Chapter 9 General Concepts

Po-Ning Chen, Professor

Institute of Communications Engineering

National Chiao Tung University

Hsin Chu, Taiwan 30010, R.O.C.

Why Introducing Random or Stochastic Processes

Oxford Dictionary

> **Random**: adj. Made, done, or happening without method or conscious decision. In *Statistics*. Governed by or involving equal chances for each item.
>
> **Stochastic**: adj. Having a random probability distribution or pattern that may be analysed statistically but may not be predicted precisely.
>
> **Process**: n. A series of actions or steps taken in order to achieve a particular end.

- Why introducing random process? For convenience of analyzing a system.
- Two models may be considered in a, e.g., communication system.
  - Deterministic model
    - **No uncertainty** about its time-dependent (exact) behavior.
  - Random or Stochastic model
    - **Uncertain** about its time-dependent (exact) behavior, but **certain** on its statistical behavior.
- Example of stochastic models
  - Channel noise and interference, or source of information such as voice

Definition (Random variable) A random variable on a probability space $(S, \mathcal{F}, P)$ (in which $\mathcal{F}$ is a σ-field and P is a probability measure for events in $\mathcal{F}$) is a real-valued function $\boldsymbol{x}(\zeta)$ (i.e., $\boldsymbol{x} : S \to \Re$) with $\{\zeta \in S : \boldsymbol{x}(\zeta) \leq x\} \in \mathcal{F}$ for every $x \in \Re$.

(page 4) 'The name "random variable" is actually a misnomer, since it is not random and not a variable ... the random variable simply maps each point (outcome) in the sample space to a number on the real line.

Richard M. Feldman and Ciriaco Valdez-Flores. *Applied Probability and Stochastic Processes.* Technology & Engineering. Springer Science & Business Media, 2 edition, 2010.

- An element of S is referred to as a *sample outcome*.
- An element of $\mathcal{F}$ is referred to as an *event*.
- An *event* is a subset of S. In other words, $\mathcal{F}$ is a non-empty collection of subsets (events) of S.
- *Probability measure* P is defined for the events in $\mathcal{F}$. In other words, all events containing in $\mathcal{F}$ should be probabilistically measurable.

Example $S = \{\Box, \triangle, \diamondsuit\}$ and $\mathcal{F} = \left\{\emptyset, \{\Box, \triangle\}, \{\diamondsuit\}, S\right\}$.

Then, a legitimate probability measure P should be defined for all events below:

$$P(\emptyset) = 0, \quad P(\{\Box, \triangle\}) = 0.7, \quad P(\{\diamondsuit\}) = 0.3, \quad P(S) = 1.$$

No specification is given or should be given for $P(\{\Box\})$ and $P(\{\triangle\})$.

Random Variables

Definition (Random variable) A random variable on a probability space $(S, \mathcal{F}, P)$ (in which $\mathcal{F}$ is a σ-field and P is a probability measure for events in $\mathcal{F}$) is a real-valued function $\boldsymbol{x}(\zeta)$ (i.e., $\boldsymbol{x} : S \to \Re$) with $\{\zeta \in S : \boldsymbol{x}(\zeta) \leq x\} \in \mathcal{F}$ for every $x \in \Re$.

- The event space $\mathcal{F}$ must be a σ-field. Why? See the next two slides.
- $\{\zeta \in S : \boldsymbol{x}(\zeta) \leq x\}$ must be an event for every $x \in \Re$.
  - Otherwise, the cumulative distribution function (cdf) of $\boldsymbol{x}$ is **not** well-defined:

$$\Pr[\boldsymbol{x} \leq x] = P\big(\{\zeta \in S : \boldsymbol{x}(\zeta) \leq x\}\big).$$

The Concept of Field/Algebra

Definition (Field/algebra) A set $\mathcal{F}$ is said to be a *field* or *algebra* of a *sample space* S if it is a nonempty collection of subsets of S with the following properties:

1. $\emptyset \in \mathcal{F}$ and $S \in \mathcal{F}$;
   - *Interpretation*: $\mathcal{F}$ should be a mechanism to determine whether the *outcome* lies in an empty set (impossible) or the sample space (certain).
2. (closure under complement action) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$;
   - *Interpretation*: "having a mechanism to determine whether the outcome lies in A" is equivalent to "having a mechanism to determine whether the outcome lies in A^c."
3. (closure under finite union) $A \in \mathcal{F}$ and $B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$.
   - *Interpretation*: If one has a mechanism to determine whether the outcome lies in A, and a mechanism to determine whether the outcome lies in B, then he can surely determine whether the outcome lies in the union of A and B.

σ-field/algebra

- To work on a *field* may result some problems when one is dealing with "*limit*".

  E.g., $S = \Re$ (the real line) and $\mathcal{F}$ is a collection of all *open*, *semi-open* and *closed* intervals whose two endpoints are rational numbers, including $\Re$ itself.

  Let
  $$A_k = [0, [\pi]_k),$$
  where $[\pi]_k \triangleq \lfloor \pi \times 10^k \rfloor / 10^k$. Does the infinite union $\cup_{i=1}^{\infty} A_i$ belong to $\mathcal{F}$? The answer is apparently negative!

- We therefore need an extension definition of field, which is named σ-field.

Definition (σ-field/σ-algebra) A set $\mathcal{F}$ is said to be a *σ-field* or *σ-algebra* of a sample space S if it is a nonempty collection of subsets of S with the following properties:

1. $\emptyset \in \mathcal{F}$ and $S \in \mathcal{F}$;
2. (closure under complement action) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$;
3. (closure under countable union) $A_i \in \mathcal{F}$ for $i = 1, 2, 3, \ldots \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Probability Measure

Definition (Probability measure) A set function P on a measurable space $(S, \mathcal{F})$ is a *probability measure*, if it satisfies:

1. $0 \leq P(\mathcal{A}) \leq 1$ for $\mathcal{A} \in \mathcal{F}$;
2. $P(\emptyset) = 0$ and $P(S) = 1$.
3. **(countable additivity)** if $\mathcal{A}_1, \mathcal{A}_2, \ldots$ is a disjoint sequence of sets in $\mathcal{F}$, then
$$P\left(\bigcup_{k=1}^{\infty} \mathcal{A}_k\right) = \sum_{k=1}^{\infty} P(\mathcal{A}_k).$$

Sufficiency of CDF

- It can be proved that we can construct a well-defined probability space $(S, \mathcal{F}, P)$ for any random variable $\boldsymbol{x}$ if its cdf $F(\cdot)$ is given.

It can be proved that any function $G(x)$ satisfying:

1. $\lim_{x\to-\infty} G(x) = 0$ and $\lim_{x\to\infty} G(x) = 1$,
2. Right-continuous,
3. Non-decreasing;

is a legitimate cdf for some random variable. It suffices to check the above three properties for $F(\cdot)$ to well-define a random variable.

† See Theorem 14.1 in [P. Billingsley, *Probability and Measure*, 3rd Edition, Wiley, 1995]

  – Hence, defining a real-valued random variable only by providing its cdf is good enough from engineering standpoint.
  – In other words, it is not necessary to mention the probability space $(S, \mathcal{F}, P)$ when defining a random variable.

- Then, why bother to introduce the probability space $(S, \mathcal{F}, P)$?

Merit of defining random variables based on $(S, \mathcal{F}, P)$

Merit 1: Good for making abstraction of something. For example, $(S, \mathcal{F}, P)$ is what truly and internally occurs but is possibly **non-observable**.

- In order to infer what really occurs for this **non-observable** random outcome ζ, an experiment that results in an **observable value** x that depends on this **non-observable** outcome must be performed.
- So, x that takes from real values is a function of $\zeta \in S$.
- Since ζ is random with respect to probability measure P, the probability of the occurrence of observation $\boldsymbol{x}$ is defined as $P(\{\zeta \in S : \boldsymbol{x}(\zeta) = x\})$.
- Some books therefore state that $\boldsymbol{x} : (S, \mathcal{F}, P) \to (\boldsymbol{x}(S), \mathcal{B}, Q)$ yields an **observation probability space** $(\boldsymbol{x}(S), \mathcal{B}, Q)$, where

$$\boldsymbol{x}(\mathcal{A}) = \{\boldsymbol{x}(\zeta) : \zeta \in \mathcal{A}\}, \quad \mathcal{B} = \{\boldsymbol{x}(\mathcal{A}) : \mathcal{A} \subset \mathcal{F}\} \text{ and } Q(\boldsymbol{x}(\mathcal{A})) = P(\mathcal{A}).$$

† See [Robert M. Gray and Lee D. Davisson, *Random Processes: A Mathematical Approach for Engineers*, Prentice Hall, 1986].

Merit of defining random variables based on $(S, \mathcal{F}, P)$

Example An atom may spin counterclockwisely or clockwisely, which is not directly observable. The original true probability space $(S, \mathcal{F}, P)$ for this atom is

$$S = \{\text{counterclockwise}, \text{clockwise}\},$$

$$\mathcal{F} = \Big\{\emptyset, \{\text{counterclockwise}\}, \{\text{clockwise}\}, \{\text{counterclockwise}, \text{clockwise}\}\Big\},$$

and

$$\begin{cases} P(\emptyset) = 0, \\ P(\{\text{counterclockwise}\}) = 0.4, \\ P(\{\text{clockwise}\}) = 0.6, \\ P(\{\text{counterclockwise}, \text{clockwise}\}) = 1. \end{cases}$$

Now an experiment that uses some advanced facilities is performed to examine the spin direction of this atom. (Suppose there is no **observation noise** in this experiment; so a 1-1 correspondence mapping from S to $\Re$ is obtained.) This results in an observable two-value random variable $\boldsymbol{x}$, namely,

$$\boldsymbol{x}(\text{counterclockwise}) = 1 \quad \text{and} \quad \boldsymbol{x}(\text{clockwise}) = -1.$$

Merit of defining random variables based on $(S, \mathcal{F}, P)$

Merit 2: $(S, \mathcal{F}, P)$ may (be too abstract and) be short of the required mathematical structure for manipulation, such as *ordering* (which is the operation required for cdf).

Example (of a random variable $\boldsymbol{x}$, whose inverse function exists)

$$
\begin{aligned}
S &= \{\blacktriangle, \blacktriangledown, \square, \blacksquare, \lozenge, \blacklozenge\} \\
\mathcal{F} &= \text{A } \sigma\text{-field collection of subsets of } S \\
P &= \text{Some assigned probability measure on } \mathcal{F}
\end{aligned}
$$

Define a random variable $\boldsymbol{x}$ on $(S, \mathcal{F}, P)$ as:

$$
\begin{aligned}
\boldsymbol{x}(\blacktriangle) &= 1 \\
\boldsymbol{x}(\blacktriangledown) &= 2 \\
\boldsymbol{x}(\square) &= 3 \\
\boldsymbol{x}(\blacksquare) &= 4 \\
\boldsymbol{x}(\lozenge) &= 5 \\
\boldsymbol{x}(\blacklozenge) &= 6
\end{aligned}
$$

Merit of defining random variables based on $(S, \mathcal{F}, P)$

Examine what subsets should be included in $\mathcal{F}$.

$$\begin{array}{lll}
\text{For } x < 1, & \{\zeta \in S : \boldsymbol{x}(\zeta) \le x\} & = \emptyset \\
\text{For } 1 \le x < 2, & \{\zeta \in S : \boldsymbol{x}(\zeta) \le x\} & = \{\blacktriangle\} \\
\text{For } 2 \le x < 3, & \{\zeta \in S : \boldsymbol{x}(\zeta) \le x\} & = \{\blacktriangle, \blacktriangledown\} \\
\text{For } 3 \le x < 4, & \{\zeta \in S : \boldsymbol{x}(\zeta) \le x\} & = \{\blacktriangle, \blacktriangledown, \square\} \\
\text{For } 4 \le x < 5, & \{\zeta \in S : \boldsymbol{x}(\zeta) \le x\} & = \{\blacktriangle, \blacktriangledown, \square, \blacksquare\} \\
\text{For } 5 \le x < 6, & \{\zeta \in S : \boldsymbol{x}(\zeta) \le x\} & = \{\blacktriangle, \blacktriangledown, \square, \blacksquare, \lozenge\} \\
\text{For } x \ge 6, & \{\zeta \in S : \boldsymbol{x}(\zeta) \le x\} & = \{\blacktriangle, \blacktriangledown, \square, \blacksquare, \lozenge, \blacklozenge\} = S
\end{array}$$

By definition, $\mathcal{F}$ must be a σ-field containing the above seven events or sets.

Note that we can **sort** $1, 2, 3, 4, 5, 6$ (to yield the cdf), but we cannot sort $\blacktriangle, \blacktriangledown, \square, \blacksquare, \lozenge, \blacklozenge$, not to mention the manipulation of $(\blacktriangle + \blacktriangledown)$ or $(\square - \blacksquare)$.

The smallest σ-field containing all subsets of the form $\{\zeta \in S : \boldsymbol{x}(\zeta) \le x\}$ is referred to as the σ-field generated by a random variable $\boldsymbol{x}$, and is usually denoted by $\sigma(\boldsymbol{x})$.

In this example, $\sigma(\boldsymbol{x})$ is the power set of S (since the inverse function of $\boldsymbol{x}(\zeta)$ exists).

Merit of defining random variables based on $(S, \mathcal{F}, P)$

Example (of a random variable $\boldsymbol{y}$ without inverse)

Define a random variable $\boldsymbol{y}$ on $(S, \mathcal{F}, P)$ as:

$$\begin{aligned} \boldsymbol{y}(\blacktriangle) = \boldsymbol{y}(\blacktriangledown) = \boldsymbol{y}(\square) &= 1 \\ \boldsymbol{y}(\blacksquare) = \boldsymbol{y}(\lozenge) = \boldsymbol{y}(\blacklozenge) &= 2 \end{aligned}$$

Examine what subsets must be included in $\mathcal{F}$.

$$\begin{aligned} \text{For } y < 1, \quad & \{\zeta \in S : \boldsymbol{y}(\zeta) \le y\} &= \emptyset \\ \text{For } 1 \le y < 2, \quad & \{\zeta \in S : \boldsymbol{y}(\zeta) \le y\} &= \{\blacktriangle, \blacktriangledown, \square\} \\ \text{For } y \ge 2, \quad & \{\zeta \in S : \boldsymbol{y}(\zeta) \le y\} &= \{\blacktriangle, \blacktriangledown, \square, \blacksquare, \lozenge, \blacklozenge\} = S \end{aligned}$$

Hence, $\mathcal{F}$ must be a σ-field containing the above three events or sets.
Thus, the smallest σ-field generated by $\boldsymbol{y}$ is

$$\sigma(\boldsymbol{y}) = \{\emptyset, \{\blacktriangle, \blacktriangledown, \square\}, \{\blacksquare, \lozenge, \blacklozenge\}, S\}.$$

The introduction of the third merit of defining random variables based on $(S, \mathcal{F}, P)$ will be deferred until the introduction of the definition of random processes.

Definition (Random vectors) A random vector on a probability space $(S, \mathcal{F}, P)$ is a real-valued function $\boldsymbol{x} : S \to \Re^k$ with $\{\zeta \in S : \boldsymbol{x}(\zeta) \le x^k\} \in \mathcal{F}$, where for two real vectors x^k and y^k, $[x^k \le y^k] \triangleq [x_1 \le y_1, x_2 \le y_2, \cdots, x_k \le y_k]$.

It is possible that $x^2 \not\le y^2$ and $x^2 \not\ge y^2$, e.g., $x^2 = (0,1)$ and $y^2 = (1,0)$. However, $\Pr[\boldsymbol{x}(1) < 0 \text{ or } \boldsymbol{x}(2) < 1]$ is well-defined because

$$\Pr[\boldsymbol{x}(1) < 0 \text{ or } \boldsymbol{x}(2) < 1] = \Pr[\boldsymbol{x}(1) < 0] + \Pr[\boldsymbol{x}(2) < 1] - \Pr[\boldsymbol{x}(1) < 0, \boldsymbol{x}(2) < 1].$$

- A random vector is a finite collection of random variables. In fact, each dimension of $\boldsymbol{x}(\zeta) = (\boldsymbol{x}(1,\zeta), \boldsymbol{x}(2,\zeta), \ldots, \boldsymbol{x}(k,\zeta))$ is itself a random variable.
- Hence, an equivalent definition of random vectors is:

  Definition (Random vectors) A random vector is a *finite* collection of random variables, each of which is defined on *the same* probability space.

- Another equivalent definition that can be seen in literatures is:

  Definition (Random vectors) A random vector is an indexed family of random variables $\{\boldsymbol{x}(i), i \in \mathcal{I}\}$, in which each $\boldsymbol{x}(i)$ is defined on the same probability space, and the index set $\mathcal{I}$ is finite.

- **Why requiring each $\boldsymbol{x}(i)$ to be defined on the same or common probability space?**

  Because, through "*defined on the same probability space*," the **joint distribution** of two (or three, four,. . ., etc) random variables can be well-defined.

$$\begin{aligned}&\Pr[\boldsymbol{x}(i) \le x_i \text{ and } \boldsymbol{x}(j) \le x_j]\\ &\quad = P\left(\{\zeta \in S : \boldsymbol{x}(i,\zeta) \le x_i \text{ and } \boldsymbol{x}(j,\zeta) \le x_j\}\right)\\ &\quad = P\left(\{\zeta \in S : \boldsymbol{x}(i,\zeta) \le x_i\} \cap \{\zeta \in S : \boldsymbol{x}(j,\zeta) \le x_j\}\right).\end{aligned}$$

  Then, it can be proved that for any x_i and x_j,

$$\begin{aligned} A_i \triangleq \{\zeta \in S : \boldsymbol{x}(i,\zeta) \le x_i\} &\in \mathcal{F} && \text{because } \boldsymbol{x}(i) \text{ is defined over } (S,\mathcal{F},P)\\ A_j \triangleq \{\zeta \in S : \boldsymbol{x}(j,\zeta) \le x_j\} &\in \mathcal{F} && \text{because } \boldsymbol{x}(j) \text{ is defined over } (S,\mathcal{F},P)\\ A_i^c &\in \mathcal{F} && \mathcal{F} \text{ closure under complement action}\\ A_j^c &\in \mathcal{F} && \mathcal{F} \text{ closure under complement action}\\ A_i^c \cup A_j^c &\in \mathcal{F} && \mathcal{F} \text{ closure under countable union}\\ (A_i^c \cup A_j^c)^c = A_i \cap A_j &\in \mathcal{F} && \mathcal{F} \text{ closure under complement action}\end{aligned}$$

  Hence, $P(A_i \cap A_j)$ is probabilistically measurable.

It can be proved from closures under complement action and countable union that $\mathcal{F}$ is closure under countable intersection.

Example

$$\begin{aligned} S &= \{\blacktriangle, \blacktriangledown, \square, \blacksquare, \Diamond, \blacklozenge\} \\ \mathcal{F} &= \text{A } \sigma\text{-field collection of subsets of } S \\ P &= \text{Some assigned probability measure on } \mathcal{F} \end{aligned}$$

Define a random vector $\{\boldsymbol{x}(i), i \in \{1,2\}\}$ as:

$$\begin{aligned} \boldsymbol{x}(1,\blacktriangle) &= 1; & \boldsymbol{x}(1,\blacksquare) &= 2; & \boldsymbol{x}(2,\blacktriangle) &= 1; & \boldsymbol{x}(2,\blacksquare) &= 2 \\ \boldsymbol{x}(1,\blacktriangledown) &= 2; & \boldsymbol{x}(1,\Diamond) &= 1; & \boldsymbol{x}(2,\blacktriangledown) &= 1; & \boldsymbol{x}(2,\Diamond) &= 2 \\ \boldsymbol{x}(1,\square) &= 1; & \boldsymbol{x}(1,\blacklozenge) &= 2; & \boldsymbol{x}(2,\square) &= 1; & \boldsymbol{x}(2,\blacklozenge) &= 2 \end{aligned}$$

Examine what subsets should be included in $\mathcal{F}$.

$$\begin{aligned} &\text{For } x_1 < 1, & \{\zeta \in S : \boldsymbol{x}(1,\zeta) \le x_1\} &= \emptyset \\ &\text{For } 1 \le x_1 < 2, & \{\zeta \in S : \boldsymbol{x}(1,\zeta) \le x_1\} &= \{\blacktriangle, \square, \Diamond\} \\ &\text{For } x_1 \ge 2, & \{\zeta \in S : \boldsymbol{x}(1,\zeta) \le x_1\} &= \{\blacktriangle, \blacktriangledown, \square, \blacksquare, \Diamond, \blacklozenge\} = S \\ &\text{For } x_2 < 1, & \{\zeta \in S : \boldsymbol{x}(2,\zeta) \le x_2\} &= \emptyset \\ &\text{For } 1 \le x_2 < 2, & \{\zeta \in S : \boldsymbol{x}(2,\zeta) \le x_2\} &= \{\blacktriangle, \blacktriangledown, \square\} \\ &\text{For } x_2 \ge 2, & \{\zeta \in S : \boldsymbol{x}(2,\zeta) \le x_2\} &= \{\blacktriangle, \blacktriangledown, \square, \blacksquare, \Diamond, \blacklozenge\} = S \end{aligned}$$

Hence, $\mathcal{F}$ must be a σ-field containing the above six sets, and both cdfs of $\boldsymbol{x}(1)$ and $\boldsymbol{x}(2)$ are well-defined.

Since $\boldsymbol{x}(1)$ and $\boldsymbol{x}(2)$ are defined on the same probability space $(S, \mathcal{F}, P)$ (in particular, $\mathcal{F}$ must contain the above six sets),

$$\Pr[\boldsymbol{x}(1) \leq x_1 \text{ and } \boldsymbol{x}(2) \leq x_2]$$

is well-defined for any x_1 and x_2. $\square$

We can further extend the random vector to a possibly *infinite* collection of random variables, all of which are defined on the same probability space.

Definition (Random process) A random process is an indexed family of random variables $\{\boldsymbol{x}(t), t \in \mathcal{I}\}$, in which each $\boldsymbol{x}(t)$ is defined on the same probability space.

- Under such a definition, all finite dimensional joint distributions are well-defined because

$$\begin{aligned} &[\boldsymbol{x}(t_1) \le x_1 \text{ and } \boldsymbol{x}(t_2) \le x_2 \text{ and } \cdots \text{ and } \boldsymbol{x}(t_k) \le x_k] \\ &= \{\zeta \in S : \boldsymbol{x}(t_1, \zeta) \le x_1 \text{ and } \boldsymbol{x}(t_2, \zeta) \le x_2 \text{ and } \cdots \text{ and } \boldsymbol{x}(t_k, \zeta) \le x_k\} \\ &= \bigcap_{i=1}^{k} \{\zeta \in S : \boldsymbol{x}(t_i, \zeta) \le x_i\} \end{aligned}$$

  is surely an event by properties of σ-field, and hence, is probabilistically measurable.

- The 3rd merit of defining random processes based on $(S, \mathcal{F}, P)$:
  - All finite (or countably infinite) dimensional joint distributions are well-defined without the tedious process of listing all of them.
- **The converse however is not true**, i.e., it is not necessarily valid that the statistical properties of a real random process are completely determined by providing all finite-dimensional joint distributions for samples.
  - See the counterexample in the next slide.

Example Define random processes $\{\boldsymbol{x}(t), t \in [0,1)\}$ and $\{\boldsymbol{y}(t), t \in [0,1)\}$ as

$$\boldsymbol{x}(t,\zeta) = \begin{cases} 1, & \zeta \neq t; \\ 0, & \zeta = t \end{cases} \quad \text{and} \quad \boldsymbol{y}(t,\zeta) = 1,$$

where $\zeta \in S = [0,1)$. Let $P(A) = \int_A d\alpha$ for any $A \in \mathcal{F}$. Then,

$$\Pr\left[\min_{t\in[0,1)} \boldsymbol{x}(t) < 1\right] \;=\; P\left(\left\{\zeta \in S : \underbrace{\min_{t\in[0,1)} \boldsymbol{x}(t,\zeta)}_{=\boldsymbol{x}(\zeta,\zeta)=0} < 1\right\}\right) = P(S) = 1,$$

but

$$\Pr\left[\min_{t\in[0,1)} \boldsymbol{y}(t) < 1\right] \;=\; P\left(\left\{\zeta \in S : \min_{t\in[0,1)} \boldsymbol{y}(t,\zeta) < 1\right\}\right) = P(\emptyset) = 0.$$

Thus, $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$ have **different** statistical properties.

However, $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$ have **exactly the same** multi-dimensional joint distribution for any samples at $t_1, t_2, \ldots, t_k$ and any k:

$$\begin{aligned}
&\Pr[\boldsymbol{x}(t_1) \le x_1 \text{ and } \boldsymbol{x}(t_2) \le x_2 \text{ and } \cdots \text{ and } \boldsymbol{x}(t_k) \le x_k] \\
&\quad = P\left(\bigcap_{i=1}^{k} \{\zeta \in S : \boldsymbol{x}(t_i, \zeta) \le x_i\}\right) \\
&\quad = \begin{cases} 1, & \min_{1\le i\le k} x_i \ge 1; \\ 0, & \text{otherwise} \end{cases} \\
&\quad = \Pr[\boldsymbol{y}(t_1) \le x_1 \text{ and } \boldsymbol{y}(t_2) \le x_2 \text{ and } \cdots \text{ and } \boldsymbol{y}(t_k) \le x_k],
\end{aligned}$$

where

$$\{\zeta \in S : \boldsymbol{x}(t_i, \zeta) \le x_i\} = \begin{cases} S, & x_i \ge 1; \\ \{t_i\}, & x_i < 1. \end{cases}$$

Notably,

$$\min_{t\in[0,1)} \boldsymbol{x}(t) \quad \text{and} \quad \min_{t\in[0,1)} \boldsymbol{y}(t)$$

are random variables defined via "uncountably infinite dimensional distributions of $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$." □

Definition (Complex random process) A complex random process is specified in terms of two real random processes defined over the same probability space.

- Note that mathematical manipulation of the complex domain, such as sorting, is undefined!
- Hence, we cannot define by letting $\mathcal{C}$ be the set of all complex number that:

  Definition (Complex random variable) A complex random variable on a probability space $(S, \mathcal{F}, P)$ is a complex-valued function $\boldsymbol{x}(\zeta)$ (i.e., $\boldsymbol{x} : S \to \mathcal{C}$) with $\{\zeta \in S : \underbrace{\boldsymbol{x}(\zeta) \le x}_{\text{undefined}}\} \in \mathcal{F}$ for every $x \in \mathcal{C}$.

- This is the reason why a complex random variable, vector or process should be treated as two real random variables, vectors or processes that are defined over the same probability space.

Calculation of Mean under $(S, \mathcal{F}, P)$

Question: Define a random variable $\boldsymbol{y}$ on $(S, \mathcal{F}, P)$ as:

$$\begin{array}{rclrcl} \boldsymbol{y}(\blacktriangle) &=& 1 & \boldsymbol{y}(\blacksquare) &=& 2 \\ \boldsymbol{y}(\blacktriangledown) &=& 1 & \boldsymbol{y}(\Diamond) &=& 2 \\ \boldsymbol{y}(\square) &=& 1 & \boldsymbol{y}(\blacklozenge) &=& 2 \end{array}$$

where

$$\begin{array}{rcl} S &=& \{\blacktriangle, \blacktriangledown, \square, \blacksquare, \Diamond, \blacklozenge\} \\ \mathcal{F} &=& \{\emptyset, \{\blacktriangle, \blacktriangledown, \square\}, \{\blacksquare, \Diamond, \blacklozenge\}, S\} \\ P &=& \{0, 1/2, 1/2, 1\} \end{array}$$

Please calculate $E[\boldsymbol{y}]$.

Answer:

$$\begin{aligned} E[\boldsymbol{y}] &= \int_S \boldsymbol{y}(\zeta)\, dP(\zeta) \\ &= \boldsymbol{y}(\blacktriangle)P(\blacktriangle) + \boldsymbol{y}(\blacktriangledown)P(\blacktriangledown) + \boldsymbol{y}(\square)P(\square) + \boldsymbol{y}(\blacksquare)P(\blacksquare) + \boldsymbol{y}(\Diamond)P(\Diamond) + \boldsymbol{y}(\blacklozenge)P(\blacklozenge) \end{aligned}$$

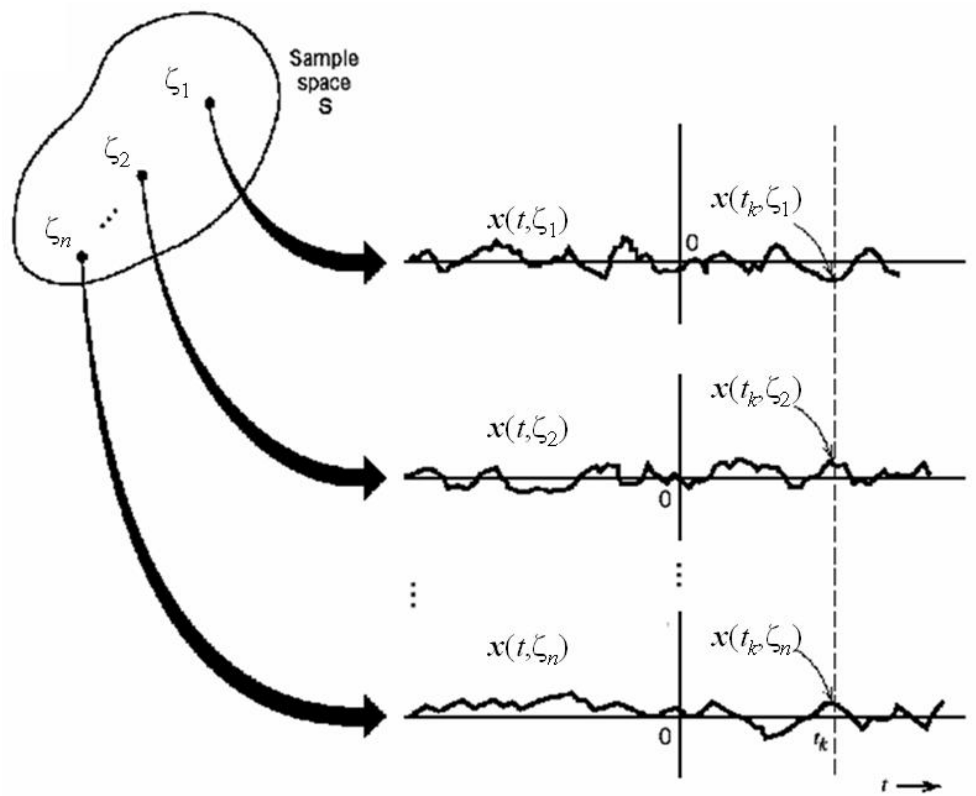
(Yet, we do not know the probability of, say, $P(\blacktriangle)$; how can we calculate $E[\boldsymbol{y}]$!)

$$\begin{aligned} &= 1 \times P(\{\blacktriangle, \blacktriangledown, \square\}) + 2 \times P(\{\blacksquare, \Diamond, \blacklozenge\}) \\ &= 1 \times (1/2) + 2 \times (1/2) = 3/2. \end{aligned}$$

□

Final note on the definition of real random process

- $\boldsymbol{x}(t, \zeta)$ is a **deterministic** function of t for fixed ζ, and is a real number for fixed t and ζ.
- while $\boldsymbol{x}(t)$ is **random** at any instant t.



Classifications of Random Processes

Classification according to $\mathcal{I}$ in $\{\boldsymbol{x}(t), t \in \mathcal{I}\}$

- $\mathcal{I} = \Re$: Continuous-time random processes.
- $\mathcal{I} =$ set of integers: Discrete-time random processes.

Classification according to *number of states* for $\boldsymbol{x}(t)$

- if $\boldsymbol{x}(t, S) \triangleq \{x \in \Re : \boldsymbol{x}(t, \zeta) = x \text{ for some } \zeta \in S\}$ is a set with countable number of elements, $\boldsymbol{x}(t)$ is a discrete-state random process.
- if $\boldsymbol{x}(t, S)$ is not countable, $\boldsymbol{x}(t)$ is a continuous-state random process.

First-order Distribution and Density

Definition (First-order distribution) The first-order distribution function of a random process $\boldsymbol{x}(t)$ is defined as $F(x;t) \triangleq \Pr[\boldsymbol{x}(t) \le x]$.

Theorem 14.1 (in Patrick Billingsley, *Probability and Measure*, 3rd Edition, Wiley, 1995) If a function $F(\cdot)$ is non-decreasing, right-continuous and satisfies $\lim_{x \downarrow -\infty} F(x) = 0$ and $\lim_{x \uparrow \infty} F(x) = 1$, then there exists a random variable and a probability space such that the cdf of the random variable defined over the probability space is equal to $F(\cdot)$.

Theorem 14.1 releases us with the burden of referring to a probability space in defining a random variable. We can indeed define a random variable $\boldsymbol{x}$ directly by its distribution, i.e., $\Pr[\boldsymbol{x} \le x]$. Nevertheless, it is better to keep in mind (and learn) that a formal mathematical notion of random variable is defined over some probability space.

Notably, Theorem 14.1 only proves the "existence" but not the "uniqueness".

First-order Distribution and Density

Remark.

Although random variables and random vectors **can be well-defined** by explicitly listing all the joint distributions without mentioning the inherited probability space (cf. Theorem 14.1 in the previous slide), random processes **cannot be well-defined** by explicitly providing all the joint distributions of finite samples **from rigorous mathematical standpoint**. The key reason is that some statistical property (e.g., $\min_{t\in[0,1)} \boldsymbol{x}(t)$) cannot be uniquely determined simply from the knowledge of joint distributions of finite samples.

Yet, **from the engineering standpoint**, as long as those statistical properties that an engineer is interested in can all be defined (e.g., mean function and autocorrelatino function), a random process is "well-defined"! An example can be found in Slide 9-47 where the Poisson process is defined without introducing its inherited probability space.

First-order Distribution and Density

Definition (First-order density) The first-order density function of a random process $\boldsymbol{x}(t)$ is defined as

$$f(x;t) \triangleq \frac{\partial F(x;t)}{\partial x},$$

provided that $F(x,t)$ is differentiable with respect to x, and $\boldsymbol{x}(t)$ has density.

- It is possible that $F(x,t)$ is not differentiable with respect to x, or $\boldsymbol{x}(t)$ has no density.

Definition (Probability density function) A random variable $\boldsymbol{x}$ and its distribution (cdf) have *density* f, if f is a non-negative function that satisfies

$$\Pr[\boldsymbol{x} \in A] = \int_A f(x)dx$$

for every $A \subset \Re$ satisfying that A can be obtained by repeating countable set-theoretic operations (mostly often, union) of open, semi-open and closed intervals.

Remarks on Borel sets

- A (in the previous definition) is called a Borel set.
- Lebesque measure λ is only defined on Borel sets.

  Definition (Lebesque measure) A Lebesque measure λ over the Borel sets is that for any Borel set A,
  $$\lambda(A) = \sum_{i=1}^{\infty} \lambda(I_i),$$
  and $\{I_i\}_{i=1}^{\infty}$ are disjoint intervals satisfying $A = \cup_{i=1}^{\infty} I_i$, and $\lambda(I)$ is equal to the right-margin of interval I minus the left-margin of the same interval.

- Hence, the largest manageable probability space is perhaps
  $$(S = [0,1), \mathcal{B}, P = \text{Lebesque measure}),$$
  where $\mathcal{B}$ is the σ-field containing all open, semi-open, closed intervals in S.

First-order Distribution and Density

Remarks on pdf

- A random variable $\boldsymbol{x}$ always has cdf F.
- If a random variable has density f, then $f(x) = \partial F(x)/\partial(x)$.
- $\partial F(x)/\partial x$ is not necessarily a density. In other words, if $f(x) = \partial F(x)/\partial x$, it may not be true that

$$\Pr[\boldsymbol{x} \in A] = \int_A f(x)dx$$

for every Borel set $A \subset \Re$.

Second-order Distribution and Density

Definition (Second-order distribution) The second-order distribution function of a random process $\boldsymbol{x}(t)$ is defined as

$$F(x_1, x_2; t_1, t_2) \triangleq \Pr[\boldsymbol{x}(t_1) \le x_1 \text{ and } \boldsymbol{x}(t_2) \le x_2].$$

Definition (Second-order density) The second-order density function of a random process $\boldsymbol{x}(t)$ is defined as

$$f(x_1, x_2; t_1, t_2) \triangleq \frac{\partial^2 F(x_1, x_2; t_1, t_2)}{\partial x_1 \partial x_2},$$

provided that $F(x_1, x_2; t_1, t_2)$ is differentiable with respect to x_1, x_2, and $\boldsymbol{x}(t)$ has second-order density at t_1, t_2.

- **Consistency condition:** For any $t_2 \neq t_1$,

  $$F(x_1; t_1) = F(x_1, \infty; t_1, t_2) \quad \text{and} \quad f(x_1; t_1) = \int_{-\infty}^{\infty} f(x_1, x_2; t_1, t_2) dx_2.$$

  This condition is always valid if a random process $\boldsymbol{x}(t)$ is defined over a probability space. However, this condition may need to be explicitly taken care of, if a random variable is defined explicitly without inherited probability space.

- The nth-order distribution and density can be defined similarly. Consistency condition should always be preserved.

First-order and Second-order Properties

Mean: The mean $\eta_x(t)$ of $\boldsymbol{x}(t)$ is $\eta_x(t) \triangleq E[\boldsymbol{x}(t)] = \int_{-\infty}^{\infty} x\, f(x;t)dx$.

Autocorrelation: The autocorrelation $R_{xx}(t_1, t_2)$ of $\boldsymbol{x}(t)$ is

$$R_{xx}(t_1, t_2) \triangleq E[\boldsymbol{x}(t_1)\boldsymbol{x}^*(t_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2^*\, f(x_1, x_2; t_1, t_2) dx_1 dx_2.$$

The autocorrelation function $R_{xx}(t_1, t_2)$ of a random process $\boldsymbol{x}(t)$ is a *positive definite* (p.d.) (non-negative definite? See the red-color note below.) function, namely,

$$\sum_i \sum_j a_i a_j^* R_{xx}(t_i, t_j) \geq 0 \quad \text{for any complex } a_i \text{ and } a_j. \qquad (9.1)$$

Proof:

$$0 \leq E\left[\left|\sum_i a_i \boldsymbol{x}(t_i)\right|^2\right] = \sum_i \sum_j a_i a_j^* E[\boldsymbol{x}(t_i)\boldsymbol{x}^*(t_j)].$$

□

The converse that any p.d. function can be the autocorrelation function of some random process is also true (cf. Existence Theorem in Slide 9-42).

Strictly speaking, p.d. = Equality for (9.1) is valid **only** when $\vec{a} = \vec{0}$.

First-order and Second-order Properties

(p. 122 in *Random Processes: A Mathematical Approach for Engineers* by Robert M. Gray and Lee D. Davisson) ... By positive definite, we mean that for any dimension k, any collection of sample times $t_0, t_1, \ldots, t_{k-1}$, and any non-zero real vector $(r_0, r_1, \ldots, r_{k-1})$ we have

$$\sum_{i=0}^{k-1}\sum_{j=0}^{k-1} r_i \cdot \Lambda(t_i, t_j) \cdot r_j > 0,$$

(where $\Lambda(t,s) = \Lambda(s,t)$ is some symmetric function.)

Average Power: The average power of $\boldsymbol{x}(t)$ at time t is

$$E[\boldsymbol{x}(t)\boldsymbol{x}^*(t)] \triangleq E[|\boldsymbol{x}(t)|^2] = R_{xx}(t,t) \geq 0.$$

Autocovariance: The autocovariance $C_{xx}(t_1, t_2)$ of $\boldsymbol{x}(t)$ is:

$$C_{xx}(t_1,t_2) \triangleq E[(\boldsymbol{x}(t_1) - \eta_x(t_1))(\boldsymbol{x}(t_2) - \eta_x(t_2))^*] = R_{xx}(t_1,t_2) - \eta_x(t_1)\eta_x^*(t_2).$$

Variance: The variance of $\boldsymbol{x}(t)$ is $E[|\boldsymbol{x}(t) - \eta_x(t)|^2] = C_{xx}(t,t)$.

Correlation Coefficient:

$$r_{xx}(t_1,t_2) \triangleq \frac{C_{xx}(t_1,t_2)}{\sqrt{C_{xx}(t_1,t_1)C_{xx}(t_2,t_2)}} \in [-1, 1].$$

Both autocovariance and correlation coefficient functions are p.d. (i.e., n.n.d.)

First-order and Second-order Properties

Cross Correlation: The cross-correlation of two processes $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$ is

$$R_{xy}(t_1, t_2) \triangleq E[\boldsymbol{x}(t_1)\boldsymbol{y}^*(t_2)].$$

Cross Covariance: The cross-covariance of two processes $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$ is

$$C_{xy}(t_1, t_2) \triangleq E[(\boldsymbol{x}(t_1) - \eta_x(t_1))(\boldsymbol{y}(t_2) - \eta_y(t_2))^*] = R_{xy}(t_1, t_2) - \eta_x(t_1)\eta_y^*(t_2).$$

General Properties

Independence: Two processes $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$ are *independent* if any finite dimensional samples of $\boldsymbol{x}(t)$ is independent of any finite dimensional samples of $\boldsymbol{y}(t)$.

Comment: Since the multi-dimensional distributions do not completely determine the statistical properties of a random process, it may be "*restricted*" to define *independence* only based on multi-dimensional samples. For example, whether or not $\min_{t\in[0,1)} \boldsymbol{x}(t)$ and $\min_{t\in[0,1)} \boldsymbol{y}(t)$ are independent is not clear under such definition!

Orthogonality: Two processes $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$ are (mutually) *orthogonal* if

$$R_{xy}(t_1, t_2) = 0$$

for every $t_1, t_2 \in \mathcal{I}$.

Uncorrelation: Two processes $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$ are *uncorrelated* if

$$C_{xy}(t_1, t_2) = 0$$

for every $t_1, t_2 \in \mathcal{I}$.

a-dependance: A random process $\boldsymbol{x}(t)$ is *a-dependent* if the two processes $\{\boldsymbol{x}(t), t < t_0\}$ and $\{\boldsymbol{x}(t), t > t_0 + a\}$ are independent for any t_0.

General Properties

Correlation a-dependence: A random process is *correlation a-dependent* if $C_{xx}(t_1, t_2) = 0$ for $|t_1 - t_2| > a$.

Strictly White: A process $\boldsymbol{x}(t)$ is *strictly white* if $\boldsymbol{x}(t_1)$ and $\boldsymbol{x}(t_2)$ are *independent* for every $t_1 \neq t_2$.

White: A process $\boldsymbol{x}(t)$ is *white* if $\boldsymbol{x}(t_1)$ and $\boldsymbol{x}(t_2)$ are *uncorrelated* for every $t_1 \neq t_2$.

Hence, $C_{xx}(t, t+\tau) = q(t)\delta(\tau)$ for a white process, which indicates that it is in general time-varying in nature (with Doppler spectrum $\int_{-\infty}^{\infty} q(t)e^{-j\lambda t}dt$).

General Properties

A few notes on white processes (Comparison with other texts)

- *J. Proakis,* Digital Communications, *McGraw-Hill, fourth edition, 2001.*

> *(p. 77) The autocorrelation function of a stochastic process $X(t)$ is*
>
> $$\phi_{xx}(\tau) = \frac{1}{2}N_0\delta(\tau).$$
>
> *Such a process is called* white noise. *. . .*

> *(p. 157) White noise is a stochastic process that is defined to have a flat (constant) power spectral density over the entire frequency range. . . .*

> *(p. 62) The function $\phi(t_1, t_2)$ is called the* autocorrelation function *of the stochastic process. . . . (p. 66) A stationary stochastic process . . .*
>
> $$\Phi(f) = \int_{-\infty}^{\infty} \phi(\tau)e^{-j2\pi f\tau}d\tau \qquad (2.2\text{-}16)$$
>
> *. . . Therefore, $\Phi(f)$ is called the* power density spectrum *of the stochastic process.*

Comments (for stationary processes)

– *It is good to define the* **power density spectrum** *as the Fourier transform of the* **autocorrelation function** *because its integration is really equal to the power.*

– *However, since for WSS processes,*

$$R_{xx}(\tau) = C_{xx}(\tau) + \mu_x^2.$$

The power density spectrum (defined based on the autocorrelation) of a white process will have an impulse $\mu_x^2\delta(f)$ *at the origin.*

– *Strictly speaking, a white process "must" be zero-mean, otherwise* $\Phi(f) =$*constant only when* $f \neq 0$*!* *(*$\Phi(0) = \infty$ *for any non-zero mean process.)*

General Properties

Robert M. Gray and Lee D. Davisson, Random Processes: A Mathematical Approach for Engineers, *Prentice-Hall, 1986.*

(p. 197) A random process $\{X_t\}$ is said to be white *if its power spectral density is a constant for all f.*

(p. 193) The power spectral density $S_X(f)$ *of the process is defined as the Fourier transform of the (auto-)covariance function; . . .*

- *The integration of its* **power spectral density** *is not the power of a* **non-zero-mean** *process.*
- *However, such a definition allows the* **existence** *of a non-zero-mean white process. Hence, the authors wrote in parentheses that:*

  (p. 197) A white process is also almost always assumed to have a zero mean, an assumption that we will make unless explicitly stated otherwise. . . .

General Properties

In our textbook:

White: *A process* $\boldsymbol{x}(t)$ *is* white *if* $\boldsymbol{x}(t_1)$ *and* $\boldsymbol{x}(t_2)$ *are* uncorrelated *for every* $t_1 \neq t_2$.

- *Hence, implicitly, a WSS process is white if its power density spectrum (defined as the Fourier transform of the autocorrelation function) is constant except at the origin.*
- *Why introducing such an* **indirect** *definition? Because it parallels the subsequent definition of* **strictly white**.

Strictly White: *A process* $\boldsymbol{x}(t)$ *is* strictly white *if* $\boldsymbol{x}(t_1)$ *and* $\boldsymbol{x}(t_2)$ *are* independent *for every* $t_1 \neq t_2$.

- *Notably, one cannot differentiate the (weakly) white process and strictly white process from their power density spectra.*

General Properties

Independent Increment: A process $\boldsymbol{x}(t)$ is a process with *independent increment* if $\boldsymbol{x}(t_2) - \boldsymbol{x}(t_1)$ and $\boldsymbol{x}(t_4) - \boldsymbol{x}(t_3)$ are independent for any $t_1 < t_2 < t_3 < t_4$.

Example. The Poisson process introduced later (cf. Slide 9-47) is a process with independent increment.

Uncorrelated Increment: A process $\boldsymbol{x}(t)$ is a process with *uncorrelated increment* if $\boldsymbol{x}(t_2) - \boldsymbol{x}(t_1)$ and $\boldsymbol{x}(t_4) - \boldsymbol{x}(t_3)$ are uncorrelated for any $t_1 < t_2 < t_3 < t_4$.

General Properties

Normal: A process $\boldsymbol{x}(t)$ is called *normal* if any finite dimensional samples of $\boldsymbol{x}(t)$ are jointly normal.

Theorem (Existence theorem) Given an arbitrary function $\eta(t)$ and a p.d. (i.e., n.n.d.) function $C(t_1, t_2)$, there exists a normal process $\boldsymbol{x}(t)$ with mean $\eta(t)$ and auto-covariance function $C(t_1, t_2)$.

- Idea behind the proof: The characteristic function of any finite dimensional samples can be given as:

$$\exp\left\{ j \sum_i \eta(t_i)\omega_i - \frac{1}{2} \sum_{i,k} C(t_i, t_k)\omega_i\omega_k \right\}.$$

Riemann Integral Stated in Example 9-3

Define $\boldsymbol{s} = \int_a^b \boldsymbol{x}(t)dt$ of a random process $\boldsymbol{x}(t)$.

"Interpreting the above as a Riemann integral" yields:

$$E[\boldsymbol{s}] = \int_a^b E[\boldsymbol{x}(t)]dt = \int_a^b \eta_x(t)dt$$

and

$$E[\boldsymbol{s}^2] = \int_a^b \int_a^b E[\boldsymbol{x}(t_1)\boldsymbol{x}(t_2)]dt_1 dt_2 = \int_a^b \int_a^b R_{xx}(t_1, t_2)dt_1 dt_2.$$

. . . .

Riemann Integral Versus Lebesque Integral

Riemann integral:

Let $s(x)$ represent a step function on $[a, b)$, which is defined as that there exists a partition $a = x_0 < x_1 < \cdots < x_n = b$ such that $s(x)$ is constant during (x_i, x_{i+1}) for $0 \leq i < n$.
If a function $f(x)$ is Riemann integrable,

$$\int_a^b f(x)dx \triangleq \sup_{\left\{s(x)\ :\ s(x)\leq f(x)\right\}} \int_a^b s(x)dx = \inf_{\left\{s(x)\ :\ s(x)\geq f(x)\right\}} \int_a^b s(x)dx.$$

Example of a non-Riemann-integrable function:
$f(x) = 0$ if x is irrational; $f(x) = 1$ if x is rational.
Then

$$\sup_{\left\{s(x)\ :\ s(x)\leq f(x)\right\}} \int_a^b s(x)dx = 0,$$

but

$$\inf_{\left\{s(x)\ :\ s(x)\geq f(x)\right\}} \int_a^b s(x)dx = (b-a).$$

Riemann Integral Versus Lebesque Integral

Lebesque integral:

Let $t(x)$ represent a **simple** function, which is defined as the linear combination of indicator functions for finitely many, mutually disjoint partitions.

For example, let $\mathcal{U}_1, \ldots, \mathcal{U}_m$ be mutually disjoint partitions of the domain $\mathcal{X}$ and $\cup_{i=1}^m \mathcal{U}_i = \mathcal{X}$. The indicator function of $\mathcal{U}_i$ satisfies $\mathbf{1}(x; \mathcal{U}_i) = 1$ if $x \in \mathcal{U}_i$, and 0, otherwise.

Then $t(x) = \sum_{i=1}^m a_i \mathbf{1}(x; \mathcal{U}_i)$ is a simple function (and $\int_{\mathcal{X}} t(x)dx = \sum_{i=1}^m a_i \cdot \lambda(\mathcal{U}_i)$, where $\lambda(\cdot)$ is a Lebesque measure).
If a function $f(x)$ is Lebesque integrable, then

$$\int_a^b f(x)dx = \sup_{\left\{t(x)\ :\ t(x) \leq f(x)\right\}} \int_a^b t(x)dx = \inf_{\left\{t(x)\ :\ t(x) \geq f(x)\right\}} \int_a^b t(x)dx.$$

The previous example is actually Lebesque integrable, and its Lebesque integral is equal to zero.

Point Processes and Renewal Processes

Point Processes: A point process is a set of random points $\boldsymbol{t}_i$ on the time axis.

Renewal Processes: A renewal process consists of the renewal intervals of a point process, namely, $\boldsymbol{z}_n = \boldsymbol{t}_n - \boldsymbol{t}_{n-1}$.

An example is that $\boldsymbol{z}_i$ is the lifetime of the ith renewed lightbulb which was replaced as soon as the $(i-1)$th renewed lightbulb failed.

Counting processes: A counting process $\boldsymbol{x}(t)$ collects the number of random points that occur during $[0, t)$.

Some Results about Poisson

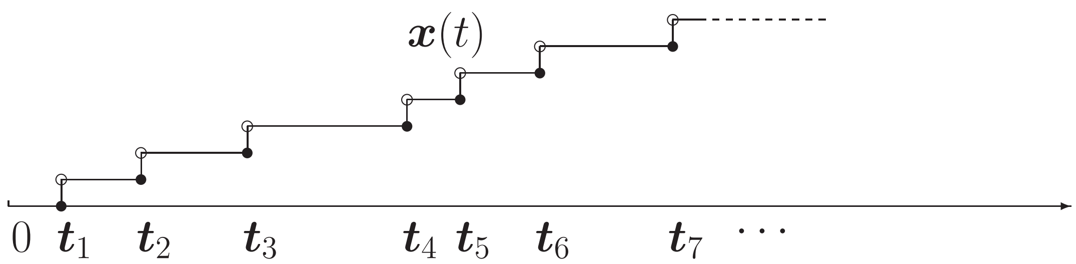
Example 9-5: Poisson process. An example to define a random process without the inherited probability space. Assume:

1. the number of Poisson point occurrences at $\{\boldsymbol{t}_1, \boldsymbol{t}_2, \boldsymbol{t}_3, \cdots\}$ in an interval $[t_1, t_2)$ is a Poisson random variable with parameter $\nu(t_1, t_2) \triangleq \int_{t_1}^{t_2} \lambda(t)dt$, i.e.,

$$\Pr\{\boldsymbol{n}[t_1, t_2) = k\} = \frac{e^{-\nu(t_1,t_2)}\left[\nu[t_1, t_2)\right]^k}{k!},$$

2. and $\boldsymbol{n}[t_1, t_2)$ and $\boldsymbol{n}[t_3, t_4)$ are independent if $[t_1, t_2)$ and $[t_3, t_4)$ are non-overlapping intervals.

Please determine the mean and autocorrelation function of $\boldsymbol{x}(t) \triangleq \boldsymbol{n}[0, t)$.



Answer:

$$\mu_x(t) = E[\boldsymbol{x}(t)] = E[\boldsymbol{n}[0,t)] = \int_0^t \lambda(t)dt.$$

For $t_1 \leq t_2$,

$$\begin{aligned}
R_{xx}(t_1,t_2) &= E[\boldsymbol{x}(t_1)\boldsymbol{x}^*(t_2)] \\
&= E[\boldsymbol{n}[0,t_1)\boldsymbol{n}[0,t_2)] \\
&= E\{\boldsymbol{n}[0,t_1)[\boldsymbol{n}[0,t_1) + \boldsymbol{n}[t_1,t_2)]\} \\
&= E[\boldsymbol{n}^2[0,t_1)] + E[\boldsymbol{n}[0,t_1)\boldsymbol{n}[t_1,t_2)] \\
&= E[\boldsymbol{n}^2[0,t_1)] + E[\boldsymbol{n}[0,t_1)]E[\boldsymbol{n}[t_1,t_2)] \quad \text{(by independence of } \boldsymbol{n}[0,t_1) \text{ and } \boldsymbol{n}[t_1,t_2)) \\
&= \left(\int_0^{t_1} \lambda(t)dt + \left(\int_0^{t_1} \lambda(t)dt\right)^2\right) + \left(\int_0^{t_1} \lambda(t)dt\right)\left(\int_0^{t_2} \lambda(t)dt - \int_0^{t_1} \lambda(t)dt\right) \\
&= \int_0^{t_1} \lambda(t)dt + \int_0^{t_1}\int_0^{t_2} \lambda(t)\lambda(s)dtds.
\end{aligned}$$

Similarly, for $t_1 > t_2$,

$$R_{xx}(t_1,t_2) = \int_0^{t_2} \lambda(t)dt + \int_0^{t_1}\int_0^{t_2} \lambda(t)\lambda(s)dtds.$$

Some Results about Poisson

Therefore,

$$R_{xx}(t_1, t_2) = \int_0^{\min\{t_1,t_2\}} \lambda(t)dt + \int_0^{t_1}\int_0^{t_2} \lambda(t)\lambda(s)dtds.$$

If $\lambda(t)$ is a constant λ, then

$$R_{xx}(t_1, t_2) = \lambda \cdot \min\{t_1, t_2\} + \lambda^2 t_1 t_2.$$

$\square$

Operational meaning of autocorrelation function: The autocorrelation function quantifies the correlation of a data point with a previous data point (or, a future data point).

$$\begin{aligned} C_{xx}(t_1, t_2) &= R_{xx}(t_1, t_2) - \mu_x(t_1)\mu_x(t_2) \\ &= \left[\lambda \cdot \min\{t_1, t_2\} + \lambda^2 t_1 t_2\right] - (\lambda t_1)(\lambda t_2) = \lambda \cdot \min\{t_1, t_2\} \end{aligned}$$

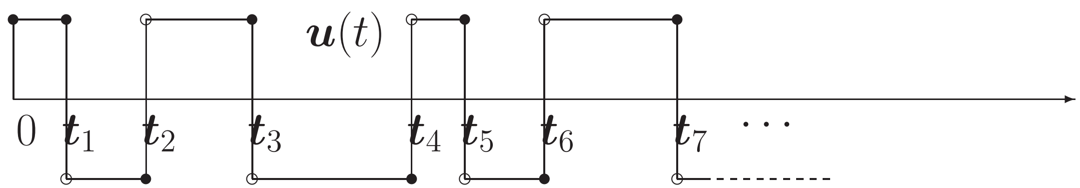
For a present point (e.g., t_1), if its autocorrelation with a distant future point (e.g., $t_2 > t_1$) does not die away, the delayed point must have a strong correlation with an earlier version of itself (e.g., $\boldsymbol{n}[0, t_2)$ is apparently affected strongly by $\boldsymbol{n}[0, t_1)$).

Example 9-6: Semirandom Telegraph Signal

Following Example 9-5 under $\lambda(t) = \lambda$, we re-define:

$$\boldsymbol{u}(t) = \begin{cases} 1, & \text{if } \boldsymbol{n}[0,t) \text{ is even;} \\ -1, & \text{if } \boldsymbol{n}[0,t) \text{ is odd.} \end{cases}$$

Please determine mean and autocorrelation functions of $\boldsymbol{u}(t)$.



Answer:

$$\begin{aligned} E[\boldsymbol{u}(t)] &= 1 \cdot \Pr[\boldsymbol{n}[0,t) = 0, 2, 4, \cdots] + (-1) \cdot \Pr[\boldsymbol{n}[0,t) = 1, 3, 5, \cdots] \\ &= 1 \cdot e^{-\lambda t}\left[1 + \frac{(\lambda t)^2}{2!} + \cdots\right] + (-1) \cdot e^{-\lambda t}\left[\lambda t + \frac{(\lambda t)^3}{3!} + \cdots\right] \\ &= e^{-\lambda t}\cosh(\lambda t) - e^{-\lambda t}\sinh(\lambda t) \\ &= e^{-\lambda t}\left(\frac{e^{\lambda t} + e^{-\lambda t}}{2}\right) - e^{-\lambda t}\left(\frac{e^{\lambda t} - e^{-\lambda t}}{2}\right) \\ &= e^{-2\lambda t}. \end{aligned}$$

Some Results about Poisson

For $t_1 \le t_2$,

$$
\begin{aligned}
&E[\boldsymbol{u}(t_1)\boldsymbol{u}^*(t_2)] \\
&= \Pr[\boldsymbol{n}[0,t_1) = \text{even} \wedge \boldsymbol{n}[0,t_2) = \text{even}] + \Pr[\boldsymbol{n}[0,t_1) = \text{odd} \wedge \boldsymbol{n}[0,t_2) = \text{odd}] \\
&\quad - \Pr[\boldsymbol{n}[0,t_1) = \text{even} \wedge \boldsymbol{n}[0,t_2) = \text{odd}] - \Pr[\boldsymbol{n}[0,t_1) = \text{odd} \wedge \boldsymbol{n}[0,t_2) = \text{even}] \\
&= \Pr[\boldsymbol{n}[0,t_1) = \text{even} \wedge \boldsymbol{n}[t_1,t_2) = \text{even}] + \Pr[\boldsymbol{n}[0,t_1) = \text{odd} \wedge \boldsymbol{n}[t_1,t_2) = \text{even}] \\
&\quad - \Pr[\boldsymbol{n}[0,t_1) = \text{even} \wedge \boldsymbol{n}[t_1,t_2) = \text{odd}] - \Pr[\boldsymbol{n}[0,t_1) = \text{odd} \wedge \boldsymbol{n}[t_1,t_2) = \text{odd}] \\
&= \Pr[\boldsymbol{n}[0,t_1) = \text{even}]\Pr[\boldsymbol{n}[t_1,t_2) = \text{even}] + \Pr[\boldsymbol{n}[0,t_1) = \text{odd}]\Pr[\boldsymbol{n}[t_1,t_2) = \text{even}] \\
&\quad - \Pr[\boldsymbol{n}[0,t_1) = \text{even}]\Pr[\boldsymbol{n}[t_1,t_2) = \text{odd}] - \Pr[\boldsymbol{n}[0,t_1) = \text{odd}]\Pr[\boldsymbol{n}[t_1,t_2) = \text{odd}] \\
&= (\Pr[\boldsymbol{n}[0,t_1) = \text{even} + \Pr[\boldsymbol{n}[0,t_1) = \text{odd}])(\Pr[\boldsymbol{n}[t_1,t_2) = \text{even}] - \Pr[\boldsymbol{n}[t_1,t_2) = \text{odd}]) \\
&= \Pr[\boldsymbol{n}[t_1,t_2) = \text{even}] - \Pr[\boldsymbol{n}[t_1,t_2) = \text{odd}] \\
&= e^{-\lambda(t_2-t_1)}\cosh[\lambda(t_2-t_1)] - e^{-\lambda(t_2-t_1)}\sinh[\lambda(t_2-t_1)] \\
&= e^{-\lambda(t_2-t_1)}\left(\frac{e^{\lambda(t_2-t_1)} + e^{-\lambda(t_2-t_1)}}{2}\right) - e^{-\lambda(t_2-t_1)}\left(\frac{e^{\lambda(t_2-t_1)} - e^{-\lambda(t_2-t_1)}}{2}\right) \\
&= e^{-2\lambda(t_2-t_1)}.
\end{aligned}
$$

Similarly, for $t_1 > t_2$,

$$E[\boldsymbol{u}(t_1)\boldsymbol{u}^*(t_2)] = e^{-2\lambda(t_1-t_2)}.$$

Some Results about Poisson

Therefore,

$$R_{uu}(t_1, t_2) = E[\boldsymbol{u}(t_1)\boldsymbol{u}^*(t_2)] = e^{-2\lambda|t_1 - t_2|}.$$

$\square$

Remarks

- $\boldsymbol{u}(t)$ is named *semirandom telegraph signal* because $\boldsymbol{u}(0) = 1$ is deterministic.
- A (fully) *random telegraph signal* can be formed by $\boldsymbol{v}(t) = \boldsymbol{a} \cdot \boldsymbol{u}(t)$, where $\boldsymbol{a}$ is independent of $\boldsymbol{u}(t)$, and $\boldsymbol{a} = +1$ and $\boldsymbol{a} = -1$ with equal probability.
- It can be shown that the mean of $\boldsymbol{v}(t)$ is zero, and the autocorrelation function of $\boldsymbol{v}(t)$ is the same as that of $\boldsymbol{u}(t)$.
- Indeed, in comparison of the statistics of $\boldsymbol{u}(t)$ and $\boldsymbol{v}(t)$,

$$\Pr[\boldsymbol{u}(t) = 1] = e^{-\lambda t}\cosh(\lambda t) = \frac{1}{2} + \frac{1}{2}e^{-2\lambda t} \quad \xrightarrow{t} \quad \Pr[\boldsymbol{v}(t) = 1] = \frac{1}{2}$$

$$\Pr[\boldsymbol{u}(t) = -1] = e^{-\lambda t}\sin(\lambda t) = \frac{1}{2} - \frac{1}{2}e^{-2\lambda t} \quad \xrightarrow{t} \quad \Pr[\boldsymbol{v}(t) = -1] = \frac{1}{2}$$

Hence, $\boldsymbol{u}(t)$ and $\boldsymbol{v}(t)$ have asymptotically equal statistics.

Sum and difference of Poisson processes

- It is easy to show that the **sum**, $\boldsymbol{z}(t) = \boldsymbol{x}_1(t) + \boldsymbol{x}_2(t)$, of two independent Poisson processes, $\boldsymbol{x}_1(t) \sim$Poisson$(\lambda_1 t)$ and $\boldsymbol{x}_2(t) \sim$Poisson$(\lambda_2 t)$, is Poisson$((\lambda_1 + \lambda_2)t)$.

- However, the **difference**, $\boldsymbol{y}(t) = \boldsymbol{x}_1(t) - \boldsymbol{x}_2(t)$, of two independent Poisson processes is not Poisson! Its statistics is computed as follows.

$$\begin{aligned}
\Pr[\boldsymbol{y}(t) = n] &= \sum_{k=\max\{0,-n\}}^{\infty} \Pr[\boldsymbol{x}_1(t) = n+k] \Pr[\boldsymbol{x}_2(t) = k] \\
&= \sum_{k=\max\{0,-n\}}^{\infty} e^{-\lambda_1 t}\frac{(\lambda_1 t)^{n+k}}{(n+k)!} e^{-\lambda_2 t}\frac{(\lambda_2 t)^k}{k!} \qquad (\text{Let } \tilde{k} = k - \max\{0,-n\}) \\
&= e^{-(\lambda_1+\lambda_2)t}\left(\frac{\lambda_1}{\lambda_2}\right)^{n/2} \sum_{\tilde{k}=0}^{\infty} \frac{\left(t\sqrt{\lambda_1\lambda_2}\right)^{n+2\max\{0,-n\}+2\tilde{k}}}{(\tilde{k}+\max\{0,-n\})!(n+\tilde{k}+\max\{0,-n\})!} \\
&= e^{-(\lambda_1+\lambda_2)t}\left(\frac{\lambda_1}{\lambda_2}\right)^{n/2} I_{|n|}\left(2t\sqrt{\lambda_1\lambda_2}\right) \quad \text{for } n = 0, \pm 1, \pm 2, \ldots,
\end{aligned}$$

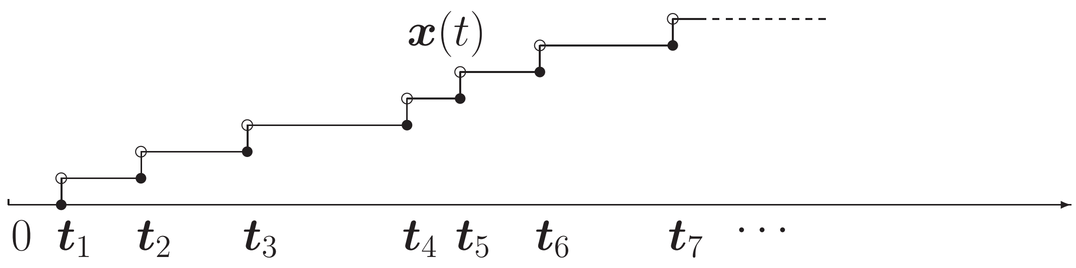
$$(\tilde{k}+\max\{0,-n\})!(n+\tilde{k}+\max\{0,-n\})! = \begin{cases} (\tilde{k}+|n|)!(\tilde{k})!, & n < 0; \\ (\tilde{k})!(|n|+\tilde{k})!, & n \geq 0 \end{cases}$$

$$|n| = n + 2\max\{0,-n\}$$

where $I_n(x) \triangleq \sum_{k=0}^{\infty} \frac{(x/2)^{n+2k}}{k!(n+k)!}$ is the modified Bessel function of order n.

Random selection of Poisson points

Let $\boldsymbol{x}(t)$ ~Poisson(λt) be formed from Poisson points $\{\boldsymbol{t}_1, \boldsymbol{t}_2, \boldsymbol{t}_3, \ldots\}$.



Suppose each occurrence $\boldsymbol{t}_i$ of $\boldsymbol{x}(t)$ gets tagged independently with probability p.

Let $\boldsymbol{y}(t)$ represent the total number of tagged events in the interval $[0, t)$.

Let $\boldsymbol{z}(t)$ represent the total number of untagged events in the interval $[0, t)$.

Claim:

$$\boldsymbol{y}(t) \sim \text{Poisson}(p\lambda t) \quad \text{and} \quad \boldsymbol{z}(t) \sim \text{Poisson}((1-p)\lambda t).$$

Random Selection of Poisson Points

Proof:

$$\begin{aligned}
\Pr[\boldsymbol{y}(t)=k] &= \sum_{n=k}^{\infty} \Pr[\boldsymbol{x}(t)=n]\,\Pr[k \text{ out of } n \text{ are tagged}|\boldsymbol{x}(t)=n] \\
&= \sum_{n=k}^{\infty} \left(e^{-\lambda t}\frac{(\lambda t)^n}{n!}\right)\left[\binom{n}{k} p^k (1-p)^{n-k}\right] \\
&= e^{-\lambda t}\frac{(p\lambda t)^k}{k!}\sum_{n=k}^{\infty}\frac{[(1-p)\lambda t]^{n-k}}{(n-k)!} \\
&= e^{-\lambda t}\frac{(p\lambda t)^k}{k!}\sum_{r=0}^{\infty}\frac{[(1-p)\lambda t]^{r}}{(r)!} \\
&= e^{-\lambda t}\frac{(p\lambda t)^k}{k!}e^{(1-p)\lambda t} \\
&= e^{-p\lambda t}\frac{(p\lambda t)^k}{k!}.
\end{aligned}$$

> This only proves the first property that defines the Poisson process! You should add the proof of the second property in Slide 9-47.

The claim on $\boldsymbol{z}(t)$ can be proved similarly. □

Remark: Given that the customer arrival forms a Poisson process, the male customer arrival also forms a Poisson process, and so does the female custom arrival.

Poisson Points and Binomial Distribution

Claim: For a Poisson process $\boldsymbol{x}(t)$ and for $t_1 < t_2$, the event $[\boldsymbol{x}(t_1) = k$ given $\boldsymbol{x}(t_2) = n]$ forms a binomial distribution $B(n, t_1/t_2)$.

Proof:

$$\begin{aligned}
\Pr[\boldsymbol{x}(t_1) = k | \boldsymbol{x}(t_2) = n] &= \frac{\Pr[\boldsymbol{x}(t_1) = k \wedge \boldsymbol{x}(t_2) = n]}{\Pr[\boldsymbol{x}(t_2) = n]} \\
&= \frac{\Pr[\boldsymbol{n}[0, t_1) = k \wedge \boldsymbol{n}[t_1, t_2) = n - k]}{\Pr[\boldsymbol{n}[0, t_2) = n]} \\
&= \frac{\Pr[\boldsymbol{n}[0, t_1) = k] \Pr[\boldsymbol{n}[t_1, t_2) = n - k]}{\Pr[\boldsymbol{n}[0, t_2) = n]} \\
&= \frac{e^{-\lambda t_1} \frac{(\lambda t_1)^k}{k!} e^{-\lambda(t_2 - t_1)} \frac{(\lambda(t_2 - t_1))^{n-k}}{(n-k)!}}{e^{-\lambda t_2} \frac{(\lambda t_2)^n}{n!}} \\
&= \binom{n}{k} \left(\frac{t_1}{t_2}\right)^k \left(1 - \frac{t_1}{t_2}\right)^{n-k} \quad \text{for } k = 0, 1, 2, \ldots, n.
\end{aligned}$$

□

Remarks

- For $0 < t_1 < T$,

$$\Pr[\boldsymbol{x}(t_1) = 1 | \boldsymbol{x}(T) = 1] = \Pr[0 \leq \boldsymbol{t}_1 < t_1 | \boldsymbol{x}(T) = 1] = \binom{1}{1}\left(\frac{t_1}{T}\right)^1 \left(1 - \frac{t_1}{T}\right)^{1-1} = \frac{t_1}{T}$$

  indicates that a Poisson arrival is equally likely to happen anywhere in an interval of length T, given that exactly one Poisson occurrence has taken place in that interval.

- In fact, the joint pdf of $\boldsymbol{t}_1, \boldsymbol{t}_2, \cdots, \boldsymbol{t}_n$ given that $\boldsymbol{x}(T) = n$ is the order statistics of $\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_n$, in which $\{\boldsymbol{s}_i\}_{i=1}^n$ are i.i.d., and each $\boldsymbol{s}_i$ is uniformly distributed over $[0, T)$.
  - "Order statistics" means

$$\{\boldsymbol{t}_1, \boldsymbol{t}_2, \cdots, \boldsymbol{t}_n\} = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_n\} \quad \text{and} \quad \boldsymbol{t}_1 \leq \boldsymbol{t}_2 \leq \cdots \leq \boldsymbol{t}_n.$$

  - Summary: A Poisson process $\boldsymbol{x}(t)$ distributes Poisson arrival points *independently and uniformly* over any finite interval $[0, T)$.

General Properties Revisited

Stationarity family

Stationarity: A random process $\boldsymbol{x}(t)$ is called *strict-sense stationary* (SSS) if its statistical properties are invariant to a shift of the origin.

Joint Stationarity: Two random processes are *jointly stationary* if their joint statistical properties are invariant to a shift of the origin.

A complex process $\boldsymbol{z}(t) = \boldsymbol{x}(t) + j\boldsymbol{y}(t)$ is stationary if the processes $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$ are jointly stationary.

Wide-Sense Stationarity: A random process $\boldsymbol{x}(t)$ is called *wide-sense stationary* (WSS) if its mean and autocorrelation functions are invariant to a shift of the origin.

As a result, the mean function $\eta_x(t)$ is a constant $\mu_x(t) = \mu_x(0) = c$, and the autocorrelation function $R_{xx}(t_1, t_2)$ only depends on the time difference $R_{xx}(t_1, t_2) = R_{xx}(t_1 - t_2, 0) = R_{xx}(t_1 - t_2)$.

Joint Wide-Sense Stationarity: Two random processes $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$ are *jointly wide-sense stationary* if their mean and autocorrelation functions, as well as their cross-correlation function, are all invariant to a shift of the origin.

General Properties Revisited

Other Forms of Stationarity

Covariance Stationarity: A process $\boldsymbol{x}(t)$ is *covariance-stationary* if the autocovariance function is invariant to a shift of the origin.

nth Order Stationarity: A process $\boldsymbol{x}(t)$ is *nth order stationary* if any n dimensional statistics is invariant to a shift of the origin.

Stationarity in an interval: A process $\boldsymbol{x}(t)$ is *stationary in an interval* if its statistical properties within that interval is invariant to a shift of the origin. Namely, $\{\boldsymbol{x}(t_i)\}_{i=1}^{n}$ and $\{\boldsymbol{x}(t_i + c)\}_{i=1}^{n}$ have the same statistics as long as all t_i and $t_i + c$ belong to that interval.

Asymptotic Stationary: A process $\boldsymbol{x}(t)$ is *asymptotic stationary* if $\boldsymbol{y}(t) = \lim_{c\to\infty} \boldsymbol{x}(t + c)$ is stationary, provided the limit exists.

General Properties Revisited

Theorem The process $\boldsymbol{x}(t) = \boldsymbol{a}\cos(\omega t) + \boldsymbol{b}\sin(\omega t)$ is SSS if, and only if, the joint density $f(a, b)$ of $\boldsymbol{a}$ and $\boldsymbol{b}$ is circularly symmetric, namely,

$$f(a, b) = g(r) \text{ for some } g(r),$$

where $r = \sqrt{a^2 + b^2}$.

Proof:

1. ***Forward*** (*Only if* part): If $\boldsymbol{x}(t)$ is SSS, then $\vec{\boldsymbol{x}}^T = [\boldsymbol{x}(0), \boldsymbol{x}(\pi/2\omega)]^T = [\boldsymbol{a}, \boldsymbol{b}]^T$ and $\vec{\boldsymbol{y}}^T = [\boldsymbol{x}(t), \boldsymbol{x}(t+\pi/2\omega)]^T$ must have the same density f. Specifically, the density of $\vec{\boldsymbol{y}} = g(\vec{\boldsymbol{x}}) = \begin{bmatrix} \cos(\omega t) & \sin(\omega t) \\ -\sin(\omega t) & \cos(\omega t) \end{bmatrix} \vec{\boldsymbol{x}}$ is equal to (cf. the next slide):

$$\begin{aligned} f_{\vec{\boldsymbol{y}}}(y_1, y_2) &= f_{\vec{\boldsymbol{x}}}\left(\begin{bmatrix} \cos(\omega t) & -\sin(\omega t) \\ \sin(\omega t) & \cos(\omega t) \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \right) \cdot \begin{vmatrix} \cos(\omega t) & -\sin(\omega t) \\ \sin(\omega t) & \cos(\omega t) \end{vmatrix} \\ &= f_{\vec{\boldsymbol{x}}}(y_1\cos(\omega t) - y_2\sin(\omega t), y_1\sin(\omega t) + y_2\cos(\omega t)). \end{aligned}$$

   The forward proof is completed by noting that $f_{\vec{\boldsymbol{x}}} = f_{\vec{\boldsymbol{y}}} = f$, and hence, $f(y_1, y_2) = f(y_1\cos(\omega t) - y_2\sin(\omega t), y_1\sin(\omega t) + y_2\cos(\omega t))$ is valid for every ωt, and every $y_1, y_2 \in \Re$.

Density for a Mapping

(P. Billingsley, *Probability and Measure*, pp. 260-261, 3rd Edition, Wiley, 1995)
Suppose that

- $\vec{\boldsymbol{x}} \in \Re^k$ has density f, and has support V that is an open set in $\Re^k$.
- g is a (one-to-one, continuously differentiable) mapping from V to U, where U is an open set in $\Re^i$. (Specifically, $g : \Re^k \to \Re^i$.)
- T is the inverse function of g, is continuously differentiable in U, and is understood as $T(\vec{y}) = (T_1(\vec{y}), T_2(\vec{y}), \cdots, T_k(\vec{y}))$, where $T_\ell : \Re^i \to \Re$, and $\vec{y} \in \Re^i$. (Specificially, $T : \Re^i \to \Re^k$.)

$T : U \to V$ continuously differentiable implies that V is open, and T^{-1}, if it exists, is also continuously differentiable.

Then, $\vec{\boldsymbol{y}} \triangleq g(\vec{\boldsymbol{x}})$ has density $\begin{cases} f(T(\vec{y})) \cdot |J(\vec{y}; T)|, & \text{for } \vec{y} \in U; \\ 0, & \text{for } \vec{y} \notin U, \end{cases}$

$$\text{where } J(\vec{y}; T) = \text{Det}\left(\begin{bmatrix} \frac{\partial T_1}{\partial y_1} & \frac{\partial T_1}{\partial y_2} & \cdots & \frac{\partial T_1}{\partial y_k} \\ \frac{\partial T_2}{\partial y_1} & \frac{\partial T_2}{\partial y_2} & \cdots & \frac{\partial T_2}{\partial y_k} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial T_k}{\partial y_1} & \frac{\partial T_k}{\partial y_2} & \cdots & \frac{\partial T_k}{\partial y_k} \end{bmatrix} (\vec{y}) \right) \neq 0 \text{ for } \vec{y} \in U.$$

Density for a Mapping

Example

- $(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \Re^2$ has density $f(x_1, x_2) = \frac{1}{2\pi} e^{-(x_1^2 + x_2^2)/2}$, and has support $V = [(-\infty, 0) \cup (0, \infty)] \times [(-\infty, 0) \cup (0, \infty)]$ that is an open set in $\Re^2$.
- g with $g(x_1, x_2) = \left(\sqrt{x_1^2 + x_2^2}, \tan^{-1}(x_2/x_1) \bmod 2\pi\right)$ is a (one-to-one, continuously differentiable) mapping from V to U, where $U = \{(y_1, y_2) : y_1 > 0 \text{ and } 0 < y_2 < 2\pi\}$ is an open set in $\Re^2$.
- T with $T(y_1, y_2) = (y_1 \cos(y_2), y_1 \sin(y_2))$ is the inverse function of g, is continuously differentiable in U, and is understood as $T(\vec{y}) = (T_1(\vec{y}) = y_1 \cos(y_2), T_2(\vec{y}) = y_1 \sin(y_2))$, where $T_\ell : \Re^2 \to \Re$, and $\vec{y} \in \Re^2$.

Then, $(\boldsymbol{y}_1, \boldsymbol{y}_2) \triangleq g(\boldsymbol{x}_1, \boldsymbol{x}_2)$ has density

$$\begin{cases} f(T(\vec{y})) \cdot |J(\vec{y}; T)|, & \text{for } \vec{y} \in U; \\ 0, & \text{for } \vec{y} \notin U \end{cases} = \begin{cases} \frac{1}{2\pi} y_1 e^{-y_1^2/2}, & \text{for } \vec{y} \in U; \\ 0, & \text{for } \vec{y} \notin U, \end{cases}$$

where $J(\vec{y}; T) = \text{Det}\left(\begin{bmatrix} \frac{\partial T_1}{\partial y_1} & \frac{\partial T_1}{\partial y_2} \\ \frac{\partial T_2}{\partial y_1} & \frac{\partial T_2}{\partial y_2} \end{bmatrix}(\vec{y})\right) = \text{Det}\begin{bmatrix} \cos(y_2) & -y_1 \sin(y_2) \\ \sin(y_2) & y_1 \cos(y_2) \end{bmatrix} = y_1 \neq 0$ for $\vec{y} \in U$.

General Properties Revisited

2. *Converse* (*If* part) :

- Define a new process for a fixed τ as:

$$\boldsymbol{x}_1(t) \triangleq \boldsymbol{a}_1 \cos(\omega t) + \boldsymbol{b}_1 \sin(\omega t),$$

  where

$$\boldsymbol{a}_1 = \boldsymbol{a}\cos(\omega\tau) + \boldsymbol{b}\sin(\omega\tau) \quad \text{and} \quad \boldsymbol{b}_1 = \boldsymbol{b}\cos(\omega\tau) - \boldsymbol{a}\sin(\omega\tau).$$

- The statistics of $\boldsymbol{x}(t)$ is completely determined by the statistics of $\boldsymbol{a}$ and $\boldsymbol{b}$. The statistics of $\boldsymbol{x}_1(t)$ is completely determined by the statistics of $\boldsymbol{a}_1$ and $\boldsymbol{b}_1$.
- However, the statistics of $(\boldsymbol{a}, \boldsymbol{b})$ and $(\boldsymbol{a}_1, \boldsymbol{b}_1)$ are completely identical because $f(a, b)$ is circular symmetric, which means that $\boldsymbol{x}(t)$ and $\boldsymbol{x}_1(t) = \boldsymbol{x}(t+\tau)$ have the same statistics for any shift τ.
- This concludes to the desired result that the statistics of $\boldsymbol{x}(t)$ is invariant to a shift of the origin. □

General Properties Revisited

Corollary The process $\boldsymbol{x}(t) = \boldsymbol{a}\cos(\omega t) + \boldsymbol{b}\sin(\omega t)$ for uncorrelated zero-mean $\boldsymbol{a}$ and $\boldsymbol{b}$ with equal variance is WSS.

Proof: The corollary is true because $E[\boldsymbol{x}(t)] = E[\boldsymbol{a}]\cos(\omega t) + E[\boldsymbol{b}]\sin(\omega t) = 0$ and

$$\begin{aligned} E[\boldsymbol{x}(t_1)\boldsymbol{x}(t_2)] &= E\{[\boldsymbol{a}\cos(\omega t_1) + \boldsymbol{b}\sin(\omega t_1)][\boldsymbol{a}\cos(\omega t_2) + \boldsymbol{b}\sin(\omega t_2)]\} \\ &= E[\boldsymbol{a}^2]\cos(\omega t_1)\cos(\omega t_2) + E[\boldsymbol{b}^2]\sin(\omega t_1)\sin(\omega t_2) = E[\boldsymbol{a}^2]\cos(\omega(t_1 - t_2)). \end{aligned}$$

□

Corollary The process $\boldsymbol{x}(t) = a\cos(\boldsymbol{\omega} t + \boldsymbol{\varphi})$ is WSS, provided that $\boldsymbol{\varphi}$ is uniformly distributed over $[-\pi, \pi)$.

Proof: The corollary is true because $E[\boldsymbol{x}(t)] = E[E[a\cos(\omega t + \boldsymbol{\varphi})|\boldsymbol{\omega} = \omega]] = 0$ and

$$\begin{aligned} E[\boldsymbol{x}(t_1)\boldsymbol{x}(t_2)] &= E\{a^2\cos(\boldsymbol{\omega} t_1 + \boldsymbol{\varphi})\cos(\boldsymbol{\omega} t_2 + \boldsymbol{\varphi})\} \\ &= E\left\{a^2\frac{\cos[\boldsymbol{\omega}(t_1 - t_2)] + \cos[\boldsymbol{\omega}(t_1 + t_2) + 2\boldsymbol{\varphi}]}{2}\right\} = \frac{a^2}{2}E[\cos(\boldsymbol{\omega}(t_1 - t_2))]. \end{aligned}$$

□

Corollary (No proof) The complex process $\boldsymbol{z}(t) = ae^{j(\boldsymbol{\omega} t + \boldsymbol{\varphi})}$ is WSS, provided that $\boldsymbol{\varphi}$ is uniformly distributed over $[-\pi, \pi)$.

General Properties Revisited

Corollary The SSS process $\boldsymbol{x}(t) = \boldsymbol{a}\cos(\omega t) + \boldsymbol{b}\sin(\omega t)$ for independent $\boldsymbol{a}$ and $\boldsymbol{b}$ is normal.

Proof:

- From the Theorem on Slide 9-60, SSS of $\boldsymbol{x}(t)$ implies that $f(a, b)$ is circularly symmetric.
- By independence of $\boldsymbol{a}$ and $\boldsymbol{b}$, $g(r) = f_{\boldsymbol{a}}(a) f_{\boldsymbol{b}}(b)$, where $r = \sqrt{a^2 + b^2}$.
- We then derive:

$$\begin{aligned}
\frac{1}{r}\frac{(\partial g(r)/\partial r)}{g(r)} &= \frac{1}{r}\frac{(\partial g(r)/\partial r)}{g(r)}\frac{(\partial r/\partial a)}{(\partial r/\partial a)} \\
&= \frac{1}{r}\frac{(\partial g(r)/\partial a)}{g(r)}\frac{1}{(\partial r/\partial a)} \\
&= \frac{1}{r}\frac{(\partial [f_{\boldsymbol{a}}(a) f_{\boldsymbol{b}}(b)]/\partial a)}{[f_{\boldsymbol{a}}(a) f_{\boldsymbol{b}}(b)]}\frac{1}{(a/r)} \\
&= \frac{1}{a}\frac{(\partial f_{\boldsymbol{a}}(a)/\partial a)}{f_{\boldsymbol{a}}(a)}.
\end{aligned}$$

General Properties Revisited

Hence, it should be true that:

$$\frac{1}{r}\frac{(\partial g(r)/\partial r)}{g(r)} = \frac{1}{a}\frac{(\partial f_{\boldsymbol{a}}(a)/\partial a)}{f_{\boldsymbol{a}}(a)} = \text{constant} \qquad \left(= -\frac{1}{\sigma^2}\right) \quad \text{(Eq. 1)},$$

because if for some α and β with $\alpha \neq \beta$,

$$\left.\frac{1}{a}\frac{(\partial f_{\boldsymbol{a}}(a)/\partial a)}{f_{\boldsymbol{a}}(a)}\right|_{a=\alpha} \neq \left.\frac{1}{a}\frac{(\partial f_{\boldsymbol{a}}(a)/\partial a)}{f_{\boldsymbol{a}}(a)}\right|_{a=\beta},$$

then as $(a,b) = (\alpha,\beta)$ and $(a,b) = (\beta,\alpha)$ yield the same $r = \sqrt{\alpha^2+\beta^2}$, a contradiction would result as:

$$\left.\frac{1}{r}\frac{(\partial g(r)/\partial r)}{g(r)}\right|_{(a,b)=(\alpha,\beta)} \neq \left.\frac{1}{r}\frac{(\partial g(r)/\partial r)}{g(r)}\right|_{(a,b)=(\beta,\alpha)}.$$

This implies (together with $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\boldsymbol{a}}(a)f_{\boldsymbol{b}}(b)dadb = 1$) that

$$g(r)|_{r=\sqrt{a^2+b^2}} = f_{\boldsymbol{a}}(a)f_{\boldsymbol{b}}(b) = \frac{1}{2\pi\sigma^2}e^{-(a^2+b^2)/(2\sigma^2)}.$$

This completes the proof that $(\boldsymbol{a},\boldsymbol{b})$ is a Gaussian random vector.

Summary: **Circular symmetry** and **independence** imply **Gaussian**.

General Properties Revisited

Normal: A process $\boldsymbol{x}(t)$ is called *normal* if any finite dimensional samples of $\boldsymbol{x}(t)$ are jointly normal.

The desired result that $\boldsymbol{x}(t) = \boldsymbol{a}\cos(\omega t) + \boldsymbol{b}\sin(\omega t)$ is a normal process can be substantiated by the observation that "linear combination of Gaussians" is still Gaussian, namely,

$$\begin{bmatrix}\boldsymbol{x}(t_1) & \boldsymbol{x}(t_2) & \cdots & \boldsymbol{x}(t_k)\end{bmatrix} = \begin{bmatrix}\boldsymbol{a} & \boldsymbol{b}\end{bmatrix} \begin{bmatrix}\cos(\omega t_1) & \cos(\omega t_2) & \cdots & \cos(\omega t_k) \\ \sin(\omega t_1) & \sin(\omega t_2) & \cdots & \sin(\omega t_k)\end{bmatrix}.$$

□

General Properties Revisited

$$\int \frac{g'(r)}{g(r)} dr = \int \left(-\frac{1}{\sigma^2} r\right) dr + \log C \;\Leftrightarrow\; \log g(r) = -\frac{r^2}{2\sigma^2} + \log C$$

$$\Leftrightarrow\; g(r) = Ce^{-r^2/(2\sigma^2)},$$

where

$$C = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(g(r)\big|_{r=\sqrt{x^2+y^2}}\right) dxdy}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(e^{-r^2/(2\sigma^2)}\big|_{r=\sqrt{x^2+y^2}}\right) dxdy} = \frac{1}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-r^2/(2\sigma^2)} dxdy} = \frac{1}{2\pi\sigma^2}.$$

Notably, it can be shown that if the constant in Eq. (1) is positive or zero, $\int_0^{\infty} g(r)dr = 1$ cannot be satisfied. Hence, we can assume that the constant is equal to $-1/\sigma^2$ for some σ.

Correlation Time

Definition (Correlation time) The correlation time τ_c of a WSS process $\boldsymbol{x}(t)$ is defined as:

$$\tau_c = \frac{1}{C_{xx}(0)} \int_0^\infty C_{xx}(\tau) d\tau.$$

- For an a-dependent WSS process $\boldsymbol{x}(t)$, $C_{xx}(\tau) = 0$ for $|\tau| > a$. Hence,

$$\begin{aligned} |\tau_c| = \left| \frac{1}{C_{xx}(0)} \int_0^\infty C_{xx}(\tau) d\tau \right| &\leq \frac{1}{C_{xx}(0)} \int_0^\infty |C_{xx}(\tau)| d\tau \\ &= \frac{1}{C_{xx}(0)} \int_0^a |C_{xx}(\tau)| d\tau \\ &\leq \int_0^a d\tau = a. \end{aligned}$$

The end of Section 9-1 Definitions

Definition A system with statistic input process $\boldsymbol{x}(t)$

($\{\boldsymbol{x}(t), t \in \mathcal{I}\}$ defined over $(S, \mathcal{F}, P)$)

is specified through an operator $\boldsymbol{T} : S \mapsto \Re^{\mathcal{X}^{\mathcal{I}} \times \mathcal{I}}$

(hence, $\{\boldsymbol{T}(x^{\mathcal{I}}, t), (x^{\mathcal{I}}, t) \in \mathcal{X}^{\mathcal{I}} \times \mathcal{I}\}$ is itself a random process defined over the same $(S, \mathcal{F}, P)$)

such that its output process $\boldsymbol{y}(t)$ is defined as $\boldsymbol{y}(t, \zeta) = \boldsymbol{T}(\{\boldsymbol{x}(s, \zeta), s \in \mathcal{I}\}, t, \zeta)$ for $t \in \mathcal{I}$ and $\zeta \in S$.

(As a result, $\{\boldsymbol{y}(t), t \in \mathcal{I}\}$ is a random process defined over the same $(S, \mathcal{F}, P)$.)

Example. $\mathcal{I} = \{1, 2, 3\}$. So, $\{\boldsymbol{x}(t), t \in \mathcal{I}\} = \{\boldsymbol{x}(1), \boldsymbol{x}(2), \boldsymbol{x}(3)\}$

- $$\begin{cases} \boldsymbol{y}(1) = \boldsymbol{T}(\{\boldsymbol{x}(1), \boldsymbol{x}(2), \boldsymbol{x}(3)\}, 1) \\ \boldsymbol{y}(2) = \boldsymbol{T}(\{\boldsymbol{x}(1), \boldsymbol{x}(2), \boldsymbol{x}(3)\}, 2) \\ \boldsymbol{y}(3) = \boldsymbol{T}(\{\boldsymbol{x}(1), \boldsymbol{x}(2), \boldsymbol{x}(3)\}, 3) \end{cases}$$

  The above system is of course **non-causal**. A causal system will have

$$\begin{cases} \boldsymbol{y}(1) = \boldsymbol{T}(\{\boldsymbol{x}(1)\}, 1) \\ \boldsymbol{y}(2) = \boldsymbol{T}(\{\boldsymbol{x}(1), \boldsymbol{x}(2)\}, 2) \\ \boldsymbol{y}(3) = \boldsymbol{T}(\{\boldsymbol{x}(1), \boldsymbol{x}(2), \boldsymbol{x}(3)\}, 3) \end{cases}$$

The above system has **memory**. A memoryless causal system will have

$$\begin{cases} \boldsymbol{y}(1) = \boldsymbol{T}(\{\boldsymbol{x}(1)\}, 1) \\ \boldsymbol{y}(2) = \boldsymbol{T}(\{\boldsymbol{x}(2)\}, 2) \\ \boldsymbol{y}(3) = \boldsymbol{T}(\{\boldsymbol{x}(3)\}, 3) \end{cases}$$

End of the example □

In usual notation, we write

$$\boldsymbol{y}(t) = \boldsymbol{T}(\{\boldsymbol{x}(s), s \in \mathcal{I}\}, t) = \boldsymbol{T}_t(\{\boldsymbol{x}(s), s \in \mathcal{I}\}),$$

where

- the second (resp. subscript) argument t in $\boldsymbol{T}(\cdot, t)$ (resp. $\boldsymbol{T}_t(\cdot)$) retains the possibility of specifying a time-varying system,
- and the first argument $\{\boldsymbol{x}(s), s \in \mathcal{I}\}$ retains the possibility of specifying a with-memory (or non-causal) system.

9-2 Systems with Stochastic Inputs

Classification of systems

- *Deterministic System:* $\boldsymbol{T}_t(x^{\mathcal{I}}, \zeta) = \boldsymbol{T}_t(x^{\mathcal{I}})$. I.e., $\boldsymbol{T}$ only depends on $x^{\mathcal{I}}$ and t, and is irrelevant to ζ.

  A random variable $\boldsymbol{z} : S \mapsto \Re$ defined over a probability space $(S, \mathcal{F}, P)$ is degenerately *deterministic* if $\boldsymbol{z}(\zeta) = z$, a constant, for all $\zeta \in S$. In such case, $\Pr[\boldsymbol{z} = z] = 1$.

- *Stochastic System:* $\boldsymbol{T}_t(x^{\mathcal{I}}, \zeta_1) \neq \boldsymbol{T}_t(x^{\mathcal{I}}, \zeta_2)$ for some $\zeta_1 \neq \zeta_2$.

  Due to the complication of a *stochastic* system, the *deterministic* system is considered mostly in the literature. However, the recent demand on research of fading channels makes necessary the consideration of a stochastic system.

Memoryless Systems

Definition (Memoryless system) A system is memoryless if $\boldsymbol{T}_t(x^{\mathcal{I}}, \zeta) = \boldsymbol{T}(x, \zeta)$.

Example ($S = \{\oplus, \ominus, \otimes, \oslash\}$, $\mathcal{F} = 2^S$, $P = \{0.1, 0.2, 0.3, 0.4\}$ resp. for S)
At some specific time t, we have

$$\boldsymbol{x}(t, \oplus) = \boldsymbol{x}(t, \ominus) = 1 \quad \text{and} \quad \boldsymbol{x}(t, \otimes) = \boldsymbol{x}(t, \oslash) = -1,$$

and the memoryless $\boldsymbol{T}$ satisfies

$$\boldsymbol{T}(1, \oplus) = 1, \quad \boldsymbol{T}(1, \ominus) = -1, \quad \boldsymbol{T}(-1, \otimes) = 1 \quad \text{and} \quad \boldsymbol{T}(-1, \oslash) = -1.$$

Then, $\begin{cases} \boldsymbol{y}(t, \oplus) &= \boldsymbol{T}(\boldsymbol{x}(t, \oplus), \oplus) = \boldsymbol{T}(1, \oplus) = 1 \\ \boldsymbol{y}(t, \ominus) &= \boldsymbol{T}(\boldsymbol{x}(t, \ominus), \ominus) = \boldsymbol{T}(1, \ominus) = -1 \\ \boldsymbol{y}(t, \otimes) &= \boldsymbol{T}(\boldsymbol{x}(t, \otimes), \otimes) = \boldsymbol{T}(-1, \otimes) = 1 \\ \boldsymbol{y}(t, \oslash) &= \boldsymbol{T}(\boldsymbol{x}(t, \oslash), \oslash) = \boldsymbol{T}(-1, \oslash) = -1 \end{cases}$

Hence, $\Pr[\boldsymbol{T}(1) = 1] = \Pr[\boldsymbol{y}(t) = 1 | \boldsymbol{x}(t) = 1] = \dfrac{P(\{\oplus\})}{P(\{\oplus, \ominus\})} = \dfrac{0.1}{0.1 + 0.2} = \dfrac{1}{3}$ and

$$\Pr[\boldsymbol{T}(1) = -1] = \frac{2}{3}, \quad \Pr[\boldsymbol{T}(-1) = 1] = \frac{3}{7}, \quad \Pr[\boldsymbol{T}(-1) = -1] = \frac{4}{7}.$$

Memoryless Systems

- It is called *memoryless* because the statistics of $\boldsymbol{y}(t)$ depends only on the statistics of $\boldsymbol{x}(t)$ and not on any other past or future values of $\boldsymbol{x}(t)$.
- A memoryless system is often denoted by $\boldsymbol{y}(t) = \boldsymbol{T}(\boldsymbol{x}(t))$, and is equivalently written as a time-independent transition probability

$$P_{\boldsymbol{y}|\boldsymbol{x}}(y|x) = \Pr[\boldsymbol{T}(x) = y].$$

- Note that $\boldsymbol{y}(n) = \boldsymbol{x}(n) + \boldsymbol{z}(n)$ for integer n, satisfying that $\boldsymbol{z}(i)$ is independent of $\boldsymbol{z}(j)$ for any $i \neq j$, may not be a memoryless system, if the statistics of $\boldsymbol{z}(i)$ is different from $\boldsymbol{z}(j)$. This is because we still need to maintain the time index n in order to know the (statistical) mapping from $\boldsymbol{x}(n)$ to $\boldsymbol{y}(n)$. "Memoryless" in its strict sense means that one **only** needs to know (the statistics of) the current input in order to determine (the statistics of) the current output (cf. Eq. (9-74) in textbook). Thus, $\{\boldsymbol{z}(n)\}$ must be i.i.d. in order to obtain a (strictly) memoryless system defined according to $\boldsymbol{y}(n) = \boldsymbol{x}(n) + \boldsymbol{z}(n)$.

In this additive system,

$$\Pr[\boldsymbol{T}(x) = y] = P_{\boldsymbol{y}|\boldsymbol{x}}(y|x) = P_{\boldsymbol{z}}(y - x).$$

Memoryless Systems

- **Independence between input $\boldsymbol{x}(t)$ and system $\boldsymbol{T}$:**

  Continue from the example in Slide 9-73.

$$\begin{aligned}\Pr[\boldsymbol{x}(t)=1 \wedge \boldsymbol{T}(1)=1] &= P(\{\oplus,\ominus\}\cap\{\oplus\}) &= 0.1\\ \Pr[\boldsymbol{x}(t)=1]\times\Pr[\boldsymbol{T}(1)=1] &= P(\{\oplus,\ominus\})\times\frac{1}{3} &= 0.1\end{aligned}$$

$$\begin{aligned}\Pr[\boldsymbol{x}(t)=1 \wedge \boldsymbol{T}(1)=-1] &= P(\{\oplus,\ominus\}\cap\{\ominus\}) &= 0.2\\ \Pr[\boldsymbol{x}(t)=1]\times\Pr[\boldsymbol{T}(1)=-1] &= P(\{\oplus,\ominus\})\times\frac{1}{3} &= 0.2\end{aligned}$$

$$\begin{aligned}\Pr[\boldsymbol{x}(t)=-1 \wedge \boldsymbol{T}(-1)=1] &= P(\{\otimes,\oslash\}\cap\{\otimes\}) &= 0.3\\ \Pr[\boldsymbol{x}(t)=-1]\times\Pr[\boldsymbol{T}(-1)=1] &= P(\{\otimes,\oslash\})\times\frac{3}{7} &= 0.3\end{aligned}$$

$$\begin{aligned}\Pr[\boldsymbol{x}(t)=-1 \wedge \boldsymbol{T}(-1)=-1] &= P(\{\otimes,\oslash\}\cap\{\oslash\}) &= 0.4\\ \Pr[\boldsymbol{x}(t)=-1]\times\Pr[\boldsymbol{T}(-1)=-1] &= P(\{\otimes,\oslash\})\times\frac{4}{7} &= 0.4\end{aligned}$$

Memoryless Systems

Lemma If the input $\boldsymbol{x}(t)$ to a memoryless system is SSS, its output $\boldsymbol{y}(t)$ is also SSS.

Proof: The statistics of

$$P\left\{\zeta \in S : \boldsymbol{y}(t+c,\zeta) \in A\right\},$$

can be rewritten as:

$$P\left\{\zeta \in S : \boldsymbol{T}(\boldsymbol{x}(t+c,\zeta),\zeta) \in A\right\},$$

which, for a given memoryless $\boldsymbol{T}$, can be replaced by:

$$P\left\{\zeta \in S : \boldsymbol{x}(t+c,\zeta) \in B\right\}, \tag{9.2}$$

where $B \triangleq \{x \in \Re : \boldsymbol{T}(x,\zeta) \in A\}$. By the SSS of $\boldsymbol{x}(t)$, (9.2) is equal to:

$$P\left\{\zeta \in S : \boldsymbol{x}(t,\zeta) \in B\right\},$$

which in turns equal

$$P\left\{\zeta \in S : \boldsymbol{T}(\boldsymbol{x}(t,\zeta),\zeta) \in A\right\} = P\left\{\zeta \in S : \boldsymbol{y}(t,\zeta) \in A\right\}.$$

Since the above proof is valid for any c, the lemma holds. $\square$

Memoryless Systems

For a non-memoryless system,

$$P\{\zeta \in S : \boldsymbol{x}(t,\zeta) \in B\} \quad \text{and} \quad P\{\zeta \in S : \boldsymbol{x}(t+c,\zeta) \in B\}$$

are still equal due to SSS of $\boldsymbol{x}(t)$.

But,

$$B_t = \{x \in \Re : \boldsymbol{T}_t(x,\zeta) \in A\}$$

may not be equal to

$$B_{t+c} = \{x \in \Re : \boldsymbol{T}_{t+c}(x,\zeta) \in A\}$$

since $\boldsymbol{T}_t(\cdot,\zeta)$ and $\boldsymbol{T}_{t+c}(\cdot,\zeta)$ may not be the same mapping.

This proof again substantiates that only i.i.d. $\{\boldsymbol{z}(n)\}$ can make $\boldsymbol{y}(n) = \boldsymbol{x}(n) + \boldsymbol{z}(n)$ a (strictly) memoryless system! (Note that if $\{\boldsymbol{z}(n)\}$ is not i.i.d., SSS $\{\boldsymbol{x}(n)\}$ may not induce SSS $\{\boldsymbol{y}(n)\}$.)

Memoryless Systems

Corollary For a memoryless system,

- if input $\boldsymbol{x}(t)$ is nth order stationary, output $\boldsymbol{y}(t)$ is also nth order stationary.
- if input $\boldsymbol{x}(t)$ is stationary in an interval, output $\boldsymbol{y}(t)$ is also stationary in the same interval.
- **however**, if input $\boldsymbol{x}(t)$ is WSS, output $\boldsymbol{y}(t)$ might **not** be WSS.

(*Counterexample*: Square-law detector $\boldsymbol{y}(t) = \boldsymbol{x}^2(t)$, where $\boldsymbol{x}(t) = \boldsymbol{a}\cos(\omega t) + \boldsymbol{b}\sin(\omega t)$ for independent zero-mean $\boldsymbol{a}$ and $\boldsymbol{b}$ with equal statistics as in Slide 9-64.)

$$\begin{aligned}
E[\boldsymbol{y}(t_1)\boldsymbol{y}(t_2)] &= E[(\boldsymbol{a}\cos(\omega t_1) + \boldsymbol{b}\sin(\omega t_1))^2(\boldsymbol{a}\cos(\omega t_2) + \boldsymbol{b}\sin(\omega t_2))^2] \\
&= \frac{E[\boldsymbol{a}^4] + E^2[\boldsymbol{a}^2]}{2} + \frac{E[\boldsymbol{a}^4] - E^2[\boldsymbol{a}^2]}{2}\cos(2\omega t_1)\cos(2\omega t_2) \\
&\quad + E^2[\boldsymbol{a}^2]\sin(2\omega t_1)\sin(2\omega t_2).
\end{aligned}$$

Hence, $\boldsymbol{y}(t)$ is WSS only when $E[\boldsymbol{a}^4] = 3E^2[\boldsymbol{a}^2]$.

Hard Limiter

Example 9-16 (Arcsine law) Define the hard limiter process $\boldsymbol{y}(t)$ as:

$$\boldsymbol{y}(t) = \begin{cases} +1, & \text{if } \boldsymbol{x}(t) > 0; \\ -1, & \text{if } \boldsymbol{x}(t) < 0, \end{cases}$$

where $\boldsymbol{x}(t)$ is a zero-mean Gaussian stationary process.
Please determine the mean and autocorrelation functions of $\boldsymbol{y}(t)$.
Answer: It is clear that $E[\boldsymbol{y}(t)] = 0$. As for $R_{yy}(t_1, t_2)$, we note that

$$\boldsymbol{y}(t_1)\boldsymbol{y}(t_2) = \begin{cases} +1, & \text{if } \boldsymbol{x}(t_1)\boldsymbol{x}(t_2) > 0; \\ -1, & \text{if } \boldsymbol{x}(t_1)\boldsymbol{x}(t_2) < 0, \end{cases}$$

Since $\boldsymbol{x}(t_1)$ and $\boldsymbol{x}(t_2)$ are jointly normal with zero mean and covariance matrix

$$\Sigma = \begin{bmatrix} R_{xx}(0) & R_{xx}(t_1 - t_2) \\ R_{xx}(t_2 - t_1) & R_{xx}(0) \end{bmatrix} = R_{xx}(0) \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$

where $\rho = R_{xx}(t_1 - t_2)/R_{xx}(0)$, we derive:

$$\begin{aligned} \Pr[\boldsymbol{x}(t_1)\boldsymbol{x}(t_2) > 0] &= 2\int_0^\infty \int_0^\infty \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}[x, y]\Sigma^{-1} \begin{bmatrix} x \\ y \end{bmatrix}\right\} dxdy \\ &= 2\int_0^\infty \int_0^\infty \frac{1}{2\pi(1-\rho^2)^{1/2} R_{xx}(0)} \exp\left\{-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)R_{xx}(0)}\right\} dxdy \\ &= \frac{1}{2} + \frac{1}{\pi}\arcsin(\rho). \end{aligned}$$

Hard Limiter

Hence,

$$E[\boldsymbol{y}(t_1)\boldsymbol{y}(t_2)] \;=\; \frac{2}{\pi}\arcsin(\rho) = \frac{2}{\pi}\arcsin\left(\frac{R_{xx}(t_1-t_2)}{R_{xx}(0)}\right).$$

□

Now, you shall know why it is named the *arcsine law.*

The below derivation is just for your reference.
Let $x = r\cos(\theta)\sqrt{1-\rho^2}$, $y = r\sin(\theta)\sqrt{1-\rho^2}$ and $u = \tan(\theta)$.

$$\begin{aligned}
&\int_0^\infty \int_0^\infty \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left\{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}\right\} dxdy \\
&= \int_0^{\pi/2}\int_0^\infty \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left\{-\frac{[1-\rho\sin(2\theta)]r^2}{2}\right\} \begin{vmatrix} \partial x/\partial r & \partial y/\partial r \\ \partial x/\partial\theta & \partial y/\partial\theta \end{vmatrix} drd\theta \\
&= \frac{(1-\rho^2)^{1/2}}{2\pi}\int_0^{\pi/2}\int_0^\infty r\exp\left\{-\frac{[1-\rho\sin(2\theta)]}{2}r^2\right\} drd\theta \\
&= \frac{(1-\rho^2)^{1/2}}{2\pi}\int_0^{\pi/2}\left(-\frac{1}{(1-\rho\sin(2\theta))}\exp\left\{-\frac{[1-\rho\sin(2\theta)]r^2}{2}\right\}\Bigg|_0^\infty\right) d\theta \\
&= \frac{1}{2\pi}\int_0^{\pi/2}\frac{(1-\rho^2)^{1/2}}{1-\rho\sin(2\theta)}d\theta \\
&= \frac{1}{\pi}\int_0^{\pi/4}\frac{(1-\rho^2)^{1/2}}{1-\rho\sin(2\theta)}d\theta = \frac{1}{\pi}\int_0^{\pi/4}\frac{(1-\rho^2)^{1/2}}{\sin^2(\theta)+\cos^2(\theta)-2\rho\sin(\theta)\cos(\theta)}d\theta
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\pi}\int_0^1 \frac{(1-\rho^2)^{1/2}}{u^2+1-2\rho u}du = \frac{1}{\pi}\int_0^1 \frac{(1-\rho^2)^{1/2}}{(u-\rho)^2+(1-\rho^2)}du = \frac{1}{\pi}\arctan\left(\frac{u-\rho}{\sqrt{1-\rho^2}}\right)\Big|_0^1 \\
&= \frac{1}{\pi}\arctan\left(\frac{1-\rho}{\sqrt{1-\rho^2}}\right) + \frac{1}{\pi}\arctan\left(\frac{\rho}{\sqrt{1-\rho^2}}\right) = \frac{1}{\pi}\arcsin\left(\sqrt{\frac{1-\rho}{2}}\right) + \frac{1}{\pi}\arcsin(\rho) \\
&= \frac{1}{\pi}\left(\frac{\pi}{4} - \frac{1}{2}\arcsin(\rho)\right) + \frac{1}{\pi}\arcsin(\rho) \qquad (9.3) \\
&= \frac{1}{4} + \frac{1}{2\pi}\arcsin(\rho),
\end{aligned}
$$

where (9.3) follows from

$$
\begin{aligned}
\sin\left[2\cdot\arcsin\left(\sqrt{\frac{1-\rho}{2}}\right)\right] &= 2\sin\left[\arcsin\left(\sqrt{\frac{1-\rho}{2}}\right)\right]\cos\left[\arcsin\left(\sqrt{\frac{1-\rho}{2}}\right)\right] \\
&= 2\left(\sqrt{\frac{1-\rho}{2}}\right)\left(\sqrt{\frac{1+\rho}{2}}\right) = \sqrt{1-\rho^2} = \sin\left(\frac{\pi}{2} - \arcsin(\rho)\right).
\end{aligned}
$$

Bussgang's Theorem

Theorem (Example 9-17: Bussgang's theorem) The cross-correlation $R_{xy}(\tau)$ of system input $\boldsymbol{x}(t)$ and system output $\boldsymbol{y}(t)$ for a stationary zero-mean Gaussian input and memoryless system is proportional to $R_{xx}(\tau)$.

Proof:

$$\begin{aligned} R_{xy}(t_1,t_2) &= E[\boldsymbol{x}(t_1)\cdot\boldsymbol{T}(\boldsymbol{x}(t_2))] \\ &= E\left[E\left[x_1\cdot\boldsymbol{T}(x_2)\right|\boldsymbol{x}(t_1)=x_1,\boldsymbol{x}(t_2)=x_2]\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} x_1\,y\,dP_{\boldsymbol{y}|\boldsymbol{x}}(y|x_2)\right)\frac{1}{2\pi|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}[x_1,x_2]\Sigma^{-1}\begin{bmatrix}x_1\\x_2\end{bmatrix}\right\}dx_1dx_2, \end{aligned}$$

where

$$\Sigma=\begin{bmatrix}R_{xx}(0) & R_{xx}(t_1-t_2)\\ R_{xx}(t_1-t_2) & R_{xx}(0)\end{bmatrix}=R_{xx}(0)\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}.$$

Hence, by letting $g(x_2)=E[\boldsymbol{T}(x_2)]$,

$$\begin{aligned} R_{xy}(t_1,t_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{x_1\cdot g(x_2)}{2\pi R_{xx}(0)(1-\rho^2)^{1/2}}\exp\left\{-\frac{x_1^2-2\rho x_1x_2+x_2^2}{2R_{xx}(0)(1-\rho^2)}\right\}dx_1dx_2 \\ &= \rho\int_{-\infty}^{\infty}\frac{x_2\cdot g(x_2)}{\sqrt{2\pi R_{xx}(0)}}\exp\left\{-\frac{x_2^2}{2R_{xx}(0)}\right\}dx_2 \\ &= \rho E[E[x_2\cdot g(x_2)|\boldsymbol{x}(t)=x_2]] = R_{xx}(t_1-t_2)\frac{E[\boldsymbol{x}(t)\cdot g(\boldsymbol{x}(t))]}{R_{xx}(0)}. \end{aligned}$$

□

Bussgang's Theorem

Following the example in Slide 9-73,

$$g(1) = E[\boldsymbol{T}(1)] = E[\boldsymbol{y}(t)|\boldsymbol{x}(t) = 1] = 1 \cdot \frac{1}{3} + (-1) \cdot \frac{2}{3} = -\frac{1}{3}$$

$$g(-1) = E[\boldsymbol{T}(-1)] = E[\boldsymbol{y}(t)|\boldsymbol{x}(t) = -1] = 1 \cdot \frac{3}{7} + (-1) \cdot \frac{4}{7} = -\frac{1}{7}$$

Special Case (a) Hard Limiter. Suppose that $\boldsymbol{T}$ is a deterministic system with $\boldsymbol{T}(x) = g(x) = \begin{cases} 1, & x \geq 0; \\ -1, & x < 0 \end{cases}$.

Then,

$$E[\boldsymbol{x}(t) \cdot g(\boldsymbol{x}(t))] \;=\; E[|\boldsymbol{x}(t)|] = \sqrt{\frac{2R_{xx}(0)}{\pi}}.$$

Hence,

$$R_{xy}(\tau) = R_{xx}(\tau)\sqrt{\frac{2}{\pi R_{xx}(0)}}.$$

Bussgang's Theorem

Special Case (b) Limiter. Suppose that $\boldsymbol{T}$ is a deterministic system with $\boldsymbol{T}(x) = g(x) = x \cdot \mathbf{1}(|x| \le c) + c \cdot \mathbf{1}(x > c) + (-c) \cdot \mathbf{1}(x < -c)$.
Then,

$$\begin{aligned} E[\boldsymbol{x}(t) \cdot g(\boldsymbol{x}(t))] &= \int_{-c}^{c} \frac{x^2}{\sqrt{2\pi R_{xx}(0)}} \exp\left\{-\frac{x^2}{2R_{xx}(0)}\right\} dx \\ &+ 2\int_{c}^{\infty} \frac{cx}{\sqrt{2\pi R_{xx}(0)}} \exp\left\{-\frac{x^2}{2R_{xx}(0)}\right\} dx \\ &= R_{xx}(0) \cdot \operatorname{erf}\left(\frac{c}{\sqrt{2R_{xx}(0)}}\right) \end{aligned}$$

Hence, $R_{xy}(\tau) = R_{xx}(\tau) \cdot \operatorname{erf}\left(\frac{c}{\sqrt{2R_{xx}(0)}}\right) = R_{xx}(\tau) \cdot \left(2G\left(\frac{c}{\sqrt{R_{xx}(0)}}\right) - 1\right)$,
where $G(\cdot)$ is the standard normal cdf.

$$\begin{aligned} \frac{E[\boldsymbol{x}(t) \cdot g(\boldsymbol{x}(t))]}{R_{xx}(0)} &= \int_{-a}^{a} \frac{y^2}{\sqrt{2\pi}} e^{-y^2/2} dy + 2a \int_{a}^{\infty} \frac{y}{\sqrt{2\pi}} e^{-y^2/2} dy \quad \text{(Let } y = x/\sqrt{R_{xx}(0)} \text{ and } a = c/\sqrt{R_{xx}(0)}.\text{)} \\ &= \left(\int_{-a}^{a} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy - \frac{y}{\sqrt{2\pi}} e^{-y^2/2}\Big|_{-a}^{a}\right) + 2a\left(-\frac{1}{\sqrt{2\pi}} e^{-y^2/2}\Big|_{a}^{\infty}\right) = \int_{-a}^{a} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy. \end{aligned}$$

Bussgang's Theorem

The text uses a different constant as $K = E[g'(\boldsymbol{x}(t))]$, which can be shown to be equal to $K = E[\boldsymbol{x}(t) \cdot g(\boldsymbol{x}(t))]/R_{xx}(0)$, and which requires the existence of $g'(\cdot)$.

If $\lim_{x\to\infty} g(x) \exp\left\{-\frac{x^2}{2R_{xx}(0)}\right\} = \lim_{x\to-\infty} g(x) \exp\left\{-\frac{x^2}{2R_{xx}(0)}\right\} < \infty$, and $g'(\cdot)$ exists, then

$$\frac{1}{R_{xx}(0)} \int_{-\infty}^{\infty} \frac{x \cdot g(x)}{\sqrt{2\pi R_{xx}(0)}} \exp\left\{-\frac{x^2}{2R_{xx}(0)}\right\} dx$$

$$= -\frac{g(x)}{\sqrt{2\pi R_{xx}(0)}} \exp\left\{-\frac{x^2}{2R_{xx}(0)}\right\}\Bigg|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{g'(x)}{\sqrt{2\pi R_{xx}(0)}} \exp\left\{-\frac{x^2}{2R_{xx}(0)}\right\} dx.$$

Special Case (a) Hard Limiter:

$$E[g'(\boldsymbol{x}(t))] = E[2\delta(\boldsymbol{x}(t))] = \int_{-\infty}^{\infty} \frac{2\delta(x)}{\sqrt{2\pi R_{xx}(0)}} \exp\left\{-\frac{x^2}{2R_{xx}(0)}\right\} dx = \sqrt{\frac{2}{\pi R_{xx}(0)}}.$$

Special Case (b) Limiter: (K in text is wrong because $g'(x) = \mathbf{1}\{-c < x < c\}$.)

$$E[g'(\boldsymbol{x}(t))] = \int_{-c}^{c} \frac{1}{\sqrt{2\pi R_{xx}(0)}} \exp\left\{-\frac{x^2}{2R_{xx}(0)}\right\} dx = 2G\left(\frac{c}{\sqrt{R_{xx}(0)}}\right) - 1.$$

Linear Systems

Definition (Linear systems) If the output due to a linear combination of the input processes is equal to the linear combination of the individually induced outputs with the same weights, then the system $\boldsymbol{T}$ is called a *linear* system.

Note that the weights for the linear combination can be random variables.

Convolutionalization of linear systems

Lemma If a first-order differentiable function f satisfies that for any $\vec{x}$ and $\vec{u}$,

$$f(\vec{x}) + f(\vec{u}) = f(\vec{x} + \vec{u}),$$

then f must be of the shape:

$$f(\vec{x}) = \sum_i \left(\left. \frac{\partial f(\vec{x})}{\partial x_i} \right|_{\vec{x}=\mathbf{0}} \right) x_i.$$

Key behind the Proof:

$$\begin{aligned}
\frac{\partial f(\vec{x})}{\partial x_i} &= \lim_{\delta\downarrow 0} \frac{f(\cdots, x_{i-1}, x_i+\delta, x_{i+1}, \cdots) - f(\cdots, x_{i-1}, x_i, x_{i+1}, \cdots)}{\delta} \\
&= \lim_{\delta\downarrow 0} \frac{[f(\cdots, x_{i-1}, x_i, x_{i+1}, \cdots) + f(\cdots, 0, \delta, 0, \cdots)] - f(\cdots, x_{i-1}, x_i, x_{i+1}, \cdots)}{\delta} \\
&= \lim_{\delta\downarrow 0} \frac{f(\cdots, 0, \delta, 0, \cdots)}{\delta} \\
&= \text{constant}
\end{aligned}$$

$\Rightarrow f(\vec{x})$ is affine from the standpoint of x_i

$\Rightarrow f$ must be of the shape:

$$f(\vec{x}) = \sum_i \left(\left.\frac{\partial f(\vec{x})}{\partial x_i}\right|_{\vec{x}=\mathbf{0}} \right) x_i + C.$$

The proof is completed by

$$f(\vec{0}) + f(\vec{0}) = f(\vec{0}+\vec{0}) \Rightarrow C + C = C \Rightarrow C = 0.$$

$\square$

Linear Systems

- Discrete-time System:

  Now define a deterministic with-memory time-varying **linear** system

$$\boldsymbol{T}_t(\{\boldsymbol{x}(s), s \in \mathcal{I}\}) = T_t(\{\boldsymbol{x}(s), s \in \{t, t-1\}\}) = T_t(\boldsymbol{x}(t), \boldsymbol{x}(t-1)).$$

  Then

$$\begin{aligned} \boldsymbol{y}(t) &= \left(\left. \frac{\partial T_t(x_1, x_2)}{\partial x_1} \right|_{x_1 = x_2 = 0} \right) \boldsymbol{x}(t) + \left(\left. \frac{\partial T_t(x_1, x_2)}{\partial x_2} \right|_{x_1 = x_2 = 0} \right) \boldsymbol{x}(t-1) \\ &= \sum_{n=0}^{1} h(n; t) \boldsymbol{x}(t-n), \end{aligned}$$

  provided that $T_t(\cdot, \cdot)$ is first-order differentiable.

  In general, for a linear system,

$$\boldsymbol{y}(t) = \sum_{n=-\infty}^{\infty} \boldsymbol{h}(n; t) \boldsymbol{x}(t-n),$$

  where $\boldsymbol{h}(n; t)$ can be a random variable.

Linear Systems

- Continuous-time System:

  It can be generalized to the continuous-time system as:

$$\boldsymbol{y}(t) = \int_{-\infty}^{\infty} \boldsymbol{h}(\tau; t)\boldsymbol{x}(t-\tau)d\tau.$$

- $\boldsymbol{h}(\tau; t)$ is usually assumed independent of $\boldsymbol{x}(t)$. If they are dependent, the statistics of $\boldsymbol{x}(t)$ (e.g., $P_{\boldsymbol{x}(t)}$) will affect the statistics of the mapping $\boldsymbol{T}_t$ (e.g., $P_{\boldsymbol{y}(t)|\boldsymbol{x}(t)}$) (cf. Slide 9-75).

- Impulse Response:

  To obtain $\boldsymbol{h}(t-s; t)$ for any specific s, just input $\boldsymbol{x}(t) = \delta(t-s)$, and the output equals

$$\boldsymbol{y}(t) \;=\; \int_{-\infty}^{\infty} \boldsymbol{h}(\tau; t)\delta((t-s)-\tau)d\tau = \boldsymbol{h}(t-s; t).$$

Time-invariant Systems

Definition (Time-invariant systems) A system is called *time-invariant* if $\boldsymbol{T}_t(x^{\mathcal{I}}, \zeta) = \boldsymbol{T}(x^{\mathcal{I}}, \zeta)$.

- If the system is time-invariant, we have $\boldsymbol{h}(\tau; t) = \boldsymbol{h}(\tau)$, which indicates that if

$$\boldsymbol{y}(t) = \int_{-\infty}^{\infty} \boldsymbol{h}(\tau)\boldsymbol{x}(t-\tau)d\tau \text{ is the output due to input } \boldsymbol{x}(t),$$

  then

$$\boldsymbol{y}(t-s) = \int_{-\infty}^{\infty} \boldsymbol{h}(\tau)\boldsymbol{x}((t-s)-\tau)d\tau \text{ is the output due to input } \boldsymbol{x}(t-s).$$

- For a linear time-invariant system,

$$\boldsymbol{y}(t) = \int_{-\infty}^{\infty} \boldsymbol{h}(\tau)\boldsymbol{x}(t-\tau)d\tau.$$

  In such case, $\boldsymbol{h}(\tau) = \boldsymbol{y}(\tau)$ (or $\boldsymbol{h}(t) = \boldsymbol{y}(t)$) with $\boldsymbol{x}(t) = \delta(t)$.

Lemma If the input $\boldsymbol{x}(t)$ to a linear time-invariant system is SSS, its output $\boldsymbol{y}(t)$ is also SSS.

Limitation of Convolutionalization of Systems

- Not all linear systems can be represented in *convolutional form* or have legitimate *impulse response*.

For example,

$$\boldsymbol{y}(t) = d\boldsymbol{x}(t)/dt = \boldsymbol{x}'(t)$$

is a linear system because

$$a\boldsymbol{y}_1(t) + b\boldsymbol{y}_2(t) = a\boldsymbol{x}_1'(t) + b\boldsymbol{x}_2'(t) = (a\boldsymbol{x}_1(t) + b\boldsymbol{x}_2(t))'.$$

It is also a time-invariant system because $\boldsymbol{y}(t-s)$ is the output due to input $\boldsymbol{x}(t-s)$ for any s. Hence, we only need to determine $\boldsymbol{h}(\tau)$ with $\boldsymbol{x}(t) = \delta(t)$. However,

$$\frac{d\boldsymbol{x}(t)}{dt} = \frac{d\delta(t)}{dt} = \lim_{\epsilon \downarrow 0} \frac{\delta(t+\epsilon) - \delta(t)}{\epsilon} = \text{undefined}.$$

Dirac Delta Function

- It is a function that exists only in principle.
- Define the Dirac delta function $\delta(t)$ as:

$$\delta(t) = \delta(-t) = \begin{cases} \infty, & t = 0; \\ 0, & t \neq 0 \end{cases} \quad \text{and} \quad \int_{-\infty}^{\infty} \delta(t)dt = \int_{-\infty}^{\infty} \delta(-t)dt = 1.$$

- **Replication Property:** Define the operation on $\delta(t)$ as for every **continuous** point of $g(t)$,

$$g(t) = \int_{-\infty}^{\infty} g(\tau)\delta(t-\tau)d\tau = \int_{-\infty}^{\infty} g(\tau)\delta(\tau - t)d\tau.$$

Dirac Delta Function

The Replication Property induces that

$$\delta(t) = 2\delta(t) \quad \text{but} \quad 1 = \int_{-\infty}^{\infty} \delta(t)dt \neq \int_{-\infty}^{\infty} 2\delta(t)dt = 2,$$

where "$g(t) = 1$" on the left-hand-side and "$g(t) = 2$" on the right-hand-side! Note that in usual operations,

$$f(t) = g(t) \text{ for } t \in \Re \text{ except for countably many points}$$

$$\Rightarrow \int_{-\infty}^{\infty} f(t)dt = \int_{-\infty}^{\infty} g(t)dt \quad \left(\text{if } \int_{-\infty}^{\infty} f(t)dt \text{ is finite}\right).$$

Hence, the multiplicative constant on $\delta(t)$ **cannot be omitted** because

$$\text{saying } \delta(0) = \infty = \infty = 2\delta(0) \text{ is tricky!}$$

Comment: $\boxed{x + a = y + a \Rightarrow x = y}$ is incorrect if $a = \infty$.
As a result, saying $\boxed{\infty = \infty}$ (or $\boxed{\delta(t) = 2\delta(t)}$) is not a "rigorously defined" statement.

Dirac Delta Function

Summary: The Dirac delta function is *meaningful* only through its *replication property*.

For example, the hard limiter in Slide 9-83:

$$f(x) = \int_{-\infty}^{x} 2\delta(\tau)d\tau = \int_{-\infty}^{\infty} (2 \cdot \mathbf{1}\{\tau < x\})\,\delta(\tau)d\tau = \begin{cases} 2 \cdot \mathbf{1}\{x > 0\}, & x \neq 0; \\ \text{undefined}, & x = 0 \end{cases}$$

is guaranteed to equal $g_{\text{Hard Limiter}}(x) + 1$ only when $x \neq 0$ because $2 \cdot \mathbf{1}\{x > 0\}$ is discontinuous at $x = 0$.

The introduction of Dirac delta function leads to a non-logical inference that $\int_{-\infty}^{x} g'(\tau)d\tau = g(x) + c$ is not always true for every x.

Example of Time-Varying Systems

Example of Time-Varying Systems (Analog modulator)

Suppose that $\boldsymbol{h}(\tau;t) = \delta(\tau)e^{j\omega_0 t}$. Then,

$$\boldsymbol{y}(t) = \int_{-\infty}^{\infty} \delta(\tau)e^{j\omega_0 t}\boldsymbol{x}(t-\tau)d\tau = \boldsymbol{x}(t)e^{j\omega_0 t}.$$

- It is definitely a linear system.
- It is time-varying because the output due to $\boldsymbol{x}(t-s)$ is not a shift of the output due to $\boldsymbol{x}(t)$.

I use a "proprietary" notation for convolution operation in the time-varying system as

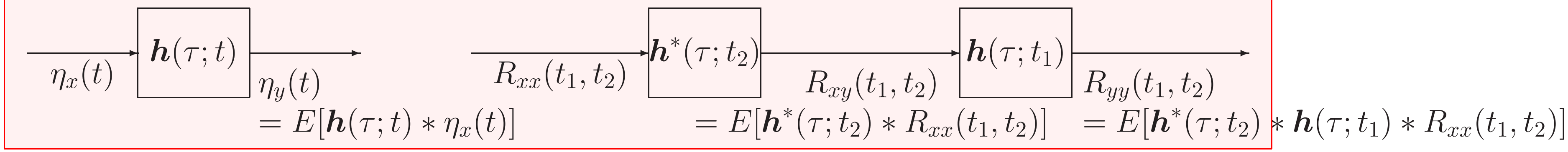
$$\boldsymbol{h}(\tau;t) * \boldsymbol{x}(t) = \int_{-\infty}^{\infty} \boldsymbol{h}(\tau;t)\boldsymbol{x}(t-\tau)d\tau.$$

For time-invariant systems, the conventional notation $\boldsymbol{h}(t) * \boldsymbol{x}(t)$ will also be used in order to be consistent with the text when no ambiguity is introduced:

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

Fundamental Theorem and Theorem 9-2

Fundamental Theorem and Theorem 9-2 For any linear system (that is defined via convolution operation),



Proof:

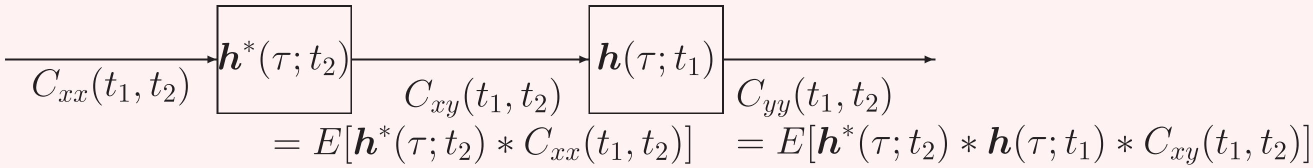
$$
\begin{aligned}
E[\boldsymbol{y}(t)] &= \int_{-\infty}^{\infty} E[\boldsymbol{h}(\tau;t)\boldsymbol{x}(t-\tau)]d\tau = \int_{-\infty}^{\infty} E[\boldsymbol{h}(\tau;t)]E[\boldsymbol{x}(t-\tau)]d\tau \\
&= \int_{-\infty}^{\infty} E[\boldsymbol{h}(\tau;t)]\eta_x(t-\tau)d\tau.
\end{aligned}
$$

$$
\begin{aligned}
R_{xy}(t_1,t_2) &= E[\boldsymbol{x}(t_1)\boldsymbol{y}^*(t_2)] = E\left[\boldsymbol{x}(t_1)\int_{-\infty}^{\infty} \boldsymbol{h}^*(\tau;t_2)\boldsymbol{x}^*(t_2-\tau)d\tau\right] \\
&= \int_{-\infty}^{\infty} E[\boldsymbol{h}^*(\tau;t_2)]E[\boldsymbol{x}(t_1)\boldsymbol{x}^*(t_2-\tau)]d\tau = \int_{-\infty}^{\infty} E[\boldsymbol{h}^*(\tau;t_2)]R_{xx}(t_1,t_2-\tau)d\tau \\
&= E\left[\int_{-\infty}^{\infty} \boldsymbol{h}^*(\tau;t_2)R_{xx}(t_1,t_2-\tau)d\tau\right].
\end{aligned}
$$

Fundamental Theorem and Theorem 9-2

$$\begin{aligned} R_{yy}(t_1,t_2) &= E[\boldsymbol{y}(t_1)\boldsymbol{y}^*(t_2)] = E\left[\int_{-\infty}^{\infty}\boldsymbol{h}(\tau;t_1)\boldsymbol{x}(t_1-\tau)d\tau\int_{-\infty}^{\infty}\boldsymbol{h}^*(s;t_2)\boldsymbol{x}^*(t_2-s)ds\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E[\boldsymbol{h}^*(s;t_2)\boldsymbol{h}(\tau;t_1)]E[\boldsymbol{x}(t_1-\tau)\boldsymbol{x}^*(t_2-s)]d\tau ds \\ &= E\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\boldsymbol{h}^*(s;t_2)\boldsymbol{h}(\tau;t_1)R_{xx}(t_1-\tau,t_2-s)d\tau ds\right]. \qquad \square \end{aligned}$$

Corollary For any linear system (that is defined via convolution operation),



Final note on Fundamental Theorem and Theorem 9-2:

- The above Fundamental Theorem, Theorem 9-2 and Corollary also apply to linear systems without legitimate convolutional forms, e.g., differentiators.
  - By treating the system as $\boldsymbol{y}(t) = \lim_{\epsilon\downarrow 0}\int_{-\infty}^{\infty}\boldsymbol{h}_\epsilon(\tau)\boldsymbol{x}(t-\tau)d\tau$ with $\boldsymbol{h}_\epsilon(\tau) = \dfrac{\delta(\tau+\epsilon)-\delta(\tau)}{\epsilon}$.

Differentiators

Definition A differentiator is a linear time-invariant (deterministic) system whose output is the derivative of the input.

By Fundamental Theorem,

$$\eta_y(t) = \frac{\partial \eta_x(t)}{\partial t}.$$

By Theorem 9-2 (regard that $\boxed{h_\epsilon(\tau) = [\delta(\tau+\epsilon) - \delta(\tau)]/\epsilon \text{ with } \epsilon \downarrow 0 \text{ is real}}$),

$$R_{xy}(t_1,t_2) = \frac{\partial R_{xx}(t_1,t_2)}{\partial t_2} \quad \text{and} \quad R_{yy}(t_1,t_2) = \frac{\partial R_{xy}(t_1,t_2)}{\partial t_1} = \frac{\partial^2 R_{xx}(t_1,t_2)}{\partial t_1 \partial t_2}$$

By the corollary on Slide 9-97,

$$C_{xy}(t_1,t_2) = \frac{\partial C_{xx}(t_1,t_2)}{\partial t_2} \quad \text{and} \quad C_{yy}(t_1,t_2) = \frac{\partial C_{xy}(t_1,t_2)}{\partial t_1} = \frac{\partial^2 C_{xx}(t_1,t_2)}{\partial t_1 \partial t_2}$$

Exercise (cf. page 403 on text) Let the input $\boldsymbol{x}(t)$ to a differentiator is a Poisson process. The resultant output $\boldsymbol{y}(t)$ is a train of Poisson impulses

$$\boldsymbol{y}(t) = \sum_i \delta(t - \boldsymbol{t}_i).$$

Find the mean and autocorrelation functions of the Poisson impulse process.

Differential Equations

Definition A deterministic differential equation with random excitation is an equation of the form:

$$a_n \boldsymbol{y}^{(n)}(t) + \cdots + a_0 \boldsymbol{y}(t) = \boldsymbol{x}(t).$$

With the assumption that the initial condition is zero, $\boldsymbol{y}(t)$ is unique, and is the output due to input $\boldsymbol{x}(t)$ onto a linear time-invariant deterministic system.

Again, by Fundamental Theorem,

$$a_n \eta_y^{(n)}(t) + \cdots + a_0 \eta_y(t) = \eta_x(t) \text{ with } \eta_y(0) = \cdots = \eta_y^{(n-1)}(0) = 0.$$

By Theorem 9-2,

$$a_n \frac{\partial^n R_{xy}(t_1,t_2)}{\partial t_2^n} + \cdots + a_0 R_{xy}(t_1,t_2) = R_{xx}(t_1,t_2) \text{ with } R_{xy}(t_1,0) = \cdots = \frac{\partial^{n-1} R_{xy}(t_1,0)}{\partial t_2^{n-1}} = 0$$

$$a_n \frac{\partial^n R_{yy}(t_1,t_2)}{\partial t_1^n} + \cdots + a_0 R_{yy}(t_1,t_2) = R_{xy}(t_1,t_2) \text{ with } R_{yy}(0,t_2) = \cdots = \frac{\partial^{n-1} R_{yy}(0,t_2)}{\partial t_1^{n-1}} = 0$$

Generalization of Theorem 9-2

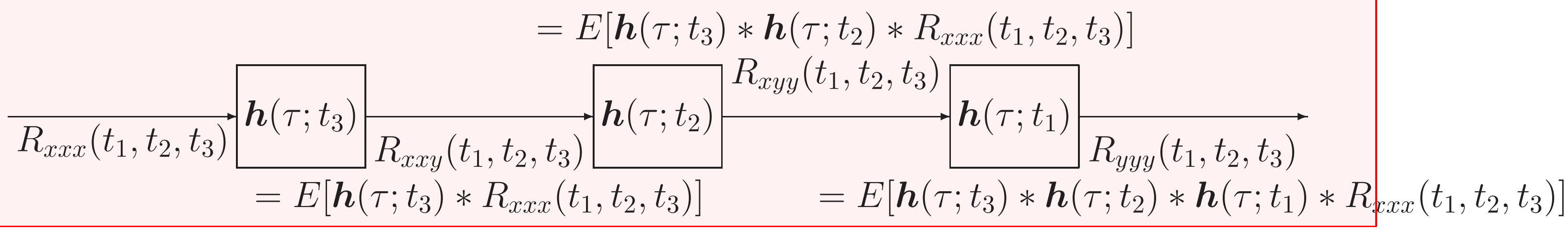
Generalization of Theorem 9-2

For a **real** system, define

$$R_{xxx}(t_1,t_2,t_3)=E[\boldsymbol{x}(t_1)\boldsymbol{x}(t_2)\boldsymbol{x}(t_3)],\quad R_{xxy}(t_1,t_2,t_3)=E[\boldsymbol{x}(t_1)\boldsymbol{x}(t_2)\boldsymbol{y}(t_3)],$$

$$R_{xyy}(t_1,t_2,t_3)=E[\boldsymbol{x}(t_1)\boldsymbol{y}(t_2)\boldsymbol{y}(t_3)],\quad R_{yyy}(t_1,t_2,t_3)=E[\boldsymbol{y}(t_1)\boldsymbol{y}(t_2)\boldsymbol{y}(t_3)].$$

For any linear system,



The end of Section 9-2 Systems with Stochastic Inputs

9-3 The Power Spectrum

Definition (Power spectrum) The *power spectrum* (or *spectral density*) of a WSS process $\boldsymbol{x}(t)$ is the Fourier transform $S_{xx}(\omega)$ of its autocorrelation $R_{xx}(\tau) = E[\boldsymbol{x}(t+\tau)\boldsymbol{x}^*(t)]$. Specifically,

$$S_{xx}(\omega) = \int_{-\infty}^{\infty} R_{xx}(\tau)e^{-j\omega\tau}d\tau.$$

- $R_{xx}(\tau) = \dfrac{1}{2\pi}\displaystyle\int_{-\infty}^{\infty} S_{xx}(\omega)e^{j\omega\tau}d\omega.$
- Since $R_{xx}(-\tau) = R_{xx}^*(\tau)$, $S_{xx}(\omega)$ is real. (It is also non-negative, which will be proved in Wiener-Khinchin Theorem in Slide 9-115.)
- If $\boldsymbol{x}(t)$ is real, $R_{xx}(\tau)$ is real and even, and so is $S_{xx}(\omega)$.
  In such case,

$$S_{xx}(\omega) = 2\int_{0}^{\infty} R_{xx}(\tau)\cos(\omega\tau)d\tau$$

$$R_{xx}(\tau) = \frac{1}{\pi}\int_{0}^{\infty} S_{xx}(\omega)\cos(\omega\tau)d\omega$$

Definition (Cross-power spectrum) The *cross-power spectrum* of two jointly WSS processes $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$ is the Fourier transform $S_{xy}(\omega)$ of their cross-correlation $R_{xy}(\tau) = E[\boldsymbol{x}(t+\tau)\boldsymbol{y}^*(t)]$. Specifically,

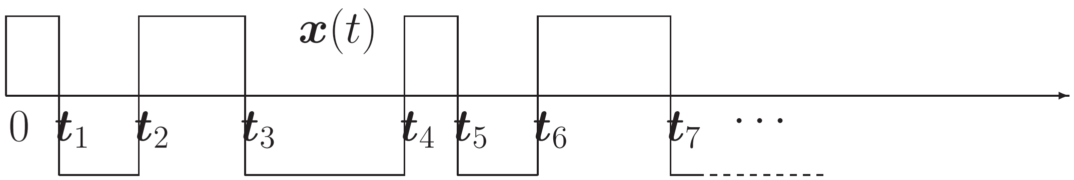
$$S_{xy}(\omega) = \int_{-\infty}^{\infty} R_{xy}(\tau)e^{-j\omega\tau}d\tau.$$

Example 9-22 (Continue from Example 9-6 on Slide 9-50): Semi-random Telegraph Signal

Following Example 9-5 under $\lambda(t) = \lambda$, we re-define

$$\boldsymbol{x}(t) = \begin{cases} 1, & \text{if } \boldsymbol{n}[0,t) \text{ is even;} \\ -1, & \text{if } \boldsymbol{n}[0,t) \text{ is odd.} \end{cases}$$

Determine the power spectrum of $\boldsymbol{x}(t)$.



Answer: We already derive that $R_{xx}(\tau) = e^{-2\lambda|\tau|}$. Hence,

$$\begin{aligned}
S_{xx}(\omega) &= \int_{-\infty}^{\infty} e^{-2\lambda|\tau|} e^{-j\omega\tau} d\tau = \int_{0}^{\infty} e^{-2\lambda\tau} e^{-j\omega\tau} d\tau + \int_{-\infty}^{0} e^{2\lambda\tau} e^{-j\omega\tau} d\tau \\
&= \int_{0}^{\infty} e^{-(2\lambda+j\omega)\tau} d\tau + \int_{0}^{\infty} e^{-(2\lambda-j\omega)\tau} d\tau \\
&= \frac{1}{2\lambda + j\omega} + \frac{1}{2\lambda - \jmath\omega} \\
&= \frac{4\lambda}{4\lambda^2 + \omega^2}.
\end{aligned}$$

□

Covariance Spectrum

Definition (Covariance spectrum) The *covariance spectrum* of a WSS process $\boldsymbol{x}(t)$ is the Fourier transform $S^c_{xx}(\omega)$ of its autocovariance $C_{xx}(\tau) = E[(\boldsymbol{x}(t+\tau)-\eta_x(t+\tau))(\boldsymbol{x}(t)-\eta_x(t))^*] = E[(\boldsymbol{x}(t+\tau)-\eta)(\boldsymbol{x}(t)-\eta)^*]$. Specifically,

$$S^c_{xx}(\omega) = \int_{-\infty}^{\infty} C_{xx}(\tau)e^{-j\omega\tau}d\tau.$$

- It can be easily shown that $S_{xx}(\omega) = S^c_{xx}(\omega) + 2\pi\eta^2\delta(\omega)$.

Exercise (Example 9-23) Let the input $\boldsymbol{x}(t)$ to a differentiator is a Poisson process. The resultant output $\boldsymbol{y}(t)$ is a train of Poisson impulses

$$\boldsymbol{y}(t) = \sum_i \delta(t - \boldsymbol{t}_i).$$

Find the covariance spectrum of the Poisson impulse process.

Existence of Processes with Specified Power Spectrum

Lemma Given an arbitrary non-negative integrable function $S(\omega)$, there exists a complex WSS process $\boldsymbol{x}(t)$ whose power spectrum is equal to $S(\omega)$.

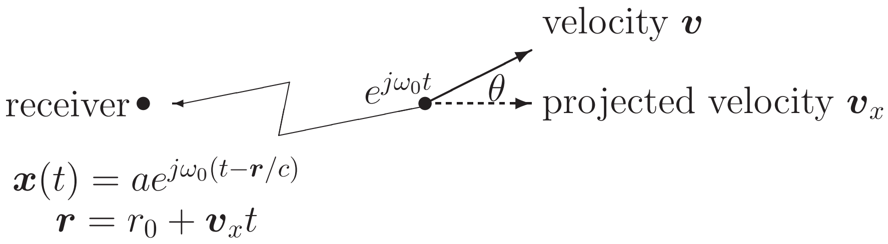
Proof: The desired complex process can be defined as $\boldsymbol{x}(t) = ae^{j(\boldsymbol{\omega}t-\boldsymbol{\varphi})}$ for $|a|^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)d\omega$, where $\boldsymbol{\omega}$ is a random variable with density $f_{\boldsymbol{\omega}}(\omega) = \frac{S(\omega)}{2\pi|a|^2}$ if $|a| > 0$, and with arbitrary density if $|a| = 0$, and $\boldsymbol{\varphi}$ is uniformly distributed over $[-\pi, \pi)$ and independent of $\boldsymbol{\omega}$. The validation of WSS of $\boldsymbol{x}(t)$ is left to you as an exercise (cf. Slide 9-64). $\square$

Lemma Given an arbitrary non-negative integrable even function $S(\omega)$, there exists a real WSS process $\boldsymbol{x}(t)$ whose power spectrum is equal to $S(\omega)$.

Proof: The desired real process can be defined as $\boldsymbol{x}(t) = a\cos(\boldsymbol{\omega}t + \boldsymbol{\varphi})$ for $a^2 = \frac{1}{\pi}\int_{-\infty}^{\infty} S(\omega)d\omega$, where $\boldsymbol{\omega}$ is a random variable with density $f_{\boldsymbol{\omega}}(\omega) = \frac{S(\omega)}{\pi a^2}$ if $|a| > 0$, and with arbitrary density if $|a| = 0$, and $\boldsymbol{\varphi}$ is uniformly distributed over $[-\pi, \pi)$ and independent of $\boldsymbol{\omega}$. The validation of WSS of $\boldsymbol{x}(t)$ is already done in Slide 9-64. $\square$

Doppler Effect

Example 9-24 (Doppler effect)



A moving transmitter transmits a harmonic oscillator signal $e^{j\omega_0 t}$ to a fixed-in-location receiver as shown above.

Assume that $\boldsymbol{v}$ is a random variable with density $f_{\boldsymbol{v}}(v)$, and $\theta \in (-\pi/2, \pi/2)$.

Hence, the received signal equals $\boldsymbol{x}(t) = ae^{j(\boldsymbol{\omega} t - \varphi)}$, where

$$\boldsymbol{\omega} = \omega_0\left(1 - \frac{\boldsymbol{v}_x}{c}\right) = \omega_0\left(1 - \frac{\boldsymbol{v}\cos(\theta)}{c}\right) \quad \text{and} \quad \varphi = \frac{r_0\omega_0}{c}.$$

Determine the power spectrum of $\boldsymbol{x}(t)$.

Answer:

The uniformity of $\boldsymbol{\varphi}$ is nothing to do with the power spectrum of $\boldsymbol{x}(t)$ in the first lemma in Slide 9-105. It is required only to fulfill the WSS requirement of $\boldsymbol{x}(t)$.

Use the lemma in the previous slide,

$$S_{xx}(\omega) = 2\pi|a|^2 f_{\boldsymbol{\omega}}(\omega) = 2\pi|a|^2 \frac{c}{\omega_0\cos(\theta)} f_{\boldsymbol{v}}\left(\frac{c}{\cos(\theta)}\left(1 - \frac{\omega}{\omega_0}\right)\right). \qquad \square$$

Doppler Effect

$$
\begin{aligned}
f_{\boldsymbol{\omega}}(\omega) &= \frac{\partial}{\partial \omega} F_{\boldsymbol{\omega}}(\omega) = \frac{\partial}{\partial \omega} \Pr[\boldsymbol{\omega} \le \omega] \\
&= \frac{\partial}{\partial \omega} \Pr\left[\omega_0 \left(1 - \frac{\boldsymbol{v} \cos(\theta)}{c}\right) \le \omega\right] = \frac{\partial}{\partial \omega} \Pr\left[\boldsymbol{v} \ge \frac{c}{\cos(\theta)} \left(1 - \frac{\omega}{\omega_0}\right)\right] \\
&= \frac{\partial}{\partial \omega} \left\{1 - F_{\boldsymbol{v}}\left(\frac{c}{\cos(\theta)} \left(1 - \frac{\omega}{\omega_0}\right)\right)\right\} = \frac{c}{\omega_0 \cos(\theta)} f_{\boldsymbol{v}}\left(\frac{c}{\cos(\theta)} \left(1 - \frac{\omega}{\omega_0}\right)\right)
\end{aligned}
$$

- Further assume that $\boldsymbol{v}$ is uniformly distributed over $[v_1, v_2]$. Then,

$$
\begin{aligned}
S_{xx}(\omega) &= 2\pi |a|^2 \frac{c}{\omega_0 \cos(\theta)} f_{\boldsymbol{v}}\left(\frac{c}{\cos(\theta)} \left(1 - \frac{\omega}{\omega_0}\right)\right) \\
&= 2\pi |a|^2 \frac{c}{\omega_0 (v_2 - v_1) \cos(\theta)} \\
&\qquad \text{for } \omega_0 \left(1 - \frac{v_2}{c} \cos(\theta)\right) \le \omega \le \omega_0 \left(1 - \frac{v_1}{c} \cos(\theta)\right).
\end{aligned}
$$

Thus, the **random** motion causes broadening of the spectrum.

Doppler Effect

- If $\Pr[\boldsymbol{v} = v] = 1$, then

$$\Pr\left[\boldsymbol{\omega} = \omega_0 \left(1 - \frac{v}{c}\cos(\theta)\right)\right] = 1,$$

which implies

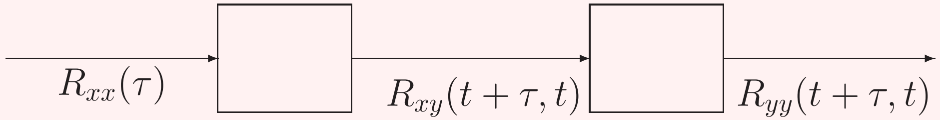
$$S_{xx}(\omega) = 2\pi|a|^2 \cdot \delta\left(\omega - \omega_0\left(1 - \frac{v}{c}\cos(\theta)\right)\right).$$

Thus, the **deterministic** motion causes a shift in spectrum frequency.

- If $\cos(\theta) \downarrow 0$, $\boldsymbol{\omega} = \omega_0$ with probability one. Hence, $S_{xx}(\omega) = 2\pi|a|^2\delta(\omega - \omega_0)$. Thus, no spectrum broadening or frequency-shift is caused by perpendicular motion, either random or deterministic.

Linear Systems Revisited

Fundamental Theorem and Theorem 9-2 Revisited For any linear system with WSS input and $\boldsymbol{h}(\tau;t) = \boldsymbol{h}_1(\tau)\boldsymbol{h}_2(t)$,



$$= E\{\boldsymbol{h}_2^*(t)[\boldsymbol{h}_1^*(-\tau) * R_{xx}(\tau)]\} = E\{\boldsymbol{h}_2(t+\tau)\boldsymbol{h}_2^*(t)[\boldsymbol{h}_1^*(-\tau) * \boldsymbol{h}_1(\tau) * R_{xx}(\tau)]\}$$

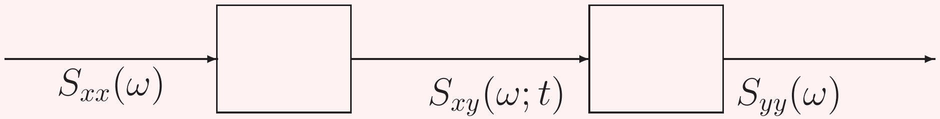
Proof: This is a consequence of

$$\begin{aligned} R_{xy}(t_1 = t+\tau, t_2 = t) &= E\left[\int_{-\infty}^{\infty} \boldsymbol{h}^*(u;t_2) R_{xx}(t_1 - t_2 + u)du\right] \\ &= E\left[\int_{-\infty}^{\infty} \boldsymbol{h}_2^*(t)\boldsymbol{h}_1^*(u) R_{xx}(\tau + u)du\right] \\ &= E\left[\boldsymbol{h}_2^*(t)\int_{-\infty}^{\infty} \boldsymbol{h}_1^*(-u) R_{xx}(\tau - u)du\right]. \end{aligned}$$

$$\begin{aligned} R_{yy}(t_1 = t+\tau, t_2 = t) &= E\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \boldsymbol{h}^*(u';t_2)\boldsymbol{h}(u;t_1) R_{xx}(t_1 - t_2 - u + u')dudu'\right] \\ &= E\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \boldsymbol{h}_2^*(t_2)\boldsymbol{h}_1^*(u')\boldsymbol{h}_2(t_1)\boldsymbol{h}_1(u) R_{xx}(\tau - u + u')dudu'\right] \\ &= E\left[\boldsymbol{h}_2(t+\tau)\boldsymbol{h}_2^*(t)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \boldsymbol{h}_1^*(-u')\boldsymbol{h}_1(u) R_{xx}(\tau - u - u')dudu'\right] \end{aligned}$$

□

Theorem 9-4 For any linear system with WSS input and $\boldsymbol{h}(\tau; t) = \boldsymbol{h}_1(\tau)e^{j\omega_0 t}$,



Proof:

$$
\begin{aligned}
S_{xy}(\omega; t) &= \int_{-\infty}^{\infty} R_{xy}(t+s, t)e^{-j\omega s}ds \\
&= e^{-j\omega_0 t}\int_{-\infty}^{\infty} E\left[\int_{-\infty}^{\infty} \boldsymbol{h}_1^*(-\tau)R_{xx}(s-\tau)d\tau\right] e^{-j\omega s}ds \\
&= e^{-j\omega_0 t}E\left[\int_{-\infty}^{\infty} \boldsymbol{h}_1^*(-\tau)e^{-j\omega\tau}\left(\int_{-\infty}^{\infty} R_{xx}(v)e^{-j\omega v}dv\right)d\tau\right], \quad v = s - \tau \\
&= e^{-j\omega_0 t}E[\boldsymbol{H}_1^*(\omega)S_{xx}(\omega)].
\end{aligned}
$$

Linear Systems Revisited

$$\begin{aligned}
S_{yy}(\omega;t) &= \int_{-\infty}^{\infty} R_{yy}(t+s,t)e^{-j\omega s}ds \\
&= \int_{-\infty}^{\infty} e^{-j\omega_0 t}e^{j\omega_0(t+s)}E\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\boldsymbol{h}_1^*(-\tau')\boldsymbol{h}_1(\tau)R_{xx}([s-\tau']-\tau)d\tau'd\tau\right]e^{-j\omega s}ds \\
&= E\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\boldsymbol{h}_1^*(\tau'')\boldsymbol{h}_1(\tau)R_{xx}(s-\tau+\tau'')e^{-j(\omega-\omega_0)s}dsd\tau d\tau''\right],\ \tau''=-\tau \\
&= E\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\boldsymbol{h}_1^*(\tau'')\boldsymbol{h}_1(\tau)R_{xx}(v)e^{-j(\omega-\omega_0)(v+\tau-\tau'')}dvd\tau d\tau''\right],\ v=s-\tau+\tau'' \\
&= S_{xx}(\omega-\omega_0)E[|\boldsymbol{H}_1(\omega-\omega_0)|^2]
\end{aligned}$$

$\square$

Linear Systems Revisited

Transfer-function form of linear systems

- In addition to *convolutionalization* of linear system, a linear time-invariant system can be represented in a spectrum form through a (random) transfer function $\boldsymbol{H}(\omega)$ as

$$\boldsymbol{Y}(\omega) = \boldsymbol{X}(\omega)\boldsymbol{H}(\omega),$$

where

$$\boldsymbol{Y}(\omega) = \int_{-\infty}^{\infty} \boldsymbol{y}(t)e^{-j\omega t}dt \quad \text{and} \quad \boldsymbol{X}(\omega) = \int_{-\infty}^{\infty} \boldsymbol{x}(t)e^{-j\omega t}dt.$$

$$a\boldsymbol{Y}_1(\omega) + b\boldsymbol{Y}_2(\omega) = [a\boldsymbol{X}_1(\omega) + b\boldsymbol{X}_2(\omega)]\boldsymbol{H}(\omega).$$

Example 9-26 The differentiator in Slide 9-98 can be represented as $\boldsymbol{Y}(\omega) = \boldsymbol{X}(\omega)\boldsymbol{H}(\omega)$ with $\boldsymbol{H}(\omega) = j\omega$.

- **Remark:** If transfer function $\boldsymbol{H}(\omega)$ has inverse Fourier transform $\boldsymbol{h}(\tau)$, then $\boldsymbol{Y}(\omega) = \boldsymbol{X}(\omega)\boldsymbol{H}(\omega)$ can be equivalently represented as $\boldsymbol{y}(t) = \boldsymbol{h}(\tau) * \boldsymbol{x}(t)$.

Extended Fourier Transform

- The Fourier transform of a function $g(t)$ exists if $\int_{-\infty}^{\infty} |g(t)| dt < \infty$, and $g(t)$ only has finite number of local maxima, minima and discontinuities in every finite interval.

Define a function $g(x) = \begin{cases} 1 & \text{if } |x| < 1 \text{ and } x \text{ is not a rational;} \\ 0 & \text{otherwise} \end{cases}$.

Then, $\int_{-1}^{1} |g(x)| dx \leq \int_{-1}^{1} dx = 2 < \infty$, but for such a function, the *conventional* Fourier transform is not defined!

- **Extended Fourier Transform:** Define the extended Fourier transform of a function $g(t)$ (that does not have Fourier transform) as $\lim_{n\to\infty} G_n(\omega)$, where $G_n(\omega)$ is the Fourier transform of $g_n(t)$ and $\lim_{n\to\infty} g_n(t) = g(t)$.

Example: $g(t) = 1$ does not have Fourier transform, but has extended Fourier transform through $g_n(t) = e^{-|t|/n}$ as

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} e^{-|t|/n} e^{-j\omega t} dt = \lim_{n\to\infty} \frac{2n}{1 + n^2\omega^2} = 2\pi\delta(\omega).$$

Adding the multiplicative constant 2π is because $\displaystyle\int_{-\infty}^{\infty} \frac{2n}{1 + n^2\omega^2} d\omega = 2\pi$.

Hilbert Transform

Definition (Hilbert transform) The system response of a *quadrature* filter

$$H(\omega) = -j\text{sgn}(\omega) = \begin{cases} -j, & \omega > 0; \\ j, & \omega < 0 \end{cases}$$

due to system input $\boldsymbol{x}(t)$ is called the *Hilbert transform* of $\boldsymbol{x}(t)$.

- $H(\omega)$ is named the *quadrature* filter because it is an all-pass filter with $\pm 90°$ phase shift.
- The extended inverse Fourier transform of $H(\omega)$ is given by

$$\begin{aligned}
\frac{1}{2\pi}\int_{-\infty}^{\infty} H_n(\omega)e^{j\omega\tau}d\omega &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(-j\text{sgn}(\omega)e^{-|\omega|/n}\right)e^{j\omega\tau}d\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{0}\left(je^{\omega/n}\right)e^{j\omega\tau}d\omega + \frac{1}{2\pi}\int_{0}^{\infty}\left(-je^{-\omega/n}\right)e^{j\omega\tau}d\omega \\
&= \frac{n}{2\pi(n\tau - j)} + \frac{n}{2\pi(n\tau + j)} \\
&= \frac{n^2\tau}{\pi(n^2\tau^2 + 1)} \quad \xrightarrow{n\to\infty} \quad \begin{cases} \frac{1}{\pi\tau} & \tau \neq 0 \\ 0 & \tau = 0 \end{cases}
\end{aligned}$$

Analytical Signal and Wiener-Khinchin Theorem

Definition (Analytical signal) The complex process of $\boldsymbol{z}(t) = \boldsymbol{x}(t) + j\hat{\boldsymbol{x}}(t)$ is called the *analytical signal* of a real process $\boldsymbol{x}(t)$, where $\hat{\boldsymbol{x}}(t)$ is the Hilbert transform of $\boldsymbol{x}(t)$.

$$\begin{aligned}
\boldsymbol{Z}(\omega) &= \boldsymbol{X}(\omega) + j\hat{\boldsymbol{X}}(\omega) \\
&= \boldsymbol{X}(\omega) + j\boldsymbol{X}(\omega)H(\omega) \\
&= \boldsymbol{X}(\omega) + j\boldsymbol{X}(\omega)[-j\text{sgn}(\omega)] \\
&= [1 + \text{sgn}(\omega)]\boldsymbol{X}(\omega) \\
&= 2\boldsymbol{X}(\omega) \cdot \mathbf{1}[\omega > 0] + \boldsymbol{X}(\omega) \cdot \mathbf{1}[\omega = 0].
\end{aligned}$$

Theorem (Wiener-Khinchin Theorem) The power spectrum for any WSS process $\boldsymbol{x}(t)$ is non-negative everywhere.

Proof: First note that $S_{xx}(\omega)$ that is defined as the (non-extended) Fourier transform of $R_{xx}(\tau)$ is continuous everywhere. Suppose $S_{xx}(\omega)$ is negative for some interval (ω_1, ω_2). Define a filter $H(\omega) = 1$ for $\omega \in (\omega_1, \omega_2)$, and zero, otherwise. Then the output power spectrum due to $\boldsymbol{x}(t)$ is equal to $S_{yy}(\omega) = S_{xx}(\omega)|H(\omega)|^2$. By

$$E[|\boldsymbol{y}(t)|^2] = \frac{1}{2\pi}\int_{\omega_1}^{\omega_2} S_{xx}(\omega)|H(\omega)|^2 d\omega = \frac{1}{2\pi}\int_{\omega_1}^{\omega_2} S_{xx}(\omega) d\omega \geq 0,$$

we obtain the desired contradiction. □

Integrated Spectrum and Covariance Spectrum

Definition (Integrated spectrum) Define the integrated spectrum of a process $\boldsymbol{x}(t)$ as:

$$F_{xx}(\omega) \triangleq \int_{-\infty}^{\omega} S_{xx}(s)ds.$$

- The role of the integrated spectrum versus the power spectrum is similar to that of the cdf versus pdf of a random variable.
- Note that the cdf alone is sufficient to well-define a random variable (cf. Slide 9-8). Also note that the pdf of a random variable may not exist (without introducing the Dirac delta functions)!
- Hence, the introduction of integrated spectrums avoids the use of singularity functions such as Dirac delta functions, and the autocorrelation function can be obtained through a Riemann-Stieltjes integral:

$$R_{xx}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{j\omega t} dF_{xx}(\omega).$$

Integrated Spectrum and Covariance Spectrum

Definition (Integrated covariance spectrum) Define the integrated covariance spectrum of a process $\boldsymbol{x}(t)$ as:

$$F_{xx}^{c}(\omega) \triangleq \int_{-\infty}^{\omega} C_{xx}(s)ds.$$

Non-negative Spectrum and P.D.

Summary

- A function $R(\tau)$ is the autocorrelation function of some WSS process $\boldsymbol{x}(t)$, if its Fourier transform $S(\omega)$ is non-negative (cf. Slide 9-105).
- If a function $R(\tau)$ has non-negative Fourier transform, we can find a process $\boldsymbol{x}(t)$ with autocorrelation function $R(\tau)$ (cf. Slide 9-105).

- There exists a process with autocorrelation $R(t_1, t_2)$ if, and only if, $R(t_1, t_2)$ is p.d. (cf. Slides 9-32 and 9-42).

- A function $R(\tau)$ has non-negative Fourier transform if, and only if, it is p.d., i.e.,

$$\sum_i \sum_j a_i a_j^* R(t_i - t_j) \geq 0 \quad \text{for any complex } a_i \text{ and } a_j.$$

Polya's Sufficient Criterion and Necessary Exclusion

How to examine whether a function $R(\tau)$ is p.d.?
Answer: *Polya's criterion.*

Lemma (Polya's sufficient criterion) A function $R(\tau)$ is p.d., if it is concave for $\tau > 0$ and it tends to a finite limit as $|\tau| \to \infty$.

Theorem 9-5 If the autocorrelation function $R_{xx}(\tau)$ of a WSS process $\boldsymbol{x}(t)$ satisfies that $R_{xx}(\tau_1) = R_{xx}(0)$ for some $\tau_1 \neq 0$, then $R_{xx}(\tau)$ is periodic with period τ_1.

Proof: By Cauchy-Schwartz's inequality,

$$|E[(\boldsymbol{x}(t+\tau+\tau_1) - \boldsymbol{x}(t+\tau))\boldsymbol{x}^*(t)]|^2 \leq E[|\boldsymbol{x}(t+\tau+\tau_1) - \boldsymbol{x}(t+\tau))|^2]E[|\boldsymbol{x}(t)|^2],$$

which is equivalent to:

$$|R_{xx}(\tau+\tau_1) - R_{xx}(\tau)|^2 \leq (2R_{xx}(0) - R_{xx}(\tau_1) - \underbrace{R_{xx}(-\tau_1)}_{=R^*_{xx}(\tau_1)=R_{xx}(0)})R_{xx}(0) = 0.$$

Therefore, $R_{xx}(\tau+\tau_1) = R_{xx}(\tau)$ for every τ. □

Polya's Sufficient Criterion and Necessary Exclusion

Similar proof to Theorem 9-5 can be used to prove the following corollary.

Corollary If the autocorrelation function $R_{xx}(\tau)$ of a WSS process $\boldsymbol{x}(t)$ is continuous at the origin, it is continuous everywhere.

Proof: By Cauchy-Schwartz's inequality,

$$|E[(\boldsymbol{x}(t+\tau+\tau_1)-\boldsymbol{x}(t+\tau))\boldsymbol{x}^*(t)]|^2 \le E[|\boldsymbol{x}(t+\tau+\tau_1)-\boldsymbol{x}(t+\tau))|^2]E[|\boldsymbol{x}(t)|^2],$$

which is equivalent to:

$$|R_{xx}(\tau+\tau_1)-R_{xx}(\tau)|^2 \le (2R_{xx}(0)-R_{xx}(\tau_1)-R_{xx}(-\tau_1))R_{xx}(0).$$

Therefore,

$$\lim_{|\tau_1|\downarrow 0} R_{xx}(\tau_1) = R_{xx}(0) \quad \text{implies} \quad \lim_{|\tau_1|\downarrow 0} R_{xx}(\tau+\tau_1) = R_{xx}(\tau) \text{ for every } \tau.$$

□

Remark

- **Necessary exclusion**: Theorem 9-5 and the followup corollary can be used to exclude those $R(\tau)$ that cannot be the autocorrelation function of some WSS process.

Example 9-30 Function $w(\tau) = \begin{cases} a^2 - \tau^2, & |\tau| < a; \\ 0, & |\tau| > a \end{cases}$ is not an autocorrelation function of any process.

This is because if $w(\tau)$ is the autocorrelation function of $\boldsymbol{x}(t)$, then the autocorrelation function of the differentiator output $\boldsymbol{y}(t)$ due to input $\boldsymbol{x}(t)$ should be:

$$R_{yy}(\tau) = -\frac{\partial^2 w(\tau)}{(\partial \tau)^2} = \begin{cases} 2, & |\tau| < a; \\ 0, & |\tau| > a \end{cases}$$

However, $R_{yy}(\tau)$ is continuous at the origin, but is not continuous at $|\tau| = a$, which indicates that $R_{yy}(\tau)$ cannot be the autocorrelation of any process.

Bound on Cross Spectrums

Lemma For any a and b,

$$\left|\int_a^b S_{xy}(\omega)d\omega\right|^2 \le \left(\int_a^b S_{xx}(\omega)d\omega\right)\left(\int_a^b S_{yy}(\omega)d\omega\right).$$

Proof: Let $\boldsymbol{z}(t)$ and $\boldsymbol{w}(t)$ be respectively the system outputs due to WSS inputs $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$ through filter $H(\omega) = 1\cdot\mathbf{1}\{a < \omega < b\}$. Then, by Cauchy-Schwartz's inequality,

$$\begin{aligned}
\left|E^2[\boldsymbol{z}(t)\boldsymbol{w}^*(t)]\right| &\le E^2\left[|\boldsymbol{z}(t)\boldsymbol{w}^*(t)|\right] \le E[|\boldsymbol{z}(t)|^2]E[|\boldsymbol{w}^*(t)|^2] \\
&= R_{zz}(0)R_{ww}(0) \\
&= \frac{1}{2\pi}\left(\int_{-\infty}^{\infty} S_{xx}(\omega)|H(\omega)|^2 d\omega\right)\frac{1}{2\pi}\left(\int_{-\infty}^{\infty} S_{yy}(\omega)|H(\omega)|^2 d\omega\right) \\
&= \frac{1}{4\pi^2}\left(\int_a^b S_{xx}(\omega)d\omega\right)\left(\int_a^b S_{yy}(\omega)d\omega\right).
\end{aligned}$$

Bound on Cross Spectrums

The proof is completed by noting that:

$$\begin{aligned}
E[\boldsymbol{z}(t)\boldsymbol{w}^*(t)] &= E\left[\int_{-\infty}^{\infty} h(\tau)\boldsymbol{x}(t-\tau)d\tau \cdot \int_{-\infty}^{\infty} h^*(s)\boldsymbol{y}^*(t-s)ds\right] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\tau)h^*(s)R_{xy}(s-\tau)d\tau ds \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\tau)h^*(s)\frac{1}{2\pi}\left(\int_{-\infty}^{\infty} S_{xy}(\omega)e^{j\omega(s-\tau)}d\omega\right) d\tau ds \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{xy}(\omega)|H(\omega)|^2 d\omega \\
&= \frac{1}{2\pi}\int_a^b S_{xy}(\omega)d\omega.
\end{aligned}$$

$\square$

Equality in the MS sense

Definition (Equality in the mean-square (MS) sense) Two processes $\{\boldsymbol{x}(t), t \in \mathcal{I}\}$ and $\{\boldsymbol{y}(t), t \in \mathcal{I}\}$ are equal in the MS sense if, and only if,

$$E[|\boldsymbol{x}(t) - \boldsymbol{y}(t)|^2] = 0 \quad \text{for every } t \in \mathcal{I}.$$

(Page 375 on text) Denote by

$$A_t \triangleq \{\zeta \in S : \boldsymbol{x}(t, \zeta) = \boldsymbol{y}(t, \zeta)\} \quad \text{and} \quad A_\infty \triangleq \bigcap_{t \in \mathcal{I}} A_t.$$

Then, the above definition requires $P(A_t) = 1$ for every specific t, and does not require $P(A_\infty) = 1$.

Note that $P(\bigcap A_t) = 1$ if $P(A_t) = 1$ that is very true for countable intersection but may not be true for uncountably infinite intersection.

Example $S = \mathcal{I} = \Re$, $A_t = S - \{t\}$ and $P(A) = \int_A \frac{1}{\sqrt{2\pi}} e^{-\alpha^2/2} d\alpha$.

Then, $P(A_\infty) = P(\bigcap_{t \in \Re} A_t) = P(\emptyset) = 0$, and still, $P(A_t) = 1$ for every t.

Exercise Is A_∞ guaranteed to be a probabilistically measurable event?

MS Periodicity

Definition A process $\boldsymbol{x}(t)$ is called *MS periodic* if

$$E[|\boldsymbol{x}(t+T) - \boldsymbol{x}(t)|^2] = 0$$

for every t.

Similar to Slide 9-124, the above definition requires $P(A_t) = 1$ for every t, where $A_t = \{\zeta \in S : \boldsymbol{x}(t+T,\zeta) = \boldsymbol{x}(t,\zeta)\}$, but does not require $P(\cap_{t\in\mathcal{I}} A_t) = 1$.

Theorem 9-1 A process $\boldsymbol{x}(t)$ is *MS periodic* if, and only if, its autocorrelation function is *doubly periodic*, namely,

$$R_{xx}(t_1 + mT, t_2 + nT) = R_{xx}(t_1, t_2) \text{ for every integer } m \text{ and } n.$$

Proof:

1. ***Forward:*** Since $\boldsymbol{x}(t)$ is MS periodic, by Cauchy-Schwartz inequality,

$$\begin{aligned} |E\{\boldsymbol{x}(t_1)\cdot(\boldsymbol{x}(t_2+T) - \boldsymbol{x}(t_2))^*\}| &\leq E\{|\boldsymbol{x}(t_1)\cdot(\boldsymbol{x}(t_2+T) - \boldsymbol{x}(t_2))^*|\} \\ &\leq E^{1/2}[|\boldsymbol{x}(t_1)|^2]E^{1/2}[|\boldsymbol{x}(t_2+T) - \boldsymbol{x}(t_2)|^2] \\ &= 0, \end{aligned}$$

which implies $R_{xx}(t_1, t_2+T) = R_{xx}(t_1, t_2)$. The forward proof is completed by repeating using the Cauchy-Schwartz inequality.

2. *Converse*:

- That $R_{xx}(t_1+mT, t_2+nT) = R_{xx}(t_1, t_2)$ for every integer m and n implies

$$R_{xx}(t+T, t+T) = R_{xx}(t+T, t) = R_{xx}(t, t+T) = R_{xx}(t, t).$$

- Hence,

$$\begin{aligned} & E[|\boldsymbol{x}(t+T) - \boldsymbol{x}(t)|^2] \\ = \ & R_{xx}(t+T, t+T) - R_{xx}(t+T, t) - R_{xx}(t, t+T) + R_{xx}(t, t) \\ = \ & 0. \end{aligned}$$

$\square$

MS Continuity

Definition (MS continuity) A process $\boldsymbol{x}(t)$ is called *MS continuous* if

$$\lim_{\epsilon \downarrow 0} E[|\boldsymbol{x}(t+\epsilon) - \boldsymbol{x}(t)|^2] = 0 \text{ for every } t.$$

- Since

  $$E[|\boldsymbol{x}(t+\epsilon)-\boldsymbol{x}(t)|^2] = R_{xx}(t+\epsilon, t+\epsilon) - R_{xx}(t+\epsilon, t) - R_{xx}(t, t+\epsilon) + R_{xx}(t,t),$$

  it turns out that a process is MS continuous if, and only if, its autocorrelation function is continuous.

  That a process $\boldsymbol{x}(t)$ is MS continuous does not imply that its sample $\boldsymbol{x}(t,\zeta)$ is continuous in t. E.g., for Poisson processes, $E[|\boldsymbol{x}(t+\epsilon)-\boldsymbol{x}(t)|^2] = E[|\boldsymbol{n}[t,t+\epsilon)|^2] = \lambda\epsilon(1+\lambda\epsilon) \xrightarrow{\epsilon \downarrow 0} 0$, but $\boldsymbol{x}(t,\zeta)$ is apparently discontinuous at $\boldsymbol{t}_i(\zeta)$.

- As far as MS periodicity is concerned (cf. Slide 9-125), since

  $$E[|\boldsymbol{x}(t+T) - \boldsymbol{x}(t)|^2] = 2 \cdot \text{Re}\{R_{xx}(0) - R_{xx}(T)\}$$

  for a WSS $\boldsymbol{x}(t)$, a WSS process is MS periodic if, and only if, the real part of its autocorrelation function is periodic.

MS Differentiability and MS Integrability

We can likewise define MS differentiability and MS integrability as follows. (Details can be found in Appendix 9A.)

Definition (MS differentiability) A process $\boldsymbol{x}(t)$ is called *MS differentiable* if

$$\lim_{\epsilon\downarrow 0} E\left[\left|\frac{\boldsymbol{x}(t+\epsilon)-\boldsymbol{x}(t)}{\epsilon} - \left(\lim_{\gamma\downarrow 0}\frac{\boldsymbol{x}(t+\gamma)-\boldsymbol{x}(t)}{\gamma}\right)\right|^2\right] = 0 \text{ for every } t.$$

- Since

$$E\left[\left|\frac{\boldsymbol{x}(t+\epsilon)-\boldsymbol{x}(t)}{\epsilon} - \boldsymbol{x}'(t)\right|^2\right] = \frac{R_{xx}(t+\epsilon,t+\epsilon)-R_{xx}(t+\epsilon,t)-R_{xx}(t,t+\epsilon)+R_{xx}(t,t)}{\epsilon^2}$$

$$-\frac{R_{xx'}(t+\epsilon,t)-R_{xx'}(t,t)}{\epsilon} - \frac{R_{x'x}(t,t+\epsilon)-R^*_{x'x}(t,t)}{\epsilon} + R_{x'x'}(t,t),$$

and (cf. Slide 9-98)

$$R_{xx'}(t_1,t_2) = \frac{\partial R_{xx}(t_1,t_2)}{\partial t_2} \quad \text{and} \quad R_{x'x'}(t_1,t_2) = \frac{\partial R_{xx'}(t_1,t_2)}{\partial t_1} = \frac{\partial^2 R_{x'x'}(t_1,t_2)}{\partial t_1 \partial t_2},$$

a process $\boldsymbol{x}(t)$ is MS differentiable if, and only if, $\partial^2 R_{xx}(t_1,t_2)/\partial t_1\partial t_2\big|_{t_1=t_2}$ exists.

MS Differentiability and MS Integrability

Definition (MS integrability) A process $\boldsymbol{x}(t)$ is called *MS (Riemann-)integrable* if

$$\lim_{\Delta\downarrow 0} E\left[\left|\int_a^b \boldsymbol{x}(t)dt - \sum_{i=0}^{\lfloor (b-a)/\Delta\rfloor} \boldsymbol{x}(a+i\Delta)\times\Delta\right|^2\right] = 0 \text{ for every } a \text{ and } b.$$

- Since, by letting $\boldsymbol{y}(t) = \int_{t-(b-a)}^{t} \boldsymbol{x}(s)ds = \int_{-\infty}^{\infty} h(\tau)\boldsymbol{x}(t-\tau)d\tau$ with $h(\tau) = 1$ for $0 \le \tau < (b-a)$,

$$E\left[\left|\boldsymbol{y}(b) - \Delta \sum_{i=0}^{\lfloor (b-a)/\Delta\rfloor} \boldsymbol{x}(a+i\Delta)\right|^2\right] = R_{yy}(b,b) + \Delta^2 \sum_{i=0}^{\lfloor (b-a)/\Delta\rfloor} \sum_{k=0}^{\lfloor (b-a)/\Delta\rfloor} R_{xx}(a+i\Delta, a+k\Delta)$$

$$-\Delta \sum_{i=0}^{\lfloor (b-a)/\Delta\rfloor} R_{xy}(a+i\Delta, b) - \Delta \sum_{i=0}^{\lfloor (b-a)/\Delta\rfloor} R_{yx}(b, a+i\Delta)$$

a process $\boldsymbol{x}(t)$ is MS (Riemann-)integrable if, and only if, $\int_a^b \int_a^b R_{xx}(t_1,t_2)dt_1dt_2$ is (Riemann-)integrable.

The end of Section 9-3 The Power Spectrum

9-4 Discrete-Time Processes

- A discrete-time process can be viewed as a sampled counterpart of a continuous-time process as
$$\boldsymbol{x}[m] = \boldsymbol{x}(m) \text{ for integer } m,$$
where in text, the index of a discrete process is specially denoted by "bracket". For convenience, we simply take the sampling period to be 1.

- For this reason, most results involving continuous-time processes can be readily extended to discrete-time processes.

- As an example, the autocorrelation function of a discrete-time process $\boldsymbol{x}[m]$ can be defined as that for integers m_1 and m_2,
$$R_{xx}[m_1, m_2] = E\{\boldsymbol{x}[m_1]\boldsymbol{x}^*[m_2]\}.$$
If it is WSS, then $R_{xx}[m_1, m_2]$ reduces to:
$$R_{xx}[\tau] = E\{\boldsymbol{x}[m+\tau]\boldsymbol{x}^*[m]\} = E\{\boldsymbol{x}(m+\tau)\boldsymbol{x}^*(m)\} = R_{xx}(\tau),$$
where $R_{xx}(\tau)$ is the autocorrelation function of the parent continuous-time process $\boldsymbol{x}(t)$.

9-4 Discrete-Time Processes

- Based on the previous observation, the power spectrum of the random process $\{\boldsymbol{x}[m], m \in \mathcal{N} = \text{set of integers}\}$ is given by:

$$S_{xx}[\omega] = \sum_{n=-\infty}^{\infty} R_{xx}[n]e^{-j\omega n} \quad \left(= \sum_{n=-\infty}^{\infty} R_{xx}[n]\left(e^{j\omega}\right)^{-n} = \sum_{n=-\infty}^{\infty} R_{xx}[n]z^{-n} \right).$$

Note that $S_{xx}[\omega]$ (which is sometimes written as $S_{xx}[z]$ with $z = e^{j\omega}$) is a function of $e^{j\omega}$, and hence, is periodic with period 2π.

$S_{xx}[\omega]$ is named the *discrete(-time) Fourier transform* of $R_{xx}[\tau]$.

- We can derive $\{R_{xx}[n]\}_{n \text{ integer}}$ from the inverse discrete(-time) Fourier transform (by integrating over one period):

$$R_{xx}[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{xx}[\omega]e^{j\omega n}d\omega \quad \text{for integer } n.$$

9-4 Discrete-Time Processes

- By noting from Slide 9-130 that for integer τ,

$$
\begin{aligned}
R_{xx}[\tau] = R_{xx}(\tau) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{xx}(\omega)e^{j\omega\tau}d\omega \\
&= \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\int_{(2k-1)\pi}^{(2k+1)\pi} S_{xx}(\omega)e^{j\omega\tau}d\omega \\
&= \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\int_{-\pi}^{\pi} S_{xx}(\omega'+2k\pi)e^{j(\omega'+2k\pi)\tau}d\omega' \quad (\omega' = \omega - 2k\pi) \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\sum_{k=-\infty}^{\infty} S_{xx}(\omega'+2k\pi)\right)e^{j\omega'\tau}d\omega',
\end{aligned}
$$

and the uniqueness of the discrete(-time) Fourier transform, we obtain:

$$
S_{xx}[\omega] = \sum_{k=-\infty}^{\infty} S_{xx}(\omega + 2k\pi).
$$

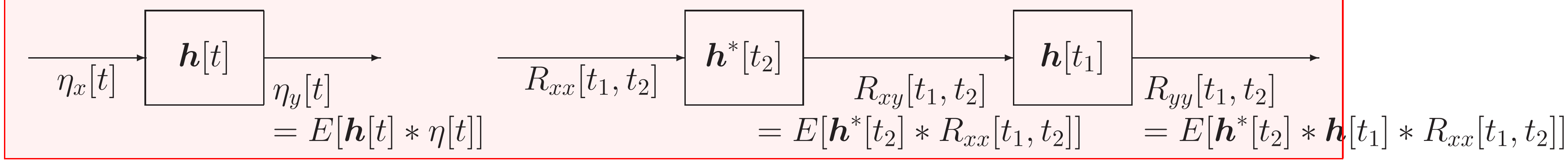
An *aliasing* in spectrums is resulted from sampling.

Convolution in Discrete-Time System

If $\boldsymbol{x}[n]$ is an input to a discrete-time system, the resulting output is the digital convolution of $\boldsymbol{x}[n]$ with $\boldsymbol{h}[n]$:

$$\boldsymbol{y}[n] = \sum_{k=-\infty}^{\infty} \boldsymbol{h}[k]\boldsymbol{x}[n-k] = \boldsymbol{h}[n] * \boldsymbol{x}[n].$$

Fundamental Theorem and Theorem 9-2 For any linear time-invariant discrete-time system (that is defined via convolution operation),



P.D. in Discrete Processes

- A function $R(\tau)$ has non-negative Fourier transform if, and only if, it is p.d., i.e.,

$$\sum_i \sum_j a_i a_j^* R(t_i - t_j) \geq 0 \quad \text{for any complex } a_i \text{ and } a_j.$$

Lemma (Schur) A (summable) discrete function $R[\tau]$, satisfying $R^*[-\tau] = R[\tau]$, has non-negative discrete(-time) Fourier transform

$$S[\omega] = \sum_{m=-\infty}^{\infty} R[m]e^{-jm\omega} = R[0] + 2\sum_{m=1}^{\infty} \text{Re}\left\{R[m]e^{-jm\omega}\right\}$$

if, and only if, the Hermitian Toeplitz matrix $\mathbb{T}_n$ is non-negative definite for every n, where

$$\mathbb{T}_n \triangleq \begin{bmatrix} R[0] & R[1] & R[2] & \cdots & R[n] \\ R^*[1] & R[0] & R[1] & \cdots & R[n-1] \\ R^*[2] & R^*[1] & R[0] & \cdots & R[n-2] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R^*[n] & R^*[n-1] & R^*[n-2] & \cdots & R[0] \end{bmatrix}.$$

Proof:

1. ***Forward:*** Suppose $S[\omega] \geq 0$. Let $\vec{a} = [a_0, a_1, \cdots, a_n]^T$.
Then,

$$\begin{aligned}
\vec{a}^\dagger \mathbb{T}_n \vec{a} &= \sum_{k=0}^{n} \sum_{m=0}^{n} a_k^* a_m R[m-k] \\
&= \sum_{k=0}^{n} \sum_{m=0}^{n} a_k^* a_m \frac{1}{2\pi} \int_{-\pi}^{\pi} S[\omega] e^{j(m-k)\omega} d\omega \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} S[\omega] \sum_{k=0}^{n} \sum_{m=0}^{n} (a_k e^{jk\omega})^* (a_m e^{jm\omega}) d\omega \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} S[\omega] \left| \sum_{m=0}^{n} a_m e^{jm\omega} \right|^2 d\omega \geq 0,
\end{aligned}$$

where "†" represents the conjugate-transpose matrix operation.

2. ***Converse:*** Suppose $\mathbb{T}_n$ is non-negative definite for every n.
Let $\vec{a} = [a_0, a_1, \cdots, a_n]^T$, where $a_m = \sqrt{1-\rho^2}\rho^m e^{jm\omega_0}$ for some $0 < \rho < 1$ and $-\pi \le \omega_0 < \pi$. Then,

$$\begin{aligned} 0 \le \vec{a}^\dagger \mathbb{T}_n \vec{a} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} (1-\rho^2) \left| \sum_{m=0}^{n} \rho^m e^{jm(\omega-\omega_0)} \right|^2 S[\omega] d\omega \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} (1-\rho^2) \left| \frac{1-\rho^{n+1}e^{j(\omega-\omega_0)(n+1)}}{1-\rho e^{j(\omega-\omega_0)}} \right|^2 S[\omega] d\omega \end{aligned}$$

Lebesque (Dominated) Convergence Theorem Suppose $|f_n(x)| \le g(x)$ on $x \in E$ for some E-integrable g, and $\lim_{n\to\infty} f_n(x) = f(x)$ for $x \in E$. Then, $\int_E \lim_{n\to\infty} f_n = \lim_{n\to\infty} \int_E f_n$.

Hence,

$$\begin{aligned} 0 &\le \lim_{n\to\infty} \frac{1}{2\pi}\int_{-\pi}^{\pi} (1-\rho^2) \left| \frac{1-\rho^{n+1}e^{j(\omega-\omega_0)(n+1)}}{1-\rho e^{j(\omega-\omega_0)}} \right|^2 S[\omega] d\omega \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \lim_{n\to\infty} (1-\rho^2) \left| \frac{1-\rho^{n+1}e^{j(\omega-\omega_0)(n+1)}}{1-\rho e^{j(\omega-\omega_0)}} \right|^2 S[\omega] d\omega. \end{aligned}$$

For the validity of the dominated convergence theorem, one only needs to examine the boundedness of the integrand in finite integral domain $[-\pi,\pi)$.

$$\left|\left|\sum_{m=0}^{n}\rho^m e^{jm(\omega-\omega_0)}\right|^2 S[\omega]\right| \le \left(\sum_{m=0}^{\infty}\left|\rho^m e^{jm(\omega-\omega_0)}\right|\right)^2\left(\sum_{\tau=-\infty}^{\infty}|R[\tau]|\right)$$
$$= \left(\frac{1}{1-\rho}\right)^2\left(\sum_{\tau=-\infty}^{\infty}|R[\tau]|\right),$$

where $|S[\omega]| = \left|\sum_{\tau=-\infty}^{\infty}R[\tau]e^{-j\omega\tau}\right| \le \sum_{\tau=-\infty}^{\infty}|R[\tau]| < \infty$ (by "summable" assumption).

$$\Rightarrow 0 \le \frac{1}{2\pi}\int_{-\pi}^{\pi}\lim_{n\to\infty}(1-\rho^2)\left|\frac{1-\rho^{n+1}e^{j(\omega-\omega_0)(n+1)}}{1-\rho e^{j(\omega-\omega_0)}}\right|^2 S[\omega]d\omega$$
$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{(1-\rho^2)}{1-2\rho\cos(\omega-\omega_0)+\rho^2}S[\omega]d\omega$$
$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{(1-\rho^2)}{1-2\rho\cos(\omega-\omega_0)+\rho^2}\varphi(e^{j\omega})d\omega,\ \text{where } \varphi(e^{j\omega}) = \sum_{\tau=-\infty}^{\infty}R[\tau](e^{j\omega})^{-\tau}$$

P.D. in Discrete Processes

Poisson's Integral Formula

$$\varphi(\rho e^{j\omega_0}) = \frac{1}{2\pi}\int_0^{2\pi} \frac{(r^2-\rho^2)}{r^2 - 2r\rho\cos(\omega-\omega_0)+\rho^2}\varphi(re^{j\omega})d\omega \quad \text{for } 0<\rho<r.$$

$$\Rightarrow \quad \varphi(\rho e^{j\omega_0}) = \sum_{\tau=-\infty}^{\infty} R[\tau](\rho e^{j\omega})^{-\tau} \geq 0 \text{ for any } 0<\rho<1$$

$$\Rightarrow \quad \text{Interior radial limit } S[\omega_0] = \lim_{\rho\uparrow 1}\varphi(\rho e^{j\omega_0}) \geq 0.$$

□

Paley-Wiener Criterion

Corollary (Paley-Wiener criterion) Following Schur's Lemma, if, in addition,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\log(S[\omega])d\omega > -\infty,$$

then $\mathbb{T}_n$ is positive definite for every n.

Proof: Suppose there exists some non-zero $\vec{a}$ such that

$$0 = \vec{a}^{\dagger}\mathbb{T}_n\vec{a} \;=\; \frac{1}{2\pi}\int_{-\pi}^{\pi} S[\omega]\left|\sum_{m=0}^{n} a_m e^{jm\omega}\right|^2 d\omega.$$

Since $S[\omega] \geq 0$, the integrand inside the above equation must be equal to zero almost everywhere. With additionally $\int_{-\pi}^{\pi}\left|\sum_{m=0}^{n} a_m e^{jm\omega}\right|^2 d\omega > 0$ (See the green box on next slide), we obtain

$$\int_{-\pi}^{\pi}\log(S[\omega])d\omega = -\infty,$$

a contradiction to Paley-Wiener criterion is obtained. □

Paley-Wiener Criterion

$$\langle e^{jn\omega}, e^{jm\omega}\rangle \triangleq \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j(n-m)\omega}d\omega = \frac{\sin((n-m)\pi)}{(n-m)\pi} = \begin{cases} 1, & n=m \\ 0, & n\neq m \end{cases}$$

Thus, $\displaystyle\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\sum_{m=0}^{n} a_m e^{jm\omega}\right|^2 d\omega = \left\langle \sum_{m=0}^{n} a_m e^{jm\omega}, \sum_{m=0}^{n} a_m e^{jm\omega}\right\rangle = \sum_{m=0}^{n}|a_m|^2 > 0.$

Let $f(\omega) \triangleq \left|\sum_{m=0}^{n} a_m e^{jm\omega}\right|^2$ and $\int_{-\pi}^{\pi} f(\omega)d\omega = \mu > 0$. Define $\mathcal{S}_\epsilon \triangleq \{\omega \in [-\pi,\pi) : f(\omega) > \epsilon\}$. Then,

$$\int_{\mathcal{S}_\epsilon} f(\omega)d\omega = \mu - \int_{[-\pi,\pi)\setminus\mathcal{S}_\epsilon} f(\omega)d\omega \geq \mu - 2\pi\epsilon = \frac{\mu}{2} \quad \text{if } \epsilon = \frac{\mu}{4\pi}.$$

Accordingly, $S[\omega] = 0$ for $\omega \in \mathcal{S}_{\mu/(4\pi)}$, which together with $|S[\omega]| < \infty$ for $\omega \in [-\pi,\pi)$ (by the "summable" assumption) implies

$$\int_{-\pi}^{\pi}\log(S[\omega])d\omega = \int_{\mathcal{S}_{\mu/(4\pi)}}\log(S[\omega])d\omega + \int_{[-\pi,\pi)\setminus\mathcal{S}_{\mu/(4\pi)}}\log(S[\omega])d\omega = -\infty,$$

where the last step holds because $\mathcal{S}_{\mu/(4\pi)}$ cannot be a set of Lebesgue measure zero.

General Properties

Corollary The process $\boldsymbol{x}[t] = \sum_{i=1}^{n} \boldsymbol{c}_i e^{j\omega_i t}$ is WSS if, and only if, $\{\boldsymbol{c}_i\}$ are uncorrelated with zero mean (provided that $\{\omega_i\}$ are all distinct and does not include zero.)

Proof: It is straightforward that $\boldsymbol{x}[t]$ is WSS when $\{\boldsymbol{c}_i\}$ are uncorrelated with zero mean.

Conversely, if $\boldsymbol{x}[t] = \sum_{i=1}^{n} \boldsymbol{c}_i e^{j\omega_i t}$ is WSS, then

$$E\{\boldsymbol{x}[t]\} = E\left\{\sum_{i=1}^{n} \boldsymbol{c}_i e^{j\omega_i t}\right\} = \sum_{i=1}^{n} E\left\{\boldsymbol{c}_i\right\} e^{j\omega_i t}$$

implies that $\{\boldsymbol{c}_i\}$ must be zero-mean (See the first green box in Slide 9-142), and

$$E\{\boldsymbol{x}[t+\tau]\boldsymbol{x}^*[t]\} = \sum_{i=1}^{n}\sum_{k=1}^{n} E\left\{\boldsymbol{c}_i\boldsymbol{c}_k^*\right\} e^{j\omega_i\tau} e^{jt(\omega_i-\omega_k)}$$

implies that $E\{\boldsymbol{c}_i\boldsymbol{c}_k^*\}_{i=1}^{n}$ must be zero for every $i \neq k$ (See the second green box in Slide 9-142). $\square$

$$\begin{bmatrix} e^{j\omega_1}-1 & e^{j\omega_2}-1 & \cdots & e^{j\omega_n}-1 \\ e^{j2\omega_1}-1 & e^{j2\omega_2}-1 & \cdots & e^{j2\omega_n}-1 \\ \vdots & \vdots & \ddots & \vdots \\ e^{jn\omega_1}-1 & e^{jn\omega_2}-1 & \cdots & e^{jn\omega_n}-1 \end{bmatrix} \begin{bmatrix} E\{\boldsymbol{c}_1\} \\ E\{\boldsymbol{c}_2\} \\ \vdots \\ E\{\boldsymbol{c}_n\} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

and with $\omega_0 = 0$,

$$\left|\text{Det}\left(\begin{bmatrix} e^{j\omega_1}-1 & e^{j\omega_2}-1 & \cdots & e^{j\omega_n}-1 \\ e^{j2\omega_1}-1 & e^{j2\omega_2}-1 & \cdots & e^{j2\omega_n}-1 \\ \vdots & \vdots & \ddots & \vdots \\ e^{jn\omega_1}-1 & e^{jn\omega_2}-1 & \cdots & e^{jn\omega_n}-1 \end{bmatrix}\right)\right| = \left|\prod_{i=0}^{n}\prod_{k=i+1}^{n}\left(e^{j\omega_i}-e^{j\omega_k}\right)\right| \neq 0$$

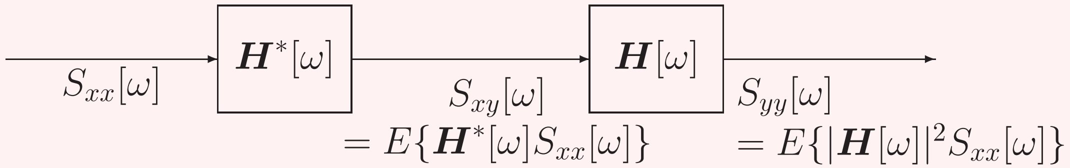
Note that if $\omega_1 = 0$, then $E\{\boldsymbol{x}[t]\}$ may be non-zero mean.

We can define $\omega_{ik} = \omega_i - \omega_k$ and use a similar proof to the above derivation to prove that $E\{\boldsymbol{c}_i\boldsymbol{c}_k^*\}$ must be zero for every $i \neq k$ if $\{\omega_{ik}\}$ are all distinct.

If, however, $\omega_{ik} = \omega_{i'k'}$ for some i, i', k and k', then we have $E\left\{\boldsymbol{c}_i\boldsymbol{c}_k^*\right\}e^{j\omega_i\tau} + E\left\{\boldsymbol{c}_{i'}\boldsymbol{c}_{k'}^*\right\}e^{j\omega_{i'}\tau} = 0$ for every integer τ, which still implies $E\left\{\boldsymbol{c}_i\boldsymbol{c}_k^*\right\} = E\left\{\boldsymbol{c}_{i'}\boldsymbol{c}_{k'}^*\right\} = 0$.

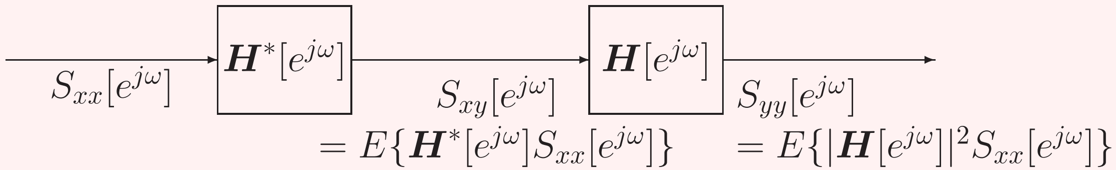
Theorem 9-4 Revisited

Theorem 9.4 (Discrete) For any linear time-invariant system with WSS input,



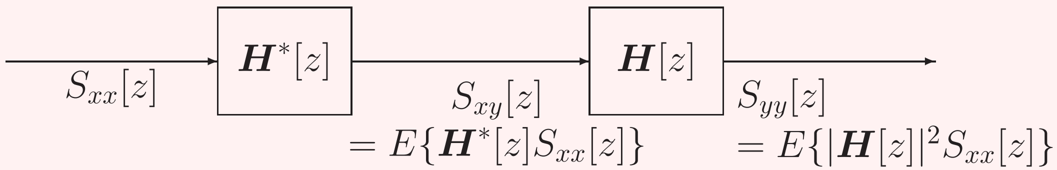
$\Downarrow$

Theorem 9.4 (Discrete) For any linear time-invariant system with WSS input,



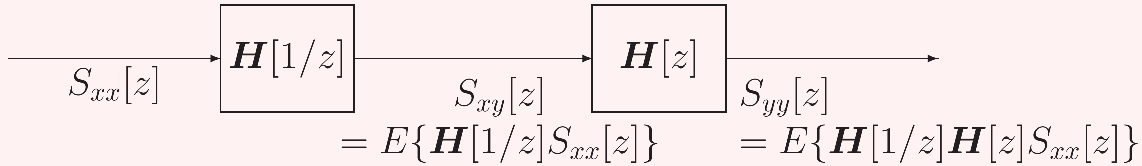
$\Downarrow$

Theorem 9.4 (Discrete with $z = e^{j\omega}$) For any linear time-invariant system with WSS input,



Theorem 9-4 Revisited

Theorem 9.4 (Discrete) For any **real** linear time-invariant system with WSS input,



Example 9-34 Define the real system through the recursion equation

$$\boldsymbol{y}[n] - \boldsymbol{a} \cdot \boldsymbol{y}[n-1] = \boldsymbol{x}[n],$$

where $\boldsymbol{a}$ is a real random variable with density $f(a)$. Then,

$$\boldsymbol{Y}[z] - \boldsymbol{a}z^{-1}\boldsymbol{Y}[z] = (1 - \boldsymbol{a}z^{-1})\boldsymbol{Y}[z] = \boldsymbol{X}[z],$$

and hence,

$$\boldsymbol{H}[z] = \frac{\boldsymbol{Y}[z]}{\boldsymbol{X}[z]} = \frac{1}{1 - \boldsymbol{a}z^{-1}}.$$

Consequently,

$$S_{yy}[z] = S_{xx}[z]E\{\boldsymbol{H}[z]\boldsymbol{H}[z^{-1}]\} = S_{xx}[z]E\left\{\frac{1}{(1 - \boldsymbol{a}z^{-1})(1 - \boldsymbol{a}z)}\right\} = S_{xx}[z]\int_{-\infty}^{\infty} \frac{f(a)}{(1 - az^{-1})(1 - az)}da.$$

The end of Section 9-4 Discrete-Time Processes