\colon R_t = s_0(t) + N_t \text{ for } 0 \leq t \leq T$$
$$H_1\colon R_t = s_1(t) + N_t \text{ for } 0 \leq t \leq T$$

- **Assumptions:**
  - N_t is whithe Gaussian noise with spectral height $\frac{N_0}{2}$.
  - N_t is independent of the transmitted signal.
- **Objective:** Determine the optimum receiver frontend.

Starting Point: KL-Expansion

- Under the i-th hypothesis, the received signal R_t can be represented over $0 \leq t \leq T$ via the expansion

$$H_i\colon R_t = \sum_{j=0}^{\infty} R_j \Phi_j(t) = \sum_{j=0}^{\infty} (s_{ij} + N_j) \Phi_j(t).$$

- **Recall:**
  - If the above representation yields *uncorrelated* coefficients R_j, then this is a Karhunen-Loeve expansion.
  - Since N_t is white, *any* orthonormal basis $\{\Phi_j(t)\}$ yields a Karhunen-Loeve expansion.
- **Insight:**
  - We can *choose* a basis $\{\Phi_j(t)\}$ that produces a low-dimensional representation for all signals $s_i(t)$.

Constructing a Good Basis

- Consider the complete, but not necessarily orthonormal, basis
$$\{s_0(t), s_1(t), \Psi_0(t), \Psi_1(t), \ldots\}.$$
where $\{\Psi_j(t)\}$ is any complete basis over $0 \leq t \leq T$ (e.g., the Fourier basis).
- Then, the Gram-Schmidt procedure is used to convert the above basis into an orthonormal basis $\{\Phi_j\}$.

Properties of Resulting Basis

- **Notice:** with this construction
  - only the first $M \leq 2$ basis functions $\Phi_j(t)$, $j < M \leq 2$ are dependent on the signals $s_i(t)$, $i \leq 2$.
    - I.e., for each $j < M$,
      $$\langle s_i(t), \Phi_j(t) \rangle \neq 0 \text{ for at least one } i = 0, 1$$
    - Recall, $M < 2$ if signals are not linearly independent.
  - The remaining basis functions $\Phi_j(t)$, $j \geq M$ are orthogonal to the signals $s_i(t)$, $i \leq 2$
    - I.e., for each $j \geq M$,
      $$\langle s_i(t), \Phi_j(t) \rangle = 0 \text{ for all } i = 0, 1$$

Back to the Decision Problem

- Our decision problem can now be written in terms of the representation

$$H_0\colon R_t = \underbrace{\sum_{j=0}^{M-1} (s_{0j} + N_j)\Phi_j(t)}_{\text{signal + noise}} + \underbrace{\sum_{j=M}^{\infty} N_j \Phi_j(t)}_{\text{noise only}}$$

$$H_1\colon R_t = \underbrace{\sum_{j=0}^{M-1} (s_{1j} + N_j)\Phi_j(t)}_{\text{signal + noise}} + \underbrace{\sum_{j=M}^{\infty} N_j \Phi_j(t)}_{\text{noise only}}$$

where

$$s_{ij} = \langle s_i(t), \Phi_j(t) \rangle$$
$$N_j = \langle N_t, \Phi_j(t) \rangle$$

- Note that N_j are independent, Gaussian random variables, $N_j \sim \mathrm{N}(0, \frac{N_0}{2})$

Vector Version of Decision Problem

- The received signal R_t and its representation $\vec{R} = \{R_j\}$ are equivalent.
  - Via the basis $\{\Phi_j\}$ one can be obtained from the other.
- Therefore, the decision problem can be written in terms of the representations

$$H_0 \colon \vec{R} = \vec{s}_0 + \vec{N}$$
$$H_1 \colon \vec{R} = \vec{s}_1 + \vec{N}$$

where

- all vectors are of infinite length,
- the elements of $\vec{N}$ are i.i.d., zero mean Gaussian,
- all elements s_{ij} with $j \geq M$ are zero.

Reducing the Number of Dimensions

- We can write the conditional pdfs for the decision problem

$$H_0\colon \vec{R} \sim \prod_{j=0}^{M-1} p_N(r_j - s_{0j}) \cdot \prod_{j=M}^{\infty} p_N(r_j)$$

$$H_1\colon \vec{R} \sim \prod_{j=0}^{M-1} p_N(r_j - s_{1j}) \cdot \prod_{j=M}^{\infty} p_N(r_j)$$

where $p_N(r)$ denotes a Gaussian pdf with zero mean and variance $\frac{N_0}{2}$.

Reducing the Number of Dimensions

- The optimal decision relies on the likelihood ratio

$$L(\vec{R}) = \frac{\prod_{j=0}^{M-1} p_N(r_j - s_{0j}) \cdot \prod_{j=M}^{\infty} p_N(r_j)}{\prod_{j=0}^{M-1} p_N(r_j - s_{1j}) \cdot \prod_{j=M}^{\infty} p_N(r_j)} = \frac{\prod_{j=0}^{M-1} p_N(r_j - s_{0j})}{\prod_{j=0}^{M-1} p_N(r_j - s_{1j})}$$

- The likelihood ratio depends only on the first M dimensions of $\vec{R}$!
  - Dimensions greater than or equal to M are *irrelevant* for the decision problem.
  - Only the the first M dimension need to be computed for optimal decisions.

Reduced Decision Problem

- The following decision problem with M dimensions is equivalent to our original decision problem (assumes $M = 2$):

$$H_0\colon \vec{R} = \begin{pmatrix} s_{00} \\ s_{01} \end{pmatrix} + \begin{pmatrix} N_0 \\ N_1 \end{pmatrix} = \vec{s}_0 + \vec{N} \sim \mathrm{N}(\vec{s}_0, \frac{N_0}{2} I)$$

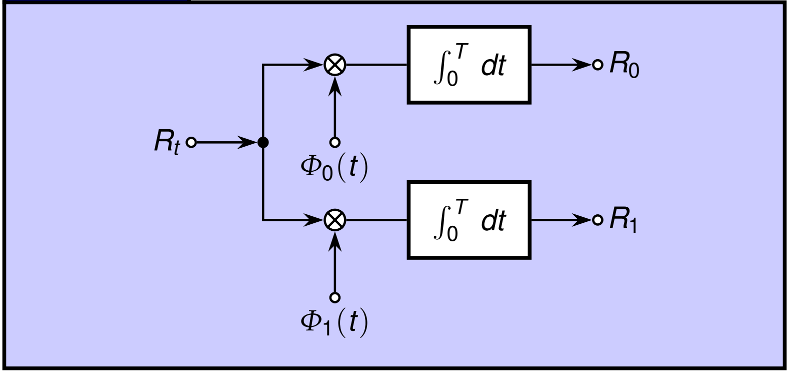
$$H_1\colon \vec{R} = \begin{pmatrix} s_{10} \\ s_{11} \end{pmatrix} + \begin{pmatrix} N_0 \\ N_1 \end{pmatrix} = \vec{s}_1 + \vec{N} \sim \mathrm{N}(\vec{s}_1, \frac{N_0}{2} I)$$

- When $s_0(t)$ and $s_1(t)$ are linearly dependent, i.e., $s_1(t) = a \cdot s_0(t)$, then $M = 1$ and the decision problem becomes one-dimensional.

Optimal Frontend - Version 1

- From the above discussion, we can conclude that an optimal frontend is given by.

Frontend 1



Optimum Receiver - Version 1

- Note that the optimum frontend projects the received signal R_t into to signal subspace spanned by the signals $s_i(t)$.
  - Recall that the first basis functions $\Phi_j(t)$, $j < M$, are obtained from the signals.
- We know how to solve the resulting, M-dimensional decision problem

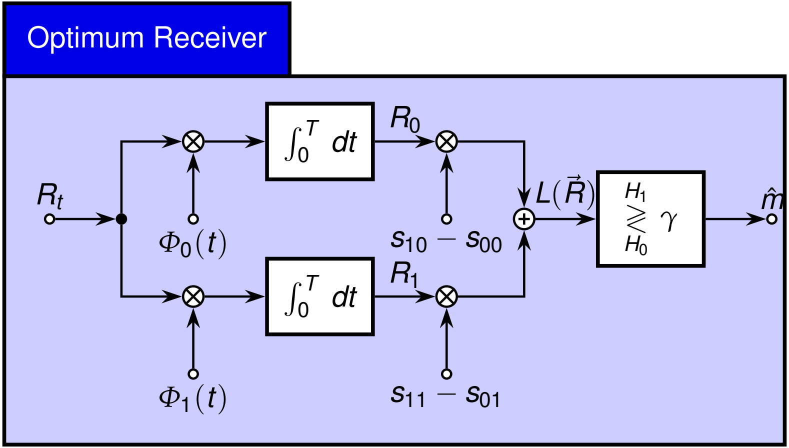
$$H_0\colon \vec{R} = \begin{pmatrix} s_{00} \\ s_{01} \end{pmatrix} + \begin{pmatrix} N_0 \\ N_1 \end{pmatrix} = \vec{s}_0 + \vec{N} \sim \mathrm{N}(\vec{s}_0, \frac{N_0}{2} I)$$

$$H_1\colon \vec{R} = \begin{pmatrix} s_{10} \\ s_{11} \end{pmatrix} + \begin{pmatrix} N_0 \\ N_1 \end{pmatrix} = \vec{s}_1 + \vec{N} \sim \mathrm{N}(\vec{s}_1, \frac{N_0}{2} I)$$

Optimum Receiver - Version 1

- MPE decision rule:
  1. Compute
  $$L(\vec{R}) = \langle \vec{R}, \vec{s}_1 - \vec{s}_0 \rangle.$$
  2. Compare to threshold:
  $$\gamma = \frac{N_0}{2} \ln(\pi_0 / \pi_1) + \frac{\|\vec{s}_1\|^2 - \|\vec{s}_0\|^2}{2}$$
  3. Decision
  $$\text{If } L(\vec{R}) > \gamma \text{ decide } s_1(t) \text{ was sent.}$$
  $$\text{If } L(\vec{R}) < \gamma \text{ decide } s_0(t) \text{ was sent.}$$

Optimum Receiver - Version 1



Probability of Error

- The probability of error for this receiver is

$$\Pr\{e\} = \pi_0 \mathrm{Q}\left(\frac{\|\vec{s}_0 - \vec{s}_1\|}{2\sqrt{\frac{N_0}{2}}} + \sqrt{\frac{N_0}{2}} \frac{\ln(\pi_0/\pi_1)}{\|\vec{s}_0 - \vec{s}_1\|} \right)$$
$$+ \pi_1 \mathrm{Q}\left(\frac{\|\vec{s}_0 - \vec{s}_1\|}{2\sqrt{\frac{N_0}{2}}} - \sqrt{\frac{N_0}{2}} \frac{\ln(\pi_0/\pi_1)}{\|\vec{s}_0 - \vec{s}_1\|} \right)$$

- For the important special case of equally likely signals:

$$\Pr\{e\} = \mathrm{Q}\left(\frac{\|\vec{s}_0 - \vec{s}_1\|}{2\sqrt{\frac{N_0}{2}}} \right) = \mathrm{Q}\left(\frac{\|\vec{s}_0 - \vec{s}_1\|}{\sqrt{2N_0}}. \right)$$

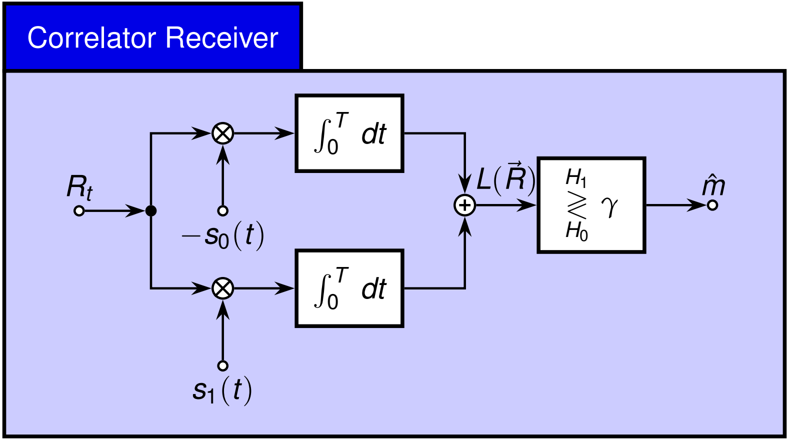
- This is the minimum probability of error achievable by *any* receiver.

Optimum Receiver - Version 2

- The optimum receiver derived above, computes the inner product
$$\langle \vec{R}, \vec{s}_1 - \vec{s}_0 \rangle.$$
- By Parseval's relationship, the inner product of the representation equals the inner product of the signals

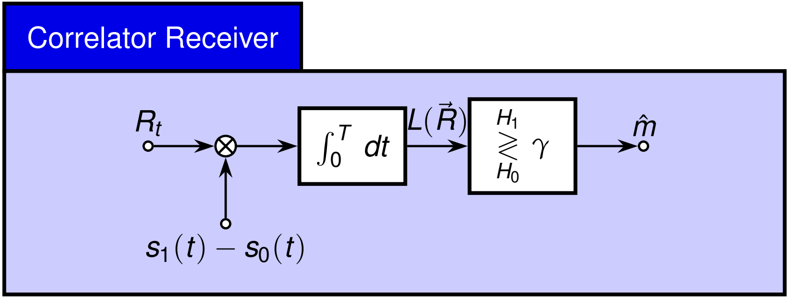
$$\begin{aligned}
\langle \vec{R}, \vec{s}_1 - \vec{s}_0 \rangle &= \langle R_t, s_1(t) - s_0(t) \rangle \\
&= \int_0^T R_t(s_1(t) - s_0(t))\, dt \\
&= \int_0^T R_t s_1(t)\, dt - \int_0^T R_t s_0(t))\, dt.
\end{aligned}$$

Optimum Receiver - Version 2



- Correlator receiver.

Optimum Receiver - Version 2a



- The two correlators can be combined into a single correlator for an even simpler frontend.

Optimum Receiver - Version 3

- Yet another, important structure for the optimum receiver frontend results from the equivalence between *correlation* and *convolution followed by sampling*.
  - Convolution:

$$y(t) = x(t) * h(t) = \int_0^T x(\tau) h(t-\tau)\, d\tau$$

  - Sample at $t = T$:

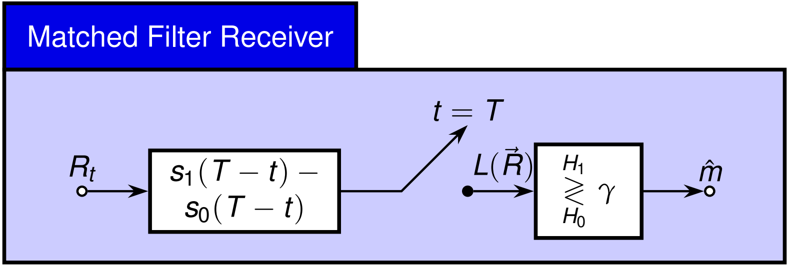
$$y(T) = x(t) * h(t)|_{t=T} = \int_0^T x(\tau) h(T-\tau)\, d\tau$$

  - Let $g(t) = h(T-t)$ (and, thus, $h(t) = g(T-t)$):

$$\int_0^T x(t) g(t)\, dt = \int_0^T x(\tau) h(T-\tau)\, d\tau = x(t) * h(t)|_{t=T}.$$

- Correlating with $g(t)$ is equivalent to convolving with $h(t) = g(T-t)$, followed by symbol-rate sampling.

Optimum Receiver - Version 3



- The filter with impulse response $h(t) = s_1(T-t) - s_0(T-t)$ is called the matched filter for $s_1(t) - s_0(t)$.

Exercises: Optimum Receiver

- For each of the following signal sets:
  1. draw a block diagram of the MPE receiver,
  2. compute the value of the threshold in the MPE receiver,
  3. compute the probability of error for this receiver for $\pi_0 = \pi_1$,
  4. find basis functions for the signal set,
  5. illustrate the location of the signals in the signal space spanned by the basis functions,
  6. draw the decision boundary formed by the optimum receiver.

On-Off Keying

- Signal set:

$$\left.\begin{aligned} s_0(t) &= 0 \\ s_1(t) &= \sqrt{\frac{E}{T}} \end{aligned}\right\} \quad \text{for } 0 \leq t \leq T$$

  - This signal set is referred to as *On-Off Keying (OOK)* or *Amplitude Shift Keying (ASK)*.

Orthogonal Signalling

- Signal set:

$$s_0(t) = \begin{cases} \sqrt{\frac{E}{T}} & \text{for } 0 \leq t \leq \frac{T}{2} \\ -\sqrt{\frac{E}{T}} & \text{for } \frac{T}{2} \leq t \leq T \end{cases}$$

$$s_1(t) = \sqrt{\frac{E}{T}} \quad \text{for } 0 \leq t \leq T$$

- Alternatively:

$$\left.\begin{aligned} s_0(t) &= \sqrt{\frac{2E}{T}} \cos(2\pi f_0 t) \\ s_1(t) &= \sqrt{\frac{2E}{T}} \cos(2\pi f_1 t) \end{aligned}\right\} \quad \text{for } 0 \leq t \leq T$$

with $f_0 T$ and $f_1 T$ distinct integers.
  - This signal set is called *Frequency Shift Keying (FSK)*.

Antipodal Signalling

- Signal set:

$$\left.\begin{aligned} s_0(t) &= -\sqrt{\frac{E}{T}} \\ s_1(t) &= \sqrt{\frac{E}{T}} \end{aligned}\right\} \quad \text{for } 0 \leq t \leq T$$

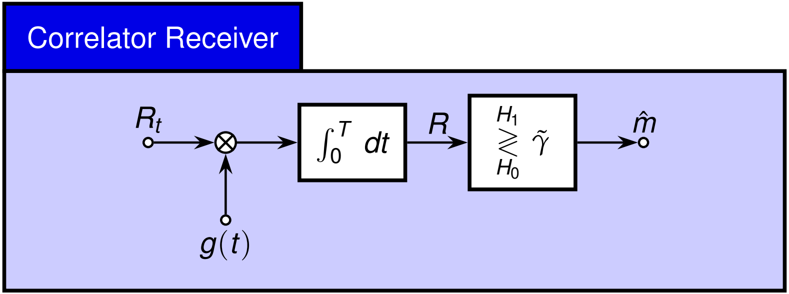
  - This signal set is referred to as *Antipodal Signalling*.
- Alternatively:

$$\left.\begin{aligned} s_0(t) &= \sqrt{\frac{2E}{T}}\cos(2\pi f_0 t) \\ s_1(t) &= \sqrt{\frac{2E}{T}}\cos(2\pi f_0 t + \pi) \end{aligned}\right\} \quad \text{for } 0 \leq t \leq T$$

  - This signal set is called *Binary Phase Shift Keying (BPSK)*.

Linear Receiver

- Consider a receiver with a "generic" linear frontend.



- We refer to these receivers as *linear receivers* because their frontend performs a linear transformation of the received signal.
  - Specifically, frontend computes $R = \langle R_t, g(t) \rangle$.

Linear Receiver

- **Objectives:**
  - derive general expressions for the conditional pdfs at the output R of the frontend,
  - derive general expressions for the error probability,
  - confirm that the optimum linear receiver correlates with $g(t) = s_1(t) - s_0(t)$,
    - i.e., the MPE receiver is also the best linear receiver.
- These results are useful for the analysis of arbitrary linear receivers.

Conditional Distributions

- Hypotheses:

$$H_0\colon R_t = s_0(t) + N_t$$
$$H_1\colon R_t = s_1(t) + N_t$$

  signals are observed for $0 \leq t \leq T$.
  - Priors are π_0 and π_1.
- Conditional distributions of $R = \langle R_t, g(t) \rangle$ are Gaussian:

$$H_0\colon R \sim \mathrm{N}(\underbrace{\langle s_0(t), g(t) \rangle}_{\mu_0}, \underbrace{\frac{N_0}{2} \|g(t)\|^2}_{\sigma^2})$$

$$H_1\colon R \sim \mathrm{N}(\underbrace{\langle s_1(t), g(t) \rangle}_{\mu_1}, \underbrace{\frac{N_0}{2} \|g(t)\|^2}_{\sigma^2})$$

MPE Decision Rule

- For the decision problem

$$H_0\colon\ R \sim \mathrm{N}(\underbrace{\langle s_0(t), g(t)\rangle}_{\mu_0}, \underbrace{\frac{N_0}{2}\|g(t)\|^2}_{\sigma^2})$$

$$H_1\colon\ R \sim \mathrm{N}(\underbrace{\langle s_1(t), g(t)\rangle}_{\mu_1}, \underbrace{\frac{N_0}{2}\|g(t)\|^2}_{\sigma^2})$$

the MPE decision rule is

$$R \underset{H_0}{\overset{H_1}{\gtrless}} \tilde{\gamma}$$

with

$$\tilde{\gamma} = \frac{\mu_0 + \mu_1}{2} + \frac{\sigma^2}{\mu_1 - \mu_0} \ln(\frac{\pi_0}{\pi_1}).$$

Probability of Error

- The probability of error, assuming $\pi_0 = \pi_1$, for this decision rule is

$$\Pr\{e\} = Q\left(\frac{\mu_1 - \mu_0}{2\sigma}\right)$$
$$= Q\left(\frac{\langle s_1(t) - s_0(t), g(t)\rangle}{2\sqrt{\frac{N_0}{2}}\|g(t)\|}\right)$$

- **Question:** Which choice of $g(t)$ minimizes the probability of error?

Best Linear Receiver

- The probability of error is minimized when

$$\frac{\langle s_1(t) - s_0(t), g(t)\rangle}{2\sqrt{\frac{N_0}{2}}\|g(t)\|}$$

is maximized with respect to $g(t)$.

- We know from the Schwartz inequality that

$$\langle s_1(t) - s_0(t), g(t)\rangle \leq \|s_1(t) - s_0(t)\| \cdot \|g(t)\|$$

with equality if and only if $g(t) = c \cdot (s_1(t) - s_0(t))$, $c > 0$.

- Hence, to minimize probability of error, choose $g(t) = s_1(t) - s_0(t)$. Then,

$$\Pr\{e\} = Q\left(\frac{\|s_1(t) - s_0(t)\|}{2\sqrt{\frac{N_0}{2}}}\right) = Q\left(\frac{\|s_1(t) - s_0(t)\|}{\sqrt{2N_0}}\right)$$

Exercise: Suboptimum Receiver

- Find the probability of error when equally likely, triangluar signals are used by the transmitter

$$s_0(t) = \begin{cases} \frac{2A}{T} \cdot t & \text{for } 0 \leq t \leq \frac{T}{2} \\ 2A - \frac{2A}{T} \cdot t & \text{for } \frac{T}{2} \leq t \leq T \\ 0 & \text{else} \end{cases} \qquad s_1(t) = -s_0(t)$$

  with $A = \sqrt{\frac{3E}{T}}$ and
- the receiver frontend simply integrates from 0 to T, i.e., $g(t) = 1$, for $0 \leq t \leq T$ and $g(t) = 0$, otherwise.
- **Answer:**

$$\Pr\{e\} = Q\left(\sqrt{\frac{3E}{2N_0}}\right)$$

Introduction

- We have focused on the problem of deciding which of *two* possible signals has been transmitted.
  - Binary Signal Sets
- We will generalize the design of optimum (MPE) receivers to signal sets with M signals.
  - M-ary signal sets.
- With binary signal sets *one* bit can be transmitted in each signal period T.
- With M-ary signal sets, $\log_2(M)$ bits are transmitted simultaneously per T seconds.
  - **Example ($M = 4$)**:

$$\begin{array}{ll} 00 \rightarrow s_0(t) & 01 \rightarrow s_1(t) \\ 10 \rightarrow s_2(t) & 11 \rightarrow s_3(t) \end{array}$$

M-ary Hypothesis Testing Problem

- We can formulate the optimum receiver design problem as a hypothesis testing problem:

$$
\begin{aligned}
H_0 &\colon R_t = s_0(t) + N_t \\
H_1 &\colon R_t = s_1(t) + N_t \\
&\vdots \\
H_{M-1} &\colon R_t = s_{M-1}(t) + N_t
\end{aligned}
$$

  with a priori probabilities $\pi_i = \Pr\{H_i\}$, $i = 0, 1, \ldots, M-1$.
- Note:
  - With more than two hypotheses, it is no longer helpful to consider the (likelihood) ratio of pdfs.
  - Instead, we focus on the hypothesis with the *maximum a posteriori (MAP)* probability or the *maximum likelihood (ML)*.

AWGN Channels

- Of most interest in communications are channels where N_t is a white Gaussian noise process.
  - Spectral height $\frac{N_0}{2}$.
- For these channels, the optimum receivers can be found by arguments completely analogous to those for the binary case.
  - Note that with M-ary signal sets, the subspace containing all signals will have up to M dimensions.
- We will determine the optimum receivers by generalizing the optimum binary receivers for AWGN channels.

Starting Point: Binary MPE Decision Rule

- We have shown, that the binary MPE decision rule can be expressed equivalently as
  - either

$$\langle R_t, (s_1(t) - s_0(t)) \rangle \underset{H_0}{\overset{H_1}{\gtrless}} \frac{N_0}{2} \ln\left(\frac{\pi_0}{\pi_1}\right) + \frac{\|s_1(t)\|^2 - \|s_0(t)\|^2}{2}$$

  - or

$$\|R_t - s_0(t)\|^2 - N_0 \ln(\pi_0) \underset{H_0}{\overset{H_1}{\gtrless}} \|R_t - s_1(t)\|^2 - N_0 \ln(\pi_1)$$

- The first expression is most useful for deriving the structure of the optimum receiver.
- The second form is helpful for interpreting the decision rule in signal space.

M-ary MPE Receiver

- The decision rule

$$\langle R_t, (s_1(t) - s_0(t)) \rangle \underset{H_0}{\overset{H_1}{\gtrless}} \frac{N_0}{2} \ln\left(\frac{\pi_0}{\pi_1}\right) + \frac{\|s_1(t)\|^2 - \|s_0(t)\|^2}{2}$$

can be rewritten as

$$Z_1 = \langle R_t, s_1(t) \rangle + \overbrace{\frac{N_0}{2} \ln(\pi_1) - \frac{\|s_1(t)\|^2}{2}}^{\gamma_1} \underset{H_0}{\overset{H_1}{\gtrless}}$$

$$\langle R_t, s_0(t) \rangle + \underbrace{\frac{N_0}{2} \ln(\pi_0) - \frac{\|s_0(t)\|^2}{2}}_{\gamma_0} = Z_0$$

M-ary MPE Receiver

- The decision rule is easily generalized to M signals:

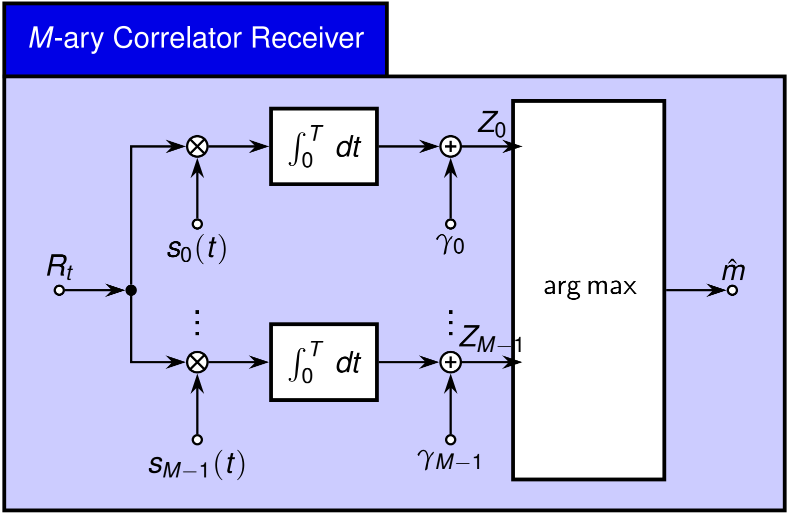
$$\hat{m} = \arg \max_{n=0,\dots,M-1} \overbrace{\langle R_t, s_n(t) \rangle + \underbrace{\frac{N_0}{2} \ln(\pi_n) - \frac{\|s_n(t)\|^2}{2}}_{\gamma_n}}^{Z_n}$$

- The optimum detector selects the hypothesis with the largest decision statistic Z_n.

M-ary MPE Receiver

- The *bias terms* γ_n account for unequal priors and for differences in signal energy $E_n = \|s_n(t)\|^2$.
- Common terms can be omitted
  - For equally likely signals,
$$\gamma_n = -\frac{\|s_n(t)\|^2}{2}.$$
  - For equal energy signals,
$$\gamma_n = \frac{N_0}{2}\ln(\pi_n)$$
  - For equally likely, equal energy signal,
$$\gamma_n = 0$$

M-ary MPE Receiver



Decision Statistics

- The optimum receiver computes the decision statistics

$$Z_n = \langle R_t, s_n(t) \rangle + \frac{N_0}{2} \ln(\pi_n) - \frac{\|s_n(t)\|^2}{2}.$$

- Conditioned on the m-th signal having been transmitted,
  - All Z_n are Gaussian random variables.
  - Expected value:

$$\mathbf{E}[Z_n | H_m] = \langle s_m(t), s_n(t) \rangle + \frac{N_0}{2} \ln(\pi_n) - \frac{\|s_n(t)\|^2}{2}$$

  - (Co)Variance:

$$\mathbf{E}[Z_j Z_k | H_m] - \mathbf{E}[Z_j | H_m] \mathbf{E}[Z_k | H_m] = \langle s_j(t), s_k(t) \rangle \frac{N_0}{2}$$

Exercise: QPSK Receiver

- Find the optimum receiver for the following signal set with $M = 4$ signals:

$$s_n(t) = \sqrt{\frac{2E}{T}} \cos(2\pi t/T + n\pi/2) \quad \text{for } 0 \leq t \leq T \text{ and } n = 0, \ldots,$$

Decision Regions

- The decision regions Γ_n and error probabilities are best understood by generalizing the binary decision rule:

$$\|R_t - s_0(t)\|^2 - N_0 \ln(\pi_0) \underset{H_0}{\overset{H_1}{\gtrless}} \|R_t - s_1(t)\|^2 - N_0 \ln(\pi_1)$$

- For *M*-ary signal sets, the decision rule generalizes to

$$\hat{m} = \arg \min_{n=0,\ldots,M-1} \|R_t - s_n(t)\|^2 - N_0 \ln(\pi_n).$$

  - This simplifies to

$$\hat{m} = \arg \min_{n=0,\ldots,M-1} \|R_t - s_n(t)\|^2$$

    for equally likely signals.
  - The optimum receiver decides in favor of the signal $s_n(t)$ that is *closest* to the received signal.

Decision Regions (equally likely signals)

- For discussing decision regions, it is best to express the decision rule in terms of the representation obtained with the orthonormal basis $\{\Phi_k\}$, where
  - basis signals Φ_k span the space that contains all signals $s_n(t)$, with $n = 0, \ldots, M-1$.
    - Recall that we can obtain these basis signals via the Gram-Schmidt procedure from the signal set.
    - There are at most M orthonormal bases.
- Because of Parseval's relationship, an equivalent decision rule is

$$\hat{m} = \arg \min_{n=0,\ldots,M-1} \|\vec{R} - \vec{s}_n\|^2,$$

where $\vec{R}$ has elements $R_k = \langle R_t, \Phi_k(t) \rangle$ and $\vec{s}_n$ has element $s_{n,k} = \langle s_n(t), \Phi_k(t) \rangle$.

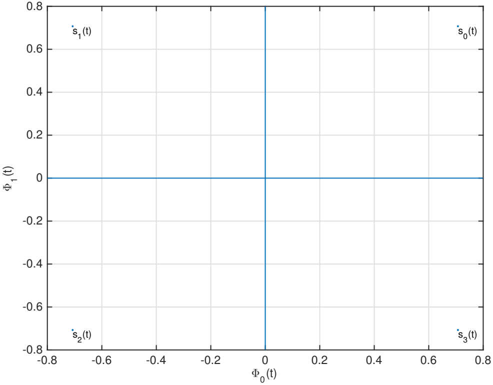
Decision Regions

- The decision region Γ_n where the detector decides that the n-th signal was sent is

$$\Gamma_n = \{\vec{r} : \|\vec{r} - \vec{s}_n\| < \|\vec{r} - \vec{s}_m\| \text{for all } m \neq n\}.$$

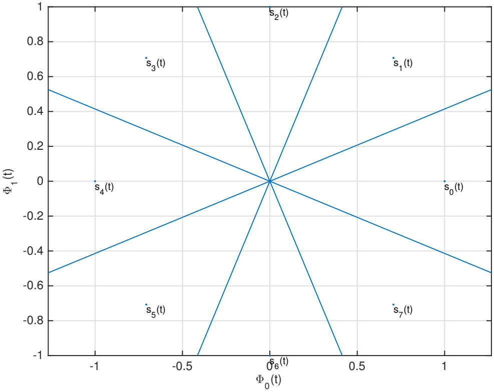
  - The decision region Γ_n is the set of all points $\vec{r}$ that are closer to $\vec{s}_n$ than to any other signal point.
- The decision regions are formed by linear segments that are *perpendicular bisectors* between pairs of signal points.
  - The resulting partition is also called a *Voronoi partition.*

Example: QPSK



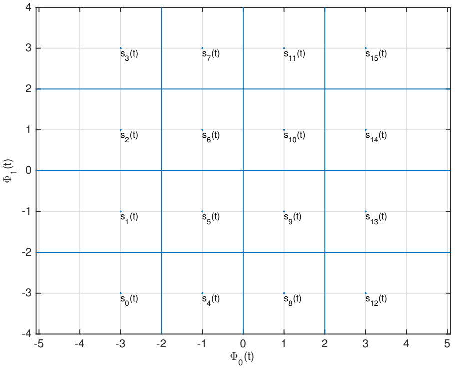
$$s_n(t) = \sqrt{2/T}\cos(2\pi f_c t + n \cdot \pi/2 + \pi/4), \text{ for } n = 0, \ldots, 3.$$

Example: 8-PSK



$$s_n(t) = \sqrt{2/T} \cos(2\pi f_c t + n \cdot \pi/4), \text{ for } n = 0, \ldots, 7.$$

Example: 16-QAM



$$s_n(t) = \sqrt{2/T}(A_I \cdot \cos(2\pi f_c t) + A_Q \cdot \sin(2\pi f_c t))$$

$$\text{with } A_I, A_Q \in \{-3, -1, 1, 3\}.$$

Symbol Energy and Bit Energy

- We have seen that error probabilities decrease when the signal energy increases.
  - Because the distance between signals increase.
- We will see further that error rates in AWGN channels depend only on
  - the signal-to-noise ratio $\frac{E_b}{N_0}$, where E_b is the average energy per bit, and
  - the geometry of the signal constellation.
- To focus on the impact of the signal geometry, we will fix either
  - the average energy per symbol $E_s = \frac{1}{M} \sum_{n=0}^{M-1} \|s_n(t)\|^2$ or
  - the average energy per bit $E_b = \frac{E_s}{\log_2(M)}$

Example: QPSK

- QPSK signals are given by

$$s_n(t) = \sqrt{\frac{2E_s}{T}} \cos(2\pi f_c t + n \cdot \pi/2 + \pi/4), \text{ for } n = 0, \ldots, 3.$$

- Each of the four signals $s_n(t)$ has energy

$$E_n = \|s_n(t)\|^2 = E_s.$$

- Hence,
  - the average symbol energy is E_s
  - the average bit energy is $E_b = \frac{E_s}{\log_2(4)} = \frac{E_s}{2}$

Example: 8-PSK

- 8-PSK signals are given by

$$s_n(t) = \sqrt{2E_s/T}\cos(2\pi f_c t + n \cdot \pi/4), \text{ for } n = 0, \ldots, 7.$$

- Each of the eight signals $s_n(t)$ has energy

$$E_n = \|s_n(t)\|^2 = E_s.$$

- Hence,
  - the average symbol energy is E_s
  - the average bit energy is $E_b = \frac{E_s}{\log_2(8)} = \frac{E_s}{3}$

Example: 16-QAM

- 16-QAM signals can be written as

$$s_n(t) = \sqrt{\frac{2E_0}{T}} \left(a_I \cdot \cos(2\pi f_c t) + a_Q \cdot \sin(2\pi f_c t) \right)$$

  with $a_I, a_Q \in \{-3, -1, 1, 3\}$.
- There are
  - 4 signals with energy $(1^2 + 1^2)E_0 = 2E_0$
  - 8 signals with energy $(3^2 + 1^2)E_0 = 10E_0$
  - 4 signals with energy $((3^2 + 3^2)E_0 = 18E_0$
- Hence,
  - the average symbol energy is $10E_0$
  - the average bit energy is $E_b = \frac{E_s}{\log_2(16)} = \frac{5E_0}{2}$

Energy Efficiency

- We will see that the influence of the signal geometry is captured by the energy efficiency

$$\eta_P = \frac{d_{\min}^2}{E_b}$$

  where $d_{\min}$ is the smallest distance between any pair of signals in the constellation.
- Examples:
  - **QPSK:** $d_{\min} = \sqrt{2E_s}$ and $E_b = \frac{E_s}{2}$, thus $\eta_P = 4$.
  - **8-PSK:** $d_{\min} = \sqrt{(2-\sqrt{2})E_s}$ and $E_b = \frac{E_s}{3}$, thus $\eta_P = 3 \cdot (2-\sqrt{2}) \approx 1.75$.
  - **16-QAM:** $d_{\min} = 2\sqrt{E_0}$ and $E_b = \frac{5E_0}{2}$, thus $\eta_P = \frac{8}{5}$.
- Note that energy efficiency decreases with the size of the constellation for 2-dimensional constellations.

Computing Probability of Symbol Error

- When decision boundaries intersect at right angles, then it is possible to compute the error probability exactly in closed form.
  - The result will be in terms of the Q-function.
  - This happens whenever the signal points form a rectangular grid in signal space.
  - Examples: QPSK and 16-QAM
- When decision regions are not rectangular, then closed form expressions are not available.
  - Computation requires integrals over the Q-function.
  - We will derive good bounds on the error rate for these cases.
  - For exact results, numerical integration is required.

Illustration: 2-dimensional Rectangle

- Assume that the *n*-th signal was transmitted and that the representation for this signal is $\vec{s}_n = (s_{n,0}, s_{n,1})'$.
- Assume that the decision region Γ_n is a rectangle

$$\begin{aligned}\Gamma_n = \{\vec{r} = (r_0, r_1)' : & s_{n,0} - a_1 < r_0 < s_{n,0} + a_2 \text{ and} \\ & s_{n,1} - b_1 < r_1 < s_{n,1} + b_2\}.\end{aligned}$$

  - Note: we have assumed that the sides of the rectangle are parallel to the axes in signal space.
  - Since *rotation* and *translation* of signal space do not affect distances this can be done without affecting the error probability.
- **Question:** What is the conditional error probability, assuming that $s_n(t)$ was sent.

Illustration: 2-dimensional Rectangle

- In terms of the random variables $R_k = \langle R_t, \Phi_k \rangle$, with $k = 0, 1$, an error occurs if

$$\overbrace{(R_0 \leq s_{n,0} - a_1 \text{ or } R_0 \geq s_{n,0} + a_2)}^{\text{error event 1}} \text{ or}$$
$$\underbrace{(R_1 \leq s_{n,1} - b_1 \text{ or } R_1 \geq s_{n,1} + b_2)}_{\text{error event 2}}.$$

- Note that the two error events are not mutually exclusive.
- Therefore, it is better to consider correct decisions instead, i.e., $\vec{R} \in \Gamma_n$:

$$s_{n,0} - a_1 < R_0 < s_{n,0} + a_2 \textbf{ and } s_{n,1} - b_1 < R_1 < s_{n,1} + b_2$$

Illustration: 2-dimensional Rectangle

- We know that R_0 and R_1 are
  - independent - because Φ_k are orthogonal
  - with means $s_{n,0}$ and $s_{n,1}$, respectively
  - variance $\frac{N_0}{2}$.
- Hence, the probability of a correct decision is

$$
\begin{aligned}
\Pr\{c|s_n\} &= \Pr\{-a_1 < N_0 < a_2\} \cdot \Pr\{-b_1 < N_1 < b_2\} \\
&= \int_{-a_1}^{a_2} p_{R_0|s_n}(r_0)\, dr_0 \cdot \int_{-b_1}^{b_2} p_{R_1|s_n}(r_1)\, dr_1 \\
&= \left(1 - Q\left(\frac{a_1}{\sqrt{N_0/2}}\right) - Q\left(\frac{a_2}{\sqrt{N_0/2}}\right)\right) \cdot \\
&\quad \left(1 - Q\left(\frac{b_1}{\sqrt{N_0/2}}\right) - Q\left(\frac{b_2}{\sqrt{N_0/2}}\right)\right).
\end{aligned}
$$

Exercise: QPSK

- Find the error rate for the signal set

$$s_n(t) = \sqrt{2E_s/T}\cos(2\pi f_c t + n \cdot \pi/2 + \pi/4), \text{ for } n = 0, \ldots, 3.$$

- **Answer:** (Recall $\eta_P = \frac{d^2_{\min}}{E_b} = 4$ for QPSK)

$$\begin{aligned}
\Pr\{e\} &= 2Q\left(\sqrt{\frac{E_s}{N_0}}\right) - Q^2\left(\sqrt{\frac{E_s}{N_0}}\right) \\
&= 2Q\left(\sqrt{\frac{2E_b}{N_0}}\right) - Q^2\left(\sqrt{\frac{2E_b}{N_0}}\right) \\
&= 2Q\left(\sqrt{\frac{\eta_P E_b}{2N_0}}\right) - Q^2\left(\sqrt{\frac{\eta_P E_b}{2N_0}}\right).
\end{aligned}$$

Exercise: 16-QAM

(Recall $\eta_P = \frac{d_{\text{min}}^2}{E_b} = \frac{8}{5}$ for 16-QAM)

- Find the error rate for the signal set ($a_I, a_Q \in \{-3, -1, 1, 3\}$)

$$s_n(t) = \sqrt{2E_0/T}a_I \cdot \cos(2\pi f_c t) + \sqrt{2E_0/T}a_Q \cdot \sin(2\pi f_c t)$$

- **Answer:** ($\eta_P = \frac{d_{\text{min}}^2}{E_b} = 4$)

$$\begin{aligned}
\Pr\{e\} &= 3Q\left(\sqrt{\frac{2E_0}{N_0}}\right) - \frac{9}{4}Q^2\left(\sqrt{\frac{2E_0}{N_0}}\right) \\
&= 3Q\left(\sqrt{\frac{4E_b}{5N_0}}\right) - \frac{9}{4}Q^2\left(\sqrt{\frac{4E_b}{5N_0}}\right) \\
&= 3Q\left(\sqrt{\frac{\eta_P E_b}{2N_0}}\right) - \frac{9}{4}Q^2\left(\sqrt{\frac{\eta_P E_b}{2N_0}}\right).
\end{aligned}$$

N-dimensional Hypercube

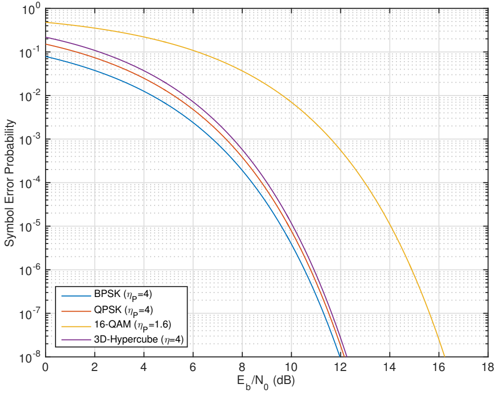
- Find the error rate for the signal set with 2^N signals of the form ($b_{k,n} \in \{-1, 1\}$):

$$s_n(t) = \sum_{k=1}^{N} \sqrt{\frac{2E_s}{NT}} b_{k,n} \cos(2\pi nt/T), \text{ for } 0 \leq t \leq T$$

- **Answer:**

$$\begin{aligned}
\Pr\{e\} &= 1 - \left(1 - Q\left(\sqrt{\frac{2E_s}{N \cdot N_0}}\right)\right)^N \\
&= 1 - \left(1 - Q\left(\sqrt{\frac{2E_b}{N_0}}\right)\right)^N \\
&= 1 - \left(1 - Q\left(\sqrt{\frac{\eta_P E_b}{2N_0}}\right)\right)^N \approx N \cdot Q\left(\sqrt{\frac{\eta_P E_b}{2N_0}}\right).
\end{aligned}$$

Comparison



- Better power efficiency η_P leads to better error performance (at high SNR).

What if Decision Regions are not Rectangular?

- **Example:** For 8-PSK, the probability of a correct decision is given by the following integral over the decision region for $s_0(t)$

$$\Pr\{c\} = \int_0^\infty \frac{1}{\sqrt{2\pi N_0/2}} \exp\left(-\frac{(x-\sqrt{E_s})^2}{2N_o/2}\right) \underbrace{\int_{-x\tan(\pi/8)}^{x\tan(\pi/8)} \frac{1}{\sqrt{2\pi N_0/2}} \exp\left(-\frac{y^2}{2N_0/2}\right) dy}_{=1-2Q\left(\frac{x\tan(\pi/8)}{\sqrt{N_0/2}}\right)} dx$$

- This integral cannot be computed in closed form.

Union Bound

- When decision boundaries do not intersect at right angles, then the error probability cannot be computed in closed form.
- An upper bound on the conditional probability of error (assuming that s_n was sent) is provided by:

$$\begin{aligned}\Pr\{e|s_n\} &\leq \sum_{k \neq n} \Pr\{\|\vec{R} - \vec{s}_k\| < \|\vec{R} - \vec{s}_n\| | \vec{s}_n\} \\ &= \sum_{k \neq n} Q\left(\frac{\|\vec{s}_k - \vec{s}_n\|}{2\sqrt{N_0/2}}\right).\end{aligned}$$

  - Note that this bound is computed from *pairwise error probabilities* between s_n and all other signals.

Union Bound

- Then, the average probability of error can be bounded by

$$\Pr\{e\} = \sum_n \pi_n \sum_{k \neq n} Q\left(\frac{\|\vec{s}_k - \vec{s}_n\|}{\sqrt{2N_0}}\right).$$

- This bound is called the union bound; it approximates the union of all possible error events by the sum of the pairwise error probabilities.

Example: QPSK

- For the QPSK signal set

$$s_n(t) = \sqrt{2E_s/T}\cos(2\pi f_c t + n \cdot \pi/2 + \pi/4), \text{ for } n = 0, \ldots, 3$$

the union bound is

$$\Pr\{e\} \leq 2Q\left(\sqrt{\frac{E_s}{N_0}}\right) + Q\left(\sqrt{\frac{2E_s}{N_0}}\right).$$

  - Recall that the exact probability of error is

$$\Pr\{e\} = 2Q\left(\sqrt{\frac{E_s}{N_0}}\right) - Q^2\left(\sqrt{\frac{E_s}{N_0}}\right).$$

"Intelligent" Union Bound

- The union bound is easily tightened by recognizing that only immediate neighbors of s_n must be included in the bound on the conditional error probability.
- Define the the neighbor set $N_{ML}(s_n)$ of s_n as the set of signals s_k that share a decision boundary with signal s_n.
- Then, the conditional error probability is bounded by

$$
\begin{aligned}
\Pr\{e|s_n\} &\leq \sum_{k \in N_{ML}(s_n)} \Pr\{\|\vec{R} - \vec{s}_k\| < \|\vec{R} - \vec{s}_n\| | \vec{s}_n\} \\
&= \sum_{k \in N_{ML}(s_n)} Q\left(\frac{\|\vec{s}_k - \vec{s}_n\|}{2\sqrt{N_0/2}}\right).
\end{aligned}
$$

"Intelligent" Union Bound

- Then, the average probability of error can be bounded by

$$\Pr\{e\} \leq \sum_n \pi_n \sum_{k \in N_{ML}(s_n)} Q\left(\frac{\|\vec{s}_k - \vec{s}_n\|}{\sqrt{2N_0}}\right).$$

- We refer to this bound as the intelligent union bound.
  - It still relies on pairwise error probabilities.
  - It excludes many terms in the union bound; thus, it is tighter.

Example: QPSK

- For the QPSK signal set

$$s_n(t) = \sqrt{2E_s/T}\cos(2\pi f_c t + n \cdot \pi/2 + \pi/4), \text{ for } n = 0, \ldots, 3$$

the intelligent union bound includes only the immediate neighbors of each signal:

$$\Pr\{e\} \leq 2Q\left(\sqrt{\frac{E_s}{N_0}}\right).$$

  - Recall that the exact probability of error is

$$\Pr\{e\} = 2Q\left(\sqrt{\frac{E_s}{N_0}}\right) - Q^2\left(\sqrt{\frac{E_s}{N_0}}\right).$$

Example: 16-QAM

- For the 16-QAM signal set, there are
  - 4 signals s_i that share a decision boundary with 4 neighbors; bound on conditional error probability: $\Pr\{e|s_i\} = 4Q(\sqrt{\frac{2E_0}{N_0}})$.
  - 8 signals s_c that share a decision boundary with 3 neighbors; bound on conditional error probability: $\Pr\{e|s_c\} = 3Q(\sqrt{\frac{2E_0}{N_0}})$.
  - 4 signals s_o that share a decision boundary with 2 neighbors; bound on conditional error probability: $\Pr\{e|s_o\} = 2Q(\sqrt{\frac{2E_0}{N_0}})$.
- The resulting intelligent union bound is

$$\Pr\{e\} \leq 3Q\left(\sqrt{\frac{2E_0}{N_0}}\right) = 3Q\left(\sqrt{\frac{4E_b}{5N_0}}\right).$$

Example: 16-QAM

- The resulting intelligent union bound is

$$\Pr\{e\} \leq 3Q\left(\sqrt{\frac{4E_b}{5N_0}}\right).$$

- Recall that the exact probability of error is

$$\Pr\{e\} = 3Q\left(\sqrt{\frac{4E_b}{5N_0}}\right) - \frac{9}{4}Q^2\left(\sqrt{\frac{4E_b}{5N_0}}\right).$$

Nearest Neighbor Approximation

- At high SNR, the error probability is dominated by terms that involve the shortest distance $d_{\min}$ between any pair of nodes.
  - The corresponding error probability is proportional to $Q(\sqrt{\frac{d_{\min}}{2N_0}})$.
- For each signal s_n, we count the number N_n of neighbors at distance $d_{\min}$.
- Then, the error probability at high SNR can be approximated as

$$\Pr\{e\} \approx \frac{1}{M} \sum_{n=0}^{M-1} N_n Q(\sqrt{\frac{d_{\min}^2}{2N_0}}) = \bar{N}_{\min} Q(\sqrt{\frac{d_{\min}^2}{2N_0}}).$$

Example: 16-QAM

- In 16-QAM, the distance between adjacent signals is $d_{\min} = 2\sqrt{E_0}$; also, $E_b = \frac{5}{2}E_0$.
- There are:
  - 4 signals with 4 nearest neighbors
  - 8 signals with 3 nearest neighbors
  - 4 signals with 2 nearest neighbors
- The average number of neighbors is $\bar{N}_{\min} = 3$.
- The error probability is approximately,

$$\Pr\{e\} \approx 3Q\left(\sqrt{\frac{2E_0}{N_0}}\right) = 3Q\left(\sqrt{\frac{4E_b}{5N_0}}\right).$$

  - same as the intelligent union bound.

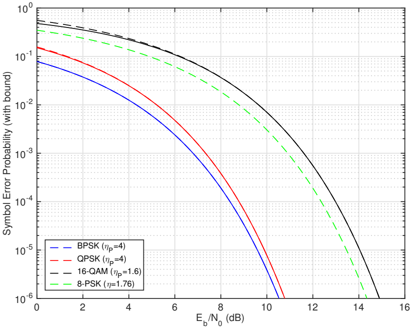
Example: 8-PSK

- For 8-PSK, each signal has 2 nearest neighbors at distance $d_{\text{min}} = \sqrt{(2-\sqrt{2})E_s}$; also, $E_b = \frac{E_s}{3}$.
- Hence, both the intelligent union bound and the nearest neighbor approximation yield

$$\Pr\{e\} \approx 2Q\left(\sqrt{\frac{(2-\sqrt{2})E_s}{2N_0}}\right) = 2Q\left(\sqrt{\frac{3(2-\sqrt{2})E_b}{2N_0}}\right)$$

  - Since, $E_b = 3E_s$.

Comparison



Solid: exact P_e, dashed: approximation. For 8PSK, only approximation is shown.

- The intelligent union bound is very tight for all cases considered here.
  - It also coincides with the nearest neighbor approximation

General Approximation for Probability of Symbol Error

- From the above examples, we can conclude that a good, general approximation for the probability of error is given by

$$\Pr\{e\} \approx \bar{N}_{\min} Q\left(\frac{d_{\min}}{\sqrt{2N_0}}\right) = \bar{N}_{\min} Q\left(\sqrt{\frac{\eta_P E_b}{2N_0}}\right).$$

- Probability of error depends on
  - signal-to-noise ratio (SNR) E_b/N_0 and
  - geometry of the signal constellation via the average number of neighbors $\bar{N}_{\min}$ and the power efficiency η_P.

Bit Errors

- So far, we have focused on symbol errors; however, ultimately we are concerned about bit errors.
- There are many ways to map groups of $\log_2(M)$ bits to the M signals in a constellation.
- **Example QPSK:** Which mapping is better?

QPSK Phase	Mapping 1	Mapping 2
$\pi/4$	00	00
$3\pi/4$	01	01
$5\pi/4$	10	11
$7\pi/4$	11	10

Bit Errors

- **Example QPSK:**

QPSK Phase	Mapping 1	Mapping 2
$\pi/4$	00	00
$3\pi/4$	01	01
$5\pi/4$	10	11
$7\pi/4$	11	10

- Note, that for Mapping 2 *nearest neighbors* differ in exactly one bit position.
  - That implies, that the most common symbol errors will induce only one bit error.
  - That is not true for Mapping 1.

Gray Coding

- A mapping of $log_2(M)$ bits to M signals is called Gray Coding if
  - The bit patterns that are assigned to nearest neighbors in the constelation
  - differ in exactly one bit position.
- With Gray coding, the most likely symbol errors induce exactly one bit error.
  - Note that there are $\log_2(M)$ bits for each symbol.
- Hence, with Gray coding the *bit error probability* is well approximated by

$$\Pr\{\text{bit error}\} \approx \frac{\bar{N}_{\min}}{\log_2(M)} Q\left(\sqrt{\frac{\eta_P E_b}{2N_0}}\right) \lesssim Q\left(\frac{d_{\min}}{\sqrt{2N_0}}\right).$$

Introduction

- We compare methods for transmitting a sequence of bits.
- We will see that the performance of these methods varies significantly.
- **New perspective:**
  - Focus on messages, i.e., sequences of bits
  - Entire message must be received correctly
- Main Result: It is possible to achieve error free communications as long as SNR is good enough and data rate is not too high.

Problem Statement

- **Problem:**
  - K bits must be transmitted in T seconds.
  - Available power is limited to P.
- **Questions:**
  - What method achieves the lowest probability of error?
  - Is error-free communications possible?

Parameters

- **Data Rate:**
$$R = \frac{K}{T} \quad \text{(bits/s)}$$
  - entire transmission takes T seconds
  - K bits are sent over T seconds
  - implicit assumption: bits are equally likely.
- **Power and energy:** transmitted signal $s(t)$ has power P and energy E
$$P = \frac{1}{T} \int_0^T |s(t)|^2 \, dt = \frac{E}{T}$$
  - Entire transmitted signal $s(t)$ is of duration T.
  - Note, bit energy is given by
$$E_b = \frac{E}{K} = \frac{PT}{K} = \frac{P}{R}.$$

Bit-by-bit Signaling

- Transmit K bit as a sequence of "one-shot" BPSK signals.
- $K = RT$ bits to be transmitted.
- Energy per bit E_b ($E_b = \frac{E}{K}$).
- Consider, signals of the form

$$s(t) = \sum_{k=0}^{K-1} \sqrt{E_b} s_k p(t - k/R)$$

  - $s_k \in \{\pm 1\}$
  - $p(t)$ is a pulse of duration $1/R = T/K$ and $\|p(t)\|^2 = 1$.
- **Question:** What is the probability that any transmission error occurs?
  - In other words, the transmission is not received without error.

Error Probability for Bit-by-Bit Signaling

- We can consider the entire message as a single K-dimensional signal set.
  - Signals are at the vertices of a K-dimensional hypercube.

$$\Pr\{e\} = 1 - \left(1 - Q\left(\frac{2E_b}{N_0}\right)\right)^K$$
$$= 1 - \left(1 - Q\left(\frac{2P}{RN_0}\right)\right)^{RT}$$

- Note, for any finite P/N_0 and R, the error rate will always tend to 1 as $T \rightarrowtail \infty$.
  - Error-free transmission is *not* possible with bit-by-bit signaling.

Block-Orthogonal Signaling

- Again,
  - $K = RT$ bits are transmitted in T seconds.
  - Energy per bit $E_b = \frac{P}{R}$.
- Signal set (Pulse-position modulation — PPM)

$$s_k(t) = \sqrt{E}p(t - kT/2^K) \quad \text{for } k = 0, 1, \ldots, 2^K - 1.$$

  where $p(t)$ is of duration $T/2^K$, $E = KE_b$, and $\|p(t)\|^2 = 1$.
- Alternative signal set (Frequency Shift Keying — FSK)

$$s_k(t) = \sqrt{\frac{2E}{T}} \cos(2\pi(f_c + k/T)t) \quad \text{for } k = 0, 1, \ldots, 2^K - 1.$$

- Signal set consists of $M = 2^K$ signals
  - each signal conveys K bits,
  - each signal occupies one of the K dimensions.

Union Bound

- The error probability for block-orthogonal signaling cannot be computed in closed form.
- At high and moderate SNR, the error probability is well approximated by the union bound.
  - Each signal has $M - 1 = 2^K - 1$ nearest neighbors.
  - The distance between neighbors is $d_{\min} = \sqrt{2E} = \sqrt{2KE_b}$.
- Union bound

$$
\begin{aligned}
\Pr\{e\} &\leq (2^K - 1) Q\left(\sqrt{\frac{KE_b}{N_0}}\right) \\
&= (2^{RT} - 1) Q\left(\sqrt{\frac{PT}{N_0}}\right)
\end{aligned}
$$

Bounding the Union Bound

- To gain further insight, we bound

$$Q(x) \leq \frac{1}{2}\exp(-x^2/2) \leq \exp(-x^2/2).$$

- Then,

$$\begin{aligned}\Pr\{e\} &\leq (2^{RT}-1)Q\left(\sqrt{\frac{PT}{N_0}}\right) \\ &\lesssim 2^{RT}\exp(-\frac{PT}{2N_0}) \\ &= \exp(-T(\frac{P}{2N_0} - R\ln 2)).\end{aligned}$$

- Hence, $\Pr\{e\} \to 0$ as $T \to \infty$!
  - As long as $R < \frac{1}{\ln 2}\frac{P}{2N_0}$.
- Error-free transmission is possible!

Reality-Check: Bandwidth

- **Bit-by-bit Signaling:** Pulse-width: $T/K = 1/R$.
  - Bandwidth is approximately equal to $B = R$.
  - Also, number of dimensions $K = RT$.
- **Block-orthogonal:** Pulse width: $T/2^K = T/2^{RT}$.
  - Bandwidth is approximately equal to $B = 2^{RT}/T$.
  - Number of dimensions is $2^K = 2^{RT}$.
- Bandwidth for block-orthogonal signaling grows exponentially with the number of bits K.
  - Not practical for moderate to large blocks of bits.

The Dimensionality Theorem

- The relationship between bandwidth B and the number of dimensions is summarized by the *dimensionality theorem:*
  - The number of dimensions D available over an interval od duration T is limited by the bandwidth B

$$D \leq B \cdot T$$

  - The theorem implies:
    - A signal occupying D dimensions over T seconds requires bandwidth

$$B \geq \frac{D}{T}$$

An Ideal Signal Set

- An ideal signal set combines the aspects of our two example signal sets:
  - $\Pr\{e\}$-behavior like block orthogonal signaling

$$\lim_{T\to\infty} \Pr\{e\} = 0.$$

  - Bandwidth behavior like bit-by-bit signaling

$$B = \frac{D}{T} = \text{constant}.$$

    - Thus, $D = BT \to \infty$ as $T \to \infty$.
- **Question:** Does such a signal set exist?

Towards Channel Capacity

- **Given:**
  - bandwidth $B = \frac{D}{T}$, where T is the duration of the transmission.
  - power P
  - Noise power spectral density $\frac{N_0}{2}$
- **Question:** What is the highest data rate R that allows error-free transmission with the above constraints?
  - We are transmitting RT bits
  - Therefore, we need $M = 2^{RT}$ signals.

Signal Set

- Our signal set consists of $M = 2^{RT}$ signals of the form

$$s_n(t) = \sum_{k=0}^{D-1} X_{n,k} p(t - kT/D)$$

  where
  - $p(t)$ are pulses of duration T/D, i.e., of bandwidth $B = D/T$.
  - Also, $\|p(t)\|^2 = 1$.
- Each signal $s_n(t)$ is defined by a length-D vector $\vec{X}_n = \{X\}_{n,k}$.
- We are looking to find $M = 2^{RT}$ length-D vectors $\vec{X}$ that lead to good error properties.
- Note that the signals $p(t - kT/D)$ form an orthonormal basis with D dimensions.

Receiver Frontend

- The receiver frontend consists of a matched filter for $p(t)$ followed by a sampler at times kT/D.
  - I.e., the frontend projects the received signal onto the orthonormal basis functions $p(t - kT/D)$.
- The vector $\vec{R}$ of matched filter outputs has elements

$$R_k = \langle R_t, p(t - kT/D) \rangle \quad k = 0, 1, \ldots, D-1$$

- Conditional on $s_n(t)$ was sent, $\vec{R} \sim N(\vec{X}_n, \frac{N_o}{2} I)$.
- The optimum receiver selects the signal s_n that's closest to $\vec{R}$.

Conditional Error Probability

- When, the signal $s_n(t)$ was sent then $\vec{R} \sim N(\vec{X}_n, \frac{N_o}{2}I)$.
- As the number of dimensions D increases, the vector $\vec{R}$ lies within a D-dimensional sphere with center $\vec{X}_k$ and radius $\sqrt{D\frac{N_0}{2}}$ with very high probability: $1 - e^{-D}$, i.e., $P_e = e^{-D}$.
  - **Important:** We allow the radius of the decoding spheres to grow with the number of dimensions D.
  - This ensures that $P_e \to 0$ as $D = BT \to \infty$.
- We call the spheres of radius $\sqrt{D\frac{N_0}{2}}$ around each signal point *decoding spheres*.
  - The decoding spheres will be part of the decision regions for each point.

Power Constraint

- The power for signal $s_n(t)$ must satisfy

$$\frac{1}{T}\int_0^T s_n^2(t)\,dt = \frac{1}{T}\sum_{k=0}^{D-1}|X_{n,k}|^2 = \frac{1}{T}\|\vec{X}_n\|^2 \leq P.$$

- Therefore, $\|\vec{X}_n\|^2 \leq PT$
- Insights:
  - The transmitted signals lie in a sphere of radius $\sqrt{PT}$.
  - The observed signals must lie in a large sphere of radius $\sqrt{PT + D\frac{N_0}{2}}$.
- **Question:** How many decoding spheres can we have and still meet the power constraint?

Capacity

- Each decoding sphere has volume $K_D(\sqrt{D\frac{N_0}{2}})^D$.
- The volume of the sphere containing the observed signals is $K_D(\sqrt{PT + D\frac{N_0}{2}})^D$
  - K_D is a constant that depends only on the number of dimensions D, e.g., $K_3 = \frac{4\pi}{3}$.
- The number of decoding spheres that fit into the the power sphere is (upper) bounded by the ratio of the volumes

$$\frac{K_D\left(\sqrt{PT + D\frac{N_0}{2}}\right)^D}{K_D\left(\sqrt{D\frac{N_0}{2}}\right)^D}$$

Capacity

- Since the number of signals $M = 2^{RT}$ equals the number of decoding spheres, it follows that error free communications is possible (in the limit as $D = BT \to \infty$) if

$$M = 2^{RT} < \frac{\left(\sqrt{PT + D\frac{N_0}{2}}\right)^D}{\left(\sqrt{D\frac{N_0}{2}}\right)^D}$$

or

$$R < \frac{D}{2T}\log_2(1 + \frac{PT}{DN_0/2}) = \frac{B}{2}\log_2(1 + \frac{P}{BN_0/2}).$$

  - Note, if we allow *complex valued* signals, then $R < B\log_2(1 + \frac{P}{BN_0})$.

Illustration: 2-bit Messages

- Consider two different ways of transmitting two bits:
  - QPSK
  - rate $2/3$ block code and BPSK modulation
- Compare the probability of at least one bit error
  - constant $\frac{E_b}{N_0}$.

QPSK

- We know that for QPSK
  - energy efficiency $\eta_u = 4$
  - (symbol) error rate

$$P_e \leq 2Q\left(\sqrt{\frac{2E_b}{N_0}}\right)$$

Benefit of a Simple Code

- The block code maps two bits to sequence of three BPSK symbols as follows:

$$00 : \{1, 1, 1\} \qquad 01 : \{1, -1, -1\}$$
$$10 : \{-1, 1, -1\} \qquad 11 : \{-1, -1, 1\}$$

- For this signal set:
  - energy efficiency $\eta_c = \frac{16}{3}$
  - (symbol) error rate

$$P_e \leq 3\mathrm{Q}\left(\sqrt{\frac{8E_b}{3N_0}}\right)$$

- Coding gain:

$$\frac{\eta_c}{\eta_u} = \frac{16/3}{4} = \frac{4}{3} \approx 1\text{dB}$$

Part IV

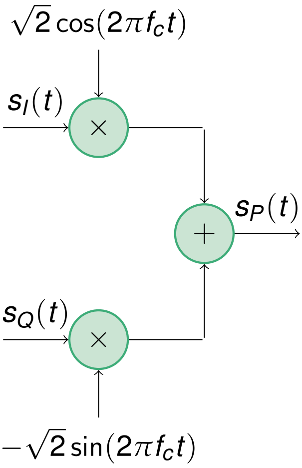
Complex Envelope and Linear Modulation

Passband Signals

- We have seen that many signal sets include both $\sin(2\pi f_c t)$ and $\cos(2\pi f_c t)$.
  - Examples include PSK and QAM signal sets.
- Such signals are referred to as passband signals.
  - Passband signals have frequency spectra concentrated around a carrier frequency f_c.
  - This is in contrast to baseband signals with spectrum centered at zero frequency.
- Baseband signals can be converted to passband signals through up-conversion.
- Passband signals can be converted to baseband signals through down-conversion.

© 2018, B.-P. Paris

Up-Conversion



- The passband signal $s_P(t)$ is constructed from two (digitally modulated) baseband signals, $s_I(t)$ and $s_Q(t)$.
  - Note that two signals can be carried simultaneously!
    - $s_I(t)$ and $s_Q(t)$ are the in-phase (I) and quadrature (Q) compenents of $s_p(t)$.
  - This is a consequence of $s_I(t)\cos(2\pi f_c t)$ and $s_Q(t)\sin(2\pi f_c t)$ being orthogonal
    - when the carrier frequency f_c is much greater than the bandwidth of $s_I(t)$ and $s_Q(t)$.

© 2018, B.-P. Paris

Exercise: Orthogonality of In-phase and Quadrature Signals

- Show that $s_I(t) \cos(2\pi f_c t)$ and $s_Q(t) \sin(2\pi f_c t)$ are orthogonal when $f_c \gg B$, where B is the bandwidth of $s_I(t)$ and $s_Q(t)$.
  - You can make your argument either in the time-domain or the frequency domain.

Baseband Equivalent Signals

- The passband signal $s_P(t)$ can be written as

$$s_P(t) = \sqrt{2} s_I(t) \cdot \cos(2\pi f_c t) - \sqrt{2} s_Q(t) \cdot \sin(2\pi f_c t).$$

- If we define $s(t) = s_I(t) + j \cdot s_Q(t)$, then $s_P(t)$ can also be expressed as

$$\begin{aligned} s_P(t) &= \sqrt{2} \cdot \Re\{s(t)\} \cdot \cos(2\pi f_c t) - \sqrt{2} \cdot \Im\{s(t)\} \cdot \sin(2\pi f_c t) \\ &= \sqrt{2} \cdot \Re\{s(t) \cdot \exp(j2\pi f_c t)\}. \end{aligned}$$

- The signal $s(t)$:
  - is called the baseband equivalent, or the complex envelope of the passband signal $s_P(t)$.
  - It contains the same information as $s_P(t)$.
  - Note that $s(t)$ is *complex-valued*.

Polar Representation

- Sometimes it is useful to express the complex envelope $s(t)$ in polar coordinates:

$$\begin{aligned} s(t) &= s_I(t) + j \cdot s_Q(t) \\ &= e(t) \cdot \exp(j\theta(t)) \end{aligned}$$

with

$$e(t) = \sqrt{s_I^2(t) + s_Q^2(t)}$$

$$\tan \theta(t) = \frac{s_Q(t)}{s_I(t)}$$

- Also,

$$\begin{aligned} s_I(t) &= e(t) \cdot \cos(\theta(t)) \\ s_Q(t) &= e(t) \cdot \sin(\theta(t)) \end{aligned}$$

© 2018, B.-P. Paris

Exercise: Complex Envelope

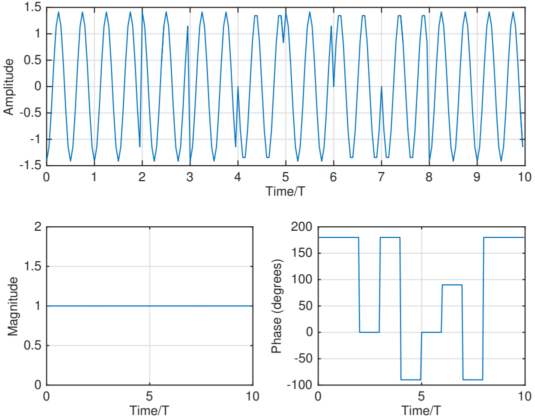
- Find the complex envelope representation of the signal

$$s_p(t) = \mathrm{sinc}(t/T)\cos(2\pi f_c t + \frac{\pi}{4}).$$

- **Answer:**

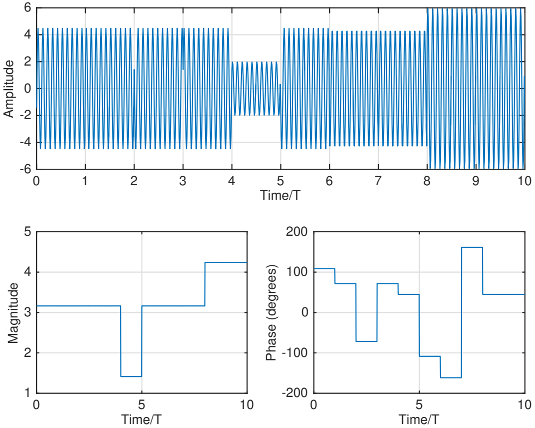
$$\begin{aligned} s(t) &= \frac{e^{j\pi/4}}{\sqrt{2}}\mathrm{sinc}(t/T) \\ &= \frac{1}{2}(\mathrm{sinc}(t/T) + j\mathrm{sinc}(t/T)). \end{aligned}$$

Illustration: QPSK with $f_c = 2/T$



- Passband signal (top): segments of sinusoids with different phases.
  - Phase changes occur at multiples of T.
- Baseband equivalent signal (bottom) is complex valued; magnitude and phase are plotted.
  - Magnitude is constant (rectangular pulses).
- Complex baseband signal shows symbols much more clearly than passband signal.

Illustration: 16-QAM with $f_c = 10/T$



- Passband signal (top): segments of sinusoids with different phases.
  - Phase and amplitude changes occur at multiples of T.
- Baseband signal (bottom) is complex valued; magnitude and phase are plotted.

Frequency Domain

- The time-domain relationships between the passband signal $s_p(t)$ and the complex envelope $s(t)$ lead to corresponding frequency-domain expressions.
- Note that

$$
\begin{aligned}
s_p(t) &= \Re\{s(t) \cdot \sqrt{2}\exp(j2\pi f_c t)\} \\
&= \frac{\sqrt{2}}{2}\left(s(t) \cdot \exp(j2\pi f_c t) + s^*(t) \cdot \exp(-j2\pi f_c t)\right).
\end{aligned}
$$

- Taking the Fourier transform of this expression:

$$
S_P(f) = \frac{\sqrt{2}}{2}\left(S(f - f_c) + S^*(-f - f_c)\right).
$$

  - Note that $S_P(f)$ has the conjugate symmetry ($S_P(f) = S_P^*(-f)$) that real-valued signals must have.

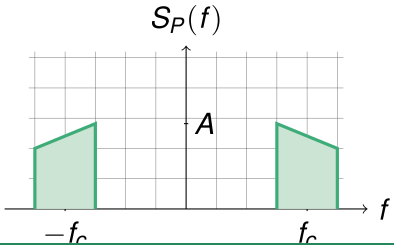
Frequency Domain

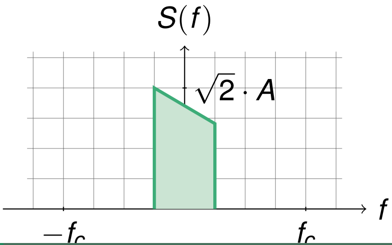
- In the frequency domain:

$$S_P(f) = \frac{\sqrt{2}}{2} \left(S(f - f_c) + S^*(-f - f_c) \right).$$

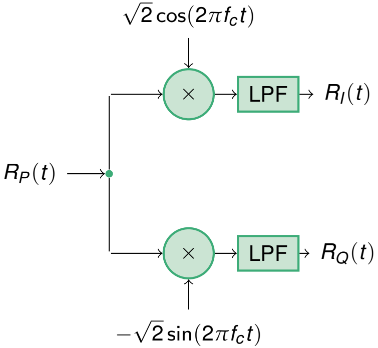
and, thus,

$$S(f) = \begin{cases} \sqrt{2} \cdot S_P(f + f_c) & \text{for } f + f_c > 0 \\ 0 & \text{else.} \end{cases}$$





Down-conversion



- The down-conversion system is the mirror image of the up-conversion system.
- The top-branch recovers the *in-phase* signal $s_I(t)$.
- The bottom branch recovers the *quadrature* signal $s_Q(t)$
  - See next slide for details.

Down-Conversion

- Let the the passband signal $s_p(t)$ be input to down-coverter:

$$s_P(t) = \sqrt{2}(s_I(t)\cos(2\pi f_c t) - s_Q(t)\sin(2\pi f_c t))$$

- Multiplying $s_P(t)$ by $\sqrt{2}\cos(2\pi f_c t)$ on the top branch yields

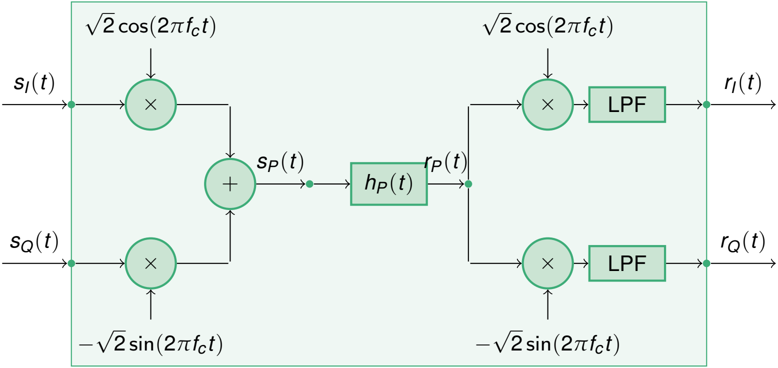
$$\begin{aligned} s_P(t)\cdot\sqrt{2}\cos(2\pi f_c t) \\ &= 2s_I(t)\cos^2(2\pi f_c t) - 2s_Q(t)\sin(2\pi f_c t)\cos(2\pi f_c t) \\ &= s_I(t) + s_I(t)\cos(4\pi f_c t) - s_Q(t)\sin(4\pi f_c t). \end{aligned}$$

- The low-pass filter rejects the components at $\pm 2f_c$ and retains $s_I(t)$.
- A similar argument shows that the bottom branch yields $s_Q(t)$.

Extending the Complex Envelope Perspective

- The baseband description of the transmitted signal is very convenient:
  - it is more compact than the passband signal as it does not include the carrier component,
  - while retaining all relevant information.
- However, we are also concerned what happens to the signal as it propagates to the receiver.
  - **Question:** Do baseband techniques extend to other parts of a passband communications system?
    - Filtering of the passband signal
    - Noise added to the passband signal

Complete Passband System



- Question: Can the pass band filtering ($h_P(t)$) be described in baseband terms?

Passband Filtering

- For the passband signals $s_P(t)$ and $R_P(t)$

$$r_P(t) = s_P(t) * h_P(t) \quad \text{(convolution)}$$

- Define a baseband equivalent impulse (complex) response $h(t)$.
- The relationship between the passband and baseband equivalent impulse response is

$$h_P(t) = \Re\{h(t) \cdot \sqrt{2}\exp(j2\pi f_c t)\}$$

- Then, the baseband equivalent signals $s(t)$ and $r(t) = r_I(t) + jr_Q(t)$ are related through

$$r(t) = \frac{s(t) * h(t)}{\sqrt{2}} \leftrightarrow R(f) = \frac{S(f)H(f)}{\sqrt{2}}.$$

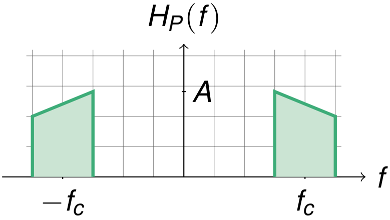
  - Note the division by $\sqrt{2}$!

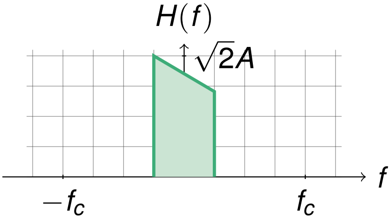
Passband and Baseband Frequency Response

- In the frequency domain

$$H(f) = \begin{cases} \sqrt{2} H_P(f + f_c) & \text{for } f + f_c > 0 \\ 0 & \text{else.} \end{cases}$$

$$H_p(f) = \frac{\sqrt{2}}{2} \left(H(f - f_c) + H^*(-f - f_c) \right)$$





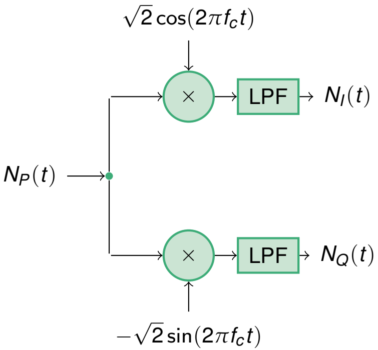
Exercise: Multipath Channel

- A multi-path channel has (pass-band) impulse response

$$h_P(t) = \sum_k a_k \cdot \delta(t - \tau_k).$$

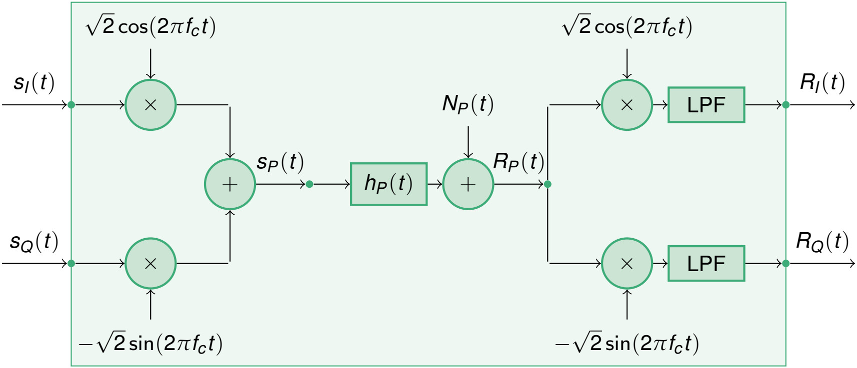
  Find the baseband equivalent impulse response $h(t)$ (assuming carrier frequency f_c) and the response to the input signal $s_p(t) = \cos(2\pi f_c t)$.

Passband White Noise



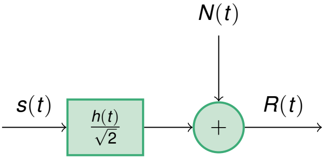
- Let (real-valued) white Gaussian noise $N_P(t)$ of spectral height $\frac{N_0}{2}$ be input to the down-converter.
- Then, each of the two branches produces indepent, white noise processes $N_I(t)$ and $N_Q(t)$ with spectral height $\frac{N_0}{2}$.
- This can be interpreted as (circular) complex noise of spectral height N_0.

Complete Passband System



- Complete pass-band system with channel (filter) and passband noise.

Baseband Equivalent System



- The passband system can be interpreted as follows to yield an equivalent system that employs only baseband signals:
  - baseband equivalent transmitted signal: $s(t) = s_I(t) + j \cdot s_Q(t)$.
  - baseband equivalent channel with complex valued impulse response: $h(t)$.
  - baseband equivalent received signal: $R(t) = R_I(t) + j \cdot R_Q(t)$.
  - complex valued, additive Gaussian noise: $N(t)$ with spectral height N_0.

Generalizing The Optimum Receiver

- We have derived all relationships for the optimum receiver for real-valued signals.
- When we use complex envelope techniques, some of our expressions must be adjusted.
  - Generalizing inner product and norm
  - Generalizing the matched filter (receiver frontend)
  - Adapting the signal space perspective
  - Generalizing the probability of error expressions

© 2018, B.-P. Paris

Inner Products and Norms

- The inner product between two complex signals $x(t)$ and $y(t)$ must be defined as

$$\langle x(t), y(t) \rangle = \int x(t) \cdot y^*(t)\, dt.$$

- This is needed to ensure that the resulting squared norm is positive and real

$$\|x(t)\|^2 = \langle x(t), x(t) \rangle = \int |x(t)|^2\, dt$$

Inner Products and Norms

- Norms are equal for passband and equivalent baseband signals.
  - Let
$$x_p(t) = \Re\{x(t)\sqrt{2}\exp(j2\pi f_c t)\}$$
$$y_p(t) = \Re\{y(t)\sqrt{2}\exp(j2\pi f_c t)\}$$
  - Then,
$$\begin{aligned}\langle x_p(t), y_p(t)\rangle &= \Re\{\langle x(t), y(t)\}\\ &= \langle x_I(t), y_I(t)\rangle + \langle x_Q(t), y_Q(t)\rangle\end{aligned}$$
  - The first equation implies
$$\|x_P(t)\|^2 = \|x(t)\|^2$$
    - Remark: the factor $\sqrt{2}$ in $x_p(t) = \Re\{x(t)\sqrt{2}\exp(j2\pi f_c t)\}$ ensures this equality.

Receiver Frontend

- Let the baseband equivalent, received signal be $R(t) = R_I(t) + jR_Q(t)$.
- Then the optimum receiver frontend for the complex signal $s(t) = s_I(t) + js_Q(t)$ will compute

$$
\begin{aligned}
R &= \langle R_P(t), s_P(t) \rangle = \Re\{\langle R(t), s(t) \rangle\} \\
&= \langle R_I(t), s_I(t) \rangle + \langle R_Q(t), s_Q(t) \rangle
\end{aligned}
$$

  - The I and Q channel are first matched filtered individually and then added together.

Signal Space

- Assume that passband signals have the form

$$s_P(t) = b_I p(t)\sqrt{2E}\cos(2\pi f_c t) - b_Q p(t)\sqrt{2E}\sin(2\pi f_c t)$$

for $0 \leq t \leq T$.
  - where $p(t)$ is a unit energy pulse waveform.
- Orthonormal basis functions are

$$\Phi_0 = \sqrt{2}p(t)\cos(2\pi f_c t) \quad \text{and} \quad \Phi_1 = \sqrt{2}p(t)\sin(2\pi f_c t)$$

- The corresponding baseband signals are

$$s(t) = b_I p(t)\sqrt{E} + j b_Q p(t)\sqrt{E}$$

- with basis functions

$$\Phi_0 = p(t) \quad \text{and} \quad \Phi_1 = jp(t)$$

Probability of Error

- Expressions for the probability of error are unchanged as long as the above changes to inner product and norm are incorporated.
- Specifically, expressions involving the distance between signals are unchanged

$$\mathrm{Q}\left(\frac{\|s_n - s_m\|}{\sqrt{2N_0}}\right).$$

- Expressions involving inner products with a suboptimal signal $g(t)$ are modified to

$$\mathrm{Q}\left(\frac{\Re\{\langle s_n - s_m, g(t)\rangle\}}{\sqrt{2N_0}\|g(t)\|}\right)$$

Summary

- The baseband equivalent channel model is much simpler than the passband model.
  - Up and down conversion are eliminated.
  - Expressions for signals do not contain carrier terms.
- The baseband equivalent signals are more tractable and easier to model (e.g., for simulation).
  - Since they are low-pass signals, they are easily sampled.
- No information is lost when using baseband equivalent signals, instead of passband signals.
- Standard, linear system equations hold (nearly)
- **Conclusion:** Use baseband equivalent signals and systems.

Introduction

- For our discussion of optimal receivers, we have focused on
  - the transmission of single symbols and
  - the signal space properties of symbol constellations.
  - We recognized the critical importance of distance between constellation points.
- The precise shape of the transmitted waveforms plays a secondary role when it comes to error rates.
- However, the spectral properties of transmitted signals depends strongly on the shape of signals.

Linear Modulation

- A digital communications signals is said to be *linearly modulated* if the transmitted signal has the form

$$s(t) = \sum_n b[n]p(t - nT)$$

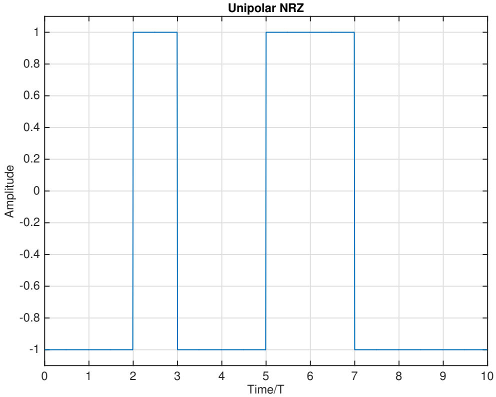
  where
  - $b[n]$ are the transmitted symbols, taking values from a fixed, finite *alphabet* $\mathcal{A}$,
  - $p(t)$ is fixed pulse waveform.
  - T is the symbol period; $\frac{1}{T}$ is the *baud rate*.
- This is referred to a linear modulation because the transmitted waveform $s(t)$ depends linearly on the symbols $b[n]$.

Illustration: Linear Modulation in MATLAB

```
function Signal = LinearModulation( Symbols, Pulse, fsT )
% LinearModulation - linear modulation of symbols with given

% initialize storage for Signal
LenSignal = length(Symbols)*fsT + (length(Pulse))-fsT;
Signal    = zeros( 1, LenSignal );

% loop over symbols and insert corresponding segment into Signal
for kk = 1:length(Symbols)
    ind_start = (kk-1)*fsT + 1;
    ind_end   = (kk-1)*fsT + length(Pulse);

    Signal(ind_start:ind_end) = Signal(ind_start:ind_end) + ...
                                Symbols(kk) * Pulse;
end
```

© 2018, B.-P. Paris

Example: Baseband Line Codes



- Unipolar NRZ (non-return-to-zero) and Manchester encoding are used for digital transmission over wired channels.

Passband Linear Modulation

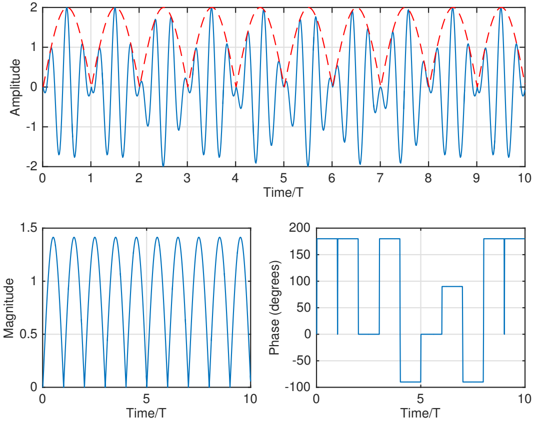
- Linearly modulated passband signals are easily described using the complex envelope techniques discussed previously.
- The baseband equivalent signals are obtained by linear modulation

$$s(t) = \sum_n b[n] p(t - nT)$$

where

  - $p(t)$ is a baseband pulse and
  - symbols $b[n]$ are complex valued.
    - For example, M-PSK is obtained when $b[n]$ are drawn from the alphabet is $\mathcal{A} = \{\exp(\frac{j2\pi n}{M})\}$, with $n = 0, 1, \ldots, M-1$.

Illustration: QPSK with $f_c = 3/T$ and Half-Sine Pulses



- Passband signal (top): segments of pulse-shaped sinusoids with different phases.
  - Phase changes occur at multiples of T.
- Baseband equivalent signal (bottom) is complex valued; magnitude and phase are plotted.
  - Magnitude reflects pulse shape.

- Pulse shape: $p(t) = \sqrt{2/T} sin(\pi t/T)$, for $0 \leq t \leq T$.

Spectral Properties of Digitally Modulated Signals

- Digitally Modulated signals are *random processes* - even though they don't look noise-like.
- The randomness is introduced by the random symbols $b[n]$.
- We know from our earlier discussion that the spectral properties of a random process are captured by its power spectral density (PSD) $S_s(f)$.
- We also know that the power spectral density is the Fourier transform of the autocorrelation function $R_s(\tau)$

$$R_s(\tau) \leftrightarrow S_s(f).$$

PSD for Linearly Modulated Signals

- An important special case arises when the symbol stream $b[n]$
  - is uncorrelated, i.e.,

$$\mathbf{E}[b[n]b^*[m]] = \begin{cases} \mathbf{E}[|b[n]|^2] & \text{when } n = m \\ 0 & \text{when } n \neq m \end{cases}$$

  - has zero mean, i.e., $\mathbf{E}[b[n]] = 0$.
- Then, the power-spectral density of the transmitted signal is

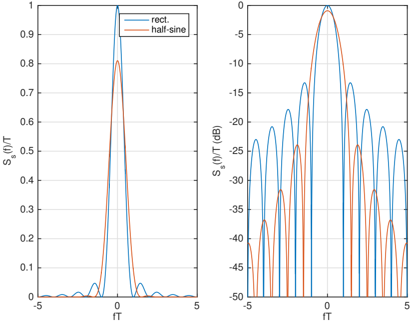
$$S_s(f) = \frac{\mathbf{E}[|b[n]|^2]}{T} |P(f)|^2$$

where $p(t) \leftrightarrow P(f)$ is the Fourier transform of the shaping pulse.
  - Note that the shape of the spectrum does not depend on the constellation.

Exercise: PSD for Different Pulses

- Assume that $\mathbf{E}[|b[n]|^2] = 1$; compute the PSD of linearly modulated signals (with uncorrelated, zero-mean symbols) when
  1. $p(t) = \sqrt{1/T}$ for $0 \leq t \leq T$. (rectangular)
  2. $p(t) = \sqrt{2/T} sin(\pi t/T)$ for $0 \leq t \leq T$. (half-sine)
- **Answers:**
  1. $S_s(f) = \text{sinc}^2(fT)$
  2. $S_s(f) = \frac{8}{\pi^2} \frac{cos^2(\pi fT)}{(1-4(fT)^2)^2}$

Comparison of Spectra



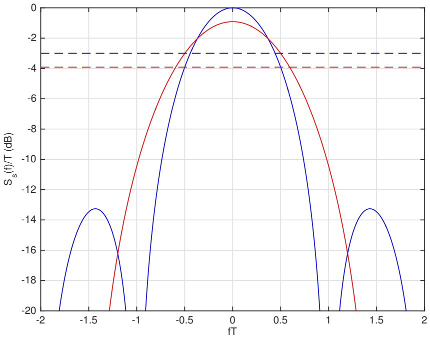
- Rectangular pulse has narrower main-lobe.
- Half-sine pulse has faster decaying sidelobes (less adjacent channel interference).
  - In general, smoother pulses have better spectra.

Measures of Bandwidth

- From the plot of a PSD, the bandwidth of the signal can be determined.
- The following three metrics are commonly used:
  1. 3dB bandwidth
  2. zero-to-zero bandwidth
  3. Fractional power containment bandwidth
- Bandwidth is measured differently for passband signals and baseband signals:
  1. For passband signals, the two-sided bandwidth is relevant.
  2. For baseband signals, the one-sided bandwidth is of interest.

© 2018, B.-P. Paris

3dB Bandwidth

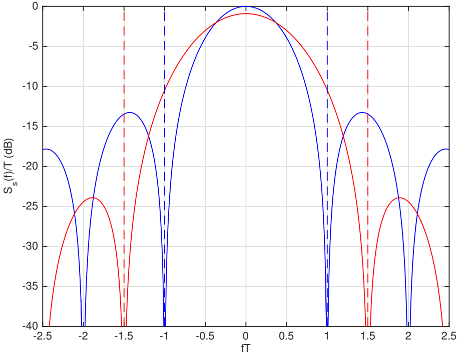


- For symmetric spectra with maximum in the center of the band ($f = 0$), the two-sided 3dB-bandwidth B_{3dB} is defined by

$$S_s\left(\frac{B_{3dB}}{2}\right) = \frac{S_s(0)}{2} = S_s\left(-\frac{B_{3dB}}{2}\right).$$

- For rectangular pulse, $B_{3dB} \approx \frac{0.88}{T}$.
- For half-sine pulse, $B_{3dB} \approx \frac{1.18}{T}$.

Zero-to-Zero Bandwidth



- The two-sided zero-to-zero bandwidth $B_{0\text{-}0}$ is the bandwidth between the two two zeros of the PSD that are closest to the peak at $f = 0$.
- In other words, for symmetric spectra

$$S_s\left(\frac{B_{0\text{-}0}}{2}\right) = 0$$
$$= S_s\left(-\frac{B_{0\text{-}0}}{2}\right).$$

- For rectangular pulse, $B_{0\text{-}0} = \frac{2}{T}$.
- For half-sine pulse, $B_{0\text{-}0} = \frac{3}{T}$.

Fractional Power-Containment Bandwidth

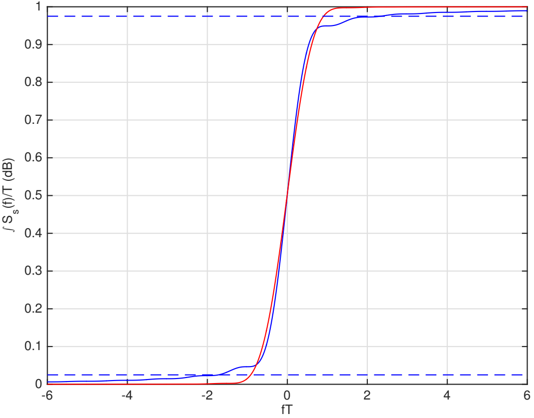
- Fractional power-containment bandwidth B_γ is the width of the smallest frequency interval that contains a fraction γ of the total signal power.
  - Total signal power

$$P_s = \frac{\mathbf{E}[|b[n]|^2]}{T} \int_0^T |p(t)|^2 \, dt = \frac{\mathbf{E}[|b[n]|^2]}{T} \int_{-\infty}^{\infty} |P(f)|^2 \, df.$$

  - For symmetric spectra, fractional power-containment bandwidth B_γ is defined through the relationship

$$\int_{-B_\gamma/2}^{B_\gamma/2} |P(f)|^2 \, df = \gamma \int_{-\infty}^{\infty} |P(f)|^2 \, df$$

Illustration: 95% Containment Bandwidth



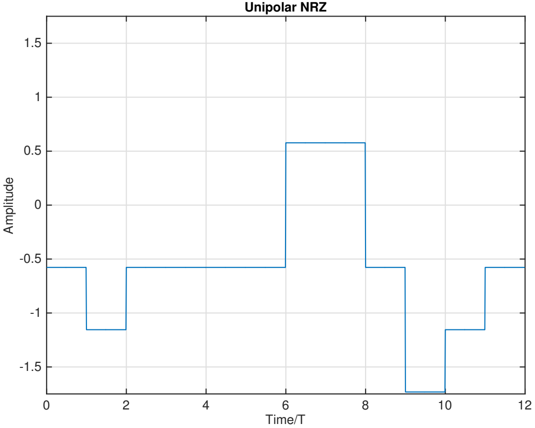
- The horizontal lines correspond to $(1-\gamma)/2$ and $1-(1-\gamma)/2$ (i.e., 2.5% and 97.5%, respectively, for $\gamma =$95%).
- For half-sine pulse, $B_{95\%}\ approx\, \frac{1.82}{T}$.
- For rectangular pulse, $B_{95\%}\ approx\, \frac{3.4}{T}$

Full-Response and Partial Response Pulses

- So far, we have considered only pulses that span exactly one symbol period T.
  - Such pulses are called *full-response pulses* since the entire signal due to the *n*-th symbol is confined to the *n*-th symbol period.
  - Recall that pulses of finite duration have infinitely long Fourier transforms.
  - Hence, full-response spectra are inherently of infinite bandwidth - the best we can hope for is to concentrate power in a narrow band.
- We can consider pulses that are longer than a symbol period.
  - Such pulses are called partial-repsonse pulses.
  - They hold promise for better spectral properties.
  - But, they cause self-interference between symbols (ISI) unless properly designed.

© 2018, B.-P. Paris

How not do partial-response signalling



- The pulse are rectangular pulses spanning 3 symbol periods.
- The transmitted information symbols are no longer obvious.
- An equalizer would be needed to “untangle” the symbol stream.

Nyquist Pulses

- To avoid interference at sampling times $t = kT$, pulses $p(t)$ must meet the Nyquist criterion

$$p(mT) = \begin{cases} 1 & \text{for } m = 0 \\ 0 & \text{for } m \neq 0 \end{cases}$$

- With this criterion, samples of the received signal at times $t = kT$ satisfy

$$s(kT) = \sum_n b[n] p(kT - nT) = b[k].$$

  - At times $t = kT$, there is **no** interference!
- Pulses satisfying the above criterion are called Nyquist pulses.

Frequency Domain Version of the Nyquist Criterion

- In the time-domain, Nyquist pulses (for transmitting at rate $1/T$) satisfy

$$p(mT) = \begin{cases} 1 & \text{for } m = 0 \\ 0 & \text{for } m \neq 0 \end{cases}$$

- An equivalent, frequency-domain criterion is

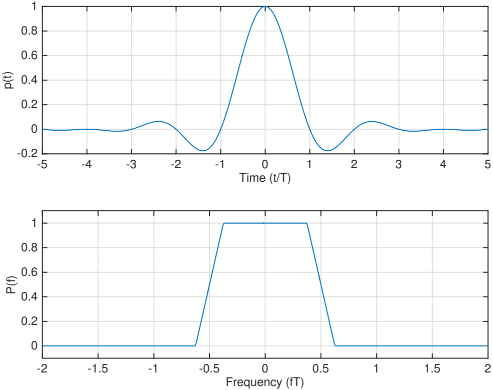
$$\sum_{k=-\infty}^{\infty} P(f + \frac{k}{T}) = T \quad \text{for all } f.$$

Example: Pulses with Trapezoidal Spectrum

- The pulse
$$p(t) = \text{sinc}(t/T) \cdot \text{sinc}(at/T)$$
is a Nyquist pulse for rate $1/T$.
  - The parameter a ($0 \leq a \leq 1$) is called the excess bandwidth.
- The Fourier transform of $p(t)$ is

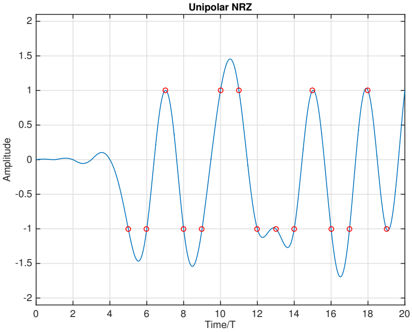
$$P(f) = \begin{cases} T & \text{for } |f| < \frac{1-a}{2T} \\ T\frac{(1+a)-2|f|T}{2a} & \text{for } \frac{1-a}{2T} \leq |f| \leq \frac{1+a}{2T} \\ 0 & \text{for } |f| > \frac{1+a}{2T} \end{cases}$$

Example: Pulses with Trapezoidal Spectrum



- The Trapezoidal Nyquist pulse has infinite duration and ist *strictly* bandlimited!

Linear Modulation with Trapezoidal Nyquist Pulses



- With the Trapezoidal Nyquist pulse, at every symbol instant $t = nT$ there is no ISI: $s(nT) = b[n]$.
- No ISI and stricly band-limited spectrum is achieved by Nyquist pulses.

Raised Cosine Pulse

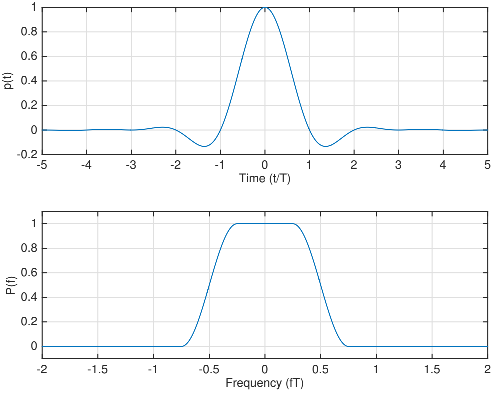
- The most widely used Nyquist pulse is the Raised Cosine Pulse:

$$p(t) = \text{sinc}(\frac{t}{T}) \frac{\cos(\pi a t/T)}{1-(2at/T)^2}.$$

with Fourier Transform

$$P(f) = \begin{cases} T & \text{for } |f| < \frac{1-a}{2T} \\ \frac{T}{2}\left[1+\cos(\frac{\pi T}{a}(|f| - \frac{1-a}{2T}))\right] & \text{for } \frac{1-a}{2T} \leq |f| \leq \frac{1+a}{2T} \\ 0 & \text{for } |f| > \frac{1+a}{2T} \end{cases}$$

Example: Pulses with Trapezoidal Spectrum



- The raised cosine pulse is strictly bandlimited!

Root-Raised Cosine Pulse

- The receiver needs to apply a matched filter.
- For linearly modulated signals, the matched filter is the pulse $p(t)$.
  - $p(t) = p(-t)$ for symmetric pulses.
- However, when the symbol stream is passed through the filter $p(t)$ twice then the Nyquist condition no longer holds.
  - $p(t) * p(t)$ is *not* a Nyquist pulse.
- The root-raised cosine filter has a Fourier transform that is the square-root of the Raised Cosine pulse's Fourier transform.
- It is strictly band-limited and the series of two root-raised-cosine filters is a Nyquist pulse.

© 2018, B.-P. Paris