Elements of a Digital Communications System

Learning Objectives and Course Outline

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 $R(t)$.

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.
    - 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: \( \pi_n \) 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 (\( M = 4 \))

Symbol Rate: The number of information symbols the source produces per second. Also called the baud rate \( R \).

- Closely related: information rate \( R_b \) indicates the number of bits the source produces per second.
- Relationship: \( R_b = R \cdot \log_2(M) \).
- Also, \( T = 1 / R \) is the symbol period.

<table>
<thead>
<tr>
<th>Bit 1</th>
<th>Bit 2</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>-3</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table: Example: How two bits can be represented in one quaternary symbol.
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 source 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 $P_N = kT_0B$.
    - $k$: Boltzmann's constant (1.38 $\cdot$ 10$^{-23}$ Ws/K).
    - $T_0$: temperature in degrees Kelvin (room temperature $\approx$ 290 K).
    - For bandwidth equal to 1 Hz, $P_N \approx 4 \cdot 10^{-21}$ 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.
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.

©2016, B.-P. Paris

Elements of a Digital Communications System

The Receiver

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

Course Outline

▶ Mathematical Prerequisites
  ▶ Basics of Gaussian Random Variables and Random Processes
  ▶ Signal space concepts
▶ Principles of Receiver Design
  ▶ Receiver frontend: the matched filter
  ▶ Optimal decision: statistical hypothesis testing
▶ 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 $\sigma^2$.
- Notation: $X \sim \mathcal{N}(m, \sigma^2)$.
- Moments:
  \[
  E[X] = \int_{-\infty}^{\infty} x \cdot p_X(x) \, dx = m
  \]
  \[
  E[X^2] = \int_{-\infty}^{\infty} x^2 \cdot p_X(x) \, dx = m^2 + \sigma^2.
  \]
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

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

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

$$\Pr \{ X > 5 \} = \int_5^{\infty} \frac{1}{\sqrt{2\pi} \cdot 2} e^{-\frac{(x-1)^2}{2 \cdot 2^2}} \, dx \quad \text{substitute } z = \frac{x-1}{2}$$

$$= \int_2^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, dz = Q(2)$$
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 $Q(x)$. 

©2016, B.-P. Paris

ECE 201: Intro to Signal Analysis
Gaussian Random Vectors

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

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

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

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

▶ Covariance matrix  

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

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

©2016, B.-P. Paris  
ECE 201: Intro to Signal Analysis  
28
Exercise: Important Special Case: N=2

▶ Consider a length-2 Gaussian random vector with

\[ \bar{m} = \bar{0} \text{ and } K = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} . \]

▶ 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 \( X_i \) comprising the random vector \( \vec{X} \) are uncorrelated (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 \( \vec{X} \sim \mathcal{N}(\bar{m}, K) \)
   ▶ Affine transformation: \( \vec{Y} = A\vec{X} + \bar{b} \)
   ▶ Then, \( \vec{Y} \sim \mathcal{N}(A\bar{m} + \bar{b}, AKAT) \)
Exercise: Generating Correlated Gaussian Random Variables

- Let \( \vec{X} \sim \mathcal{N}(\vec{m}, \mathbf{K}) \), with
  \[
  \vec{m} = \vec{0} \quad \text{and} \quad \mathbf{K} = \sigma^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.
  \]

  - 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 \( f_c = 1 \text{GHz} \).
  - 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 communication system.
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 $\omega$ of the sample space $\Omega$ to a function of time (i.e., a signal).
- Notation: $X_t(\omega)$ - we will frequently omit $\omega$ to simplify notation.
- Observations:
  - We will be interested in both real and complex valued random processes.
  - Note, for a given random outcome $\omega_0$, $X_t(\omega_0)$ 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 $\omega_0$, $X_t(\omega_0)$ 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:

\[
\begin{align*}
X_t(\omega_1) &= \cos(2\pi f_0 t) & X_t(\omega_2) &= -\sin(2\pi f_0 t) \\
X_t(\omega_3) &= -\cos(2\pi f_0 t) & X_t(\omega_4) &= \sin(2\pi f_0 t)
\end{align*}
\]
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 \( 2\pi \), \( \Theta(\omega) \sim U[0, 2\pi] \).
  - For both of these processes, the complete statistical description of the random process can be found.
Exercise: 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 $p_{X_t X_t}(x_1, x_2)$ for this process.
  - Note, since the phase values are discrete the above pdfs must be expressed with the help of $\delta$-functions.
  - Alternatively, one can derive a probability mass function.

Answer: Discrete Phase Process

- First-order density function:
  
  $$p_{X_t}(x) = \frac{1}{4} \left( \delta(x - \cos(2\pi f_0 t)) + \delta(x + \sin(2\pi f_0 t)) + \right. \\
  \left. \delta(x + \cos(2\pi f_0 t)) + \delta(x - \sin(2\pi f_0 t)) \right)$$

- Second-order density function:
  
  $$p_{X_t X_t}(x_1, x_2) = \frac{1}{4} \left( \delta(x_1 - \cos(2\pi f_0 t_1)) \cdot \delta(x_2 - \cos(2\pi f_0 t_2)) + \\
  \delta(x_1 + \sin(2\pi f_0 t_1)) \cdot \delta(x_2 + \sin(2\pi f_0 t_2)) + \\
  \delta(x_1 + \cos(2\pi f_0 t_1)) \cdot \delta(x_2 + \cos(2\pi f_0 t_2)) + \\
  \delta(x_1 - \sin(2\pi f_0 t_1)) \cdot \delta(x_2 - \sin(2\pi f_0 t_2)) \right)$$
Exercise: Continuous Phase Process

- **Continuous Phase Process**: Let $\Theta(\omega)$ be a random variable that is uniformly distributed between 0 and $2\pi$, $\Theta(\omega) \sim [0, 2\pi)$. 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 $p_{X_t X_{t'}}(x_1, x_2)$ for this process.

Answer: Continuous Phase Process

- First-order density:
  
  $$p_{X_t}(x) = \frac{1}{\pi \sqrt{1 - x^2}} \quad \text{for } |x| \leq 1.$$  
  
  Notice that $p_{X_t}(x)$ does **not** depend on $t$.

- Second-order density:
  
  $$p_{X_t X_{t'}}(x_1, x_2) = \frac{1}{\pi \sqrt{1 - x_2^2}} \cdot \frac{1}{2} \cdot \delta(x_1 - \cos(2\pi f_0 (t_1 - t_2) + \arccos(x_2))) +$$

  $$\delta(x_1 - \cos(2\pi f_0 (t_1 - t_2) - \arccos(x_2))))$$
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 $X_t$
    \[
    X_t(\omega) = \sum_k X_k(\omega)p(t - nT)
    \]

First Order Density

- 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 Processes 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 $t_1, \cdots, t_n$
    
    $$
    \vec{X} = \begin{pmatrix}
    X_{t_1} \\
    \vdots \\
    X_{t_n}
    \end{pmatrix}
    $$
    
    is a Gaussian random vector for all $t_1, \cdots, t_n$.

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 and Autocorrelation Functions

- 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:
    \[
    E[X_t] = m_X(t) = \int_{-\infty}^{\infty} x \cdot p_X(x) \, dx
    \]
    - Note, that the mean of a random process is a deterministic signal.
    - The mean is computed from the first order density function.

  - **Definition:** The autocorrelation function of a random process is defined as:
    \[
    R_X(t, u) = E[X_tX_u] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \cdot p_{X_t, X_u}(x, y) \, dx \, dy
    \]
    - Autocorrelation is computed from second order density

Autocovariance Function

- Closely related: autocovariance function:
  \[
  C_X(t, u) = E[(X_t - m_X(t))(X_u - m_X(u))] = R_X(t, u) - m_X(t)m_X(u)
  \]
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 $2\pi$, $\Theta(\omega) \sim [0, 2\pi)$. 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:
    $$\text{Var}(X_t) = \mathbb{E}[X_t^2] = R_X(0) = \frac{N_0}{2} \delta(0) = \infty.$$ 
  - Also, for $t \neq u$: $\mathbb{E}[X_t X_u] = R_X(t, u) = R_X(t-u) = 0$.

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 Riemann and the random variable $Z$ satisfy:

Given $\epsilon > 0$, there exists a $\delta > 0$ so that

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

with:

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

Mean Square Convergence - Why We Care

▶ It can be shown that the integral converges if

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

▶ Important: When the integral converges, then the order of integration and expectation can be interchanged, e.g.,

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

▶ 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 $R_W(t, u) = \sigma^2 \min(t, u)$
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 $p_{X_tY_u}(x, y)$ 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:

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

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

Clearly, \( X_t \) and \( Y_t \) are jointly defined random processes.

Standard LTI system – convolution:

\[
Y_t = \int h(t - \tau) X_\tau \, d\tau = h(t) * X_t
\]

Recall: this convolution is “well-behaved” if

\[
\int \int R_X(\tau, \nu) h(t - \tau) h(t - \nu) \, d\tau \, d\nu < \infty
\]

E.g.: \( \int \int R_X(\tau, \nu) \, d\tau \, d\nu < \infty \) and \( h(t) \) stable.

Second Order Description of Output: Mean

The expected value of the filter’s output \( Y_t \) is:

\[
E[Y_t] = E[\int h(t - \tau) X_\tau \, d\tau]
\]

\[
= \int h(t - \tau) E[X_\tau] \, d\tau
\]

\[
= \int h(t - \tau) m_X(\tau) \, d\tau
\]

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

\[
E[Y_t] = m_Y(t) = m_X \int h(\tau) \, d\tau
\]

is also constant.
## Crosscorrelation of Input and Output

- The crosscorrelation between input and output signals is:
  \[ R_{XY}(t, u) = E[X_t Y_u] = E[X_t \int h(u - \tau) X_\tau \, d\tau] \]
  \[ = \int h(u - \tau) E[X_t X_\tau] \, d\tau \]
  \[ = \int h(u - \tau) R_X(t, \tau) \, d\tau \]

- For a wss input process:
  \[ R_{XY}(t, u) = \int h(u - \tau) R_X(t, \tau) \, d\tau = \int h(v) R_X(t, u - v) \, dv \]
  \[ = \int h(v) R_X(t - u + v) \, dv = R_{XY}(t - u) \]

- Input and output are jointly stationary.

## Autocorrelation of Output

- The autocorrelation of \( Y_t \) is given by
  \[ R_Y(t, u) = E[Y_t Y_u] = E[\int h(t - \tau) X_\tau \, d\tau \int h(u - \nu) X_\nu \, d\nu] \]
  \[ = \int \int h(t - \tau) h(u - \nu) R_X(\tau, \nu) \, d\tau \, d\nu \]

- For a wss input process:
  \[ R_Y(t, u) = \int \int h(t - \tau) h(u - \nu) R_X(\tau, \nu) \, d\tau \, d\nu \]
  \[ = \int \int h(\lambda) h(\lambda - \gamma) R_X(t - \lambda, u - \lambda + \gamma) \, d\lambda \, d\gamma \]
  \[ = \int \int h(\lambda) h(\lambda - \gamma) R_X(t - u - \gamma) \, d\lambda \, d\gamma = R_Y(t - u) \]

- 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:

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

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 $\Delta f$
  - 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 $R_X(t)$, 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 representation 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

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

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

$$E[|X_t|^2] = R_X(0) = \int_{-\infty}^{\infty} S_X(f) df.$$ 

- $S_X(f)$ 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 $f(\cdot)$.
Filtering of Random Processes

- Random process $X_t$ with autocorrelation $R_X(\tau)$ and PSD $S_X(f)$ is input to LTI filter with impulse 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 $R_Y(\tau) = R_X(\tau) \ast C_h(\tau)$,
- where $C_h(\tau) = h(\tau) \ast h(-\tau)$.
- In frequency domain: $S_Y(f) = S_X(f) \cdot \mathcal{F}\{C_h(\tau)\}$
- With
  $$\mathcal{F}\{C_h(\tau)\} = \mathcal{F}\{h(\tau) \ast h(-\tau)\}$$
  $$= \mathcal{F}\{h(\tau)\} \cdot \mathcal{F}\{h(-\tau)\}$$
  $$= H(f) \cdot H^*(f) = |H(f)|^2.$$  

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:
  $$S_Y(f) = \left| \frac{1}{1 + j2\pi fRC} \right|^2 \frac{N_0}{2}.$$
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 disregard 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 \( S \) is a set of scalars. If \( a, b \) are scalars, then for any \( x, y \in S \) the following properties hold:
  1. \( a \cdot x \) is defined and \( a \cdot x \in S \).
  2. \( a \cdot (b \cdot x) = (a \cdot b) \cdot x \)
  3. Let 1 and 0 denote the multiplicative and additive identities of the field of scalars, then \( 1 \cdot x = x \) and \( 0 \cdot x = \vec{0} \) for all \( x \in S \).
  4. Associate properties:
    \[
    a \cdot (x + y) = a \cdot x + a \cdot y
    \]
    \[
    (a + b) \cdot x = a \cdot x + b \cdot y
    \]

Running Examples
- The space of length-\( N \) vectors \( \mathbb{R}^N \)
  \[
  \begin{pmatrix}
  x_1 \\
  \vdots \\
  x_N
  \end{pmatrix} + \begin{pmatrix}
  y_1 \\
  \vdots \\
  y_N
  \end{pmatrix} = \begin{pmatrix}
  x_1 + y_1 \\
  \vdots \\
  x_N + y_N
  \end{pmatrix}
  \]
  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).
Subspaces

- **Definition:** Let $S$ be a linear vector space. The space $L$ is a **subspace** of $S$ if
  1. $L$ is a subset of $S$ and
  2. $L$ is closed.
- **Example:** Let $S$ be $L^2(T_a, T_b)$. Define $L$ 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.

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 $x, y \in S$ 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. $\langle x, y \rangle = \langle y, x \rangle$
  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.

- \( x(t), y(t) \in L^2(a, b) \) with
  \[
  \langle x(t), y(t) \rangle = \int_{a}^{b} x(t)y(t) \, dt
  \]

  - **Answer:** Yes; this is the continuous-time equivalent of the

Exercise: Valid Inner Products?

- \( x, y \in \mathbb{C}^N \) with
  \[
  \langle x, y \rangle = \sum_{n=1}^{N} x_n y_n^\ast
  \]

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

  with \( K \) an \( N \times N \)-matrix

  - **Answer:** Only if \( K \) is positive definite (i.e., \( x^T K x > 0 \) for all \( x \neq 0 \)).
Norm of a Vector

**Definition:** The norm of vector \( x \in S \) 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

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

**Illustration:**

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

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 \( x, y \in S \), where \( S \) 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 \( a = -\langle x, y \rangle \frac{1}{\|y\|^2} \).

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

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

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

\[
w_m(t) = \frac{1}{T} e^{j2\pi mt/T}.
\]
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:

\[ d^2(x, y) = \|x - y\|^2 \]
\[ = \|(x - z) + (z - y)\|^2 \]
\[ = (\langle x - z, z - y \rangle, (x - z) + (z - y)) \]

\[ d^2(x, y) = \langle x - z, x - z \rangle + 2\langle x - z, z - y \rangle + \langle z - y, z - y \rangle \]
\[ \leq \langle x - z, x - z \rangle + 2|\langle x - z, z - y \rangle| + |\langle z - y, z - y \rangle| \]
\[ \leq (x - z, x - z) + 2\|x - z\| \cdot \|z - y\| + \langle z - y, z - y \rangle (Schwartz) \]
\[ = (d(x, z) + (d(y, z))|^2 \]

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.

**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)
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,

\[
\|z\| = \|x - y\| = \min_{v \in \mathcal{L}} \|x - v\|.
\]

- \( y \) is called the projection of \( x \) onto \( \mathcal{L} \).
- The distance between \( x \) and 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 \( v_n(t) = A_n \cos(2\pi nt / T) \) for a fixed \( n \).
- 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 performs the projection of \( x(t) \) onto \( \mathcal{L} \).