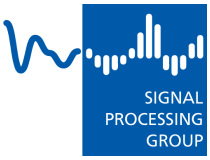
Fundamentals of Signal Processing

Prof. Dr.-Ing. Abdelhak M. Zoubir

Manuscript

Revised and updated in Spring 2023

TECHNISCHE UNIVERSITÄT DARMSTADT



Imprint

Fundamentals of Signal Processing
Manuscript
Revised and updated in Spring 2023

Author:
Prof. Dr.-Ing. Abdelhak M. Zoubir
Technische Universität Darmstadt
Institute of Telecommunications
Signal Processing Group
Merckstr. 25
64283 Darmstadt
Germany

Contact:
E-Mail: office@spg.tu-darmstadt.de
Phone: +49 6151-16 21341
Fax: +49 6151-16 21342
Web: www.spg.tu-darmstadt.de

Preface

The objective of this manuscript is to introduce basic concepts of statistical methods for signal processing. In particular, the course focuses on linear, time-invariant systems and the interaction of random (stochastic) signals in such systems. A system is defined as a combination and interaction of several components to perform a desired task. This task might be the transmission of information from a transmitter to a receiver. Our focus will be on linear, time-invariant systems. A system is said to be linear if it fulfills the homogenity and additivity properties. A time-invariant system has the property that a shifted input in time leads to the corresponding output signal, shifted by the same time shift. Although real systems may deviate from the above properties, linear, time-invariant systems are a good approximation for real-life systems which may be either non-linear and/or time-variant. Also, the theory of linear, time-invariant systems is well understood and established.
Random signals are encountered in many physical systems. Noise is the most obvious example. It could be sensor noise, which is mainly white noise, or interference, which is more structured - such as clutter in radar. It is of great importance to understand the interaction of random signals and linear, time-invariant systems. Of particular interest is optimal system design such as the Wiener filter, which is covered in this manuscript. The goal is to design a linear, time-invariant system that is capable of filtering noise, for example. The resulting flter is obtained by minimizing the mean-squared error bewteen the desired signal and its estimated counterpart.
First, we will introduce the concepts of probability theory. It is clear that this topic deserves a treatment on its own, but we will concisely cover the tools needed later in the manuscript. In chapter 2, we will review Shannon's sampling theorem and discuss the fundamentals of signal reconstruction. This review is necessary as the remainder of the manuscript is a treatment of discrete-time signals and systems. Chapter 3 introduces the topic of stochastic processes and Chapter 4 deals with the power spectral density of such processes. Chapter 5 treats the interaction of random signals with linear, time-invariant systems. In this chapter, we will introduce finite impulse response (FIR) and infinite impulse response (IIR) filters. A special class of filtered random processes is the Autoregressive Moving Average (ARMA) processes which can reduce the Autoregressive (AR) or Moving Average (MA) process. This topic is treated in Chapter 6. Chapter 7 concerns basics of detection and estimation based on the Matched Filter, the method of least squares and the Wiener Filter.
A great attempt has been made to introduce difficult statistical signal processing concepts in a clear and lucid way, without compromising rigor. We hope that the manuscript will provide an aid to grasp the concepts discussed in the classroom. There exist many texts on the subject and the reader of this manuscript should consult these should they need further details on a particular topic. We recommend, in particular, the book by Papoulis ("Probability, Random Variables and Stochastic Processes") which has been a wide-spread and outstanding companion for students and researchers working on this topic for many decades.
We hope that this manuscript will assist you in understanding the first fundamentals of statistical signal processing, and that this will spark your interest to further studies so as to solve challenging and ever timely problems.
Useful references, which have been used for the preparation of this text include:

1. Böhme, J.F. *S*tochastische Signale, Teubner Studienbücher, 1998.
2. Papoulis, A. *P*robability, Random Variables and Stochastic Processes, McGraw-Hill, Inc., 1991
3. Hänsler, E. *S*tatistische Signale, Springer Verlag, 3. Auflage, 2001.
4. Leon-Garcia, A. *P*robability and Random Processes for Electrical Engineering, Addison-Wesley, Inc., 1994
5. P. Z. Peebles, J. *P*robability, Random Variables and Random Signal Principles, McGraw-Hill, Inc., 2001
6. Rao, C. R. *L*inear Statistical Inference and Its Applications John Wiley & Sons, Inc., 1973
7. Yates, R. & Goodman, D. *P*robability and Stochastic Processes John Wiley & Sons, Inc., 1999

Contents

1 Probability Theory and Random Variables

1.1 Introduction to Probability Theory

Probability theory is fundamental to understanding stochastic processes, which form a large part of this text. It allows to study and describe non-deterministic events or measured quantities, which may, at first glance appear to be random. It also serves to, e.g., discover patterns within measured data, predict future events or construct probability models. For example, an archeologist can gain knowledge about what most likely happened thousands of years ago. An engineer analyzes measurements to ensure with a given probability that a new design is highly reliable. These examples illustrate that probability theory is central to any kind of scientific research. In the sequel some important concepts are defined and illustrated with simple examples to enhance the understanding.

1.1.1 Set Theory

A *set* is an arbitrary collection of objects. These objects, termed *elements* of the set, may represent anything depending on the application to which set theory is applied. In image processing, a set can refer, for example, to all black pixels of a binary image. In general, we will use a capital letter to denote a set, and the Greek letter ζ ("zeta") to denote an element of a set. Braces $\{\cdot\}$ are used to define the contents of a set. A set can be specified by listing all its elements, for example $A = \{a, b, c\}$, or by specifying their properties, for example $B =$ { "all even numbers" }. If every element of a set A is also an element of another set B, then A is said to be a *subset* of B, denoted by $A \subset B$. If a set consists of n elements, then the total number of its subsets equals 2^n. We define a *universal set* or *space* by $\mathcal{S}$ ("calligraphic S") to be the set containing all objects under consideration. The *empty* or *null* set is denoted by $\emptyset$ and contains no elements.

Example 1.1.1 *We shall denote by ζ_i the faces of a die. These faces are the elements of the set $\mathcal{S} = \{\zeta_1, \ldots, \zeta_6\}$. In this case, $\mathcal{S}$ has $2^6 = 64$ subsets, which are*

$$\{\zeta_1\}, \ldots, \{\zeta_1, \zeta_2\}, \ldots, \{\zeta_1, \zeta_2, \zeta_3\}, \ldots, \mathcal{S}.$$

Example 1.1.2 *Suppose that a coin is tossed twice. The resulting outcomes are the four objects { hh, ht, th, tt } forming the set:*

$$\mathcal{S} = \{\text{hh, ht, th, tt}\}$$

where hh *is an abbreviation for the element* "heads-heads". *The set $\mathcal{S}$ has $2^4 = 16$ subsets. For example,*

$$\begin{array}{rcccc} A & = & \{ \text{"head at the first tossing"} \} & = & \{\text{hh,ht}\} \\ B & = & \{ \text{"only one head shown"} \} & = & \{\text{ht, th}\} \\ C & = & \{ \text{"head shows at least once"} \} & = & \{\text{hh, ht, th}\} \end{array}$$

In the first equality, the sets A, B and C are represented by their properties; in the second, in terms of their elements.

Example 1.1.3 *Let $\mathcal{S}$ be the set of all points in the square of Figure 1.1. Its elements are all ordered pairs of numbers (x, y) where*

$$0 \le x \le T,\ 0 \le y \le T$$

The shaded area is a subset A of $\mathcal{S}$ consisting of all points (x, y) such that $-b \le x - y \le a$. The notation

$$A = \{-b \le x - y \le a\}$$

describes A in terms of the properties of x and y.

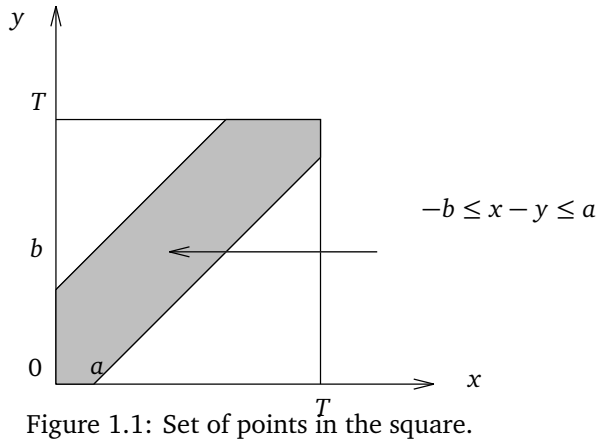


Figure 1.1: Set of points in the square.

$\bar{A} = \mathcal{S} - A$	Complement of A in $\mathcal{S}$
$A \cup B$	Union of A and B
$A \cap B$	Intersection of A and B
$A - B = A \cap \bar{B}$	Complement of B in A
$A = B$	A equals B
$A \subset B, A = A \cap B$	A is a subset of B
$A \cap B = \emptyset$	A and B are mutually exclusive
$\mathcal{S}$	Universal set or space
$\emptyset$	Empty set, null set or zero set
ζ	Element of a set

Table 1.1: Summary of set notation.

The union of two sets A and B is defined as the set $C = A \cup B$, which contains all the elements in A or B or both. The intersection of the two sets is defined as the set $C = A \cap B$ and contains all elements common to both A and B. The complement (in $\mathcal{S}$) of a set A is defined as the set $\bar{A}$ consisting of all elements in $\mathcal{S}$ that are not in A. For a set $A \subset B$ the difference $B - A$ is defined as the elements of B that are not in A. A summary of set notation is provided in Table 1.1. Set operations and set relations can be illustrated by the use of *Venn diagrams*. Figure 1.2 shows some basic set operations, using Venn diagrams.

Example 1.1.4 *The universal set $\mathcal{S}$ is the binary image of flowers shown in Figure 1.3. $A =$ {"all black pixels"}, $B =$ {"all white pixels"} are subsets of $\mathcal{S}$. Clearly, A and B are mutually exlusive and $B = \bar{A}$ is the complement of A in $\mathcal{S}$.*

1.1.2 Probability Space

Fundamental to the study of probability theory is the concept of an *experiment*. Each performance of this experiment is termed a *trial* and results in some measurable *outcome*. The sample space of the experiment is the set $\mathcal{S}$ of all possible outcomes ζ_i. An *event* A is defined as some subset of $\mathcal{S}$ and it is said to occur if the result of a trial is an element of A. The sample space $\mathcal{S}$ is also termed the certain event as it must always occur. The null set $\emptyset$ is termed the impossible event and can never occur.

The sample space $\mathcal{S}$ is called a *discrete sample space* if the outcomes ζ_i are countable, i.e. can be put into a one-to-one correspondence with integers. $\mathcal{S}$ is termed a *continuous sample space* if the outcomes are not countable. Examples of random experiments are given in Example 1.1.5. The corresponding discrete or continuous sample spaces are given in Example 1.1.6.

Example 1.1.5 *Leon-Garcia [1994]*
Experiment E_1: *Select a ball from an urn containing balls numbered 1 to 50. Note the number of the ball.*
Experiment E_2: *Select a ball from an urn containing balls numbered 1 to 4. Suppose that balls 1 and 2 are black and balls 3 and 4 are white. Note the number and color of the ball.*
Experiment E_3: *Count the number of voice packets containing only silence produced from a group of N speakers in a* 10*-ms period*
Experiment E_4: *A block of information is transmitted repeatedly over a noisy channel until an error-free block arrives*

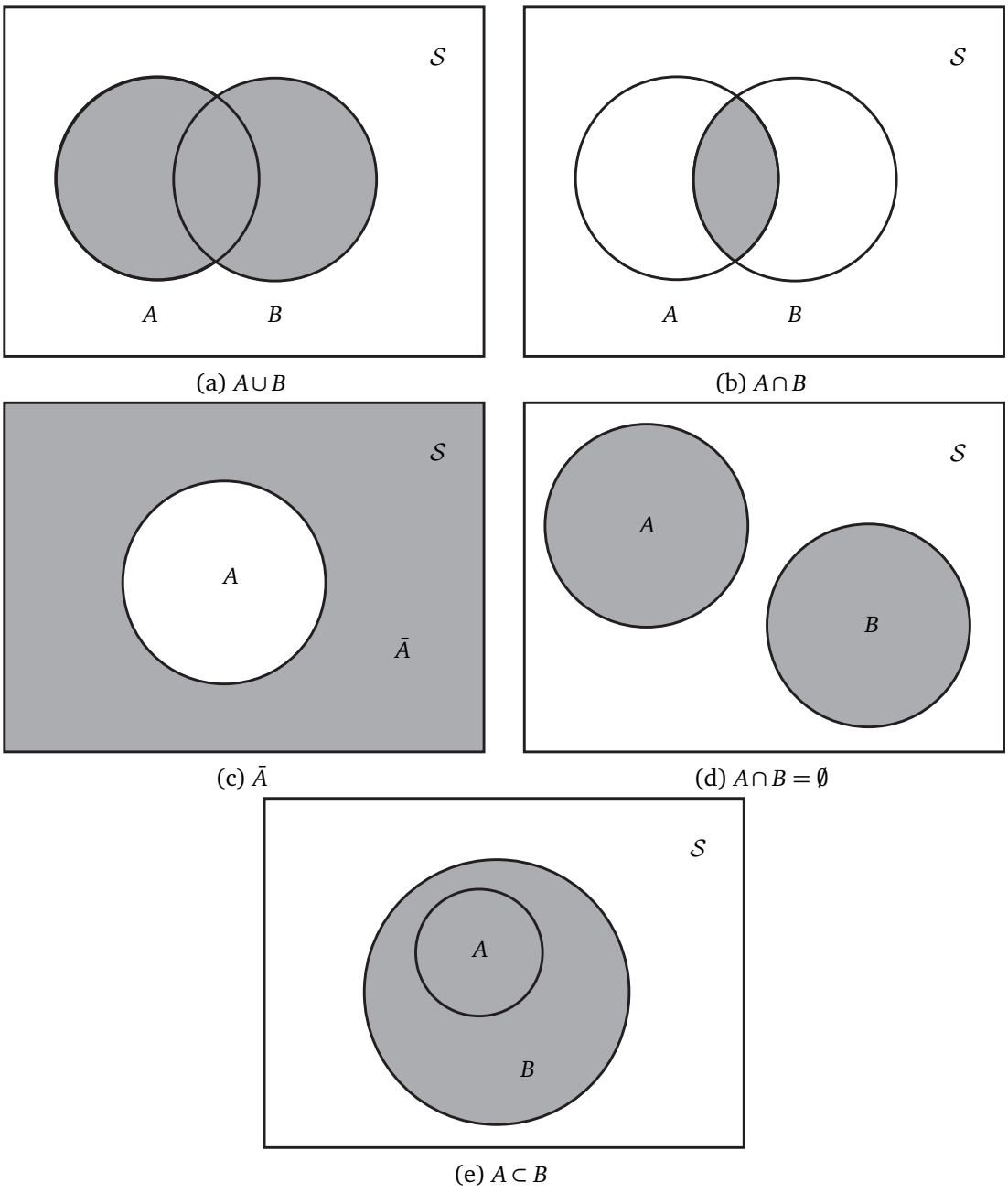


Figure 1.2: Venn diagrams, illustrating basic set operations and set relations.

Figure 1.3: Binary Image.

at the receiver. Count the number of transmissions required.
Experiment E_5: *Measure the time between two message arrivals at a message center.*
Experiment E_6: *Measure the lifetime of a given computer memory chip in a specified environment.*
Experiment E_7: *Determine the value of a voltage waveform at a time* t_1.
Experiment E_8: *A system component is installed at time* $t = 0$. *For* $T \geq 0$ *let* $X(t) = 1$ *as long as the component is functioning, and let* $X(t) = 0$ *after the component fails.*

Example 1.1.6 *Leon-Garcia [1994]*
The sample spaces corresponding to the experiments defined in Example 1.1.5 are given below using set notation:

$$\begin{aligned}
\mathcal{S}_1 &= \{1, 2, \ldots, 50\} \\
\mathcal{S}_2 &= \{(1, b), (2, b), (3, w), (4, w)\} \\
\mathcal{S}_3 &= \{0, 1, \ldots, N\} \\
\mathcal{S}_4 &= \{1, 2, 3, \ldots\} \\
\mathcal{S}_5 &= \{t | t \geq 0\} = [0, \infty) \\
\mathcal{S}_6 &= \{t | t \geq 0\} = [0, \infty) \\
\mathcal{S}_7 &= \{v | -\infty < v < \infty\} = (-\infty, \infty) \\
\mathcal{S}_8 &= \textit{set of functions } X(t) \textit{ for which } X(t) = 1 \textit{ for } 0 \leq t < t_0 \textit{ and } X(t) = 0 \textit{ for } t \geq t_0, \textit{ where} \\
&\quad t_0 > 0 \textit{ is the time when the component fails.}.
\end{aligned}$$

1.1.3 Probability

We need to be able to assess the outcome of an experiment in a quantative manner. This is done using *probability*. Historically, a number of approaches have been taken to define the concept of a probability. The most notable of these approaches are discussed below.

Classical Definition

For a given random experiment, the probability $\mathsf{P}(A)$ of an event A is defined as the ratio

$$\mathsf{P}(A) = \frac{N_A}{N}$$

where N is the number of possible outcomes and N_A is the number of outcomes that are favorable to the event A. Using this definition, probabilities are assigned to events *a priori*, without actual experimentation. It is also assumed that all events are equally likely to occur.

Example 1.1.7 *In the die experiment, there are 6 possible outcomes and the outcomes that are favorable to the event* {"even"} *are 3, hence* P*{"even"}* $= 3/6 = 1/2$*, provided the die is fair.*

A problem with this definition, highlighted in the following example, is the significance of the numbers N and N_A.

Example 1.1.8 *Papoulis [1991]*
We roll two dice and we want to find the probability p *that the sum of the numbers that show equals 7.*

(a) We could consider as possible outcomes the 11 sums 2,3,...,12. Of these only one is favorable, hence $p = 1/11$. *This is wrong!*

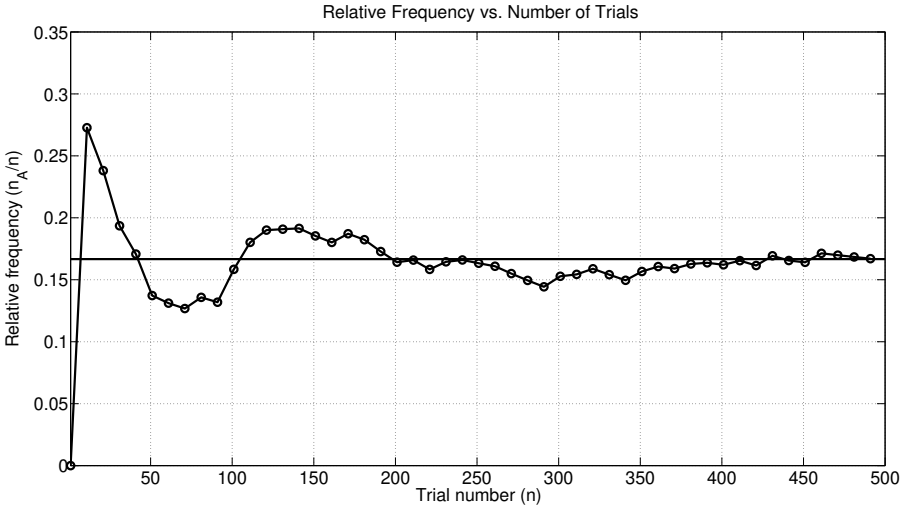


Figure 1.4: Relative frequency of the experiment of Example 1.1.8 and 1.1.9 as the number of trials increases.

(b) We could count as possible outcomes all pairs of numbers not distinguishing between the first and the second die. We have now 21 outcomes of which the pairs (3,4),(5,2), and (6,1) are favorable. Thus, N_A=3 and N=21 and hence $p = 3/21$. This is also wrong.

(c) To get the correct result, we need to count all pairs of numbers distinguishing between the first and the second die. The total of outcomes is now 36 and the favorable outcomes are the six pairs (3,4), (4,3), (5,2), (2,5), (6,1) and (1,6), hence $p = 6/36 = 1/6$.

Relative Frequency Definition

The relative frequency definition is based on *a posteriori* assignment of probability to an event from experimental outcomes. If a random experiment is conducted n times and the event A occurs n_A times, the probability $\mathsf{P}(A)$ is the limit

$$\mathsf{P}(A) = \lim_{n\to\infty} \frac{n_A}{n}$$

This can be a problem as n_A and n might be large but are finite and their ratio cannot, therefore, be equated, even approximately, to a limit.

Example 1.1.9 *Consider again the experiment of Example 1.1.8. The sample space is $\mathcal{S} = \{\zeta_1, \ldots, \zeta_{11}\}$ where the elementary events ζ_i correspond to the 11 possible sums resulting from rolling two dice and adding the result. The event of interest is defined as $A = \{\zeta_6\}$ where ζ_6 is the outcome* "sum equals 7". *Figure 1.4 shows the relative frequency of the event A as the number of trials n increases. It is observed that the frequency approaches $1/6 \approx 0.167$.*

Axiomatic Definition

The axiomatic approach to probability theory provides a logical structure for deductively determining the probability of one event, given the probability of another. This approach is based solely on three axioms.

Let A and B be events, which lie on a sample space $\mathcal{S}$. We assign a number $\mathsf{P}(A)$ which we call the probability of the event A. This number is so chosen as to satisfy the following

1. $\mathsf{P}(A) \geq 0$.
2. $\mathsf{P}(\mathcal{S}) = 1$.
3. $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B)$, if A and B are mutually exclusive.

The third axiom applies to N events $A_n, n = 1, 2, \ldots, N$, where N may possibly be infinite, defined on a sample space $\mathcal{S}$. If $A_m \cap A_n = \emptyset$ for all $m \neq n$, then

$$\mathsf{P}\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{N} \mathsf{P}(A_n)$$

Example 1.1.10 *Peebles, Jr. [2001]*

(a) We consider the experiment of obtaining a number x by spinning the pointer on a "fair" wheel of chance that is labeled from 0 to 100 points. The sample space $\mathcal{S} = \{0 < x \leq 100\}$. Since the wheel is fair, the probability of the pointer falling between any two numbers $x_2 \geq x_1$ should be $(x_2 - x_1)/100$. Axiom 1 is satisfied for the event $A = \{x_1 \leq x \leq x_2\}$ for all x_1 and x_2. When $x_2 = 100$ and $x_1 = 0$, then axiom 2 is also satisfied.

If we break the wheel's circumference into N contiguous segments

$$A_n = \{x_{n-1} < x \leq x_n\}; \quad x_n = 100 \cdot \frac{n}{N}; \quad n = 1, 2, \ldots, N; \quad x_0 = 0.$$

Then $\mathsf{P}(A_n) = 1/N$ and for any N,

$$\begin{aligned} \mathsf{P}\left(\bigcup_{n=1}^{N} A_n\right) &= \sum_{n=1}^{N} \mathsf{P}(A_n) \\ &= \sum_{n=1}^{N} \frac{1}{N} = 1 = \mathsf{P}(\mathcal{S}) \end{aligned}$$

from axiom 3.

(b) Now, consider the case when the interval $(x_n - x_{n-1}) \to 0$. The probability $\mathsf{P}(A_n) \to \mathsf{P}(x_n)$; that is $\mathsf{P}(A_n)$ becomes the probability of the pointer falling exactly on the point x_n. Since $N \to \infty$ this means $\mathsf{P}(A_n) \to 0$.

Example 1.1.3b highlights an important concept. An event with probability 0 may occur, and is distinct from the null event $\emptyset$ which has probability 0 but can never occur. Similarly an event with probability 1 might not occur in a trial and is distinct from the certain event $\mathcal{S}$, which must always occur. Such situations typically arise when defining discrete events on continuous sample spaces.

1.1.4 Conditional Probability

In some experiments, it may be that some events are not mutually exclusive because of common elements in the sample space. These elements correspond to the joint occurrence of the non-exclusive events. The conditional probability of an event A assuming B, denoted by $\mathsf{P}(A|B)$, is by definition the ratio

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

where we assume that $\mathsf{P}(B) > 0$. The probability $\mathsf{P}(A \cap B)$ is the joint probability for two events A and B which intersect in the sample space. One can show that

$$\mathsf{P}(A \cap B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cup B)$$

as it can be readily shown from the Figure 1.5. Equivalently

$$\begin{aligned} \mathsf{P}(A \cup B) &= \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B) \\ &\leq \mathsf{P}(A) + \mathsf{P}(B) \end{aligned}$$

If $A \cap B = \emptyset$, then $\mathsf{P}(A \cap B) = \mathsf{P}(\emptyset) = 0$. From the definition of $\mathsf{P}(A|B)$ the following properties follow:
If $B \subset A$ then $\mathsf{P}(A|B) = 1$ because $A \cap B = B$. Similarly if $A \subset B$, then

$$\mathsf{P}(A|B) = \frac{\mathsf{P}(A)}{\mathsf{P}(B)} \geq \mathsf{P}(A)$$

Conditional probabilities are indeed probabilities, i.e. they satisfy the axioms:

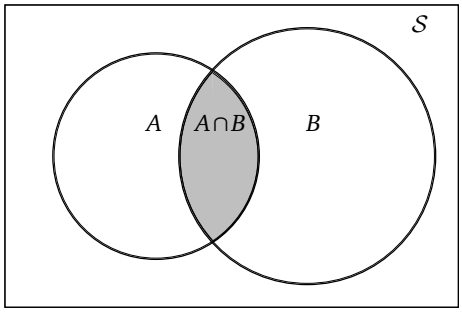


Figure 1.5: Joint probability.

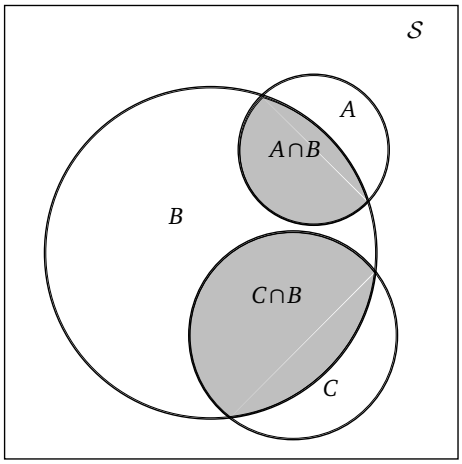


Figure 1.6: Mutually exclusive events.

(i) $\mathsf{P}(A|B) \geq 0$ because $\mathsf{P}(A \cap B) \geq 0$ and $\mathsf{P}(B) > 0$

(ii) $\mathsf{P}(\mathcal{S}|B) = 1$ because $B \subset \mathcal{S}$ and $\mathsf{P}(\mathcal{S} \cap B) = \mathsf{P}(B)$

(iii) If A and C are mutually exclusive, then $A \cap B$, and $C \cap B$ are also mutually exclusive (see Figure 1.6).

Hence,

$$\begin{aligned} \mathsf{P}(A \cup C|B) &= \frac{\mathsf{P}((A \cup C) \cap B)}{\mathsf{P}(B)} \\ &= \frac{\mathsf{P}(A \cap B) + \mathsf{P}(C \cap B)}{\mathsf{P}(B)} \\ &= \frac{\mathsf{P}(A \cap B)}{P(B)} + \frac{\mathsf{P}(C \cap B)}{P(B)} \\ &= \mathsf{P}(A|B) + \mathsf{P}(C|B) \end{aligned}$$

From the above, it follows that all 3 axioms hold also for conditional probabilities.

Example 1.1.11 *In the fair die experiment, we shall determine the conditional probability of the event* $\{\zeta_2\}$ *("die a 2") assuming that the event* {"even"} *occurred. With* $A = \{\zeta_2\}$, $B = \{\text{"even"}\} = \{\zeta_2, \zeta_4, \zeta_6\}$. *We have* $P(A) = 1/6$ *and* $P(B) = 3/6 = 1/2$ *and with,* $A \cap B = A$, *we have*

$$P(\{\zeta_2 \mid \{\text{ "even" }\}\}) = \frac{P(\{\zeta_2\})}{P(\{\text{ "even" }\})} = \frac{1}{3} \geq \mathsf{P}(\{\zeta_2\})$$

This probability equals the relative frequency of the occurrence of the event $\{\zeta_2\}$ *in the sequence of the experiment whose outcomes are even numbers.*

Example 1.1.12 *Papoulis [1991]*
A box contains three white balls w_1, w_2, w_3 *and two red balls* r_1, r_2. *We remove at random two balls in succession. What is the probability that the first removed ball is white and the second red?*

Since the box contains 3 white and 2 red balls, the probability of the event $W_1 =$ {"white first"} equals 3/5. If a white ball is removed, there remains two white and two red balls, hence, the conditional probability $P(\{R_2|W_1\})$ of the event $R_2 =$ {"red second"} assuming {"white first"} equals 2/4. From this and the formula given above

$$P(W_1 \cap R_2) = P(R_2|W_1) \cdot P(W_1) = \frac{2}{4} \times \frac{3}{5} = \frac{6}{20}$$

where $W_1 \cap R_2$ is the event {"white first", "red second"}.

Alternatively, we can use the frequency of occurrence to find this result, i.e., we count the ordered pairs (w_1w_2), (w_1w_3), (w_1r_1),$(w_2r_2)\ldots$, (r_2w_1),(r_2w_2),(r_2w_3),(r_2r_1) and find 20. The event {"white first", "red second"} *consists of six outcomes and hence the probability of the event white first and red second is $\frac{6}{20}$.*

1.1.5 Bayes Theorem

Given a partition $\mathcal{U} = [A_1, \ldots, A_n]$ of $\mathcal{S}(A_i \cap A_j = \emptyset, i \neq j)$ and an arbitrary event B, then

$$\mathsf{P}(B) = \mathsf{P}(B|A_1)\mathsf{P}(A_1) + \ldots + \mathsf{P}(B|A_n)\mathsf{P}(A_n) \tag{1.1}$$

Proof 1.1.1 *The events $B \cap A_i$ and $B \cap A_j$ are mutually exclusive because the events A_i and A_j are mutually exclusive. Hence*

$$P(B) = P(B \cap A_1) + \ldots + P(B \cap A_n),$$

which is known as the total probability theorem. Because $\mathsf{P}(B \cap A_i) = P(B|A_i)P(A_i)$ and $P(B \cap A_i) = P(A_i \cap B)$, we conclude that

$$P(A_i|B) = P(B|A_i) \cdot \frac{P(A_i)}{P(B)} \tag{1.2}$$

Using (1.1) and (1.2), we obtain Bayes' theorem that states

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + \ldots + P(B|A_n)\mathsf{P}(A_n)}.$$

Example 1.1.13 *Papoulis [1991]*
We have four boxes. Box 1 contains 2000 components of which 5 % are defective. Box 2 contains 500 components of which 40 % are defective. Boxes 3 and 4 contain 1000 each with 10 % defective (see Figure 1.7 and Figure 1.8).

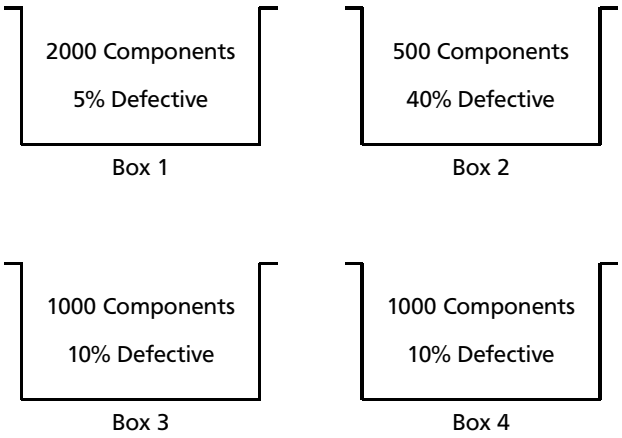


Figure 1.7: Defective components.

We select at random one of the four boxes and we remove at random a single component. We wish to answer the question "what is the probability that the selected component is defective?"

We denote by B_i the event consisting of all components in the ith box and by D the event consisting of all defective components. Clearly

$$P(B_1) = P(B_2) = P(B_3) = P(B_4) = \frac{1}{4}$$

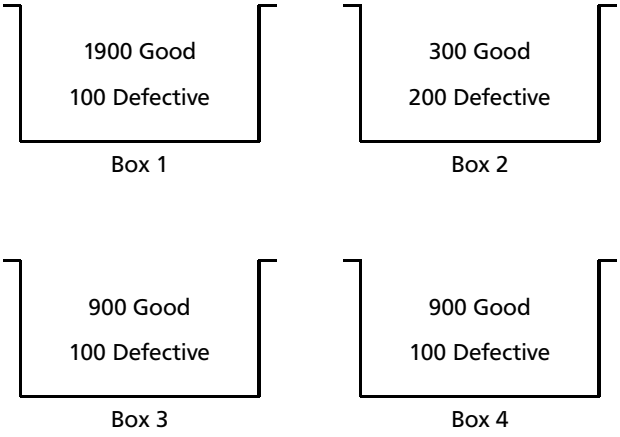


Figure 1.8: Mutually exclusive events.

because the boxes are selected at random. The probability that a component taken from a specific box is defective equals the ratio of the defective to the total number of components in that box. This means that

$$P(D|B_1) = \frac{100}{2000} = 0.05, \quad P(D|B_2) = \frac{200}{500} = 0.4$$

$$P(D|B_3) = \frac{100}{1000} = 0.1, \quad P(D|B_4) = \frac{100}{1000} = 0.1$$

B_1, B_2, B_3 and B_4 form a partition of $\mathcal{S}$. Therefore,

$$P(D) = \sum_{i=0}^{4} P(D|B_i) \cdot P(B_i) = 0.05 \times \frac{1}{4} + 0.4 \times \frac{1}{4} + 0.1 \times \frac{1}{4} + 0.1 \times \frac{1}{4} = 0.1625$$

which is the probability that the selected component is defective.

1.1.6 Independent Events

A and B are called two independent events if

$$\mathsf{P}(A \cap B) = \mathsf{P}(A) \cdot \mathsf{P}(B)$$

Independence means that the event A occurs has no influence on the probability that B occurs.
Then, we have

$$\mathsf{P}(A|B) = \mathsf{P}(A)$$

and

$$\mathsf{P}(B|A) = \mathsf{P}(B).$$

The events A_1, A_2 and A_3 are independent if they are independent in pairs,

$$\mathsf{P}(A_i \cap A_j) = \mathsf{P}(A_i) \cdot \mathsf{P}(A_j), \ i \neq j$$

and

$$\mathsf{P}(A_1 \cap A_2 \cap A_3) = \mathsf{P}(A_1)\mathsf{P}(A_2)\mathsf{P}(A_3)$$

Note that three events might be independent in pairs but not independent.

Example 1.1.14

$$P(A) = P(B) = P(C) = \frac{1}{5}$$

and $p = P(A \cap B) = P(A \cap C) = P(B \cap C) = P(A \cap B \cap C)$

(a) If $p = 1/25$, then these events are independent in pairs but they are not independent because

$$P(A \cap B \cap C) \neq P(A) \cdot P(B) \cdot P(C)$$

(b) If $p = 1/125$, $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$ but the events are not independent because

$$P(A \cap B) \neq P(A) \cdot P(B)$$

1.1.7 Combined Experiments

Example 1.1.15 *We are given two experiments. The first experiment is the rolling of a fair die*

$$\mathcal{S}_1 = \{\zeta_1, \ldots, \zeta_6\},\ P_1(\zeta_i) = \frac{1}{6}.$$

The second experiment is the tossing of a fair coin

$$\mathcal{S}_2 = \{\text{h,t}\} \qquad P_2(\text{h}) = P_2(\text{t}) = \frac{1}{2}$$

where h *and* t *stand for heads and tails respectively. We wish to answer the question: What is the probability that we get* "two" *on the die and* "heads" *on the coin?*

The Cartesian product of two experiments on $\mathcal{S}_1$ and $\mathcal{S}_2$ is a new experiment on $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2$ whose events are all Cartesian products of the form [1] *$A \times B$, where A is an event of $\mathcal{S}_1$ and B is an event of $\mathcal{S}_2$, and their unions and intersections.*

Example 1.1.16 *We flip a coin twice, each flip being one experiment, then*

$$\begin{aligned} \mathcal{S}_1 &= \{\text{h,t}\} \\ \mathcal{S}_2 &= \{\text{h,t}\} \end{aligned}$$

$$\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2 = \{(\text{h,h}),(\text{h,t}),(\text{t,h}),(\text{t,t})\}$$

Example 1.1.17 *Papoulis [1991]*
Suppose

$$\begin{aligned} A &= \{x_1 \le x \le x_2\} \\ B &= \{y_1 \le y \le y_2\} \end{aligned}$$

the $A \times B$ is the rectangle shown in Figure 1.9. In the experiment $\mathcal{S}$, the probabilities of the event $A \times \mathcal{S}_2$ and $\mathcal{S}_1 \times B$ are

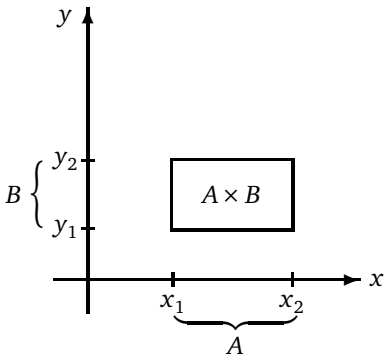


Figure 1.9: Cartesian product of 2 events.

such that

$$P(A \times \mathcal{S}_2) = P_1(A) \quad \textit{and} \quad P(\mathcal{S}_1 \times B) = P_2(B)$$

where $P_1(A)$ is the probability of the event A in the experiments on $\mathcal{S}_1$, and $P_2(B)$ is the probability of the event B in the experiment on $\mathcal{S}_2$. The event $A \times \mathcal{S}_2$ of the experiment $\mathcal{S}$ occurs if the event A of $\mathcal{S}_1$ occurs, irrespective of the outcome of $\mathcal{S}_2$. Similarly, the event $\mathcal{S}_1 \times B$ of $\mathcal{S}$ occurs if B of the experiment $\mathcal{S}_2$ occurs, irrespective of the outcome of $\mathcal{S}_1$. This justifies the equations of $P(A \times \mathcal{S}_2)$ and $P(\mathcal{S}_1 \times B)$ (see Figure 1.10). Those equations determine only the probabilities of the event $A \times \mathcal{S}_2$ and $\mathcal{S}_1 \times B$. To find the probabilities of $A \times B$ and that of their unions and intersections, we need additional information about $\mathcal{S}_1$ and $\mathcal{S}_2$, such as independence. In many applications $A \times \mathcal{S}_2$ and $\mathcal{S}_1 \times B$ of the combined experiment are independent for any A and B because

$$(A \times \mathcal{S}_2) \cap (\mathcal{S}_1 \times B) = A \times B$$

[1] If A is a subset of $\mathcal{S}_1$ and B is a subset of $\mathcal{S}_2$ then $A \times B$ consists of all pairs (ζ_1, ζ_2), where $\zeta_1 \in A$ and $\zeta_2 \in B$, is a subset of $\mathcal{S}$.

and

$$P(A \times B) = P(A \times \mathcal{S}_2)P(\mathcal{S}_1 \times B) = P_1(A)P_2(B)$$

Note that the elementary event $\{\zeta_1, \zeta_2\}$ can be written as $\{\zeta_1\} \times \{\zeta_2\}$ of the elementary events $\{\zeta_1\}$ and $\{\zeta_2\}$ of $\mathcal{S}_1$ and $\mathcal{S}_2$.

$$P(\{\zeta_1, \zeta_2\}) = P(\{\zeta_1\}) \cdot P(\{\zeta_2\})$$

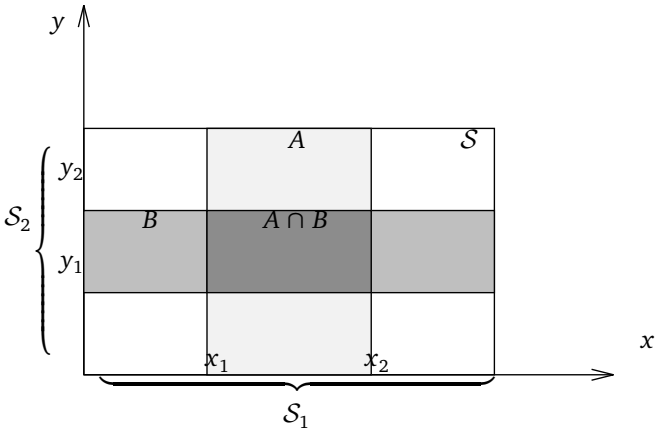


Figure 1.10: Independent events.

Example 1.1.18 *Papoulis [1991]*
A box B_1 contains 10 white and 5 red balls and a box B_2 contains 20 white and 20 red balls. A ball is drawn from each box. What is the probability that the ball from B_1 will be white and the ball from B_2 red?

- *$\mathcal{S}_1$ is the drawing from B_1. $\mathcal{S}_2$ is the drawing from B_2.*
- *$\mathcal{S}_1$ has 15 elements : 10 white and 5 red balls*

$$P(W_1 = \{\text{“all white balls in } B_1\text{”}\}) = \frac{10}{15}$$

- *$\mathcal{S}_2$ has 40 elements : 20 white and 20 red balls.*

$$P(R_2 = \{\text{“all red balls”}\}) = \frac{20}{40}$$

- *$\mathcal{S}_1 \times \mathcal{S}_2$ has 40×15 elements: all possible pairs that can be drawn.*

We want

$$\begin{aligned} P(W_1 \times R_2) &= P(\{\text{“white from } B_1 \text{ and red from } B_2\text{”}\}) \\ &= P(W_1)P(R_2) = \frac{10}{15} \times \frac{20}{40} \end{aligned}$$

1.2 Random Variables

A random variable is a number $X(\zeta)$ assigned to every outcome $\zeta \in \mathcal{S}$ of an experiment. This could be the gain in a game of chance, the voltage of a random source, or the cost of a random component.

Given $\zeta \in \mathcal{S}$, the mapping $\zeta \to X(\zeta) \in \mathbb{R}$ is a random variable, if for all $x \in \mathbb{R}$ the set $\{\zeta : X(\zeta) \leq x\}$ is also an event

$$\{X \leq x\} := \{\zeta : X(\zeta) \leq x\}$$

with probability

$$\mathsf{P}(\{X \leq x\}) := \mathsf{P}(\{\zeta : X(\zeta) \leq x\})$$

Example 1.2.1 *The phase angle ϕ of a voltage takes values between $\pm\pi$. $\mathcal{S}$ in this case is $[-\pi, \pi]$. Let the outcome of a measurement of the phase angle be ζ. An example for a random variable is the mapping $\zeta \to X(\zeta) = 1$ for positive phase angles, $\zeta \to X(\zeta) = 0$ for $\zeta = 0$ and $\zeta \to X(\zeta) = -1$ for negative phase angles.*

1.2.1 The Distribution Function and its Properties

The mapping $x \in \mathbb{R} \rightarrow F_X(x) = \mathsf{P}(\{X \leq x\})$ is called the probability distribution function of the random variable X. Note that for conveniance, we omit the independent variable ζ in $X(\zeta)$. We will deal throughout the manuscript with continuous random variables.

The probability distribution function has the following properties

1. $$0 \leq F_X(x) \leq 1$$

   $$F_X(\infty) = 1$$

   $$F_X(-\infty) = 0$$

2. $F_X(x)$ is continuous from the right i.e.,

   $$\lim_{\epsilon \rightarrow 0} F_X(x+\epsilon) = F_X(x)$$

3. If $x_1 < x_2$ then $F_X(x_1) \leq F_X(x_2)$, i.e., $F_X(x)$ is a non-decreasing function of x

   $$\mathsf{P}(\{x_1 < X \leq x_2\}) = F_X(x_2) - F_X(x_1) \geq 0$$

4. For a continuous variable X, $\mathsf{P}(x_1 < X < x_2) = \mathsf{P}(x_1 < X \leq x_2)$ because

   $$\mathsf{P}(\{x\}) = \mathsf{P}((-\infty, x]) - \mathsf{P}((-\infty, x)) = F_X(x) - F_X(x-0) = 0$$

1.2.2 The Probability Density Function and its Properties

The probability density function is defined as:

$$f_X(x) = \frac{dF_X(x)}{dx}.$$

If it exists has the property:

1. $$f_X(x) \geq 0.$$

2. To find $F_X(x)$, we integrate $f_X(x)$, i.e.,

   $$\mathsf{P}(\{X \leq x\}) = F_X(x) = \int_{-\infty}^{x} f_X(u)du$$

3. Because $F_X(\infty) = 1$,

   $$\int_{-\infty}^{\infty} f_X(x)dx = 1.$$

4. Further,

   $$\mathsf{P}(\{x_1 < X \leq x_2\}) = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)\, dx$$

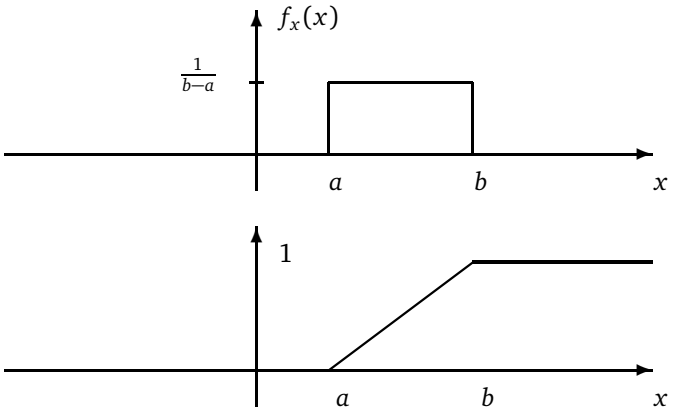


Figure 1.11: Density function $f_X(x)$ and distribution $F_X(x)$ of a uniform random variable on $[a, b]$.

1.2.3 Examples of Distribution Functions

Uniform Distribution

X is said to be uniformly distributed on $[a, b]$ if it admits the probability density function

$$f_X(x) = \frac{1}{b-a}\left[u(x-a) - u(x-b)\right],\ a < b$$

and probability distribution function

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du = \frac{1}{b-a}\left[(x-a)u(x-a) - (x-b)u(x-b)\right]$$

Herein,

$$u(x) = \begin{cases} 1, & x > 0 \\ \frac{1}{2}, & x = 0 \\ 0, & x < 0 \end{cases}$$

Both $F_X(x)$ and $f_X(x)$ are displayed in Figure 1.11.

Examples of uniform distributions are given in Example 1.2.2.

Example 1.2.2

- *The noise caused by quantizing the pixels of an image to discrete levels has approximately a uniform distribution.*
- *Consider, for example the task of locating a whale with an underwater microphone based on the sounds they make for communicating with each other. If nothing is known about the occurrences of the whales in the area, the angle of arrival of the whale signal impinging onto the microphone can be assumed to be uniformly distributed between* $0°$ *and* $360°$.
- *Consider the phase of a communications signal, which depends on the distance to the receiver. The phase angle is often assumed to be uniformly distributed between* 0 *and* 2π.

Exponential Distribution

Let X be exponentially distributed with parameter λ ($X \sim \exp(\lambda)$). Then,

$$f_X(x) = \lambda e^{-\lambda x}u(x), \quad \lambda > 0$$

and

$$F_X(x) = \left(1 - e^{-\lambda x}\right)u(x).$$

The exponential distribution is useful in many applications in engineering, for example, to describe the life-time X of a transistor. The breakdown rate of a transistor is defined by

$$\lim_{\Delta \to 0^+} \frac{1}{\Delta} \frac{\mathsf{P}\{x < X \leq x + \Delta\}}{\mathsf{P}\{x < X\}}$$

and is given by

$$\frac{f_X(x)}{1 - F_X(x)} = \lambda = \text{constant}$$

Example 1.2.3 *Yates and Goodman [1999]*
The probability that a telephone call lasts no more than t minutes is often described by an exponential distribution. $F_T(t)$ is given by:

$$F_T(t) = \begin{cases} \left(1 - \mathrm{e}^{-\frac{t}{3}}\right), & t \geq 0 \\ 0, & \textit{otherwise.} \end{cases}$$

Consequently, the probability density function $f_T(t)$ of the duration in minutes can be described by:

$$f_T(t) = \frac{dF_T(t)}{dt} = \begin{cases} (1/3)\,\mathrm{e}^{-\frac{t}{3}}, & t \geq 0 \\ 0, & \textit{otherwise.} \end{cases}$$

The probability of a telephone call lasting, e.g., between 2 and 4 minutes is:

$$\mathsf{P}(2 \leq T \leq 4) = F_T(4) - F_T(2) = \mathrm{e}^{-\frac{2}{3}} - \mathrm{e}^{-\frac{4}{3}} = 0.250.$$

Normal (Gaussian) Distribution

X is normally distributed or equivalently Gaussian distributed with mean μ and variance σ^2, also written as $X \sim \mathcal{N}(\mu, \sigma^2)$, if

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad \sigma^2 > 0, -\infty < \mu < \infty$$

A closed-form expression for $F_X(x)$ does not exist. For $\mu = 0,\ \sigma^2 = 1,\ X$ is said to be standard normally distributed, and this distribution function is denoted by $\Phi_X(x)$, i.e., we have

$$\Phi_X(\frac{x-\mu}{\sigma}) = F_X(x)$$

Both $F_X(x)$ and $f_X(x)$ are displayed in Figure 1.10.

The Normal or Gaussian distribution is very popular in engineering to model noise. Examples are:

Example 1.2.4

- *The additive Gaussian noise channel model in communications.*
- *The additive Gaussian noise caused by thermal noise in resistors.*
- *Images contain additive Gaussian noise due to the amplifiers in the cameras.*
- *In radar, sensor noise is also very well modeled by additive Gaussian noise.*

Example 1.2.5 *Figure 1.13 shows various examples of a probability density function of a normal distribuion with different μ and σ.*

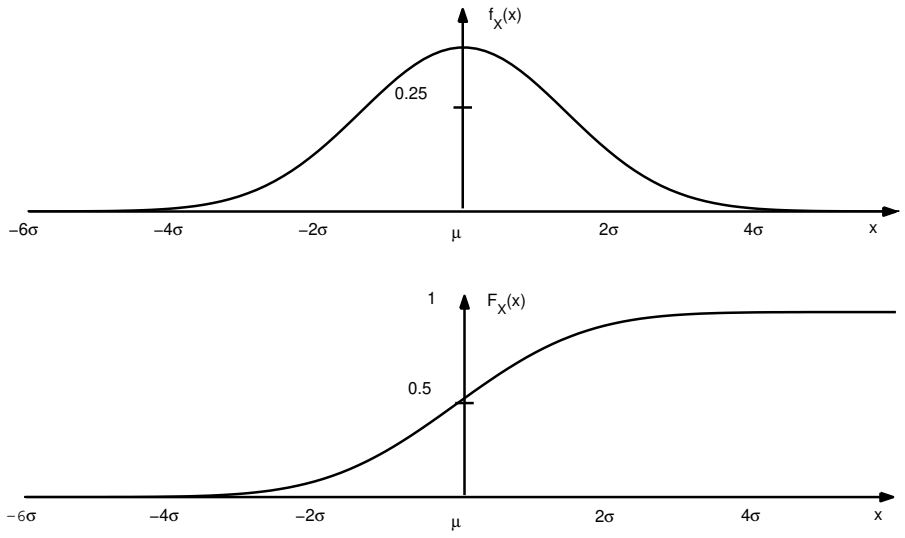


Figure 1.12: Density function $f_X(x)$ and distribution $F_X(x)$ of a Gaussian random variable.

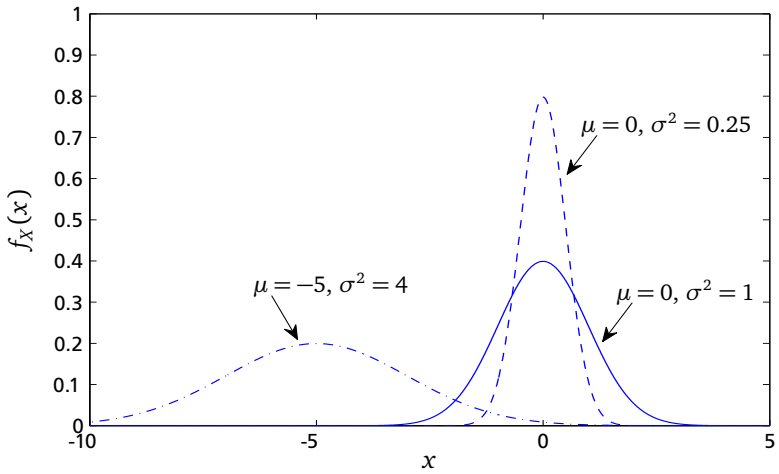


Figure 1.13: Probability density function of a normal distribution for different μ and σ values.

Example 1.2.6 *Show that for the normal distribution* $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$.

$$\int_{-\infty}^{\infty} f_X(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\,dx$$

If we let $s = \frac{x-\mu}{\sigma}$, *then*

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2}\,ds$$

Because $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} > 0$, *we calculate*

$$\begin{aligned}\left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx\right)^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}\,dx\,dy \\ &= \int_0^{2\pi}\int_0^{\infty} \frac{1}{2\pi} e^{-\frac{1}{2}r^2} r\,dr\,d\varphi\end{aligned}$$

and

$$\int_0^{2\pi}\int_0^{\infty}\frac{1}{2\pi}e^{-\frac{1}{2}r^2}r\,dr\,d\varphi=\int_0^{\infty}e^{-\frac{1}{2}r^2}r\,dr=\int_0^{\infty}e^{-s}\,ds=1$$

where $r=\sqrt{x^2+y^2}$ *and* $\varphi=\arctan\left(\frac{y}{x}\right)$.

Example 1.2.7 *Papoulis [1991]*
A random variable $X\sim\mathcal{N}(1000,50^2)$. *Find the probability that* X *is between 900 and 1050. We have*

$$P(\{900\le X\le 1050\})=F_X(1050)-F_X(900)$$

Numerical values of $\Phi_X(x)$ *over a range of* x *are shown in Table 1. Using Table 1, we conclude that*

$$P(\{900\le X\le 1050\})=\Phi_X(1)+\Phi_X(2)-1=0.819.$$

since

$$\Phi_X(-x)=1-\Phi_X(x).$$

Often, we define the following as the error-function. Note that we use $\text{erf}(x)=\Phi_X(x)-\frac{1}{2}$, though there are other definitions in the literature.

$$\text{erf}(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-x^2/2}du=\Phi_X(x)-\frac{1}{2}$$

Table 1: Numerical values for $\text{erf}(x)$ function.

x	$\text{erf}(x)$	x	$\text{erf}(x)$	x	$\text{erf}(x)$	x	$\text{erf}(x)$
0.05	0.01994	0.80	0.28814	1.55	0.43943	2.30	0.48928
0.10	0.03983	0.85	0.30234	1.60	0.44520	2.35	0.49061
0.15	0.05962	0.90	0.31594	1.65	0.45053	2.40	0.49180
0.20	0.07926	0.95	0.32894	1.70	0.45543	2.45	0.49286
0.25	0.09871	1.00	0.34134	1.75	0.45994	2.50	0.49379
0.30	0.11791	1.05	0.35314	1.80	0.46407	2.55	0.49461
0.35	0.13683	1.10	0.36433	1.85	0.46784	2.60	0.49534
0.40	0.15542	1.15	0.37493	1.90	0.47128	2.65	0.49597
0.45	0.17364	1.20	0.38493	1.95	0.47441	2.70	0.49653
0.50	0.19146	1.25	0.39495	2.00	0.47726	2.75	0.49702
0.55	0.20884	1.30	0.40320	2.05	0.47982	2.80	0.49744
0.60	0.22575	1.35	0.41149	2.10	0.48214	2.85	0.49781
0.65	0.24215	1.40	0.41924	2.15	0.48422	2.90	0.48813
0.70	0.25804	1.45	0.42647	2.20	0.48610	2.95	0.49841
0.75	0.27337	1.50	0.43319	2.25	0.48778	3.00	0.49865

1.2.4 Multiple Random Variables

The Joint Distribution Function and its Properties

The probabilities of two events $A=\{X_1\le x_1\}$ and $B=\{X_2\le x_2\}$ have already been defined as functions of x_1 and x_2, respectively, called probability distribution functions

$$\begin{aligned}F_{X_1}(x_1) &= \mathsf{P}(\{X_1\le x_1\})\\ F_{X_2}(x_2) &= \mathsf{P}(\{X_2\le x_2\})\end{aligned}$$

We must introduce a new concept to include the probability of the joint event $\{X_1\le x_1,X_2\le x_2\}=\{(X_1,X_2)\in D\}$, where x_1 and x_2 are two arbitrary real-valued numbers and D is the quadrant shown in Figure 1.14.

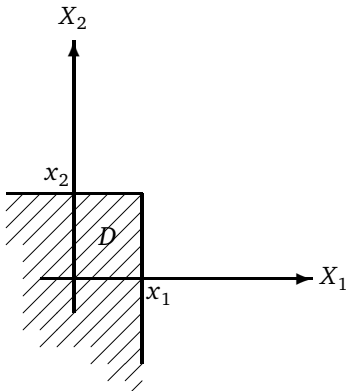


Figure 1.14: The quadrant D of two arbitrary real-valued numbers x and y.

The joint distribution $F_{X_1X_2}(x_1,x_2)$ of two random variables X_1 and X_2 is the probability of the event $\{X_1 \leq x_1, X_2 \leq x_2\}$,

$$F_{X_1X_2}(x_1,x_2) = \mathsf{P}(\{X_1 \leq x_1, X_2 \leq x_2\}).$$

For n random variables $X_1,\ldots,X_n$, the joint distribution function is denoted by

$$F_{X_1X_2\ldots X_n}(x_1,\ldots,x_n) = \mathsf{P}(\{X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n\}) \tag{1.3}$$

The multivariate probability distribution function has the following properties.

1)

$$\lim_{x_1\to\infty,\ldots,x_n\to\infty} F_{X_1X_2\ldots X_n}(x_1,\ldots,x_n) = F_{X_1X_2\ldots X_n}(\infty,\infty,\ldots,\infty) = 1$$
$$\lim_{x_i\to-\infty} F_{X_1X_2\ldots X_n}(x_1,\ldots,x_{i-1},x_i,x_{i+1},\ldots,x_n) = F_{X_1X_2\ldots X_n}(x_1,\ldots,x_{i-1},-\infty,x_{i+1},\ldots,x_n) = 0$$
$$\lim_{x_i\to\infty} F_{X_1X_2\ldots X_n}(x_1,\ldots,x_{i-1},x_i,x_{i+1},\ldots,x_n) = F_{X_1\ldots X_{i-1}X_{i+1}\ldots X_n}(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)$$

2) F is continuous from the right in each component x_i when the remaining components are fixed.

3) F is a non-decreasing function of each component x_i when the remaining components are fixed.

4) For the nth difference

$$\Delta_{\mathbf{a}}^{\mathbf{b}}F = \sum_{k\in\{0,1\}^n} (-1)^{\left[n-\sum_{i=1}^n k_i\right]} F(b_1^{k_1}a_1^{1-k_1},\ldots,b_n^{k_n}a_n^{1-k_n}) \geq 0$$

where $\mathbf{a} = (a_1,\ldots,a_n)^T$, $\mathbf{b} = (b_1,\ldots,b_n)^T$ with $a_i < b_i$, $i = 1,\ldots,n$

where ^{T} denotes transpose. For $n = 1$, $\Delta_{\mathbf{a}}^{\mathbf{b}}F$ is equivalent to $F(b) - F(a)$, and for $n = 2$, $\Delta_{\mathbf{a}}^{\mathbf{b}}F$ is equivalent to $F(b_1,b_2) - F(b_1,a_2) - F(a_1,b_2) + F(a_1,a_2)$. To show that (1)–(3) are not sufficient to define a probability, one should discuss the function $F_{X_1X_2}(x_1,x_2)$, defined as:

$$F_{X_1X_2}(x_1,x_2) = \begin{cases} 0, & x_1 + x_2 < 0, \\ 1, & \text{elsewhere,} \end{cases}$$

which is depicted in Figure 1.15.

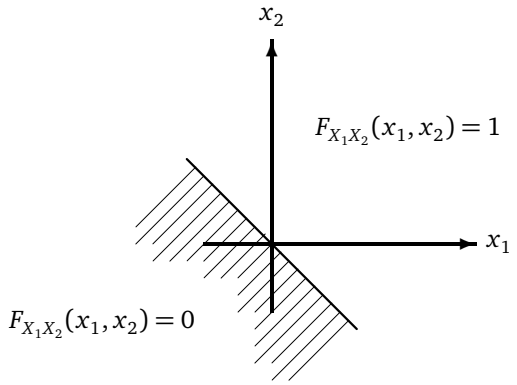


Figure 1.15: The Function $F_{X_1X_2}(x_1,x_2)$ is defined above.

Joint Density Function and its Properties

For the random variables $X_1,\ldots,X_n$, the joint density function is defined by the nth derivative of the joint distribution function, wherever it exists[2]:

$$f_{X_1\ldots X_n}(x_1,\ldots,x_n)=\frac{\partial^n F_{X_1\ldots X_n}(x_1,\ldots,x_n)}{\partial x_1\cdots\partial x_n} \tag{1.4}$$

The properties of the joint density function follow from its definition (1.4) and the properties of the joint distribution function (1.3):

1.

$$f_{X_1\ldots X_n}(x_1,\ldots,x_n)\geq 0$$

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{X_1\ldots X_n}(x_1,\ldots,x_n)dx_1\cdots dx_n=1$$

2. The joint distribution function may be obtained from the joint density by direct integration

$$F_{X_1\ldots X_n}(x_1,\ldots,x_n)=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n} f_{X_1\ldots X_n}(\alpha_1,\ldots,\alpha_n)d\alpha_1\cdots d\alpha_n$$

3. The marginal distribution and density functions are obtained, respectively, according to

$$\begin{aligned} F_{X_i}(x_i) &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\int_{-\infty}^{x_i} f_{X_1\ldots X_n}(x_1,\ldots,x_{i-1},\alpha,x_{i+1},\ldots,x_n)d\alpha\, dx_1\cdots dx_{i-1}dx_{i+1}\cdots dx_n \\ f_{X_i}(x_i) &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{X_1\ldots X_n}(x_1,\ldots,x_n)dx_1\cdots dx_{i-1}dx_{i+1}\cdots dx_n \end{aligned}$$

Demonstration of property 1 is sufficient to show that a given function is a valid joint density.

Example 1.2.8 *From the joint distribution of the amplitudes and periods of sea waves we can infer for example, that higher waves tend to be more regular than smaller waves.*

[2]In the case of discrete random variables, the joint distribution function will have step discontinuities, at which the derivate is undefined. However, it is still possible to define the joint density function at these points by using impulse (Dirac delta) functions.

Conditional Distribution and Density

The conditional distribution function of a random variable X given that some event A occurs is

$$F_X(x|A) = \mathsf{P}(X \leq x|A) = \frac{\mathsf{P}(X \leq x \cap A)}{\mathsf{P}(A)},$$

for an event A with non-zero probability. The conditional density function is defined as the derivative w.r.t x, where it exists, of the joint distribution function:

$$f_X(x|A) = \frac{\partial F_X(x|A)}{\partial x}.$$

It is often of interest to determine the conditional distribution of a random variable X_1 given that another random variable, X_2, takes on a particular value x_2. Consider the event

$$A = \{x_2 - \Delta < X_2 \leq x_2 + \Delta\}$$

where $\Delta > 0$. The conditional distribution function is given by

$$F_{X_1}(x_1|x_2 - \Delta < X_2 \leq x_2 + \Delta) = \frac{\int_{x_2-\Delta}^{x_2+\Delta} \int_{-\infty}^{x_1} f_{X_1X_2}(\alpha_1, \alpha_2) d\alpha_1 \alpha_2}{\int_{x_2-\Delta}^{x_2+\Delta} f_{X_2}(\alpha) d\alpha}. \tag{1.5}$$

For X_1 and X_2, both continuous random variables, the denominator of Equation (1.5) becomes zero in the limit as $\Delta \to 0$, however, the conditional distribution function may still exist. For a small value of Δ, Equation (1.5) may be expressed as

$$F_{X_1}(x_1|x_2 - \Delta < X_2 \leq x_2 + \Delta) \approx \frac{2\Delta \int_{-\infty}^{x_1} f_{X_1X_2}(\alpha, x_2) d\alpha}{2\Delta f_{X_2}(x_2)}$$

and in the limit as $\Delta \to 0$ we have

$$F_{X_1}(x_1|X_2 = x_2) = \frac{\int_{-\infty}^{x_1} f_{X_1X_2}(\alpha, x_2) d\alpha}{f_{X_2}(x_2)} \tag{1.6}$$

for all x_2 where $f_{X_2}(x_2) \neq 0$.

The conditional density function is obtained by differentiating Equation (1.6) w.r.t x_1:

$$f_{X_1}(x_1|x_2) \triangleq f_{X_1}(x_1|X_2 = x_2) = \frac{f_{X_1X_2}(x_1, x_2)}{f_{X_2}(x_2)}.$$

In the general case, we may obtain the joint conditional density of random variables $X_1, \ldots, X_n$ given that the random variables $X_{n+1}, \ldots, X_m$ take on the values $x_{n+1}, \ldots, x_m$, respectively:

$$\begin{aligned} f_{X_1 \ldots X_n}(x_1, \ldots, x_n|x_{n+1}, \ldots, x_m) &\triangleq f_{X_1 \ldots X_n}(x_1, \ldots, x_n|X_{n+1} = x_{n+1}, \ldots, X_m = x_m) \\ &= \frac{f_{X_1 \ldots X_m}(x_1, \ldots, x_m)}{f_{X_{n+1} \ldots X_m}(x_{n+1}, \ldots, x_m)}. \end{aligned}$$

Example 1.2.9

- *In radar applications, an example of a conditional distribution is that of missed detection, i.e. the probability distribution of deciding that a target is absent conditioned on the presence of a target. Similarly, the case of false alarm is a conditional distribution, i.e. the probability distribution of deciding that a target is present conditioned on the absence of the target.*
- *In binary communications, an example of a conditional distribution is the probability distribution of receiving a "0" when a "1" has been sent.*

Statistical Independence

Suppose that in a combined experiment the result of an experiment i depends on the first to $(i-1)$th experiments. With the density function $f_{X_1}(x_1)$ for the first experiment, one constructs the density function of the total experiment

$$\begin{aligned} f_{X_1\ldots X_n}(x_1,\ldots,x_n) &= f_{X_1}(x_1)\prod_{i=2}^{n} f_{X_i}(x_i|x_1,\ldots,x_{i-1}) \\ &= f_{X_1}(x_1)\prod_{i=2}^{n} f_{X_i}(x_i|X_1=x_1,\ldots,X_{i-1}=x_{i-1}) \end{aligned}$$

If the ith experiment is independent of the previous ones, then

$$f_{X_i}(x_i|x_1,\ldots,x_{i-1}) = f_{X_i}(x_i)$$

and

$$f_{X_1\ldots X_n}(x_1,\ldots,x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

or equivalently

$$F_{X_1\ldots X_n}(x_1,\ldots,x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$$

which means that the events $\{X_1 \le x_1\},\ldots,\{X_n \le x_n\}$ are independent. For the special case of $n=2$, we have the following properties:

1)

$$\begin{aligned} F_{X_1X_2}(\infty,\infty) &= \mathsf{P}(X_1 \le \infty, X_2 \le \infty) = 1 \\ F_{X_1X_2}(-\infty,x_2) &= \mathsf{P}(X_1 \le -\infty, X_2 \le x_2) = 0 = F_{X_1X_2}(x_1,-\infty) \end{aligned}$$

because $\{X_1=-\infty, X_2 \le x_2\} \subset \{X_1=-\infty\}, \{X_1 \le x_1, X_2=-\infty\} \subset \{X_2=-\infty\}$.

2)

$$\lim_{\epsilon\to 0} F_{X_1X_2}(x_1+\epsilon,x_2) = \lim_{\epsilon\to 0} F_{X_1X_2}(x_1,x_2+\epsilon) = F_{X_1X_2}(x_1,x_2).$$

3) The event $\{x_{11} < X_1 \le x_{12}, X_2 \le x_2\}$ consists of all the points (X_1,X_2) in D_1, and the event $\{X_1 \le x_1, x_{21} < X_2 \le x_{22}\}$ consists of all the points in D_2. From Figure 1.16, we see that

$$\{X_1 \le x_{12}, X_2 \le x_2\} = \{X_1 \le x_{11}, X_2 \le x_2\} \cup \{x_{11} < X_1 \le x_{12}, X_2 \le x_2\}.$$

The two summands are mutually exclusive, hence

$$\mathsf{P}(\{X_1 \le x_{12}, X_2 \le x_2\}) = \mathsf{P}(\{X_1 \le x_{11}, X_2 \le x_2\}) + \mathsf{P}(\{x_{11} < X_1 \le x_{12}, X_2 \le x_2\})$$

and therefore,

$$\mathsf{P}(\{x_{11} < X_1 \le x_{12}, X_2 \le x_2\}) = F_{X_1X_2}(x_{12},x_2) - F_{X_1X_2}(x_{11},x_2).$$

Similarly,

$$\mathsf{P}(\{X_1 < x_1, x_{21} < X_2 \le x_{22}\}) = F_{X_1X_2}(x_1,x_{22}) - F_{X_1X_2}(x_1,x_{21})$$

4)

$$\begin{aligned} &\mathsf{P}(\{x_{11} < X_1 \le x_{12}, x_{21} < X_2 \le x_{22}\}) = \\ &\quad F_{X_1X_2}(x_{12},x_{22}) - F_{X_1X_2}(x_{11},x_{22}) - F_{X_1X_2}(x_{12},x_{21}) + F_{X_1X_2}(x_{11},x_{21}) > 0 \end{aligned}$$

This is a probability that (X_1,X_2) is in the rectangle D_3 (see Figure 1.17).

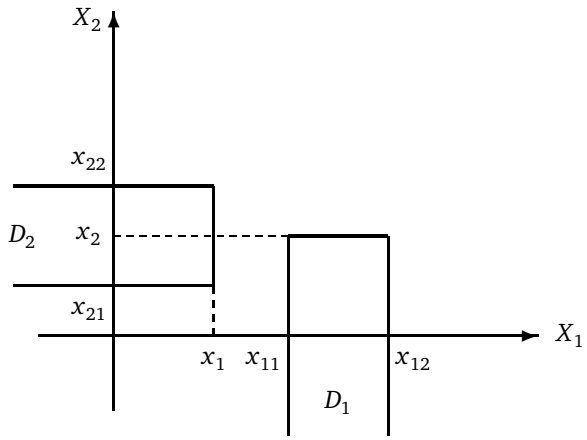


Figure 1.16: Regions of definition D_1 and D_2 for a bivariate distribution.

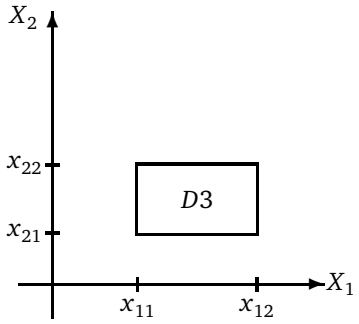


Figure 1.17: Probability that (X_1, X_2) is in the rectangle D_3.

In practical applications, the assumption of independence of random variables is often made. For example, the noise generated by hydrophones in an array signal processing context is independent from sensor to sensor.

Example 1.2.10 *The bivariate normal distribution*

$$f_{X_1X_1}(x_1,x_2) = \frac{1}{2\pi\sigma_{X_1}\sigma_{X_2}\sqrt{1-\rho^2}} \times \exp\left[-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_{X_1}}{\sigma_{X_1}}\right)^2 - 2\rho\frac{(x_1-\mu_{X_1})(x_2-\mu_{X_2})}{\sigma_{X_1}\sigma_{X_2}} + \left(\frac{x_2-\mu_{X_2}}{\sigma_{X_2}}\right)^2\right]\right]$$

where μ_{X_1} and μ_{X_2} are respectively the means of X_1 and X_2, $\sigma^2_{X_1} > 0$ and $\sigma^2_{X_2} > 0$ the corresponding variances, and $-1 < \rho < 1$ is the correlation coefficient.

$\mathcal{N}(\mu_{X_1}, \sigma^2_{X_1})$ and The marginal distributions are denoted by $\mathcal{N}(\mu_{X_2}, \sigma^2_{X_2})$.

The conditional density function $f_{X_2}(x_2|x_1)$ of X_2 when $X_1 = x_1$ is $N(\mu_{X_2} + \rho\frac{\sigma_{X_2}}{\sigma_{X_2}}(x_1-\mu_{X_2}), \sigma^2_{X_2}(1-\rho^2))$.

If X_1 and X_2 are independent, then $\rho = 0$, and clearly $f_{X_1X_2}(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$.

1.2.5 Operations on Random Variables

The concept of a random variable was previously introduced as a means of providing a definition of events defined on a sample space. It forms a mathematical model for describing characteristics of some real-valued, physical-world random variables, which are mostly based on a single concept – expectation.

Expectation

The mathematical expectation of a continuous random variable X,

$$\mathsf{E}[X]$$

which may be read "the expected value of X" or "the mean of X" is defined as

$$\mathsf{E}[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) dx$$

where $f_X(x)$ is the probability density function of X. If X happens to be discrete with N possible values x_i having probabilities $\mathsf{P}(x_i)$ of occurrence, then

$$\begin{aligned} f_X(x) &= \sum_{i=1}^{N} \mathsf{P}(x_i) \cdot \delta(x - x_i) \\ \mathsf{E}[X] &= \sum_{i=1}^{N} x_i \mathsf{P}(x_i) \end{aligned}$$

Example 1.2.11

i) Normal Distribution: If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left\{-\frac{1}{2}(x-\mu)^2/\sigma^2\right\}$$

$$\begin{aligned} \mathsf{E}[X] &= \int_{-\infty}^{\infty} x f_X(x) dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x \exp\{-\frac{1}{2}(x-\mu)^2/\sigma^2\} dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma u + \mu) \exp\{-\frac{1}{2}u^2\} \sigma du \\ &= \frac{1}{\sqrt{2\pi}} \left[\sigma \int_{-\infty}^{\infty} u \cdot \exp\{-\frac{1}{2}u^2\} du + \mu \int_{-\infty}^{\infty} \exp\{-\frac{1}{2}u^2\} du\right] \\ &= \frac{1}{\sqrt{2\pi}} \left[\sigma \cdot 0 + \mu\sqrt{2\pi}\right] \\ &= \mu \end{aligned}$$

ii) Uniform distribution: If X is uniformly distributed on the interval $[a, b]$, then

$$f_X(x) = \frac{1}{b-a} \cdot \mathrm{rect}_{[a,b]}(x)$$

where

$$\mathrm{rect}_{[a,b]} = \begin{cases} 1, & a \le x \le b \\ 0, & \text{else.} \end{cases}$$

$$\begin{aligned} \mathsf{E}[X] &= \int_{-\infty}^{\infty} x f_X(x) dx = \frac{1}{b-a} \int_a^b x dx \\ &= \frac{1}{b-a} \frac{1}{2}(b^2 - a^2) \\ &= \frac{(b-a)(b+a)}{2(b-a)} \\ &= \frac{a+b}{2} \end{aligned}$$

which can be verified graphically.

iii) Exponential Distribution: If X is exponentially distributed, then

$$f_X(x) = \frac{1}{\lambda}\exp\left[-\frac{x-a}{\lambda}\right]u(x-a)$$

$$\begin{aligned}
\mathsf{E}[X] &= \int_{-\infty}^{\infty} x \cdot f_X(x)dx \\
&= \int_{-\infty}^{\infty} x \cdot \frac{1}{\lambda}\exp\left[-\frac{x-a}{\lambda}\right]u(x-a)dx \\
&= \frac{1}{\lambda}\int_{a}^{\infty} x \cdot \exp\left[-\frac{x-a}{\lambda}\right]dx \\
&= \frac{1}{\lambda}\int_{a}^{\infty} x \cdot \exp\left[-\frac{x}{\lambda}\right]dx \exp\left[\frac{a}{\lambda}\right] \\
\mathsf{E}[X] &= a + \lambda
\end{aligned}$$

Note: If a random variable's density is symmetrical about a line $x = a$, then $\mathsf{E}[X] = a$, i.e.

$$\mathsf{E}[X] = a, \text{ if } f_X(x-a) = f_X(-x+a)$$

Expected Value of a Function of Random Variables

Suppose that we are interested in the mean of the random variable $Y = g(X)$

$$\mathsf{E}[Y] = \int_{-\infty}^{\infty} y \cdot f_Y(y)dx$$

and we are given $f_X(x)$. Then,

$$\mathsf{E}[Y] = \mathsf{E}[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)dx$$

Example 1.2.12 *The average power in a 1 Ω resistor can be found as*

$$\mathsf{E}[V^2] = \int_{-\infty}^{\infty} v^2 f_V(v)dv$$

where V and $f_V(v)$ are respectively the random voltage and its probability density function.

In particular, if we apply the transformation $g(\cdot)$ to the random variable X, defined as

$$g(X) = X^n, \ n = 0, 1, 2, \ldots,$$

the expectation of $g(X)$ is known as the nth order moment of X,

$$\mathsf{E}[X^n] = \int_{-\infty}^{\infty} x^n f_X(x)dx$$

and is denoted by μ_n. It is also of importance to use central moments of X around the mean. These are defined as

$$\mu_n' = \mathsf{E}(X - \mathsf{E}[X])^n, \ n = 0, 1, 2\ldots$$

$$\int_{-\infty}^{\infty} (x - \mathsf{E}[X])^n \cdot f_X(x)dx = \int (x - \mu_1)^n f_X(x)dx$$

Clearly the first order central moment is zero.

The Variance

The variance of a random variable X is by definition

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mathsf{E}[X])^2 \cdot f_X(x) dx$$

The positive constant σ, is called the standard deviation of X. From the definition it follows that σ^2 is the mean of the random variable $(X - \mathsf{E}[X])^2$, i.e. $\sigma^2 = \mathsf{E}(X - \mathsf{E}X)^2$. Thus,

$$\begin{aligned}
\sigma^2 &= \mathsf{E}[(X - \mathsf{E}[X])^2] = \mathsf{E}[X^2 - 2X \cdot (\mathsf{E}[X]) + (\mathsf{E}[X])^2] \\
&= \mathsf{E}[X^2] - 2\mathsf{E}[X\mathsf{E}[X]] + (\mathsf{E}[X])^2 \\
&= \mathsf{E}[X^2] - 2(\mathsf{E}[X])^2 + (\mathsf{E}[X])^2 \\
\sigma^2 &= \mathsf{E}[X^2] - (\mathsf{E}[X])^2
\end{aligned}$$

The variance can be understood as the dispersion of the random variable relative to the expected value.

Example 1.2.13 *We will now find the variance of three common distributions.*

i) Normal distribution:

Let $X \sim \mathcal{N}(\mu, \sigma^2)$. The probability density function of X is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right].$$

Clearly

$$\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\} dx = \sigma\sqrt{2\pi}$$

Differentiating with respect to σ, we obtain

$$\int_{-\infty}^{\infty} \frac{(x-\mu)^2}{\sigma^3} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\} dx = \sqrt{2\pi}$$

Multiplying both sides by $\sigma^2/\sqrt{2\pi}$, we conclude that

$$\mathsf{E}[(X-\mu)^2] = \mathsf{E}[(X - \mathsf{E}[X])^2] = \sigma^2.$$

ii) Uniform distribution:

$$\begin{aligned}
\mathsf{E}\left[(X - \mathsf{E}[X])^2\right] &= \mathsf{E}[X^2] - (\mathsf{E}[X])^2 \\
&= \int_{-\infty}^{\infty} x^2 f_X(x) dx - \left(\frac{a+b}{2}\right)^2 \\
&= \int_a^b \frac{x^2}{b-a} dx - \left(\frac{a+b}{2}\right)^2 \\
&= \frac{(b-a)^2}{12}.
\end{aligned}$$

iii) Exponential distribution:

$$f_X(x) = \lambda \exp\{-\lambda x\} u(x), \ \lambda > 0$$

It can be easily shown that

$$\mathsf{E}[X] = \int x \cdot f_X(x) dx = \int_0^{\infty} x\lambda \exp -\lambda x dx = \frac{1}{\lambda}$$

and that

$$\sigma^2 = \int (x - \mathsf{E}[X])^2 \cdot f_X(x) dx = \int_0^{\infty} (x - \frac{1}{\lambda})^2 \lambda \exp -\lambda x dx = \frac{1}{\lambda^2}$$

Transformation of a Random Variable

In practice, one may wish to transform one random variable X into a new random variable Y by means of a transformation

$$Y = g(X),$$

where the probability density function $f_X(x)$ or distribution function of X is known. We showed how one calculate $\mathsf{E}[g(X)]$ given $f_X(x)$. We turn our attention to finding the probability density function of $f_Y(y)$ or distribution function $F_Y(y)$ of Y. Let us assume that $g(\cdot)$ is continuous and differentiable at all values of x for which $f_X(x) \neq 0$. Furthermore, assume that $g(\cdot)$ is monotone for which the inverse $g^{-1}(\cdot)$ exists. Then,

$$F_Y(y) = \mathsf{P}[\{Y \leq y\}] = \mathsf{P}[\{g(X) \leq y\}] = \mathsf{P}[\{X \leq g^{-1}(y)\}] = \int_{-\infty}^{g^{-1}(y)} f_X(x)dx$$

holds. The density function of $Y = g(X)$ is obtained via differentiation, which leads to

$$f_Y(y) = f_X\left(g^{-1}(y)\right) \cdot \left|\frac{dg^{-1}(y)}{dy}\right|$$

Example 1.2.14 *Let*

$$Y = aX + b.$$

To find the probability density function of Y knowing $F_X(x)$ or $f_X(x)$, we calculate

$$\begin{aligned} F_Y(y) &= P(\{Y \leq y\}) = \mathsf{P}(\{aX + b \leq y\}) \\ &= P\left(\left\{X \leq \frac{y-b}{a}\right\}\right),\ a > 0 \\ F_Y(y) &= F_X\left(\frac{y-b}{a}\right), \end{aligned}$$

and

$$F_Y(y) = 1 - F_X\left(\frac{y-b}{a}\right),\ a < 0$$

By differentiating $F_Y(y)$, we obtain the probability density function

$$f_Y(y) = \frac{1}{|a|} \cdot f_X\left(\frac{y-b}{a}\right)$$

Example 1.2.15 *Let*

$$Y = X^2$$

To find the probability density function of Y knowing $F_X(x)$ or $f_X(x)$, we calculate

$$F_Y(y) = P\left(\{X^2 \leq y\}\right) = \begin{cases} 0, & y < 0 \\ P(\{-\sqrt{y} \leq X \leq \sqrt{y}\}), & y \geq 0 \end{cases}$$

$$F_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

By differentiating $F_Y(y)$, we obtain the probability density function

$$f_Y(y) = \begin{cases} 0, & y < 0 \\ \dfrac{f_X(\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}}, & y > 0 \end{cases}$$

$F_Y(y)$ is non differentiable at $y = 0$. We may choose arbitrarily $f_Y(y) = 0$ at $y = 0$.

Example 1.2.16 *The probability density function $f_Z(z)$ of a random variable Z is given by:*

$$f_Z(z) = \begin{cases} \lambda \exp\{-\lambda z\}, & \forall\, z \geq 0 \\ 0, & \textit{otherwise} \end{cases}$$

Let $W = \sinh^{-1}(Z)$, then the probability density function $f_W(w)$ of W can be calculated as follows:

$$f_W(w) = f_Z(g^{-1}(w)) \left| \frac{dg^{-1}(w)}{dw} \right|.$$

Here,

$$\begin{aligned} w &= \sinh^{-1}(z) \\ z &= g^{-1}(w) = \sinh(w) \\ \frac{dg^{-1}(w)}{dw} &= \cosh(w) \end{aligned}$$

Hence, it follows that:

$$\begin{aligned} f_W(w) &= f_Z(\sinh(w)) \cdot \cosh(w) \\ &= \lambda \exp(-\lambda \sinh(w)) \cdot \cosh(w). \end{aligned}$$

It remains to find the region in which $f_W(w)$ is defined, as $f_Z(z)$ is a piecewise continuous function. This is done by inserting the boundaries in which $f_Z(z)$ is defined in $w = \sinh^{-1}(z)$.

$$\begin{aligned} \text{for } z = 0: &\quad w = \sinh^{-1}(0) = 0 \\ \text{for } z \to \infty: &\quad w \to \infty. \end{aligned}$$

Hence,

$$f_W(w) = \begin{cases} \lambda \exp(-\lambda \sinh(w)) \cdot \cosh(w), & 0 \leq w \leq \infty \\ 0, & \textit{otherwise} \end{cases}$$

The Characteristic Function

The characteristic function (cf) of a random variable X is defined as follows:

$$\begin{aligned} \Phi_X(\omega) &= \mathsf{E}\left[e^{j\omega X}\right] && (1.7) \\ &= \int_{-\infty}^{\infty} f_X(x) e^{j\omega x} dx. && (1.8) \end{aligned}$$

This can be interpreted as the expectation of the function $g(X) = e^{j\omega X}$, or as the Fourier transform (with reversed sign of the exponent) of the probability density function (pdf) of X. Considering Equation (1.8) as a Fourier transform, we can readily apply all the properties of Fourier transforms to the characteristic function. Most importantly, the inverse relationship of Equation (1.8) is given by

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_X(\omega) e^{-j\omega x} d\omega.$$

From the definition above $|\Phi_X(\omega)| \leq 1$, and thus unlike the density function it always exists.

Example 1.2.17 *We shall derive the cf of a Gaussian distributed random variable. First, we consider $Z \sim \mathcal{N}(0,1)$.*

$$\Phi_Z(\omega) = \int f_Z(z) e^{j\omega z} dz$$

$$\begin{aligned}\int f_Z(z)e^{j\omega z}dz &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2}e^{j\omega z}dz \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-j\omega)^2+\frac{1}{2}(j\omega)^2}dz \\ &= e^{-\frac{1}{2}\omega^2}\int f_Z(z-j\omega)dz \\ &= e^{-\frac{1}{2}\omega^2}\end{aligned}$$

Now we consider the more general case of $X = \sigma Z + \mu'$ which is distributed as $\mathcal{N}(\mu, \sigma^2)$. Then,

$$\begin{aligned}\Phi_X(\omega) &= \mathsf{E}\left[e^{j\omega(\sigma Z+\mu)}\right] \\ &= e^{j\omega\mu}\Phi_Z(\sigma\omega) \\ &= e^{j\omega\mu-\frac{1}{2}\sigma^2\omega^2}\end{aligned}$$

The cf is also known as the moment-generating function as it may be used to determine the nth order moment of a random variable. This can be seen from the following derivation. Consider the power series expansion of an exponential term for $x < \infty$:

$$e^x = \sum_{n=0}^{\infty}\frac{x^n}{n!}.$$

Substituting this into Equation (1.8) yields

$$\begin{aligned}\Phi_X(\omega) &= \int_{-\infty}^{\infty} f_X(x)\left\{1 + j\omega x + \frac{(j\omega x)^2}{2!} + \cdots\right\}dx \\ &= 1 + j\omega\mathsf{E}[X] + \frac{(j\omega)^2}{2!}\mathsf{E}[X^2] + \cdots \qquad (1.9)\end{aligned}$$

Differentiating the above expression n times and evaluating it at the origin, gives

$$\left(\frac{d}{d\omega}\right)^n \Phi_X(\omega)\Big|_{\omega=0} = j^n\mathsf{E}[X^n]$$

which means that the nth order moment of X can be determined according to

$$\mathsf{E}[X^n] = \frac{1}{j^n}\left(\frac{d}{d\omega}\right)^n \Phi_X(\omega)\Big|_{\omega=0}.$$

This is known as the *moment theorem*. We note that when the power series expansion of Equation (1.9) converges, the characteristic function and hence the pdf of X are completely determined by the moments of X.

1.3 Covariance

If the statistical properties of the two random variables X_1 and X_2 are described by their joint probability density function $f_{X_1X_2}(x_1, x_2)$, then we have

$$\mathsf{E}[g(X_1, X_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1, x_2)\cdot f_{X_1X_2}(x_1, x_2)dx_1dx_2$$

For $g(X_1, X_2) = X_1^nX_2^k$

$$\mathsf{E}[X_1^nX_2^k] = \mu_{n,k} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1^nx_2^k f_{X_1X_2}(x_1, x_2)dx_1\,dx_2, \;\; n, k = 0, 1, 2, \ldots$$

are called the joint moments of X_1 and X_2. Clearly $\mu_{0k} = \mathsf{E}[X_2^k]$, while $\mu_{n0} = \mathsf{E}[X_1^n]$. We denote[3] the second order moment $\mathsf{E}[X_1X_2]$ of X_1 and X_2 by $r_{X_1X_2}$,

$$r_{X_1X_2} = \mathsf{E}[X_1X_2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2 f_{X_1X_2}(x_1,x_2)dx_1dx_2$$

If $r_{X_1X_2} = \mathsf{E}[X_1X_2] = \mathsf{E}[X_1]\cdot\mathsf{E}[X_2]$, then X_1 and X_2 are known to be uncorrelated. If $r_{X_1X_2} = 0$, then X_1 and X_2 are orthogonal.

The moments μ'_{nk}, $k = 0,1,2,\ldots$ are known to be joint central moments and defined by

$$\begin{aligned}\mu'_{nk} &= \mathsf{E}[(X_1-\mathsf{E}[X_1])^n(X_2-\mathsf{E}[X_2])^k] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_1-\mathsf{E}[X_1])^n(x_2-\mathsf{E}[X_2])^k f_{X_1X_2}(x_1,x_2)dx_1dx_2,\ n,k=0,1,2,\ldots\end{aligned}$$

The second order central moments

$$\begin{aligned}\mu'_{20} &= \sigma^2_{X_1} \\ \mu'_{02} &= \sigma^2_{X_2}\end{aligned}$$

are the variances of X_1 and X_2, respectively. The joint moment μ'_{11} is called covariance of X_1 and X_2 and is denoted by $c_{X_1X_2}$

$$\begin{aligned}c_{X_1X_2} &= \mathsf{E}[(X_1-\mathsf{E}[X_1])(X_2-\mathsf{E}[X_2])] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_1-\mathsf{E}[X_1])(x_2-\mathsf{E}[X_2])\cdot f_{X_1X_2}(x_1,x_2)dx_1dx_2.\end{aligned}$$

One can easily show that

$$c_{X_1X_2} = r_{X_1X_2} - \mathsf{E}[X_1]\mathsf{E}[X_2].$$

If two random variables are independent, i.e. $f_{X_1X_2}(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$, then $c_{X_1X_2} = 0$. The converse however is not true (only for the Gaussian case). The normalized second-order moment $\rho = \mu'_{11}/\sqrt{\mu'_{20}\mu'_{02}} = \frac{c_{X_1X_2}}{\sigma_{X_1}\sigma_{X_2}}$ is known as the correlation coefficient of X_1 and X_2. It can be shown that

$$-1 \le \rho \le 1$$

A correlation coefficient of ± 1 implies that a linear equation describes the relationship between X_1 and X_2 perfectly. A positive value implies that X_2 increases as X_1 increases while a negative value implies that X_2 decreases as X_1 increases. The case $\rho = 0$ means that there is no linear correlation between the variables, but does not say anything about a non-linear correlation. We will show that two uncorrelated random variables are not necessarily independent.

Example 1.3.1 *U is uniformly distributed on $[-\pi,\pi)$*

$$X_1 = \cos U,\ X_2 = \sin U$$

X_1 and X_2 are not independent because $X_1^2 + X_2^2 = 1$

$$\begin{aligned}c_{X_1X_2} &= \mathsf{E}[(X_1-\mathsf{E}[X_1])(X_2-\mathsf{E}[X_2])] \\ &= \mathsf{E}[X_1X_2] = \mathsf{E}[\cos U \sin U] \\ &= \tfrac{1}{2}\mathsf{E}[\sin 2U] \\ &= 0\end{aligned}$$

Thus X_1 and X_2 are uncorrelated but not independent.

[3] In the engineering literature $r_{X_1X_2}$ is often called correlation. We reserve this term to describe the normalized covariance as it is common in the statistical literature.

Example 1.3.2 *Let X_i, $i = 1,\ldots,N$ be random variables and α_i, $i = 1,\ldots,N$ be scalar values. We calculate the mean and the variance of*

$$Y = \sum_{i=1}^{N} \alpha_i X_i$$

given means and variances (and covariances) of X_i.

$$\mathsf{E}[Y] = \sum_{i=1}^{N} \alpha_i \mathsf{E}[X_i]$$

$$\sigma_Y^2 = \mathsf{E}[(Y - \mathsf{E}[Y])^2] = \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j c_{X_i X_j}$$

Only for the special case $c_{X_i X_j} = 0$, $i \neq j$

$$\sigma_Y^2 = \sum_{i=1}^{N} \alpha_i^2 \sigma_{X_i}^2.$$

1.4 The Multivariate Gaussian Random Distribution

We will now consider a set of real-valued random variables $X_1, X_2, \ldots, X_n$. If, for each random variable $X_i \leq x_i \in \mathcal{S}$ for all $x_i \in \mathbb{R}$, then:

$$X_1 \leq x_1, \ldots, X_n \leq x_n = X_1 \leq x_1 \cap \ldots \cap X_n \leq x_n \in \mathcal{S}.$$

The vector containing these random variables is called a random vector and is given by:

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix}$$

The N-dimensional probability distribution function of the random vector $\mathbf{X}$ is defined as:

$$F_X(\mathbf{x}) = \mathbf{F}_{\mathbf{X_1},\ldots,\mathbf{X_n}}(\mathbf{x_1},\ldots,\mathbf{x_n}) = \mathsf{P}(\{\mathbf{X_1} \leq \mathbf{x_1},\ldots,\mathbf{X_n} \leq \mathbf{x_n}\}),$$

where $\mathbf{x} = [\mathbf{x_1},\ldots,\mathbf{x_n}]^{\mathbf{T}} \in \mathbb{R}^{\mathbf{n}}$. The N-dimensional probability denisty function of the random variable is thus:

$$f_X(\mathbf{x}) = \mathbf{f}_{\mathbf{X_1},\ldots,\mathbf{X_n}}(\mathbf{x_1},\ldots,\mathbf{x_n}).$$

Definition 1.4.1 *A random vector $\mathbf{X}$ is said to be Gaussian if its elements $X_1, X_2, \ldots, X_N$ have an N-dimensional Gaussian probability density function.*

Let $\mathbf{X}$ be the column vector denoting the N random variables:

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix}$$

The N-dimensional Gaussian probability density function is:

$$f_X(\mathbf{x}) = \frac{1}{(2\pi)^{N/2} \cdot |\mathbf{C}|^{1/2}} e^{-\frac{1}{2}\left[(\mathbf{x}-\mu)^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{x}-\mu)\right]}$$

where the mean vector is:

$$\mu = \mathsf{E}[\mathbf{X}] = \begin{bmatrix} \mathsf{E}[X_1] \\ \mathsf{E}[X_2] \\ \vdots \\ \mathsf{E}[X_N] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_N \end{bmatrix},$$

$\mu_i = \mathsf{E}[X_i], i = 1, \ldots, N$, and $(\mathbf{x} - \mu)^{\mathrm{T}}$ denotes the transpose of the column vector $(\mathbf{x} - \mu)$. In the equation above $|\mathbf{C}|$ is the determinant of the matrix $\mathbf{C}$, and $\mathbf{C}^{-1}$ is its inverse. The covariance matrix $\mathbf{C}$ is defined by:

$$\mathbf{C} = \begin{pmatrix} c_{X_1X_1} & c_{X_1X_2} & \cdots & c_{X_1X_N} \\ c_{X_2X_1} & c_{X_2X_2} & \cdots & c_{X_2X_N} \\ \vdots & \vdots & \ddots & \vdots \\ c_{X_NX_1} & c_{X_NX_2} & \cdots & c_{X_NX_N} \end{pmatrix}$$

where its elements are:

$$c_{X_iX_j} = \mathsf{E}\left[(X_i - \mu_i) \cdot (X_j - \mu_j)\right]$$

for $i, j = 1, \ldots, N$. The elements of the covariance matrix are given by:

$$c_{X_iX_j} = r_{X_iX_j} - \mu_i\mu_j$$

If, in addition, the X_i happen to be uncorrelated, such that $\mathsf{E}[X_iX_j] = \mathsf{E}[X_i] \cdot \mathsf{E}[X_j] = \mu^2$ for $i \neq j$, then the covariance matrix becomes diagonal, i.e.,

$$\mathbf{C} = \begin{pmatrix} \sigma^2 & 0 & \ldots & 0 \\ 0 & \sigma^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma^2 \end{pmatrix} = \sigma^2\mathbf{I}$$

where $\sigma^2 = \mathsf{E}[X^2] = r_{XX} - \mu^2$, assuming all variances of the X_i are equal.

1.5 Central Limit Theorem

Under certain conditions, a sum of independently and identically distributed (iid) random variables can be shown to have an asymptotically Gaussian distribution. This has motivated the use of the Gaussian distribution as a suitable model for random phenomena encountered in many engineering problems. The following theorem applies to the case of independently and identically distributed (iid) random variables.

Theorem 1.5.1 *Rao [1973] Lindeberg-Levy Theorem:*
Let $X_n, n = 1, 2, \ldots, N$ be a sequence of iid random variables, such that $\mathsf{E}[X_n] = \mu$ and $\mathsf{Var}[X_n] = \sigma^2 \neq 0$ exist. The random variable Y is defined as

$$Y = \sqrt{N}(\bar{X} - \mu)/\sigma,$$

where $\bar{X} = N^{-1}\sum_{n=1}^{N} X_n$. Then the distribution function of Y tends to that of a normal Gaussian distribution as $N \to \infty$.

The proof of Theorem 1.5.1 given here follows that of Rao [1973]. Let $Z_n = X_n - \mu$. Since the first two moments exist,

$$\Phi_Z(\omega) = 1 - \frac{1}{2}\sigma^2\omega^2 + o(\omega^2)$$

where the notation $o(\omega^2)$ implies that

$$\lim_{\omega \to \infty} \frac{o(\omega^2)}{\omega^2} = 0$$

e.g. the term $o(\omega^2)$ may consist of powers of ω less than two. The cf of Y is then,

$$\Phi_Y(\omega) \quad = \quad \left[\Phi_Z\left(\frac{\omega}{\sigma\sqrt{N}}\right)\right]^N$$

$$= \left[1 - \frac{\omega^2}{2N} + o\left(\frac{\omega^2}{N}\right)\right]^N$$

$$\log[\Phi_Y(\omega)] = N \log\left[1 - \frac{\omega^2}{2N} + o\left(\frac{\omega^2}{N}\right)\right]$$

which tends to

$$-\frac{\omega^2}{2} \quad \text{as} \quad N \to \infty.$$

Therefore,

$$\Phi_Y(\omega) \to e^{-\frac{\omega^2}{2}} \quad \text{as} \quad N \to \infty.$$

Recall from Example 1.2.17 that the cf of a standard Gaussian random variable W is given by

$$\Phi_W(\omega) = e^{-\frac{\omega^2}{2}}$$

and the cf forms a unique 'Fourier transform' pair with the probability distribution function. This result implies that the pdf of $\bar{X}$ can be approximated by a Gaussian distribution with mean μ and variance σ^2/N as N becomes large. This is known as the *Central Limit Theorem*. The following theorem extends this result to the case of independent random variables with different means and variances.

Theorem 1.5.2 *Rao [1973] Lyapunov Theorem:*
Let $X_n, n = 1, 2, \ldots, N$ be a sequence of independent random variables, such that $E[X_n] = \mu_n$ and $\mathsf{Var}[X_n] = \sigma_n^2 \neq 0$ and $\mathsf{E}|X_n - \mu_n|^3 = \beta_n$ exist for each n. Furthermore, let

$$B_N = \left(\sum_{n=1}^{N} \beta_n\right)^{\frac{1}{3}}, \qquad C_N = \left(\sum_{n=1}^{N} \sigma_n^2\right)^{\frac{1}{2}}.$$

Then if $\lim(B_N/C_N) = 0$ as $N \to \infty$, the distribution function of

$$Y = \frac{\sum_{n=1}^{N}(X_n - \mu_n)}{C_N}$$

tends to that of the normal Gaussian distribution.

The proof of Theorem 1.5.2 is similar to that of Theorem 1.5.1.

Bibliography

A. Leon-Garcia. *Probability and Random Processes for Electrical Engineering*. Addison-Wesley, Inc., second edition, 1994.

A. Papoulis. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, Inc., third edition, 1991.

P. Z. Peebles, Jr. *Probability, Random Variables and Random Signal Principles*. McGraw-Hill, Inc., fourth edition, 2001.

C. R. Rao. *Linear Statistical Inference and Its Applications*. John Wiley & Sons, Inc., second edition, 1973.

R.D. Yates and D.J. Goodman. *Probability and Stochastic Processes*. John Wiley & Sons, Inc., 1999.

2 Digital Processing of Continuous-Time Signals

Discrete-time signals arise in many ways such as the monthly rainfall or the weekly DAX stock index. However, most commonly discrete-time signals occur in representations of continuous-time signals. This is partly due to the fact that the processing of continuous-time signals is often carried out on the discrete-time sequences obtained by sampling the continuous-time signal. It is remarkable that under reasonable constraints a continuous-time signal can be adequately represented by a sequence. In this chapter we discuss the process of periodic sampling.

2.1 Periodic Sampling

The process of obtaining a sequence of of samples from a continuous-time signal is referred to as sampling. If the sampling is carried out at periodic intervals the process is denoted *periodic sampling*.
Let $x_c(t)$ be a continuous-time signal and its Fourier transform be denoted by $X_c(j\Omega)$

$$X_c(j\Omega) = \int_{-\infty}^{+\infty} x_c(t) e^{-j\Omega t} dt\,.$$

The signal $x_c(t)$ is obtained from $X_c(j\Omega)$ by

$$x_c(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X_c(j\Omega) e^{+j\Omega t} d\Omega\,.$$

The typical method of obtaining a discrete-time representation of a continuous-time signal is through periodic sampling, wherein a sequence of samples $x(n)$ is obtained from the continuous-time signal $x_c(t)$ according to

$$x(n) = x_c(t)|_{t=nT_s} = x_c(nT_s), \qquad n = 0, \pm 1, \pm 2, \ldots, \qquad T_s > 0.$$

Herein, T_s is the sampling period in seconds, and its reciprocal, $f_s = 1/T_s$, is the sampling frequency, in samples per second and $\Omega_s = 2\pi f_s$ is the sampling frequency in rad/sec. Intuitevly, you can imagine the continuous-time signal $x_c(t)$ passing through a switch that opens and closes at sampling times nT. When the switch closes we have a value $x_c(nT)$ at the output, when the switch is open the output is equal to zero. Thus, the total sequence at the output is given at discrete times and can be represented by $x(n)$ as in the equation above. This equation is the input-output relationship of an ideal A/D converter. It is generally not invertible since many continuous-time signals can produce the same output sequence of samples. Fortunately, this ambiguity can be removed by restricting the class of input signals to the sampler.

Example 2.1.1 *Real life examples of periodically sampled signals can be seen in Figures 2.1.1 and 2.1.1:*

- *Audio signal sampled periodically with sampling frequency of $f_s = 44.1$ kHz, which corresponds to a sampling period of $T_s = 22.7\mu s$.*
- *Electrocardiogramm (ECG) signal sampled with sampling frequency 256 Hz , which corresponds to a sampling period of $T_s = 3.9$ ms.*

It is convenient to represent the sampling process in two stages, as shown in Figure 2.3. This consists of an impulse train modulator which is followed by conversion of the impulse train into a discrete sequence. The system G in Figure 2.3 represents a system that converts an impulse train into a discrete-time sequence. The modulation signal $p(t)$ is a periodic impulse train,

$$p(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_s),$$

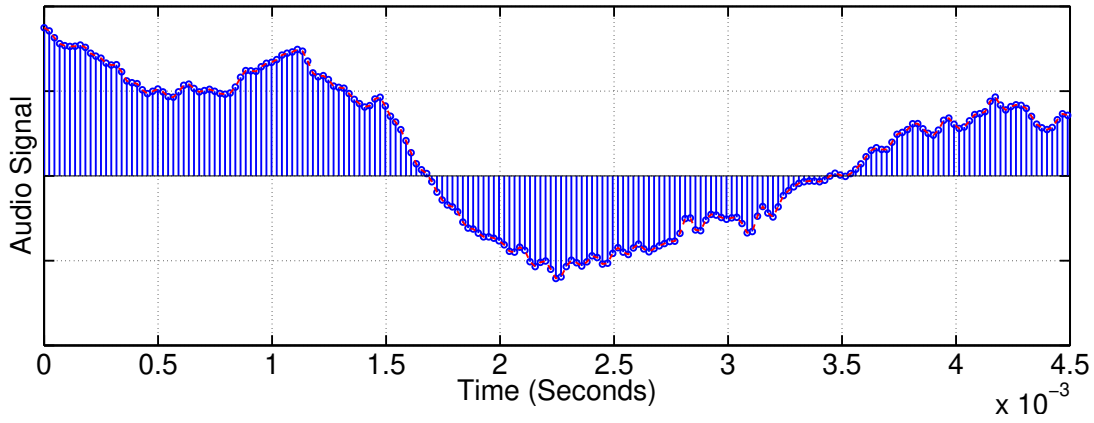


Figure 2.1: Audio signal sampled with a sampling frequency of 44.1 Hz.

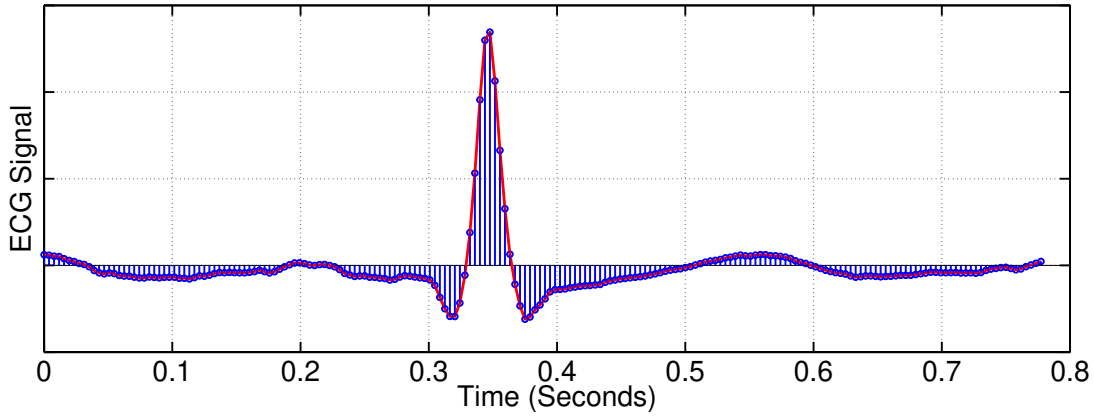


Figure 2.2: ECG signal sampled with sampling frequency 256 Hz.

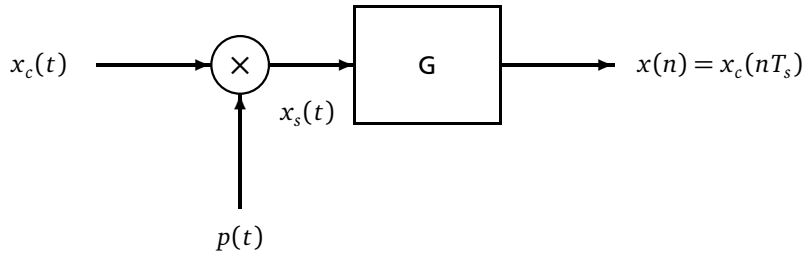


Figure 2.3: Continuous-Time to digital converter.

where $\delta(t)$ is called Dirac's delta function. Consequently, we have

$$x_s(t) = x_c(t) \cdot p(t) = x_c(t) \cdot \sum_{n=-\infty}^{\infty} \delta(t - nT_s) = \sum_{n=-\infty}^{\infty} x_c(nT_s)\delta(t - nT_s) = \sum_{n=-\infty}^{\infty} x(n)\delta(t - nT_s).$$

The Fourier transform of $x_s(t), X_s(j\Omega)$, is obtained by convolving $X_c(j\Omega)$ and $P(j\Omega)$. The Fourier transform of a periodic impulse train is a periodic impulse train, i.e.,

$$P(j\Omega) = \frac{2\pi}{T_s} \sum_{k=-\infty}^{\infty} \delta(\Omega - \frac{2\pi k}{T_s}) = \frac{2\pi}{T_s} \sum_{k=-\infty}^{\infty} \delta(\Omega - k\Omega_s).$$

Since

$$X_s(j\Omega) = \frac{1}{2\pi} X_c(j\Omega) * P(j\Omega),$$

it follows that

$$X_s(j\Omega) = \frac{1}{T_s}\sum_{k=-\infty}^{\infty} X_c\left(j\left(\Omega - \frac{2\pi k}{T_s}\right)\right) = \frac{1}{T_s}\sum_{k=-\infty}^{\infty} X_c\left(j\Omega - kj\Omega_s\right).$$

Figure 2.4 depicts the frequency domain representation of impulse train sampling. Figure 2.4(a) represents the Fourier transform of a band-limited signal, where the highest non-zero frequency component in $X_c(j\Omega)$ is at Ω_B. Figure 2.4(b) shows the periodic impulse train $P(j\Omega)$, Figure 2.4(c) shows $X_s(j\Omega)$, which is the result of convolving $X_c(j\Omega)$ with $P(j\Omega)$. Here ideal sampling is seen, in the frequency domain, to produce replications of the band-limited $X_c(j\Omega)$ at integer multiples of the sampling frequency Ω_s, thus $X_s(j\Omega)$ is periodic with period Ω_s.

From Figure 2.4(c) it is evident that if

$$\Omega_s - \Omega_B > \Omega_B\,, \text{ or } \Omega_s > 2\Omega_B$$

the replica of $X_c(j\Omega)$ do not overlap, and therefore $x_c(t)$ can be recovered from $x_s(t)$ with an ideal low pass filter. However, in Figure 2.4(d) the condition $\Omega_s > 2\Omega_B$ does not hold and the replicas of $X_c(j\Omega)$ overlap. As a result the original signal cannot be recovered by ideal low pass filtering. The reconstructed signal is related to the original continuous-time signal through a distortion referred to as *aliasing*.

If $x_c(t)$ is a low-pass limited signal with no components above Ω_B rad/sec, then the sampling frequency has to be equal to or greater than $2\Omega_B$ rad/sec. This is known as the Nyquist criterion and $2\Omega_B$ is known as the Nyquist frequency.

2.1.1 Non-Ideal Sampling

In practical situations, sampling cannot be performed using Dirac's delta trains since they are not physically realisable. Instead it is necessary to use a signal which closely approximates Dirac's delta train. In this case we express $p(t)$ as

$$p(t) = \sum_{n=-\infty}^{\infty} q(t - nT_s) \tag{2.1}$$

where $q(t) = 0$ for $|t| > \frac{\tau}{2}$. We see that the Fourier transform $P(j\Omega)$ of $p(t)$ can be written as

$$P(j\Omega) = \Omega_s \sum_{n=-\infty}^{\infty} Q(jn\Omega_s)\delta(\Omega - n\Omega_s)$$

where $Q(j\Omega) = \mathcal{F}\{q(t)\}$. We then use the convolution theorem of the Fourier transform to obtain

$$\begin{aligned} X_s(j\Omega) &= \frac{\Omega_s}{2\pi} X_c(j\Omega) * \sum_{n=-\infty}^{\infty} Q(jn\Omega_s)\delta(\Omega - n\Omega_s) \\ X_s(j\Omega) &= \frac{1}{T_s}\sum_{n=-\infty}^{\infty} Q(jn\Omega_s)[X_c(j\Omega) * \delta(\Omega - n\Omega_s)] \\ &= \frac{1}{T_s}\sum_{n=-\infty}^{\infty} Q(jn\Omega_s) X_c(j\Omega - jn\Omega_s) \end{aligned} \tag{2.2}$$

For the case of an ideal A/D converter we have $q(t) = \delta(t)$ so that $Q(j\Omega) = 1$.

Example 2.1.2 (Sampling) *In this example we consider an approximation to the ideal sampler which uses Dirac's delta train. The simplest approximation uses the rectangular pulse train shown in Figure 2.5. We wish to find the Fourier transform of the signal $x_s(t)$ when $p(t)$ is a rectangular pulse train.*

We can express the rectangular pulse train $p(t)$ as shown in (2.1) with

$$q(t) = \frac{1}{\tau}\text{rect}\left(\frac{t}{\tau}\right), \qquad 0 < \tau < \frac{T_s}{2}.$$

The Fourier transform tables can be used to obtain,

$$q(t) = \frac{1}{\tau}\text{rect}\left(\frac{t}{\tau}\right) \longleftrightarrow Q(j\Omega) = \text{sinc}\left(\frac{\Omega\tau}{2\pi}\right),$$

where $\text{sinc}(x) = \frac{\sin \pi x}{\pi x}$. *We can then find* $X_s(j\Omega)$ *using (2.2),*

$$X_s(j\Omega) = \frac{1}{T_s} \sum_{n=-\infty}^{\infty} \text{sinc}\left(\frac{n\Omega_s \tau}{2\pi}\right) X_c(j\Omega - jn\Omega_s), \qquad \Omega_s = \frac{2\pi}{T_s}.$$

Thus $X_s(j\Omega)$ *is composed of periodic replications of* $X_c(j\Omega)$ *where the nth period is scaled by the value of* $\text{sinc}\left(\frac{\Omega\tau}{2\pi}\right)$ *at* $\Omega = n\Omega_s$. *For small values of* τ *the rectangular functions of Figure 2.5 will closely approximate delta functions. This is reflected in the expression for* $X_s(j\Omega)$ *since for small values of* τ *the* sinc *function has large bandwidth and* $X_s(j\Omega)$ *approximates the ideal case. More formally, it can be seen that,*

$$\lim_{\tau \to 0} q(t) = \delta(t)$$

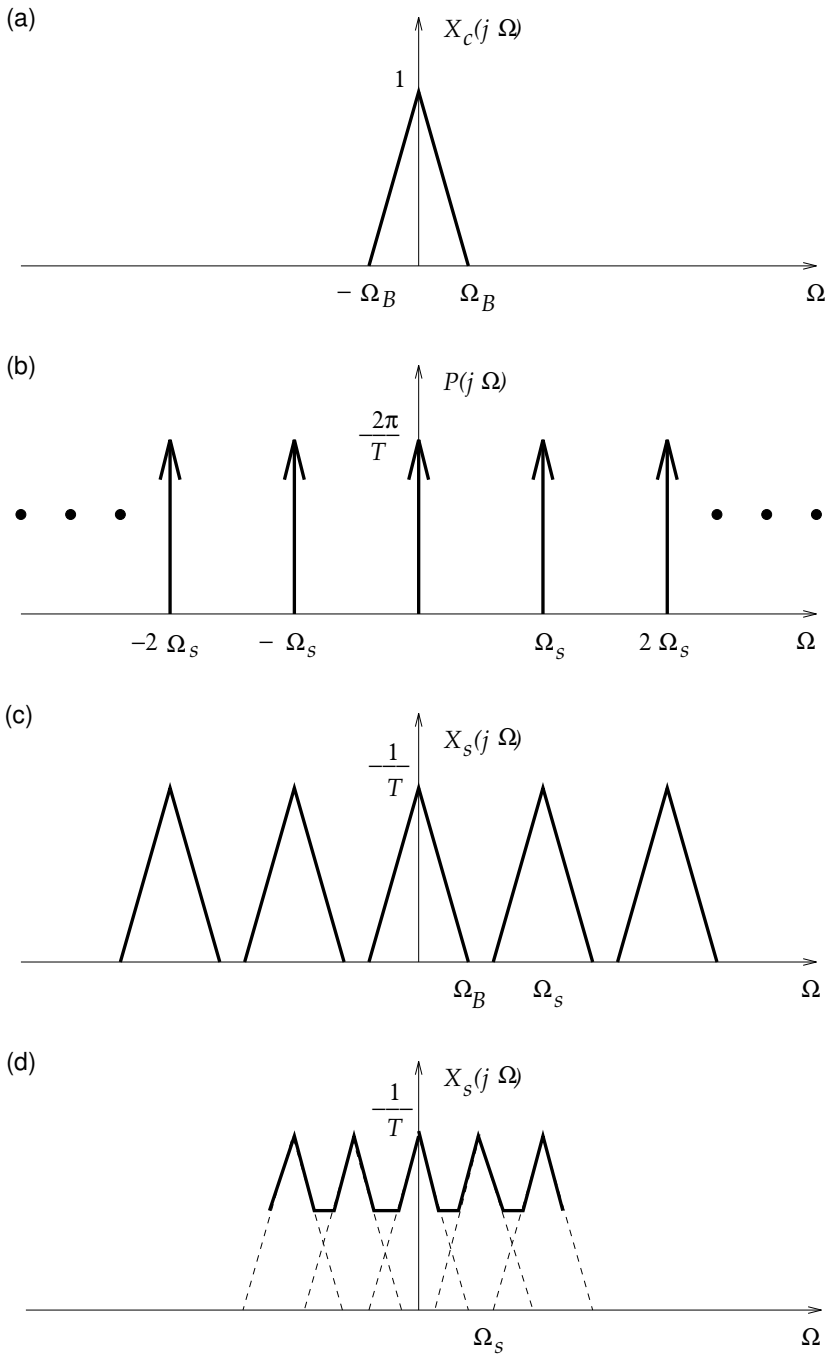


Figure 2.4: Effect in the frequency domain of sampling in the time domain: (a) Spectrum of the continuous-time signal, (b) Spectrum of the sampling function, (c) Spectrum of the sampled signal with $\Omega_s > 2\Omega_B$, (d) Spectrum of the sampled signal with $\Omega_c < 2\Omega_B$.

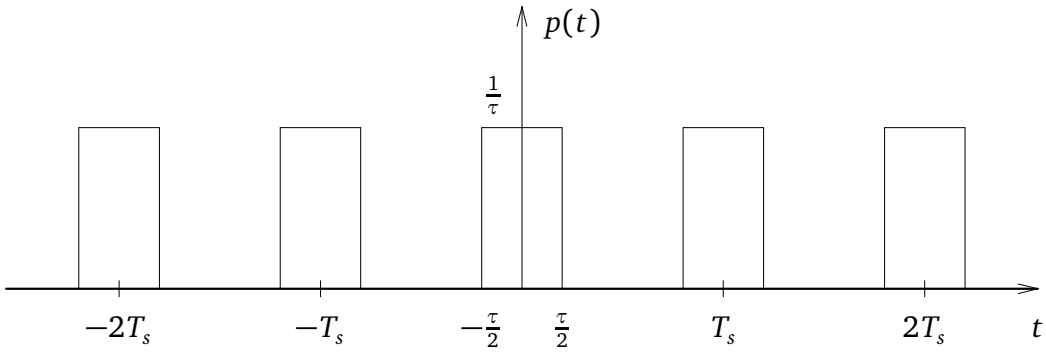


Figure 2.5: Rectangular pulse train.

so that as τ approaches zero, sampling using the rectangular pulse train becomes the same as sampling using Dirac's delta train.

For the remainder of this chapter we shall assume that ideal sampling has taken place.

2.2 Reconstruction of Band-Limited Signals

Seeing Figure 2.4, it is evident that if we are in pocession of the signal $x_s(t)$ or respectively its frequency response $X_s(j\Omega)$ and want to reconstruct the continuous-time signal $x_c(t)$ or its frequency response $X_c(j\Omega)$ we need only to pass it through a low-pass filter that passes only those frequencies contained in $x_c(t)$. If $x(n)$ is the input to an ideal low pass filter with frequency response $H_r(j\Omega)$ and impulse response $h_r(t)$, then the output of the filter will be

$$x_r(t) = \sum_{n=-\infty}^{\infty} x(n)h_r(t-nT_s).$$

where T_s is the sampling interval. The reconstruction filter commonly has a gain of T_s and a cutoff frequency of $\Omega_c = \Omega_s/2 = \pi/T_s$. This choice is appropriate for any relationship between Ω_s and Ω_B that avoids aliasing, i.e., so long as $\Omega_s > 2\Omega_B$. The impulse response of such a filter is given by

$$h_r(t) = \frac{\sin(\pi t/T_s)}{\pi t/T_s}.$$

Consequently, the relation between $x_r(t)$ and $x(n)$ is given by

$$x_r(t) = \sum_{n=-\infty}^{\infty} x(n)\frac{\sin[\pi(t-nT_s)/T_s]}{\pi(t-nT_s)/T_s},$$

with $x_c(nT) = x(n)$.

It is important to note that the continuous-time signal can only be reconstructed using a low-pass filter if the Nyquist criterion is satisfied. This is clearly illustrated in Figure 2.4(c) and (d) which show the effect on $X_s(j\Omega)$ of sampling above and below the Nyquist criterion. Since, in (d) we cannot separate $X_c(j\Omega)$ it is not possible to reconstruct the continuous-time signal using a low-pass filter.

The fact that the sampling frequency must exceed a certain value in order for the continuous-time signal to be reconstructed is intuitively satisfying. Loosely speaking, it means that if we don't take enough samples from the original signal we will not have enough information to reconstruct the original signal from the samples.

2.3 Discrete-Time Processing of Continuous-Time Signals.

A continuous-time signal is processed by digital signal processing techniques using analog-to-digital (A/D) and digital-to-analog (D/A) converters before and after processing. This concept is illustrated in Figure 2.6. The low pass filter limits the bandwidth of the continuous-time signal to reduce the effect of aliasing. The spectral characteristics (magnitude) of the low pass filter is shown in Figure 2.7.

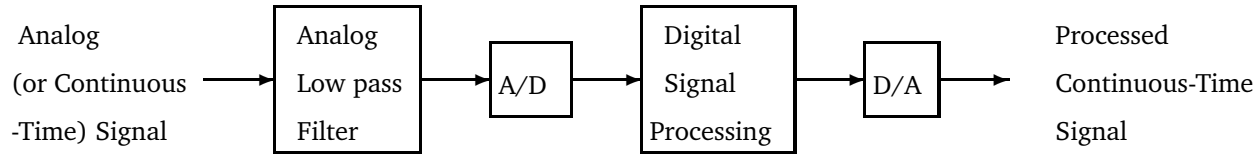


Figure 2.6: Discrete-time processing of continuous-time signals.

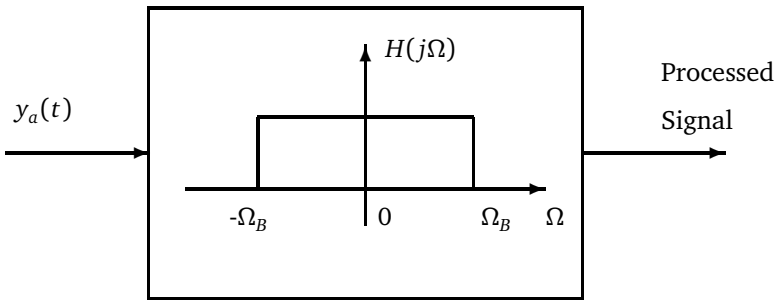


Figure 2.7: Continuous-Time Filter.

Example 2.3.1 *Assume that we want to restore the signal $s_a(t)$ from $y_a(t) = s_a(t) + n_a(t)$, where $n_a(t)$ is a background noise. The spectral characteristics of these signals are shown in Figure 2.8.*

We can solve this problem using either continuous-time or digital techniques. Using a continuous-time approach, we would filter the signal $s_a(t)$ using a continuous-time low-pass filter. Alternatively, we could adopt a digital approach, as shown in Figure 2.6. We would first low-pass filter $y_a(t)$ to prevent aliasing, and then convert the continuous-time signal into a discrete-time signal using an A/D converter. Any processing is then performed digitally on the discrete-time signal. The processed signal is then converted back to continuous form using an D/A converter.

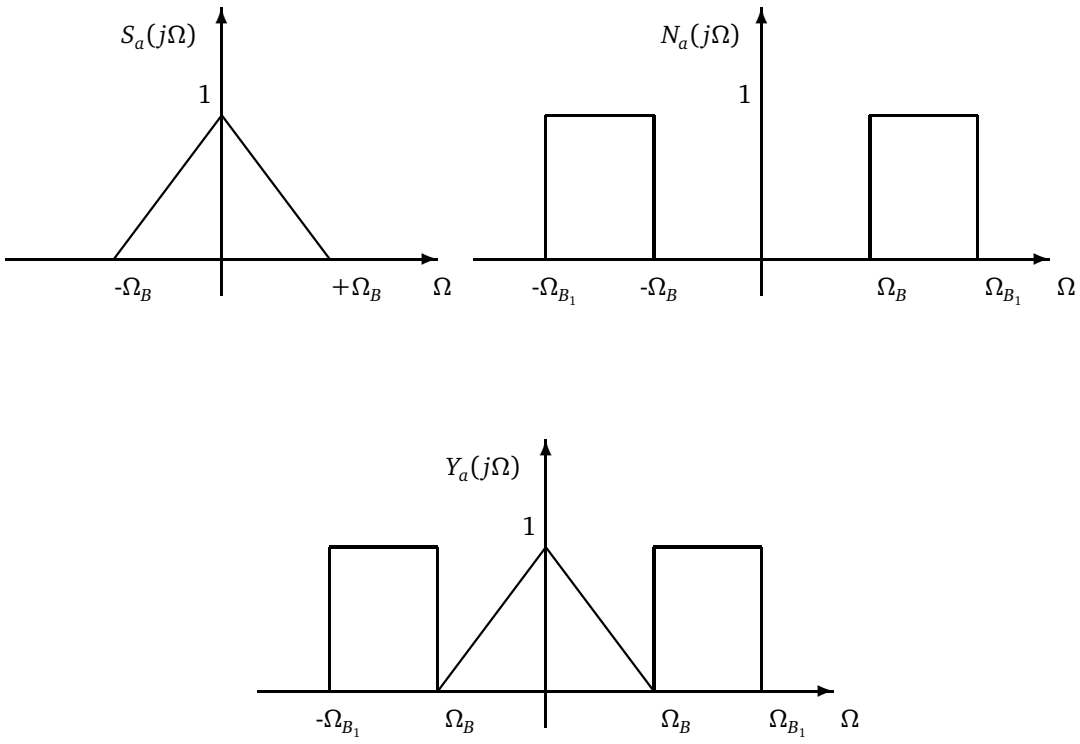


Figure 2.8: Frequency spectra.

The advantages of the digital approach compared with the continuous-time approach are:

1. Flexibility, if we want to change the continuous-time filter because of change in signal and noise characteristics, we would have to change the hardware components. Using the digital approach, we only need to modify

the software.

2. Better control of accuracy requirements. Tolerances in continuous-time circuit components make it extremely difficult for the system designer to control the accuracy of the system.
3. The signals can be stored without deterioration or loss of signal quality.
4. Lower cost of the digital implementation.

2.4 Band-Pass Limited Signals

We consider now a special case of low-pass limited sampling. We characterise a band-pass limited signal $x_c(t)$ by the spectrum $X_c(j\Omega)$ shown in Figure 2.9 where $\Omega_B > \Omega_{-B}$.

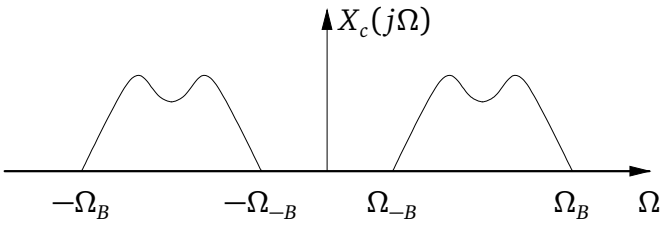


Figure 2.9: Spectrum of a band-pass limited signal $x_c(t)$.

It can be seen from Figure 2.9 that the signal is no longer limited to the spectral band $|\Omega| \leq \Omega_B$ but now satisfies

$$X_c(j\Omega) = 0 \text{ for } |\Omega| < \Omega_{-B} \text{ and } |\Omega| > \Omega_B$$

Next we consider sampling the signal with a sampling frequency $\Omega_s = \frac{2\pi}{T_s}$. If we choose $\Omega_s > 2\Omega_B$ as recommended then the sampled signal has the spectrum shown in Figure 2.10.

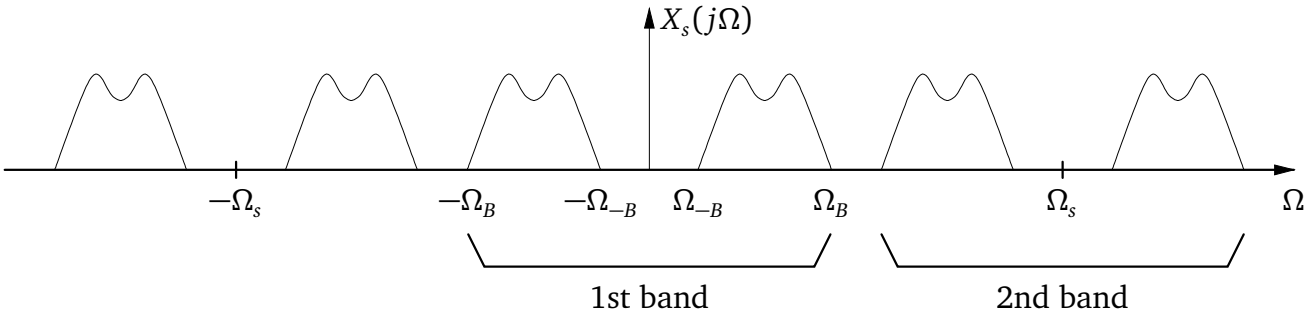


Figure 2.10: Spectrum of a sampled band-pass limited signal $x_s(t)$.

Figure 2.10 shows that if the sampling frequency Ω_s is greater than the Nyquist frequency $2\Omega_B$ there is no spectral overlap. However the band-pass nature of the signal $x_c(t)$ allows us to sample at a frequency below the Nyquist frequency and still avoid spectral overlap. In fact, for band-pass limited signals the requirement to avoid aliasing may be a sampling frequency $\Omega_s < \Omega_B$. The example below considers this issue.

Example 2.4.1 (Spectrum of a band-pass limited signal) *A continuous-time signal $x_c(t)$ is sampled with sampling period T_s. The spectrum of $x_c(t)$ is shown in Figure 2.11.*

We wish to obtain the Fourier transform of $x_s(t)$ for the case where the sampling period is given by $T_s = \frac{2\pi}{\Omega_0}$. Note that the sampling frequency $\Omega_s = \frac{2\pi}{T_s} = \Omega_0$ is less than the Nyquist frequency $2\Omega_0$. The intermediate signal $x_s(t)$ is given by

$$x_s(t) = x_c(t) \cdot \sum_{n=-\infty}^{\infty} \delta(t - nT_s)$$

We obtain the Fourier transform $X_s(j\Omega)$ of $x_s(t)$ as

$$X_s(j\Omega) = \frac{1}{T_s} \sum_{n=-\infty}^{\infty} X_c\left(j\Omega - j\frac{2\pi n}{T_s}\right)$$

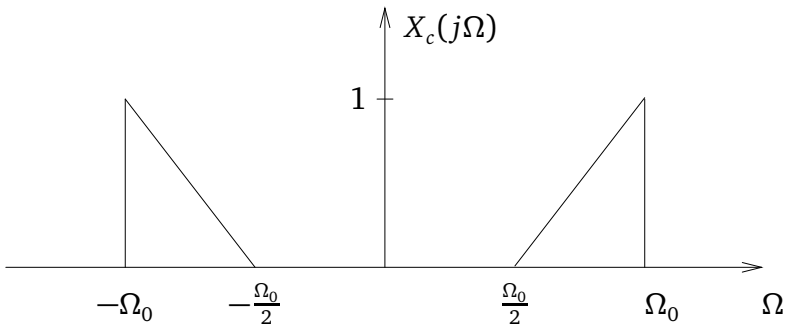


Figure 2.11: Spectrum of continuous-time signal.

In this case we have $T_s = \frac{2\pi}{\Omega_0}$ *so that*

$$X_s(j\Omega) = \frac{\Omega_0}{2\pi} \sum_{n=-\infty}^{\infty} X_c(j\Omega - jn\Omega_0)$$

This spectrum is shown in Figure 2.12. We see that even though $\Omega_s < 2\Omega_0$ *there is no spectral overlap. This can be attributed to the fact that the continuous-time signal* $x_c(t)$ *is a band-pass limited signal.*

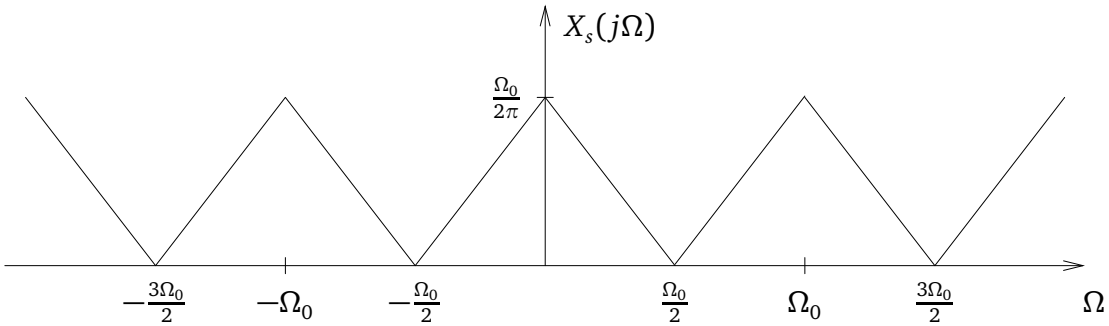


Figure 2.12: Spectrum of the sampled signal.

3 Stochastic Processes (Random Processes)

Random processes naturally arise in many engineering applications. For example, the voltage waveform of a noise source or the bit stream of a binary communications signal are continuous-time random processes. The monthly rainfall values or the daily stock market index are examples of discrete-time random processes. However, random sequences are most commonly encountered when sampling continuous-time processes, as discussed in the preceding chapter. The remainder of the manuscript focuses solely on discrete-time random processes. However the results presented herein apply in a similar fashion to the continuous- time case. Using probabilistic methods, we want to find a model for such processes. The aim is not necessarily to describe an observed signal but to describe the class of possible outcomes of a system under the same conditions.

Definition 3.0.1 *A real-valued random process is an indexed set of real functions of time that has certain statistical properties. We characterize a real-valued random process by a set of real functions and associated probability description.*

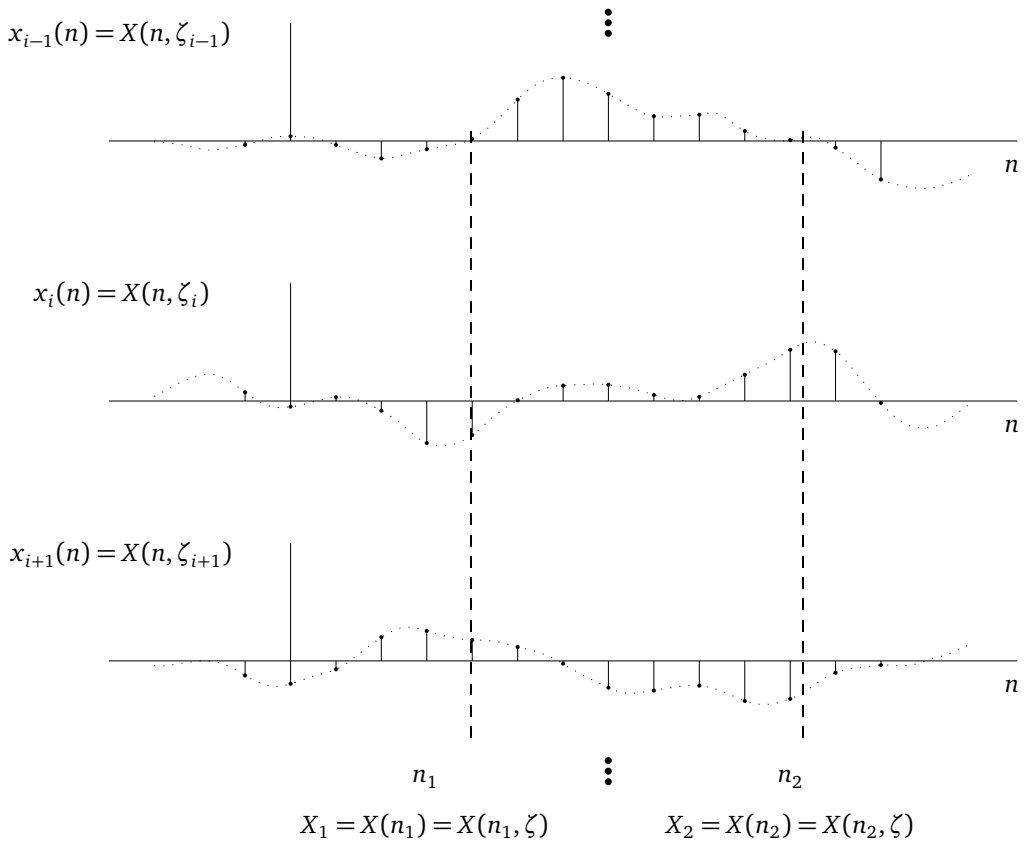


Figure 3.1: Voltage waveforms emitted from a noise source.

In general, $x_i(n) = X(n, \zeta_i)$, $n \in \mathbb{Z}$, $\zeta_i \in \mathcal{S}$, denotes the waveform that is obtained when the event ζ_i of the process occurs. $x_i(n)$ is called a sample function of the process (see Figure 3.1). The set of all possible sample functions $\{X(n, \zeta_i)\}$ is called the ensemble and defines the random process $X(n)$. $X(n)$ can describe, for instance, a noise source. A random process can be seen as a function of two variables, time n and elementary event ζ. When n is fixed, $X(n, \zeta)$ is simply a random variable. When ζ is fixed, $X(n, \zeta) = x(n)$ is a function of time known as "sample function" or realization. If the noise source has a Gaussian distribution, then any of the random variables would be

described by

$$f_{X_i}(x_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \cdot \mathrm{e}^{-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}}$$

where $X_i = X(n_i, \zeta)$. Also, all joint distributions are multivariate Gaussian. At each point in time, the value of the waveform is a random variable described by its probability density function $f_X(x;n)$ at that time n.

The probability that the process $X(n)$ will have a value in the range $[a, b]$ at n is given by

$$P(\{a \le X(n) \le b\}) = \int_a^b f_X(x;n)dx$$

The moments of a random process $X(n)$ describe the nature of its distribution. They are essential in determining some of the processes main stochastic characteristics.

The first moments of $X(n)$ are

$$\begin{aligned}
\mu_1(n) &= \mathsf{E}[X(n)] &&= \int_{-\infty}^{\infty} x\, f_X(x;n)dx \\
\mu_2(n) &= \mathsf{E}\left[X(n)^2\right] &&= \int_{-\infty}^{\infty} x^2 f_X(x;n)dx \\
&\vdots &&\vdots \\
\mu_l(n) &= \mathsf{E}\left[X(n)^l\right] &&= \int_{-\infty}^{\infty} x^l f_X(x;n)dx,
\end{aligned}$$

where $\mu_1(n)$ is the mean of $X(n)$, and $\mu_2(n) - \mu_1(n)^2$ is the variance of $X(n)$. Because in general, $f_X(x;n)$ depends on n, the moments will depend on time.
The probability that a process $X(n)$ lies in the interval $[a, b]$ at time n_1 and in the interval $[c, d]$ at time n_2 is given by the joint density function $f_{X_1X_2}(x_1, x_2; n_1, n_2)$,

$$\mathsf{P}\left(\{a \le X(n_1) \le b,\ c \le X(n_2) \le d\}\right) = \int_a^b \int_c^d f_{X_1X_2}(x_1, x_2; n_1, n_2)dx_1dx_2.$$

A complete description of a random process requires knowledge of all such joint densities for all possible point locations and orders $f_{X_1X_2\ldots X_N}(x_1, \ldots, x_N; n_1, \ldots, n_N)$.

- A continuous random process consists of a random process with associated continuously distributed random variables, $X(n), n \in \mathbb{Z}$. For example, noise in communication circuits at a given time; the voltage measured can have any value.
- A discrete random process consists of a random process with associated discrete random variables, e.g. the output of an ideal limiter.

3.1 Stationarity and Ergodicity

Stationarity

As stated in the previous section the moments of a random process are typically time dependent, because $f_X(x;n)$ depends on n. Stationarity deals specifically with this dependency, i.e. the dependency of a statistical characterization of a random process on the observation time. Assume that a random process is statistically characterized at different time intervals. If the statistical properties are independent of the observation time the process is stationary, otherwise it is non stationary.

Definition 3.1.1 *A random process $X(n)$, $n \in \mathbb{Z}$ is said to be stationary to the order N if, for any n_1, n_2, … n_N, and $x_1, x_2, \ldots, x_N$, and any n_0*

$$
\begin{aligned}
F_X(x_1,\ldots,x_N;n_1,\ldots,n_N) &= \mathsf{P}(\{X(n_1)\le x_1, X(n_2)\le x_2,\ldots,X(n_N)\le x_N\}) \\
&= \mathsf{P}(\{X(n_1+n_0)\le x_1, X_2(n_2+n_0)\le x_2,\ldots,X(n_N+n_0)\le x_N\}) \\
&= F_X(x_1,\ldots,x_N;n_1+n_0,\ldots,n_N+n_0)
\end{aligned}
$$

holds or equivalently:

$$
f_X(x_1,x_2,\ldots,x_N;n_1,\ldots,n_N) = f_X(x_1,x_2,\ldots,x_N;n_1+n_0,\ldots,n_N+n_0).
$$

The process is said to be strictly stationary if it is stationary to the infinite order. If a process is stationary, it can be translated in time without changing its statistical description.

Ergodicity

Ergodicity is a topic dealing with the relationship between statistical averages and time averages. Suppose that, for example, we wish to determine the mean $\mu_1(n)$ of a process $X(n)$. For this purpose, we observe a large number of samples $X(n,\zeta_i)$ and we use their ensemble average as the estimate of $\mu_1(n)=\mathsf{E}[X(n)]$

$$
\hat{\mu}_1(n) = \frac{1}{L}\sum_{i=1}^{L} X(n,\zeta_i)
$$

Suppose, however, that we have access only to a single sample $x(n)=X(n,\zeta)$ of $X(n)$ for $-N\le n\le N$. Can we use then the time average

$$
\overline{X(n)} = \lim_{N\to\infty}\frac{1}{2N+1}\sum_{n=-N}^{N} X(n,\zeta)
$$

as the estimate of $\mu_1(n)$? This is, of course, impossible if $\mu_1(n)$ depends on n. However, if $\mu_1(n)=\mu_1$ is a constant, then under the general conditions, $\overline{X(n)}$ approaches μ_1.

Definition 3.1.2 *A random process is said to be ergodic if all time averages of any sample function are equal to the corresponding ensemble averages (expectation).*

Example 3.1.1 *dc and rms values are defined in terms of time averages, but if the process is ergodic, they may be evaluated by use of ensemble averages. The dc value of $X(n)$ is:*

$$
X_{dc} = \mathsf{E}[X(n)]
$$

$$
X_{rms} = \sqrt{\mathsf{E}[X(n)^2]}.
$$

If $X(n)$ is ergodic, the time average is equal to the ensemble average, that is

$$
X_{dc} = \lim_{N\to\infty}\frac{1}{2N+1}\sum_{n=-N}^{N} X(n) = \mathsf{E}[X(n)] = \int_{-\infty}^{\infty} x f_X(x;n)dx = \mu_X
$$

Similarly, we have for the rms value:

$$
X_{rms} = \sqrt{\lim_{N\to\infty}\frac{1}{2N+1}\sum_{n=-N}^{N} X(n)^2} = \sqrt{\mathsf{E}[X(n)^2]} = \sqrt{\sigma_X^2+\mu_X^2} = \int_{-\infty}^{\infty} x^2 f_X(x;n)dx
$$

where σ_X^2 is the variance of $X(n)$.

3.1.1 Ergodicity and Stationarity

If a process is ergodic, all time and ensemble averages are interchangeable. The time average is by definition independent of time. As it equals the ensemble averages (such as moments), the latter are also independent of time. Thus, an ergodic process must be stationary. But not all stationary processes are ergodic.

Example 3.1.2 *An Ergodic Random Process*

Let the random process be given by: $X(n) = A\cos(\omega_0 n + \Phi)$*, where* A *and* ω_0 *are constants and* Φ *is a random variable that is uniformly distributed over* $[0, 2\pi)$*. First we evaluate some ensemble averages (expectation).*

$$\begin{aligned} \mathsf{E}[X(n)] &= \int_{-\infty}^{\infty} x(\varphi) f_{\Phi}(\varphi)\, d\varphi \\ &= \int_{0}^{2\pi} A\cos(\omega_0 n + \varphi)\, \frac{1}{2\pi}\, d\varphi \\ &= 0. \end{aligned}$$

Furthermore,

$$\begin{aligned} \mathsf{E}\left[X(n)^2\right] &= \int_{0}^{2\pi} A^2 \cos(\omega_0 n + \varphi)^2\, \frac{1}{2\pi}\, d\varphi \\ &= \frac{A^2}{2} \end{aligned}$$

In this example, the time parameter n disappeared when the ensemble averages were evaluated. The process is stationary up to the second order.

Now, evaluate the corresponding time averages using a sample function of the random process. $X(n, \zeta_1) = A\cos\omega_0 n$ that occurs when $\Phi = 0$. The zero phase $\Phi = 0$ corresponds to one of the events. The time average for any of the sample functions can be calculated by letting Φ be any value between 0 and 2π. The time averages are:

- First moment: $\overline{X(n)} = \lim\limits_{N\to\infty} \dfrac{1}{2N+1} \sum\limits_{n=-N}^{N} A\cos(\omega_0 n + \varphi) = 0$
- Second moment: $\overline{X(n)^2} = \lim\limits_{N\to\infty} \dfrac{1}{2N+1} \sum\limits_{n=-N}^{N} [A\cos(\omega_0 n + \varphi)]^2 = \dfrac{A^2}{2}$

In this example, φ disappears when the time average is calculated. We see that the time average is equal to the ensemble average for the first and second moments.

We have not yet proved that the process is *ergodic* because we have to evaluate all the orders of moments and averages. In general, it is difficult to prove that a process is ergodic. In this manuscript we will focus on stationary processes.

Example 3.1.3 *Suppose that the amplitude in the previous example is now random. Let us assume that* A *and* Φ *are independent. In this case, the process is stationary but not ergodic because if* $\mathsf{E}[A^2] = \sigma^2$*, then*

$$\mathsf{E}[X(n)^2] = \mathsf{E}[A^2] \cdot \mathsf{E}[\cos(\omega_0 n + \Phi)^2] = \frac{\sigma^2}{2}$$

but

$$\lim_{T\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} X(n)^2 = \frac{1}{2}A^2$$

Summary:

- An ergodic process has to be stationary.
- However, if a process is stationary, it may or may not be ergodic.

3.2 Complex Random Processes

Definition 3.2.1 *A complex random process is* $X(n) \triangleq Y(n) + j\,Z(n)$ *where* $Y(n)$ *and* $Z(n)$ *are real-valued random processes.*

A complex process is stationary if $Y(n)$ and $Z(n)$ are jointly stationary, i.e.

$$f_X(y_1,\ldots,y_N,z_1,\ldots,z_N;n_1,\ldots,n_{2N}) = f_X(y_1,\ldots,y_N,z_1,\ldots,z_N;n_1+n_0,\ldots,n_{2N}+n_0)$$

for all $y_1,\ldots,y_N$, $z_1,\ldots,z_N$ and all $n_1,\ldots,n_{2N}$, n_0 and N. The mean of a complex random process is defined as:

$$\mathsf{E}[X(n)] = \mathsf{E}[Y(n)] + j\,\mathsf{E}[Z(n)] = \mu_Y + j\,\mu_Z$$

3.3 Second-Order Moment Functions and Wide-Sense Stationarity

3.3.1 Second-Order Moment Function (SOMF)

The second-order moment function can be visualized as a measure of similarity between a random process and a time-shifted version of it for different time shifts. The higher the value of the SOMF at a certain time-shift, the larger the similarity between the process and its time-shifted version.

Definition 3.3.1 *The second-order moment function (SOMF) for a complex random process* $X(n)$ *is defined as:*

$$\begin{aligned} r_{XX}(n_1,n_2) &= \mathsf{E}[X(n_1)\cdot X(n_2)^*] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2^* f_{X_1X_2^*}(x_1,x_2^*;n_1,n_2)\,dx_1\,dx_2^* \end{aligned}$$

where $X_1 = X(n_1)$ *and* $X_2 = X(n_2)$.

If the process $X(n)$ is stationary to the second order, the SOMF is only a function of the time difference $\kappa = n_2 - n_1$, and

$$r_{XX}(\kappa) = \mathsf{E}[X(n+\kappa)\cdot X(n)^*]$$

Definition 3.3.2 *The SOMF of a real-valued stationary process* $Y(n)$ *is defined as:*

$$r_{YY}(n_1,n_2) = \mathsf{E}[Y(n_1)\cdot Y(n_2)]$$

Definition 3.3.3 *A (complex) random process* $X(n)$ *is said to be* wide-sense stationary *if:*

1. $\mathsf{E}[X(n)]$ *is a (complex) constant.*
2. $r_{XX}(n_1,n_2) = r_{XX}(\kappa)$*, where* $\kappa = n_2 - n_1$

A process that is stationary to order 2 or greater is certainly wide-sense stationary. However, a finite order stationary process is not necessarily stationary.

Properties of the SOMF of a Wide-Sense Stationary Process

1. $r_{XX}(0) = \mathsf{E}[X(n)\cdot X(n)^*] = \sigma_X^2 + |\mu_X|^2$ is the second order moment
   - $r_{XX}(0) = \sigma_X^2 + \mu_X^2$ if $X(n)$ is a real-valued random process
2. $r_{XX}(\kappa)^* = r_{XX}(-\kappa)$
   - $r_{XX}(\kappa) = r_{XX}(-\kappa)$ if $X(n)$ is a real-valued random process
3. $r_{XX}(0) \geq |r_{YY}(\kappa) + r_{ZZ}(\kappa)|$, where $X(n) = Y(n) + j\,Z(n)$
   - $r_{XX}(0) \geq |r_{XX}(\kappa)|$ if $X(n)$ is a real-valued random process

Proofs for 1 and 2 follow from the definition. The third property holds because

$$\begin{aligned} \mathsf{E}[(X(n+\kappa) \pm X(n))\cdot(X(n+\kappa)\pm X(n))^*] &\geq 0 \\ \mathsf{E}[X(n+\kappa)\cdot X(n+\kappa)^*] \pm \mathsf{E}[X(n+\kappa)\cdot X(n)^*] \pm \mathsf{E}[X(n+\kappa)^*\cdot X(n)] + \mathsf{E}[X(n)\cdot X(n)^*] &\geq 0 \\ 2\,r_{XX}(0) \pm (r_{XX}(\kappa) + r_{XX}(\kappa)^*) &\geq 0 \\ r_{XX}(0) \pm (r_{YY}(\kappa) + r_{ZZ}(\kappa)) &\geq 0 \end{aligned}$$

3.3.2 Cross-Second-Order Moment Function

Definition 3.3.4 *The "cross-SOMF" for two complex random processes $X_1(n)$ and $X_2(n)$ is:*

$$r_{X_1X_2}(n_1, n_2) = \mathsf{E}\left[X_1(n_1) \cdot X_2(n_2)^*\right]$$

If the complex random processes are jointly wide-sense stationary, the cross-SOMF becomes:

$$r_{X_1X_2}(n_1, n_2) = r_{X_1X_2}(\kappa) \text{ where } \kappa = n_2 - n_1$$

Note that for two processes $X_1(n)$ and $X_2(n)$ to be jointly wide-sense stationary, they have to be stationary to the second order. This implies that the conditions for wide-sense stationarity of the individual processes has to be examined in addition to the one stated above.

Definition 3.3.5 *The "cross-SOMF" for two real-valued processes $Y(n)$ and $Z(n)$ is:*

$$r_{YZ}(n_1, n_2) = \mathsf{E}[Y(n_1) \cdot Z(n_2)]$$

If $Y(n)$ and $Z(n)$ are jointly stationary, the cross-SOMF becomes:

$$r_{YZ}(n_1, n_2) = r_{YZ}(n_1 - n_2) = r_{YZ}(\kappa)$$

where $\kappa = n_1 - n_2$.

Properties of the Cross-SOMF of Jointly Stationary Real Processes

Consider two real-valued processes $X(n)$ and $Y(n)$. For $X(n)$ and $Y(n)$, the following relations hold.

1. $r_{XY}(-\kappa) = r_{YX}(\kappa)$
2. $|r_{XY}(\kappa)| \leq \sqrt{r_{XX}(0) \cdot r_{YY}(0)}$.
3. $|r_{XY}(\kappa)| \leq \frac{1}{2}\left[r_{XX}(0) + r_{YY}(0)\right]$

Property 2 holds because

$$\mathsf{E}[(X(n+\kappa) - aY(n))^2] \geq 0 \forall a \in \mathbb{R}$$

$$\mathsf{E}[X(n+\kappa)^2] + a^2\mathsf{E}[Y(n)^2] - 2a\mathsf{E}[X(n+\kappa)Y(n)] \geq 0 \ \forall a \in \mathbb{R}$$

$$r_{XX}(0) + a^2 r_{YY}(0) - 2a r_{XY}(\kappa) \geq 0$$

This quadratic form (in a) is valid if the roots of a are not real (except as a double real root). Therefore

$$r_{XY}(\kappa)^2 - r_{XX}(0) r_{YY}(0) \leq 0$$

$$r_{XY}(\kappa)^2 \leq r_{XX}(0) r_{YY}(0) \Rightarrow |r_{XY}(\kappa)| \leq \sqrt{r_{XX}(0) r_{YY}(0)}$$

Sometimes

$$\frac{r_{XY}(\kappa)}{\sqrt{r_{XX}(0) r_{YY}(0)}}$$

is denoted by $\rho_{XY}(\kappa)$.
Property 2 constitutes a tighter bound than that of property 3, because the geometric mean of two positive numbers cannot exceed the arithmetic mean, that is

$$\sqrt{r_{XX}(0) \cdot r_{YY}(0)} \leq \frac{1}{2}\left[r_{XX}(0) + r_{YY}(0)\right]$$

- Two random processes $X(n)$ and $Y(n)$ are said to be uncorrelated if:

$$r_{XY}(\kappa) = \mathsf{E}[X(n+\kappa)] \cdot \mathsf{E}\left[Y(n)\right] = \mu_X \cdot \mu_Y \ \forall \kappa \in \mathbb{Z}$$

- Two random processes $X(n)$ and $Y(n)$ are orthogonal if:

$$r_{XY}(\kappa) = 0, \ \forall \kappa \in \mathbb{Z}$$

- If $X(n) \equiv Y(n)$, the cross-SOMF becomes the SOMF.
- If the random processes $X(n)$ and $Y(n)$ are jointly ergodic, the time average may be used to replace the ensemble average.

  If $X(n)$ and $Y(n)$ are jointly ergodic

$$r_{XY}(\kappa) \stackrel{\triangle}{=} \mathsf{E}[X(n+\kappa) \cdot Y(n)] = \mathsf{E}[X(n) \cdot Y(n-\kappa)] \equiv \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} X(n)Y(n-\kappa)$$

  In this case, cross-SOMFs and auto-SOMFs may be measured by using a system composed of a delay element, a multiplier and an accumulator. This is illustrated in Figure 3.2, where the block $z^{-\kappa}$ is a delay line of order κ.

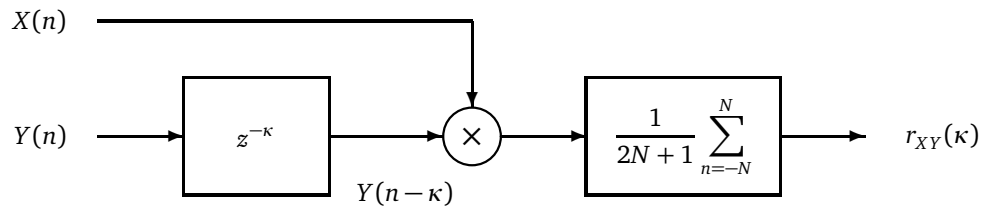


Figure 3.2: Measurement of SOMF.

Example 3.3.1 *Consider the two random processes $S_1(n)$ and $S_2(n)$, given by:*

$$S_1(n) = A\cos(\omega_1 n + \phi_1)$$
$$S_2(n) = B\cos(\omega_2 n + \phi_2)$$

ϕ_1 and ϕ_2 are two random variables that are uniformly distributed over the interval $(0, 2\pi)$. A and B are two normally distributed, zero-mean random variables with variances σ_A^2 and σ_B^2, respectively. Moreover, the random variables A, B, ϕ_1 and ϕ_2 are statistically independent. To assess whether the two random processes are wide sense stationary we first analyze $E\{S_1(n)\}$ and $E\{S_2(n)\}$. Clearly, $E\{S_1(n)\} = E\{S_2(n)\} = 0$. Next, we calculate the second-order moment function of each process:

$$\begin{aligned}
r_{S_1S_1}(n+k,n) &= E\{S_1(n+k)S_1(n)\} \\
&= E\{A^2\cos(\omega_1(n+k)+\phi_1)\cdot\cos(\omega_1 n+\phi_1)\} \\
&= E\{A^2/2[\cos(\omega_1 k)+\cos(\omega_1(2n+k)+2\phi_1)]\} \\
&= \frac{\sigma_A^2}{2}\cos(\omega_1 k) + \frac{\sigma_A^2}{2}\underbrace{E\{\cos(\omega_1(2n+k)+2\phi_1)\}}_{=0} \\
&= \frac{\sigma_A^2}{2}\cos(\omega_1 k) \\
&= \frac{\sigma_B^2}{2}\cos(\omega_2 k)
\end{aligned}$$

The processes $S_1(n)$ and $S_2(n)$ are clearly wide sense stationary as $E\{S_1(n)\} = E\{S_2(n)\} = 0$ and their second-order moment functions are only a function of the time-difference k.
Furthermore, the processes are jointly stationary as the cross-SOMF is not time-dependent:

$$\begin{aligned}
r_{S_1S_2}(n+k,n) &= E\{A\cos(\omega_1(n+k)+\phi_1)\cdot B\cos(\omega_1 n+\phi_1)\} \\
&= \underbrace{E\{A\}}_{=0}\cdot E\{B\}\cdot\cdots = 0.
\end{aligned}$$

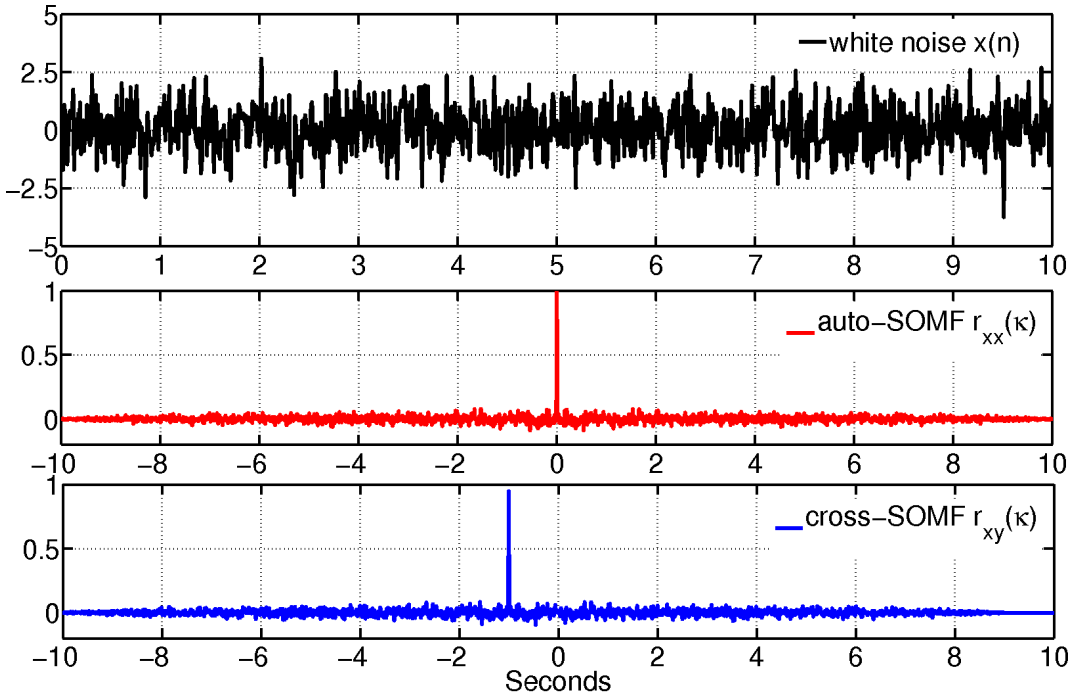


Figure 3.3: Auto- and cross-second-order moment function of white noise.

Example 3.3.2 *Figure 3.3 shows the realization of Gaussian white noise process $x(n)$ and its auto- and cross-second-order moment functions, $r_{XX}(\kappa)$ and $r_{XY}(\kappa)$. Gaussian white noise is characterized by a zero-mean normal distribution with variance σ^2. $y(n)$ is a delayed version of $x(n)$ with a delay of 1s at a sampling rate of $f_s = 100Hz$. The SOMF $r_{XX}(\kappa)$ shows a peak for $\kappa = 0$ and the cross-SOMF shows a peak at the time-shift between $x(n)$ and $y(n)$. This illustrates that the only significant correlation between a white Gaussian noise process and itself, occurs when there is no time-shift between them. The cross-SOMF $r_{XY}(\kappa)$ can easily be used to determine the time-shift between the two processes by determining κ for which $r_{XY}(\kappa)$ has its maximum. Intuitively, one can imagine that $y(n)$ is compared to $x(n)$ at time-shift $\kappa = 0$. Then $y(n)$ is shifted by κ and compared again. This is repeated for all time shifts κ and results in a function $r_{XY}(\kappa)$ that describes the measure of similarity between both realizations at all time-shifts.*
Similarly, Figure 3.4 shows a realization of a recorded Glockenspiel signal $x(n)$ and its auto- and cross-second-order moment functions, $r_{XX}(\kappa)$ and $r_{XY}(\kappa)$. $y(n)$ is a delayed version of $x(n)$ with a delay of 1000ms at a sampling rate of $f_s = 44.1kHz$. Here, we see a different behaviour of the auto- and cross- SOMF when compared to the white Gaussian noise case. Both the auto- and cross- SOMF show significant values for time-shifts other than $\kappa = 0$ and $\kappa = 1$s, respectively. This is an expected behaviour, as music signals often contain periodicities.

3.3.3 Central Second-Order Moment Function (Covariance Function)

The central second-order moment function is called the covariance function and is defined by

$$c_{XX}(n+\kappa, n) = \mathsf{E}\left[(X(n+\kappa) - \mathsf{E}[X(n+\kappa)])(X(n) - \mathsf{E}[X(n)])^*\right]$$

which can be re-written in the form

$$c_{XX}(n+\kappa, n) = r_{XX}(n+\kappa, n) - \mathsf{E}[X(n+\kappa)]\mathsf{E}[X(n)^*]$$

The covariance function for two processes $X(n)$ and $Y(n)$ is defined by

$$c_{XY}(n+\kappa, n) = \mathsf{E}[(X(n+\kappa) - \mathsf{E}[X(n+\kappa)])(Y(n) - \mathsf{E}[Y(n)])^*]$$

or equivalently

$$c_{XY}(n+\kappa, n) = r_{XY}(n+\kappa, n) - \mathsf{E}[X(n+\kappa)]\mathsf{E}[Y(n)^*]$$

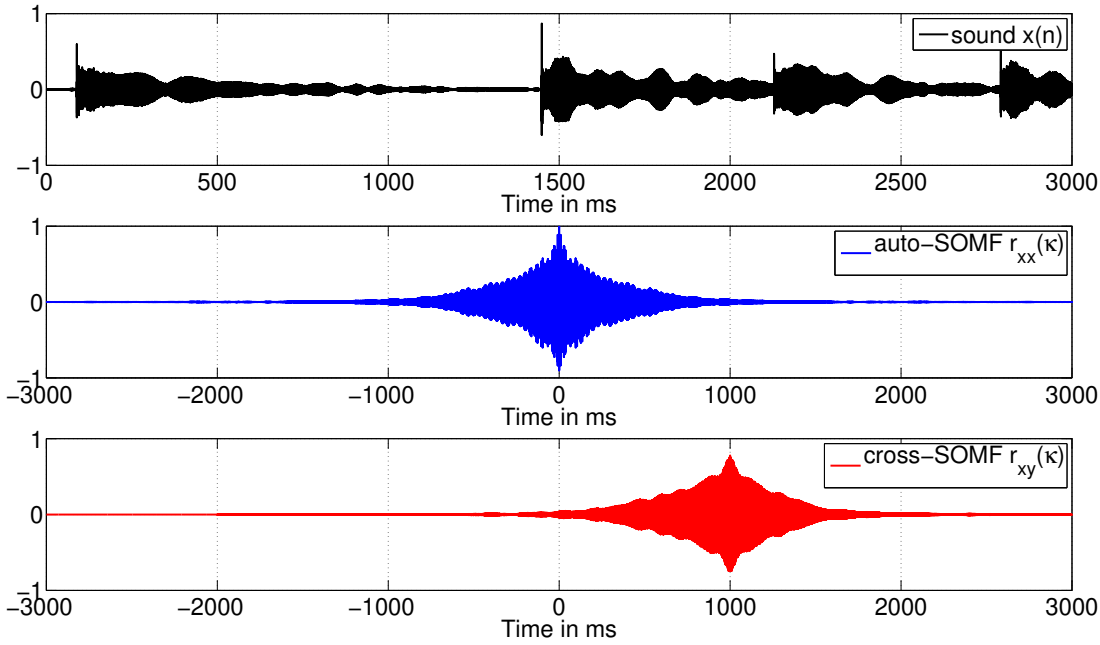


Figure 3.4: Auto- and cross-second-order moment function of a glockenspiel signal.

For processes that are at least jointly wide sense stationary, we have

$$c_{XX}(\kappa) = r_{XX}(\kappa) - |\mathsf{E}[X(n)]|^2$$

and

$$c_{XY}(\kappa) = r_{XY}(\kappa) - \mathsf{E}[X(n)]\mathsf{E}[Y(n)^*]$$

The variance of a random process $X(n)$ is obtained from $c_{XX}(\kappa)$ by setting $\kappa = 0$. If the process is stationary

$$\begin{aligned} \sigma_X^2 &= \mathsf{E}[(X(n) - \mathsf{E}[X(n)])\,(X(n) - \mathsf{E}[X(n)])^*] \\ &= r_{XX}(0) - \mathsf{E}[X(n)]\mathsf{E}[X(n)^*] \\ &= r_{XX}(0) - |\mu_X|^2 \end{aligned}$$

For two random processes, if $c_{YX}(n, n+\kappa) = 0$, then they are called uncorrelated, which is equivalent to $r_{XY}(n+\kappa, n) = \mathsf{E}[X(n+\kappa)]\mathsf{E}[Y(n)^*]$. They are orthogonal if $r_{XY}(n+\kappa, n) = 0$.

Example 3.3.3 *Here we consider a practical example, as depicted in Figure 3.5, which may occur e.g. in audio signal processing. Two spatially separated microphones record the waveform $S(n)$ from the same audio source. The microphone output signals are denoted by $X_1(n)$ and $X_2(n)$, and can be modeled as*

$$\begin{aligned} X_1(n) &= S(n) + V_1(n) \\ X_2(n) &= S(n - n_0) + V_2(n) \end{aligned}$$

where $S(n)$ is a real-valued WSS random process with zero mean and covariance function $c_{SS}(\kappa)$. The quantities $V_1(n)$ and $V_2(n)$ are real-valued independent white noise processes with zero mean and variances σ^2. We now want to determine the time-delay $n_0 \in \mathbb{Z}$, which is realised by the filter z^{-n_0}, which can be used e.g. to determine the angle-of-arrival of the audio source.

This can be done by calculating the covariance function between $X_1(n)$ and $X_2(n)$, as follows

$$\begin{aligned} c_{X_1X_2}(\kappa) &= \mathsf{E}\left[(S(n+\kappa) + V_1(n+\kappa))(S(n-n_0) + V_2(n))\right] \\ &= \mathsf{E}[S(n+\kappa)S(n-n_0)] + \mathsf{E}[S(n+\kappa)V_2(n)] + \mathsf{E}[V_1(n+\kappa)S(n-n_0)] + \mathsf{E}[V_1(n+\kappa)V_2(n)] \end{aligned}$$

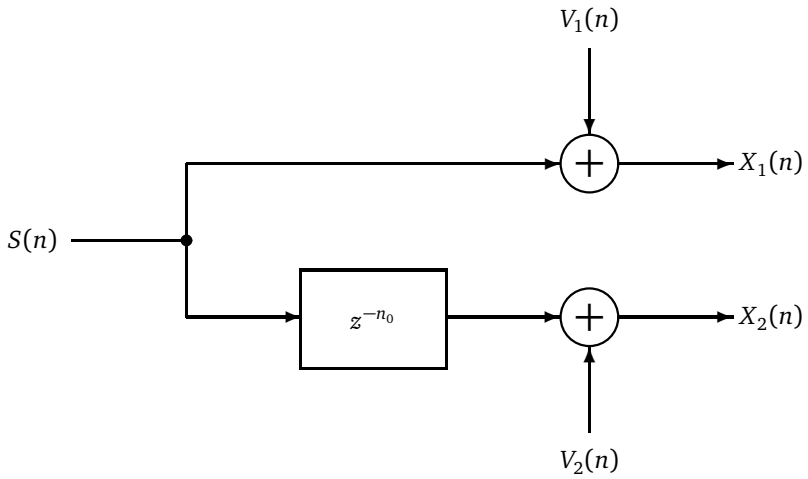


Figure 3.5: Practical use of the cross-covariance for time-delay estimation.

$$= \quad c_{SS}(\kappa + n_0)$$

where the remaining terms are zero, because $S(n)$, $V_1(n)$ and $V_2(n)$ are assumed to be mutually uncorrelated.

Consider for instance, $S(n) = \cos(\omega_0 n + \phi)$ where ω_0 is a constant and ϕ is uniformly distributed on $[0, 2\pi)$. The covariance function is $c_{SS}(\kappa) = 1/2 \cdot \cos(\omega_0 \kappa)$ which gives

$$c_{X_1X_2}(\kappa) = 1/2 \cdot \cos(\omega_0(\kappa + n_0))$$

A simple algorithm to determine the time-delay would be to search for the first maximum of $c_{X_1X_2}(\kappa)$.

Example 3.3.4 *A complex process $Z(n)$ is comprised of a sum of M complex signals*

$$Z(n) = \sum_{m=1}^{M} A_m e^{j(\omega_0 n + \Phi_m)}$$

Here ω_0 is constant, A_m is a zero-mean random amplitude of the mth signal and Φ_m is its random phase. We assume all variables A_m and Φ_m for $m = 1, 2, \ldots, M$ to be statistically independent and the random phases uniformly distributed on $[0, 2\pi)$. Find the mean and SOMF. Is the process wide sense stationary ?

$$\begin{aligned}
\mathsf{E}[Z(n)] &= \sum_{m=1}^{M} \mathsf{E}[A_m] e^{j\omega_0 n} \mathsf{E}\left[e^{j\Phi_m}\right] \\
&= \sum_{m=1}^{M} \mathsf{E}[A_m] e^{j\omega_0 n} \{\mathsf{E}[\cos(\Phi_m)] + j\mathsf{E}[\sin(\Phi_m)]\} \\
&= 0
\end{aligned}$$

$$\begin{aligned}
r_{ZZ}(n+\kappa, n) &= \mathsf{E}[Z(n+\kappa)Z(n)^*] \\
&= \mathsf{E}\left[\sum_{m=1}^{M} A_m e^{j\omega_0(n+\kappa)} e^{j\Phi_m} \cdot \sum_{l=1}^{M} A_l e^{-j\omega_0 n} e^{-j\Phi_l}\right] \\
&= \sum_{m=1}^{M}\sum_{l=1}^{M} \mathsf{E}[A_m A_l] \cdot e^{j\omega_0 \kappa} \cdot \mathsf{E}\left[e^{j(\Phi_m - \Phi_l)}\right]
\end{aligned}$$

Since

$$\mathsf{E}\left[e^{j(\Phi_m - \Phi_l)}\right] = \begin{cases} 1 & m = l \\ 0 & m \neq l \end{cases}$$

The former holds because for $m \neq l$,

$$\mathsf{E}\left[e^{j\Phi_m - \Phi_l}\right] = \mathsf{E}\left[e^{j\Phi_m}\right]\mathsf{E}\left[e^{-j\Phi_l}\right] = 0$$

Thus if $m = l$,

$$r_{ZZ}(n+\kappa, n) = \sum_{m=1}^{M} \mathsf{E}[A_m^2] \cdot e^{j\omega_0 \kappa} = e^{j\omega_0 \kappa} \sum_{m=1}^{M} \mathsf{E}[A_m^2] = r_{ZZ}(\kappa)$$

Therefore, the complex process $Z(n)$ is wide-sense stationary.

Bibliography

J.F. Böhme. *Stochastische Signale*. Teubner Studienbücher, 1998.

A. Papoulis. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, Inc., third edition, 1991.

P. Z. Peebles, Jr. *Probability, Random Variables and Random Signal Principles*. McGraw-Hill, Inc., fourth edition, 2001.

4 Power Spectral Density

In the last chapter, we have discussed random processes in the time-domain. We now turn to the frequency-domain representation of random processes, in analogy to the Fourier analysis of deterministic discrete-time signals. Note that the discrete-time Fourier transform (DTFT) $X(e^{j\omega})$ of some signal $x(n)$ is only defined if the sum

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x(n)e^{-j\omega n}$$

converges. A sufficient condition for the sum to converge uniformly to a continuous function of ω is given if $x(n)$ is absolutely summable, i.e. we have $\sum_{n=-\infty}^{\infty} |x(n)| < \infty$.

In the case of a random process $X(n)$, which represents an ensemble of discrete-time signals, this condition is not fulfilled and consequently, the DTFT of a random process does not exist. However, a frequency domain representation, namely the power spectral density (PSD) or the spectrum of a random process can be found by taking the Fourier transform of the SOMF or the covariance function, respectively. Both, the PSD and the spectrum are important concepts for the analysis of signals in the frequency domain.

4.1 Power Spectral Density

4.1.1 Definition of Power Spectral Density

Let $X(n, \zeta_i)$, $n \in \mathbb{Z}$, represent a sample function of a real-valued random process $X(n, \zeta) \equiv X(n)$. We define a truncated version of this sample function as follows:

$$x_N(n) = \begin{cases} X(n, \zeta_i), & |n| \leq M,\ M \in \mathbb{Z}^+ \\ 0, & \text{elsewhere} \end{cases}$$

where $N = 2M + 1$ is the length of the truncation window. As long as M is finite, we presume that $x_N(n)$ will satisfy

$$\sum_{n=-M}^{M} |x_N(n)| < \infty$$

Then the Fourier transform of $x_N(n)$ becomes:

$$\begin{aligned} X_N(\mathrm{e}^{\mathrm{j}\omega}) &= \sum_{n=-\infty}^{\infty} x_N(n)\,\mathrm{e}^{-\mathrm{j}\omega n} \\ &= \sum_{n=-M}^{M} X(n, \zeta_i)\,\mathrm{e}^{-\mathrm{j}\omega n} \end{aligned}$$

The normalized energy in the time interval $[-M, M]$ is

$$E_N = \sum_{n=-M}^{M} x_N(n)^2 = \int_{-\pi}^{\pi} \left|X_N(\mathrm{e}^{\mathrm{j}\omega})\right|^2 \frac{d\omega}{2\pi}$$

where Parseval's theorem has been used to obtain the integral. By dividing the expression for E_N by N we obtain the average power P_N in $x_N(n)$ over the interval $[-M, M]$:

$$P_N = \frac{1}{2M+1} \sum_{n=-M}^{M} x_N(n)^2 = \int_{-\pi}^{\pi} \frac{|X_N(\mathrm{e}^{\mathrm{j}\omega})|^2}{2M+1} \frac{d\omega}{2\pi} \tag{4.1}$$

From Equation (4.1), we see that the average power P_N, in the truncated ensemble member $x_N(n)$, can be obtained by the integration of $|X_N(\mathrm{e}^{\mathrm{j}\omega})|/N$. To obtain the average power for the entire sample function, we must consider the limit as $M \to \infty$. P_N provides a measure of power for a single sample function. If we wish to determine the average power of the *random process*, we must also take the average (expected value) over the entire ensemble of sample functions. The average power of the process $X(n, \zeta)$ is defined as

$$P_{XX} = \lim_{M\to\infty} \frac{1}{2M+1} \sum_{n=-M}^{M} \mathsf{E}[X(n, \zeta)^2] = \lim_{M\to\infty} \int_{-\pi}^{\pi} \frac{\mathsf{E}[|X_N(\mathrm{e}^{\mathrm{j}\omega}, \zeta)|^2]}{2M+1} \frac{d\omega}{2\pi}$$

Remark 1: The average normalized power P_{XX} in a random process $X(n)$ is given by the time average of its second moment.

Remark 2: If $X(n)$ is wide-sense stationary, $\mathsf{E}[X(n)^2]$ is a constant and $P_{XX} = \mathsf{E}[X(n)^2]$. P_{XX} can be obtained by a frequency domain integration. If we define the power spectral density (PSD) of $X(n)$ by

$$S_{XX}(e^{j\omega}, \zeta) = \lim_{M\to\infty} \frac{\mathsf{E}\left[\left|X_N(e^{j\omega}, \zeta)\right|^2\right]}{2M+1}$$

where

$$X_N(e^{j\omega}, \zeta) = \sum_{n=-M}^{M} X(n, \zeta)\, e^{-j\omega n}$$

then the normalized average power is given by

$$P_{XX} = \int_{-\pi}^{\pi} S_{XX}(e^{j\omega}) \frac{d\omega}{2\pi} = r_{XX}(0) = \mathsf{E}[X(n)^2]$$

Example 4.1.1 *Consider the random process*

$$X(n) = A \cdot \cos(\omega_0 n + \Phi)$$

where A and ω_0 are real-valued constants and Φ is a random variable uniformly distributed on $[0, \pi/2)$. We shall find P_{XX} for $X(n)$ using

$$P_{XX} = \lim_{M\to\infty} \sum_{n=-M}^{M} \frac{\mathsf{E}\left[X(n)^2\right]}{2M+1}.$$

The mean-squared value of $X(n)$ is

$$\begin{aligned}
\mathsf{E}[X(n)^2] &= \mathsf{E}\left[A^2 \cos(\omega_0 n + \Phi)^2\right] \\
&= \mathsf{E}\left[\frac{A^2}{2} + \frac{A^2}{2}\cos(2\omega_0 n + 2\Phi)\right] \\
&= \frac{A^2}{2} + \frac{A^2}{2}\int_0^{\pi/2} \cos(2\omega_0 n + 2\varphi)\frac{2}{\pi}\, d\varphi \\
&= \frac{A^2}{2} + \frac{A^2}{2}\left[\frac{\sin(2\omega_0 n + 2\varphi)}{2}\right]_0^{\pi/2}\frac{2}{\pi} \\
\mathsf{E}[X(n)^2] &= \frac{A^2}{2} + \frac{A^2}{2}\left[\frac{\sin(2\omega_0 n + \pi) - \sin(2\omega_0 n)}{2}\right]\frac{2}{\pi} \\
&= \frac{A^2}{2} + \frac{A^2}{2}\frac{2}{\pi}\frac{-2\sin(2\omega_0 n)}{2} \\
&= \frac{A^2}{2} - \frac{A^2}{\pi}\sin(2\omega_0 n)
\end{aligned}$$

$X(n)$ is non stationary because $\mathsf{E}\left[X(n)^2\right]$ is a function of time. The time average of the above expression is

$$\lim_{M\to\infty} \frac{1}{2M+1} \sum_{n=-M}^{M} \left[\frac{A^2}{2} - \frac{A^2}{\pi}\sin(2\omega_0 n)\right] = \frac{A^2}{2}$$

Using

$$P_{XX} = \lim_{M\to\infty} \int_{-\pi}^{\pi} \frac{\mathsf{E}\left[\left|X_N(e^{j\omega})^2\right|\right]}{2M+1} \frac{d\omega}{2\pi}$$

yields $P_{XX} = \frac{A^2}{2}$ which agrees with the above, which should be shown by the reader as an exercise.

The results derived for the PSD can be obtained similarly for two processes $X(n)$ and $Y(n)$, say. The cross-power density spectrum is then given by

$$S_{XY}(e^{j\omega}) = \lim_{M\to\infty} \frac{E\left[X_N(e^{j\omega})Y_N(e^{j\omega})^*\right]}{2M+1}$$

and thus the total average cross power is

$$P_{XY} = \int_{-\pi}^{\pi} S_{XY}(e^{j\omega}) \frac{d\omega}{2\pi}$$

Example 4.1.2 *Figure 4.1 shows an electrocardiogram signal in the time domain and its power spectral density (PSD) estimate. The first peak in the PSD is at 1.3 Hz which corresponds to 78 heartbeats per minute, on average. This is information is much easier to obtain when observing the power spectral density rather than the corresponding time signal. The power spectral density helps us, thus, to infer information on the existing frequency components in a random process and also the periodicities therein.*

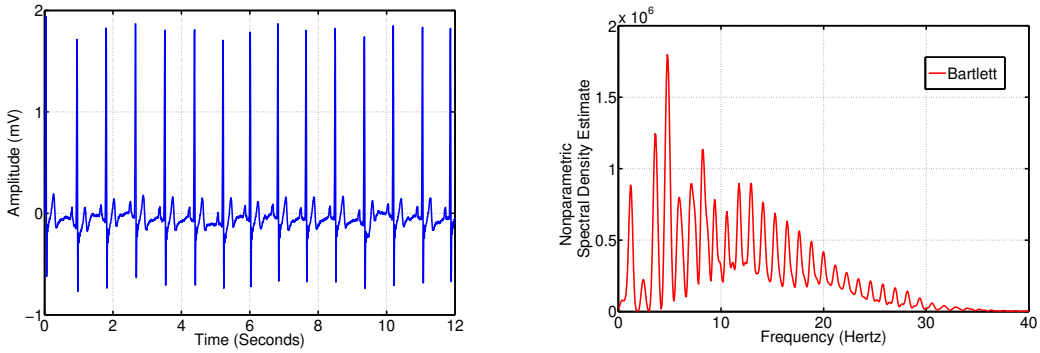


Figure 4.1: Electrocardiogram (ECG) signal in the time domain (left). Power spectral density estimate of the ECG signal (right).

4.1.2 Relationship Between the PSD and SOMF: Wiener-Khintchine Theorem

Theorem 4.1.1 *Wiener-Khintchine Theorem*

When $X(n)$ is a wide-sense stationary process, the PSD can be obtained from the Fourier Transform of the SOMF:

$$S_{XX}(e^{j\omega}) = \mathcal{F}\{r_{XX}(\kappa)\} = \sum_{\kappa=-\infty}^{\infty} r_{XX}(\kappa)e^{-j\omega\kappa}$$

and conversely,

$$r_{XX}(\kappa) = \mathcal{F}^{-1}\left\{S_{XX}(e^{j\omega})\right\} = \int_{-\pi}^{\pi} S_{XX}(e^{j\omega})e^{j\omega\kappa}\frac{d\omega}{2\pi}$$

Consequence: there are two distinctly different methods that may be used to evaluate the PSD of a random process:

1. The PSD is obtained directly from the definition

$$S_{XX}(e^{j\omega}) = \lim_{M\to\infty} \frac{E\left[\left|X_N(e^{j\omega})\right|^2\right]}{2M+1}$$

where

$$X_N(e^{j\omega}) = \sum_{n=-M}^{M} X(n)e^{-j\omega n}$$

2. The PSD is obtained by evaluating the Fourier transform of $r_{XX}(\kappa)$, where $r_{XX}(\kappa)$ has to be obtained first.

4.1.3 Properties of the PSD

The PSD has as number of important properties:

1. The PSD is always real-valued, even if $X(n)$ is complex

$$\begin{aligned} S_{XX}(\mathrm{e}^{\mathrm{j}\omega})^* &= \sum_{\kappa=-\infty}^{\infty} r_{XX}(\kappa)^* \mathrm{e}^{+\mathrm{j}\omega\kappa} \\ &= \sum_{\kappa=-\infty}^{\infty} r_{XX}(-\kappa)\mathrm{e}^{+\mathrm{j}\omega\kappa} \\ &= \sum_{\kappa=-\infty}^{\infty} r_{XX}(\kappa')\mathrm{e}^{-\mathrm{j}\omega\kappa'} \\ &= S_{XX}(\mathrm{e}^{\mathrm{j}\omega}) \end{aligned}$$

2. $S_{XX}(\mathrm{e}^{\mathrm{j}\omega}) \geq 0$, even if $X(n)$ is complex.
3. When $X(n)$ is real-valued, $S_{XX}(\mathrm{e}^{-\mathrm{j}\omega}) = S_{XX}(\mathrm{e}^{\mathrm{j}\omega})$.
4. $\int_{-\pi}^{\pi} S_{XX}(\mathrm{e}^{\mathrm{j}\omega}) \frac{d\omega}{2\pi} = P_{XX}$ is the total normalised average power.
5. $S_{XX}(\mathrm{e}^{\mathrm{j}\omega})$ is 2π-periodic.

Example 4.1.3

$$X(n) = A\sin(\omega_0 n + \Phi)$$

where A, ω_0 are constants and Φ is uniformly distributed on $[0, 2\pi)$. The mean of $X(n)$ is given by

$$\mathsf{E}[X(n)] = \int_0^{2\pi} \frac{1}{2\pi} A\sin(\omega_0 n + \varphi)\, d\varphi = 0$$

and its SOMF is

$$r_{XX}(\kappa) = \mathsf{E}[X(n+\kappa)X(n)] = \frac{A^2}{2}\cos(\omega_0\kappa)$$

By taking the Fourier transform of $r_{XX}(\kappa)$, we obtain the PSD

$$S_{XX}(\mathrm{e}^{\mathrm{j}\omega}) = \mathcal{F}\{r_{XX}(\kappa)\} = \frac{A^2}{2}\pi\left\{\eta(\omega-\omega_0) + \eta(\omega+\omega_0)\right\}$$

where

$$\eta(\omega) = \sum_{k=-\infty}^{\infty} \delta(\omega + k2\pi)$$

is Dirac's delta comb.

Example 4.1.4 *Suppose that the random variables A_i are uncorrelated with zero mean and variance σ_i^2, $i = 1, \ldots, P$. Define*

$$Z(n) = \sum_{i=1}^{P} A_i \exp\{\mathrm{j}\omega_i n\}.$$

It follows that

$$r_{ZZ}(\kappa) = \sum_{i=1}^{P} \sigma_i^2 \exp\{\mathrm{j}\omega_i \kappa\}$$

and

$$S_{ZZ}(\mathrm{e}^{\mathrm{j}\omega}) = \sum_{i=1}^{P} \sigma_i^2 \eta(\omega - \omega_i)$$

Example 4.1.5 *Suppose now that the random variables A_i and B_i, $i = 1, \ldots, P$ are pairwise uncorrelated with zero mean and common variance*

$$\mathsf{E}[A_i^2] = \mathsf{E}[B_i^2] = \sigma_i^2, \quad i = 1, \ldots, P.$$

Define

$$X(n) = \sum_{i=1}^{P} [A_i \cos(\omega_i n) + B_i \sin(\omega_i n)].$$

We find

$$r_{XX}(\kappa) = \sum_{i=1}^{P} \sigma_i^2 \cos(\omega_i \kappa)$$

and therefore

$$S_{XX}(\mathrm{e}^{\mathrm{j}\omega}) = \pi \sum_{i=1}^{P} \sigma_i^2 [\eta(\omega - \omega_i) + \eta(\omega + \omega_i)]$$

4.1.4 White Noise Process

Definition 4.1.1 *A random process $X(n)$ is said to be a "white noise process" if the PSD is constant over all frequencies:*

$$S_{XX}(\mathrm{e}^{\mathrm{j}\omega}) = \sigma_X^2.$$

The SOMF for a white process is obtained by taking the inverse Fourier transform of $S_{XX}(\mathrm{e}^{\mathrm{j}\omega})$.

$$r_{XX}(\kappa) = \sigma_X^2 \, \delta(\kappa)$$

where $\delta(\kappa)$ is the Kronecker delta function given by

$$\delta(\kappa) = \begin{cases} 1, & \kappa = 0 \\ 0, & \kappa \neq 0 \end{cases}$$

4.1.5 Cross-Power Spectral Density

Suppose that we have observations $x_N(n)$ and $y_N(n)$ of a random process $X(n)$ and $Y(n)$, respectively. Let $X_N(\mathrm{e}^{\mathrm{j}\omega})$ and $Y_N(\mathrm{e}^{\mathrm{j}\omega})$ be their respective Fourier transforms, i.e.

$$X_N(\mathrm{e}^{\mathrm{j}\omega}) = \sum_{n=-M}^{M} x_N(n) \mathrm{e}^{-\mathrm{j}\omega n}$$

and

$$Y_N(\mathrm{e}^{\mathrm{j}\omega}) = \sum_{n=-M}^{M} y_N(n) \mathrm{e}^{-\mathrm{j}\omega n}$$

The cross-spectral density function is given as

$$S_{XY}(\mathrm{e}^{\mathrm{j}\omega}) = \lim_{M \to \infty} \frac{\mathsf{E}\left[X_N(\mathrm{e}^{\mathrm{j}\omega}) Y_N(\mathrm{e}^{\mathrm{j}\omega})^*\right]}{2M+1}$$

Substituting $X_N(\mathrm{e}^{\mathrm{j}\omega})$ and $Y_N(\mathrm{e}^{\mathrm{j}\omega})$, gives

$$S_{XY}(\mathrm{e}^{\mathrm{j}\omega}) = \lim_{M \to \infty} \frac{1}{2M+1} \sum_{n_1=-M}^{M} \sum_{n_2=-M}^{M} r_{XY}(n_1, n_2) \mathrm{e}^{-\mathrm{j}\omega(n_1 - n_2)}$$

Taking the inverse Fourier transform of both sides

$$\int_{-\pi}^{\pi} S_{XY}(\mathrm{e}^{\mathrm{j}\omega}) \mathrm{e}^{\mathrm{j}\omega\kappa} \frac{d\omega}{2\pi} = \int_{-\pi}^{\pi} \lim_{M \to \infty} \frac{1}{2M+1} \sum_{n_1=-M}^{M} \sum_{n_2=-M}^{M} r_{XY}(n_1, n_2) \mathrm{e}^{-\mathrm{j}\omega(n_1 - n_2)} \mathrm{e}^{\mathrm{j}\omega\kappa} \frac{d\omega}{2\pi}$$

$$\begin{aligned}\int_{-\pi}^{\pi} S_{XY}(e^{j\omega})e^{j\omega\kappa}\frac{d\omega}{2\pi} &= \lim_{M\to\infty}\frac{1}{2M+1}\sum_{n_1=-M}^{M}\sum_{n_2=-M}^{M} r_{XY}(n_1,n_2)\int_{-\pi}^{\pi} e^{j\omega(n_2+\kappa-n_1)}\frac{d\omega}{2\pi}\\ &= \lim_{M\to\infty}\frac{1}{2M+1}\sum_{n_1=-M}^{M}\sum_{n_2=-M}^{M} r_{XY}(n_1,n_2)\delta(n_2+\kappa-n_1)\\ &= \lim_{M\to\infty}\frac{1}{2M+1}\sum_{n_2=-M}^{M} r_{XY}(n_2+\kappa,n_2) \qquad (4.2)\end{aligned}$$

provided $-M \le n_2+\kappa \le M$ so that the delta function is in the range of the summation. This condition can be relaxed in the limit as $M\to\infty$. The above result shows that the time average of the SOMF and the PSD form a Fourier transform pair. Clearly, if $X(n)$ and $Y(n)$ are jointly stationary, then this result also implies

$$S_{XY}(e^{j\omega}) = \sum_{\kappa=-\infty}^{\infty} r_{XY}(\kappa)e^{-j\omega\kappa}$$

Definition 4.1.2 *For two jointly stationary processes $X(n)$ and $Y(n)$, the cross-power spectral density is given by*

$$S_{XY}(e^{j\omega}) = \mathcal{F}\{r_{XY}(\kappa)\} = \sum_{\kappa=-\infty}^{\infty} r_{XY}(\kappa)e^{-j\omega\kappa}$$

and is in general complex.

Properties:

1. $S_{XY}(e^{j\omega}) = S_{YX}(e^{j\omega})^*$. In addition, if $X(n)$ and $Y(n)$ are real-valued, then $S_{XY}(e^{j\omega}) = S_{YX}(e^{-j\omega})$.
2. $\mathsf{Re}\{S_{XY}(e^{j\omega})\}$ and $\mathsf{Re}\{S_{YX}(e^{j\omega})\}$ are even functions of ω if $X(n)$ and $Y(n)$ are real-valued.
3. $\mathsf{Im}\{S_{XY}(e^{j\omega})\}$ and $\mathsf{Im}\{S_{YX}(e^{j\omega})\}$ are odd functions of ω if $X(n)$ and $Y(n)$ are real-valued.
4. $S_{XY}(e^{j\omega}) = 0$ and $S_{YX}(e^{j\omega}) = 0$ if $X(n)$ and $Y(n)$ are orthogonal.

Property 1 results from the fact that:

$$\begin{aligned}S_{YX}(e^{j\omega}) &= \sum_{\kappa=-\infty}^{\infty} r_{YX}(\kappa)e^{-j\omega\kappa}\\ &= \sum_{\kappa=-\infty}^{\infty} r_{XY}(-\kappa)^* e^{-j\omega\kappa}\\ &= \left[\sum_{\kappa=-\infty}^{\infty} r_{XY}(\kappa)e^{-j\omega\kappa}\right]^*\\ &= S_{XY}(e^{j\omega})^*\end{aligned}$$

Example 4.1.6 *Let the SOMF function of two processes $X(n)$ and $Y(n)$ be*

$$r_{XY}(n+\kappa,n) = \frac{AB}{2}\left[\sin(\omega_0\kappa) + \cos(\omega_0[2n+\kappa])\right]$$

where A, B and ω_0 are constants. We find

$$\lim_{M\to\infty}\frac{1}{2M+1}\sum_{n=-M}^{M} r_{XY}(n+\kappa,n) = \frac{AB}{2}\sin(\omega_0\kappa) + \underbrace{\frac{AB}{2}\lim_{M\to\infty}\frac{1}{2M+1}\sum_{n=-M}^{M}\cos(\omega_0[2n+\kappa])}_{=0}$$

so using the result of Equation (4.2)

$$\begin{aligned}S_{XY}(e^{j\omega}) &= \mathcal{F}\left\{\frac{AB}{2}\sin(\omega_0\kappa)\right\}\\ &= \frac{\pi}{j}\frac{AB}{2}\left[\eta(\omega-\omega_0) - \eta(\omega+\omega_0)\right]\end{aligned}$$

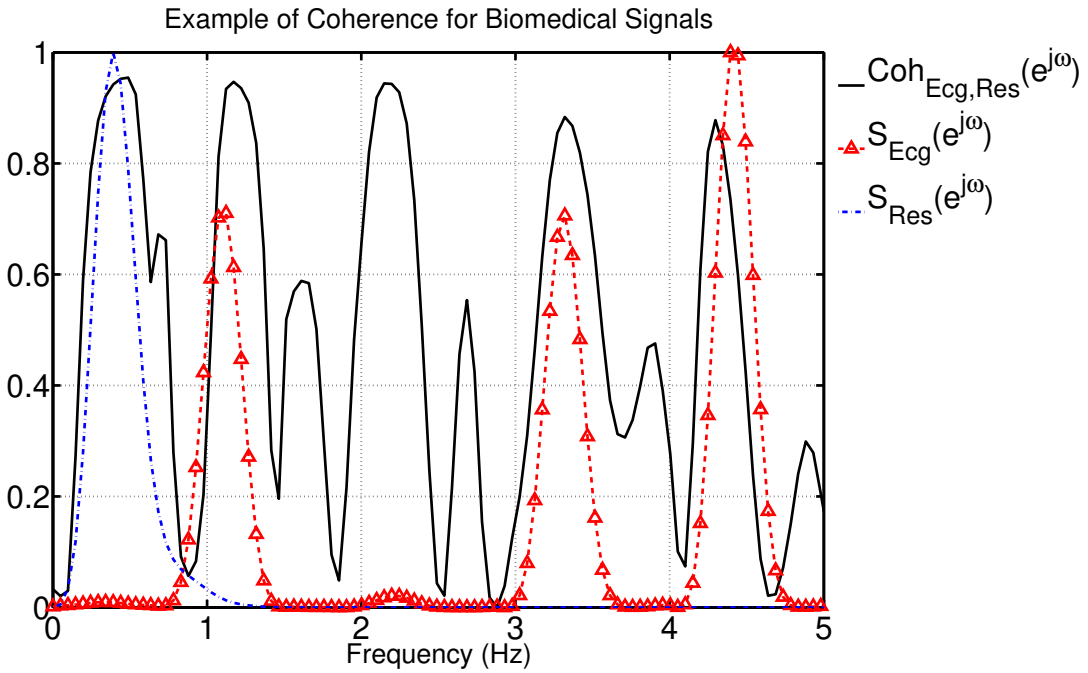


Figure 4.2: Coherence of an ECG signal and a respiration signal.

Example 4.1.7 *The coherence between two random processes $X(n)$ and $Y(n)$ is a measure of how well $X(n)$ corresponds to $Y(n)$ at a certain frequency ω. It is defined as:*

$$Coh_{XY}(e^{j\omega}) = \frac{|S_{XY}(e^{j\omega})|^2}{S_{XX}(e^{j\omega})S_{YY}(e^{j\omega})}.$$

and can only take values between 0 and 1.
Figure 4.1.7 shows an example of the coherence of biomedical signals. Here we can see the power spectral denisty of a respiration signal and that of an ECG signal. The respiration signal shows its first peak at $0.4Hz$ and the ECG signal shows its first peak at $1.1Hz$. Clearly visible is the high coherence between both signals at the pulse frequency bands and at the respiration frequency bands and their harmonics.

4.1.6 DC and RMS Values for Random Processes

The DC value, the RMS value, and the average power are usually defined in terms of time average operations. For ergodic random processes, these time averages are equivalent to ensemble averages. We present here a summary of these relationships. $X(n)$ is an ergodic random process that may correspond to either a sampled voltage across or current through a 1-Ω impedance.

1. DC value:

$$X_{DC} = \lim_{M\to\infty} \frac{1}{2M+1} \sum_{n=-M}^{M} X(n) = \mathsf{E}[X(n)] = \mu_X$$

2. Normalised DC power:

$$P_{DC} = \left[\lim_{M\to\infty} \frac{1}{2M+1} \sum_{n=-M}^{M} X(n) \right]^2 = (\mathsf{E}[X(n)])^2 = X_{DC}^2$$

3. RMS value:

$$X_{RMS} = \sqrt{\lim_{M\to\infty} \frac{1}{2M+1} \sum_{n=-M}^{M} X(n)^2} = \sqrt{r_{XX}(0)} = \sqrt{\int_{-\pi}^{\pi} S_{XX}(e^{j\omega}) \frac{d\omega}{2\pi}}$$

4. RMS value of the AC part:

$$\begin{aligned}(X_{RMS})_{AC} &= \sqrt{\lim_{M\to\infty}\frac{1}{2M+1}\sum_{n=-M}^{M}(X(n)-X_{DC})^2} \\ &= \sqrt{(\mathsf{E}[X(n)-X_{DC}])^2} \\ &= \sqrt{\mathsf{E}[X(n)^2]-X_{DC}^2} \\ &= \sqrt{\int_{-\pi}^{\pi} S_{XX}(e^{j\omega})\frac{d\omega}{2\pi}-X_{DC}^2}\end{aligned}$$

5. Normalised total average power:

$$P_{XX} = \lim_{M\to\infty}\frac{1}{2M+1}\sum_{n=-M}^{M}X(n)^2 = r_{XX}(0) = \int_{-\pi}^{\pi} S_{XX}(e^{j\omega})\frac{d\omega}{2\pi}$$

6. Normalised average power of the AC part:

$$\begin{aligned}P_{AC} &= \lim_{M\to\infty}\frac{1}{2M+1}\sum_{n=-M}^{M}(X(n)-X_{DC})^2 \\ &= \mathsf{E}[(X(n)-X_{DC})^2] \\ &= \mathsf{E}[X(n)^2]-X_{DC}^2 \\ &= \int_{-\pi}^{\pi} S_{XX}(e^{j\omega})\frac{d\omega}{2\pi}-X_{DC}^2 \\ &= P_{XX}-P_{DC}\end{aligned}$$

4.2 Definition of the Spectrum

The PSD of a stationary random process has been defined as the Fourier transform of the SOMF. The 'spectrum' of a stationary random process $X(n)$ is defined as the Fourier transform of the covariance function (central SOMF):

$$C_{XX}(e^{j\omega}) = \sum_{n=-\infty}^{\infty} c_{XX}(n)e^{-j\omega n}$$

If $\sum_n |c_{XX}(n)| < \infty$, $C_{XX}(e^{j\omega})$ exists and is bounded and uniformly continuous. The inversion is given by

$$c_{XX}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} C_{XX}(e^{j\omega})e^{j\omega n}\,d\omega$$

$C_{XX}(e^{j\omega})$ is real-valued, 2π periodic and non-negative.

The cross-spectrum of two jointly stationary random processes $X(n)$ and $Y(n)$ is defined by

$$C_{XY}(e^{j\omega}) = \sum_{n=-\infty}^{\infty} c_{XY}(n)e^{-j\omega n}$$

with the property

$$C_{XY}(e^{j\omega}) = C_{YX}(e^{j\omega})^*$$

The inversion is given by

$$c_{XY}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} C_{XY}(e^{j\omega})e^{j\omega n}\,d\omega$$

If $X(n)$ and $Y(n)$ are real-valued,

$$\begin{aligned}C_{XX}(e^{j\omega}) &= C_{XX}(e^{-j\omega}) \\ C_{XY}(e^{j\omega}) = C_{XY}(e^{-j\omega})^* &= C_{YX}(e^{-j\omega}) = C_{YX}(e^{j\omega})^*\end{aligned}$$

The spectrum of a real-valued process is completely specified on the interval $[0,\pi]$.

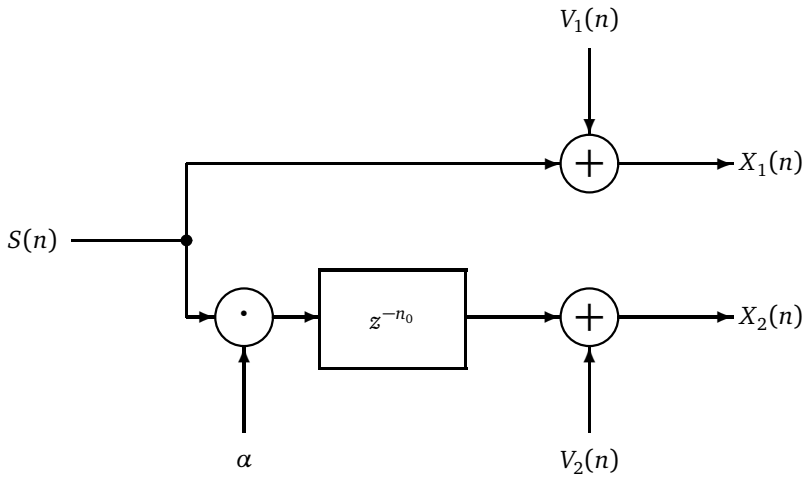


Figure 4.3: Practical use of the cross-spectrum for time-delay estimation.

Example 4.2.1 *$X_1(n)$ and $X_2(n)$ in Figure 4.3 are two random processes received at two sensors. Using the cross-spectrum we want to determine the time-delay between the processes.*
$V_1(n)$ and $V_2(n)$ are two zero-mean, stationary and uncorrelated white random processes with variance σ^2. The reference signal $S(n)$ is given by:

$$S(n) = \sin(\omega_0 n + \phi)$$

ϕ is uniformly distributed on $[0, 2\pi)$, α is a random variable with expectation value 1 and variance A^2. The parameters ω_0 and n_0 are unknown, real-valued constants. Furthermore, ϕ, α, $V_1(n)$ and $V_2(n)$ are pairwise independant for all n.
First, we calculate the cross-covariance function $c_{X_1X_2}(\kappa)$:

$$\begin{aligned} c_{X_1X_2}(\kappa) &= \mathrm{E}[(S(n+\kappa) + V_1(n+\kappa))(\alpha S(n-n_0) + V_2(n))] \\ &= c_{SS}(\kappa + n_0) = \tfrac{1}{2}\cos(\omega_0(\kappa + n_0)). \end{aligned}$$

Using $c_{X_1X_2}(\kappa)$, we can calculate the cross-spectrum $C_{X_1X_2}$:

$$C_{X_1X_2}(\mathrm{e}^{\mathrm{j}\omega}) = \tfrac{\pi}{2}(\eta(\omega - \omega_0) + \eta(\omega + \omega_0))\, e^{j\omega n_0}.$$

where

$$\eta(\omega) = \sum_{k=-\infty}^{\infty} \delta(\omega + k2\pi)$$

The phase of the cross-spectrum is $\angle C_{X_1X_2}(\mathrm{e}^{\mathrm{j}\omega}) = \omega n_0$. The time-delay n_0 can, thus, be determined by:

1. *Determining the frequency ω_0 using the position of the peak in $|C_{X_1X_2}(\mathrm{e}^{\mathrm{j}\omega})|$.*
2. *Calculating n_0 using the phase of the cross-spectrum at ω_0.*

$$n_0 = \frac{\angle C_{X_1X_2}(\mathrm{e}^{\mathrm{j}\omega})|_{\omega=\omega_0}}{\omega_0}$$

We have determined the time-delay between two random processes arriving at two sensors using the cross-spectrum.

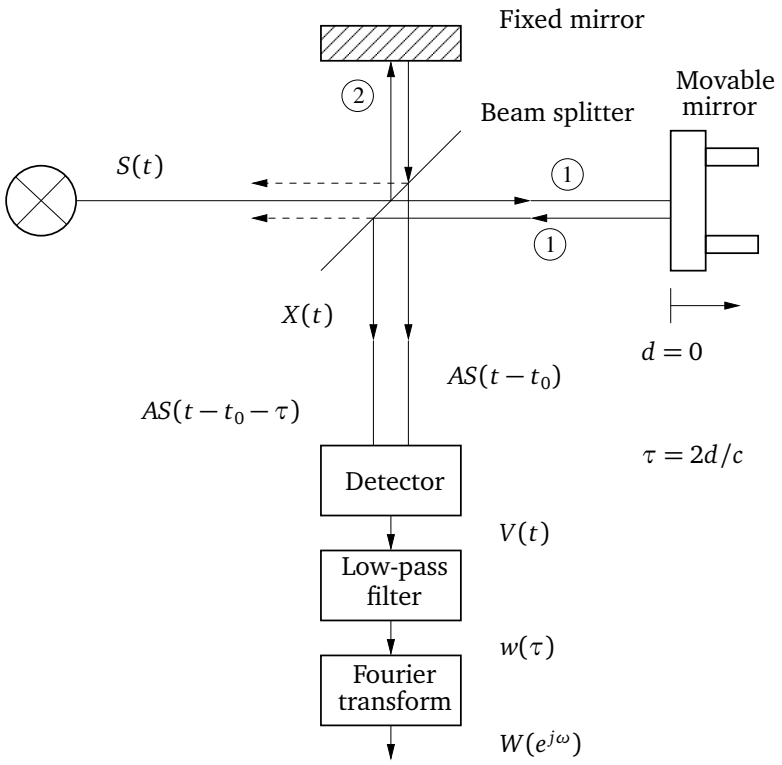


Figure 4.4: Measuring the spectrum of a light source using a Michelson interferometer.

4.3 The Michelson Interferometer

In spectroscopy we wish to investigate the distribution of power in a radiation field as a function of frequency (or wavelength/wavenumber). Consider, for example, the situation shown in Figure 4.4 where a Michelson interferometer is used to measure the spectrum of a light source. Let $S(t)$ be a zero-mean stochastic stationary process which may be complex-valued. It can be seen that

$$X(t) = A(S(t-t_0) + S(t-t_0-\tau))$$

and

$$V(t) = |X(t)|^2 = A^2|S(t-t_0) + S(t-t_0-\tau)|^2 .$$

Using the notation $\tilde{t}$ for $t-t_0$, we get

$$V(t) = A^2\left[S(\tilde{t})S(\tilde{t})^* + S(\tilde{t})S(\tilde{t}-\tau)^* + S(\tilde{t}-\tau)S(\tilde{t})^* + S(\tilde{t}-\tau)S(\tilde{t}-\tau)^*\right].$$

After low-pass filtering, we obtain (ideally)

$$\begin{aligned} w(\tau) = \mathsf{E}[V(t)] &= A^2\left[c_{SS}(0) + c_{SS}(\tau) + c_{SS}(-\tau) + c_{SS}(0)\right] \\ &= A^2\left[2c_{SS}(0) + 2\mathsf{Re}\{c_{SS}(\tau)\}\right] \end{aligned}$$

Performing a Fourier transform leads to

$$W(e^{j\omega}) = \mathcal{F}\{w(\tau)\} = A^2(2c_{SS}(0)\delta(\omega) + C_{SS}(e^{j\omega}) + C_{SS}(e^{-j\omega}))$$

Thus, the output of a Michelson interferometer is proportional to the spectrum of the light source. One of the objectives in spectroscopic interferometry is to get a good estimate of the spectrum of $S(t)$ without moving the mirror a large d.

Bibliography

J.F. Böhme. *Stochastische Signale*. Teubner Studienbücher, 1998.

A. Papoulis. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, Inc., third edition, 1991.

P. Z. Peebles, Jr. *Probability, Random Variables and Random Signal Principles*. McGraw-Hill, Inc., fourth edition, 2001.

5 Random Signals and Linear Time-Invariant Systems: FIR and IIR Filters

5.1 Convergence of Random Variables

The determination of asymptotic properties of deterministic variables is a well known concept. However, how do we approach convergence of random variables? Convergence of random variables can be defined in several ways, in the sequel, we consider different definitions of convergence of a sequence of random variables X_k, $k = 0, 1, 2, \ldots$. These are given in order of power.

1. **Convergence with probability one or almost surely convergence**
   A variable X_k converges to X with probability 1 if

$$\mathsf{P}\left(\lim_{k\to\infty} |X_k - X| = 0\right) = 1$$

2. **Convergence in the Mean Square Sense**
   A variable X_k converges to X in the MSS if

$$\lim_{k\to\infty} \mathsf{E}\left[|X_k - X|^2\right] = 0$$

3. **Convergence in Probability**
   X_k converges in probability to X if $\forall\ \epsilon > 0$

$$\lim_{k\to\infty} \mathsf{P}(|X_k - X| > \epsilon) = 0$$

4. **Convergence in Distribution**
   X_k converges in distribution to X (or X_k is asymptotically F_X distributed) if

$$\lim_{k\to\infty} F_{X_k}(x) = F_X(x)$$

   for all continuous points of F_X.

The convergences 1 to 4 are not equally weighted.

i) Convergence with probability 1 implies convergence in probability.

ii) Convergence with probability 1 implies convergence in the MSS, provided second order moments exist.

iii) Convergence in the MSS implies convergence in probability.

iv) Convergence in probability implies convergence in distribution.

With the concepts of convergence one may define continuity, differentiability in the stochastic sense and so on.

5.2 Introduction to Digital Filtering

5.2.1 General Concept of Filtering

Generally, filtering reduces or enhances certain aspects of a signal. In contrast to analog filter, digital filters work with sampled, discrete-time signals. There are two major types of digital filters: Finite impulse response (FIR) filters and infinite impulse response (IIR) filters. The frequency domain behavior of a filter is described mathematically in terms of its transfer function

$$H(e^{j\omega}) = \frac{Y(e^{j\omega})}{X(e^{j\omega})},$$

where $X(e^{j\omega})$ and $Y(e^{j\omega})$ are the Fourier transforms of deterministic signals; for wide-sense stationary signals, the frequency domain behavior of a filter is described mathematically in terms of the cross-spectrum and the spectrum

$$H(e^{j\omega}) = \frac{C_{YX}(e^{j\omega})}{C_{XX}(e^{j\omega})}.$$

Filtering is an essential tool that is applied everywhere, e.g., in speech and audio signal processing, such as, in radios, cellphones, etc., in image processing for denoising and in many more applications.

For $h(n)$ being real, it follows that $H(e^{j\omega}) = H(e^{-j\omega})^*$ so $H(e^{j\omega})$ is completely specified for $0 \leq \omega < \pi$.

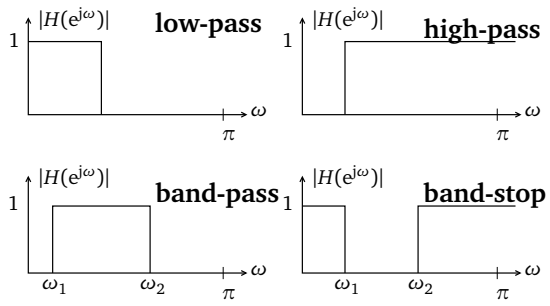


Figure 5.1: Four general types of filters.

An ideal low-pass filter has only a passband and a stopband. But in practice we also have a transition band and have to specify a tolerance in passband and stopband. Filter design is covered in 'Digital Signal Processing'.

5.2.2 Linear Constant Coefficient Difference Equation (LCCDE)

A linear time-invariant discrete-time system can be characterised by a **Linear Constant Coefficient Difference Equation** (LCCDE):

$$\sum_{k=0}^{M} a_k y(n-k) = \sum_{r=0}^{N-1} b_r x(n-r) \tag{5.1}$$

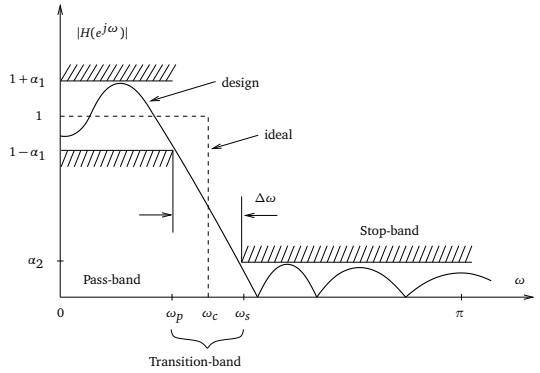


Figure 5.2: Specification of a low pass filter.

We distinguish between two different system responses:

FIR: Finite impulse response system has $M = 0$
$y(n) = \frac{1}{a_0} \sum_{r=0}^{N-1} b_r x(n-r)$

IIR: Infinite impulse response system has $M > 0$
$\sum_{k=0}^{M} a_k y(n-k) = \sum_{r=0}^{N-1} b_r x(n-r)$

5.3 Finite Impulse Response (FIR) Filter

For FIR filters ($M = 0$), the LCCDE simplifies to

$$y(n) = \frac{1}{a_0} \sum_{r=0}^{N-1} b_r x(n-r). \tag{5.2}$$

A comparison with the convolution sum

$$y(n) = \sum_{k=-\infty}^{\infty} h(k)x(n-k)$$

shows that

$$h(n) = \begin{cases} \frac{b_n}{a_0}, & n = 0, \ldots, N-1 \\ 0, & \text{otherwise.} \end{cases}$$

The input-output relationship is given by an FIR filter through the convolution sum

$$y(n) = \sum_{k=-\infty}^{\infty} h(k)x(n-k) = \sum_{k=0}^{N-1} h(k)x(n-k)$$

represented in the signal flow graph in Figure 5.3.

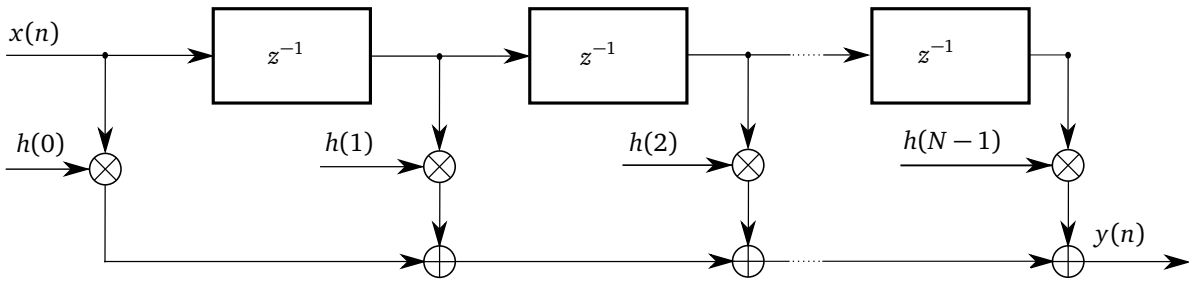


Figure 5.3: Signal flow graph of a FIR filter.

A digital filter with unit sample response $h(n)$ has a linear phase if

$$H(e^{j\omega}) = H_M(e^{j\omega})e^{-j\omega\alpha} \qquad |\omega| < \pi$$

with $H_M(e^{j\omega}) \in \mathbb{R}$ and α being a constant. With $H_M(e^{j\omega}) = 1$ follows $y(n) = h(n) * x(n) = x(n-\alpha)$, i.e. a linear phase preserves the shape of the signal.

5.3.1 Linear Phase Filters

Generalized linear phase system is characterized by

$$H(e^{j\omega}) = H_M(e^{j\omega})e^{-j\omega\alpha}e^{j\beta} = H_M(e^{j\omega})e^{-j\phi(\omega)},$$

where the phase can be expressed by $\phi(\omega) = -\beta + \omega\alpha$.

Which conditions must then hold true for $H_M(e^{j\omega}) \in \mathbb{R}$?

$$\begin{aligned} H_M(e^{j\omega}) &= H(e^{j\omega})e^{+j\phi(\omega)} \\ &= \sum_{n=-\infty}^{\infty} h(n)e^{-j\omega n + j\phi(\omega)} \\ &= \sum_{n=-\infty}^{\infty} h(n)[\cos(-\omega n - \beta + \omega\alpha) + j\sin(-\omega n - \beta + \omega\alpha)] \end{aligned}$$

It can easily be seen that the condition is met with

$$\sum_{n=-\infty}^{\infty} h(n)\sin[\omega(n-\alpha)+\beta] = 0 \;\; \forall\, \omega \in \mathbb{R}.$$

It is a necessary but not a sufficient condition on $h(n)$, α, and β for $H_M(e^{j\omega}) \in \mathbb{R}$. The condition in (5.3.1) can be achieved with either

$$\beta = 0 \;\text{ or }\; \pi, \;\; 2\alpha = N-1, \; h(2\alpha - n) = h(n)$$

or

$$\beta = \tfrac{\pi}{2} \;\text{ or }\; \tfrac{3\pi}{2}, \;\; 2\alpha = N-1, \;\; h(2\alpha - n) = -h(n).$$

There are two possible causal FIR Systems which meet these conditions:

- $h(n) = \begin{cases} h(N-1-n), & 0 \le n \le N-1, \\ 0, & \text{otherwise}, \end{cases}$

  then $H(e^{j\omega}) = H_e(e^{j\omega})e^{-j\omega(N-1)/2}$
- $h(n) = \begin{cases} -h(N-1-n), & 0 \le n \le N-1, \\ 0, & \text{otherwise}, \end{cases}$

  then $H(e^{j\omega}) = jH_o(e^{j\omega})e^{-j\omega(N-1)/2} = H_o(e^{j\omega})e^{-j\omega(N-1)/2+j\pi/2}$

5.3.2 Generalized Linear Phase for FIR Filters

We classify FIR filters into four types, depending on whether $h(n)$ is symmetric ($h(n) = h(N-1-n)$) or anti-symmetric ($h(n) = -h(N-1-n)$), and whether the length of the unit sample response is odd or even.

Type I: symmetric unit sample response and N is odd.
Type II: symmetric unit sample response and N is even.
Type III: antisymmetric unit sample response and N is odd.
Type IV: antisymmetric unit sample response and N is even.

Table 5.1 shows an overview of all four types and Figure 5.4 shows an example for each one.
In the following, we only investigate **Type I** filters

5.3.3 Implementation of FIR Filters

The input-output relationship is given by a FIR filter through the convolution sum

$$y(n) = \sum_{k=-\infty}^{\infty} h(k)x(n-k) = \sum_{k=0}^{N-1} h(k)x(n-k),$$

represented in the signal flow graph.

Type	Symmetry of $h(n)$	N	β	Form of $H_M(e^{j\omega})$
I	$h(n)=h(N-1-n)$ symmetric	odd	0 π	$\sum_{n=0}^{\frac{N-1}{2}} a(n)\cdot\cos\omega n$
II	$h(n)=h(N-1-n)$ symmetric	even	0 π	$\sum_{n=1}^{\frac{N}{2}} b(n)\cdot\cos\omega(n-\frac{1}{2})$
III	$h(n)=-h(N-1-n)$ anti-symmetric	odd	$\pi/2$ $3\pi/2$	$\sum_{n=1}^{\frac{N-1}{2}} c(n)\cdot\sin\omega n$
IV	$h(n)=-h(N-1-n)$ anti-symmetric	even	$\pi/2$ $3\pi/2$	$\sum_{n=1}^{\frac{N}{2}} d(n)\cdot\sin\omega(n-\frac{1}{2})$

Table 5.1: Overview of generalized linear phase FIR filters.

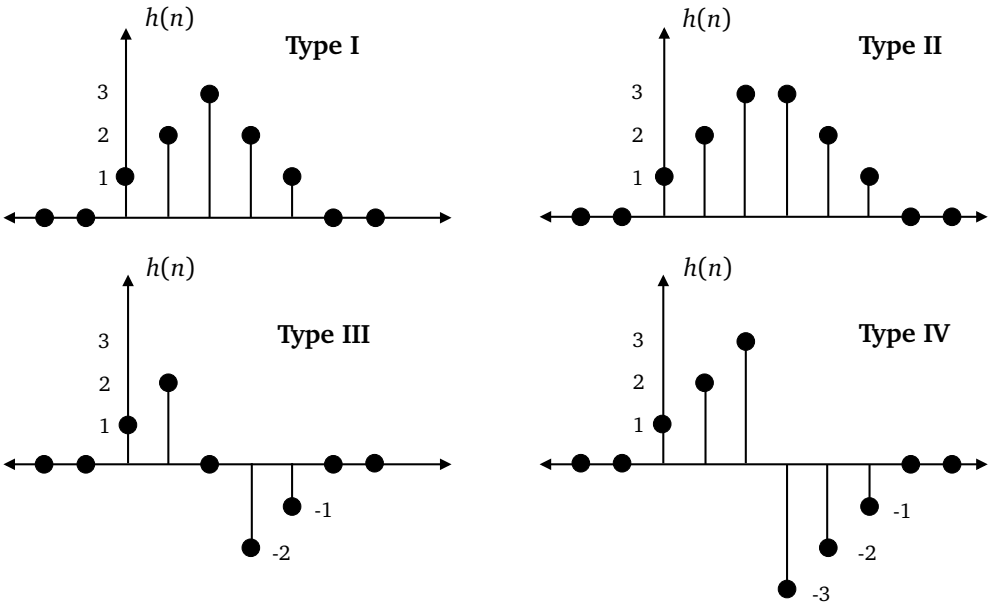
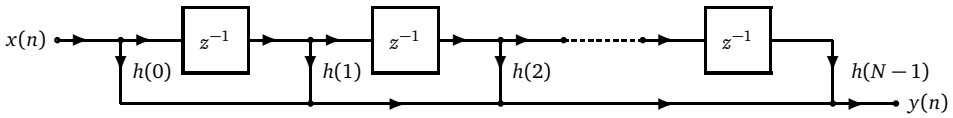


Figure 5.4: Examples of generalized linear phase FIR filters.



5.3.4 Type I FIR linear Phase Filter

The **Type I** FIR linear phase filter has a symmetric unit sample response and its length N is odd,

$$h(n)=h(N-1-n), \qquad 0\le n\le N-1.$$

Its frequency response is given by

$$H(e^{j\omega})=\sum_{n=0}^{N-1}h(n)e^{-j\omega n}=e^{-j\omega\frac{N-1}{2}}\sum_{k=0}^{(N-1)/2}a(k)\cos(\omega k),$$

where $a(0)=h(\frac{N-1}{2})$ and $a(k)=2h(\frac{N-1}{2}-k)$, $k=1,2,\ldots,\frac{N-1}{2}$.

Example 5.3.1 *Example of a type I FIR filter.*

Given a filter with unit sample response

$$h(n) = \begin{cases} 1, & 0 \le n \le 4 \\ 0, & otherwise \end{cases},$$

we can evaluate the frequency response to be

$$H(e^{j\omega}) = \sum_{n=0}^{4} e^{-j\omega n} = \frac{1-e^{-j5\omega}}{1-e^{-j\omega}} = e^{-j2\omega}\frac{\sin\left(5\frac{\omega}{2}\right)}{\sin\left(\frac{\omega}{2}\right)}.$$

Note:

$$\sum_{n=0}^{N-1} q^n = \frac{1-q^N}{1-q} \text{ for } q \neq 1$$

Figure 5.5 shows the magnitude and phase of the unit sample response.

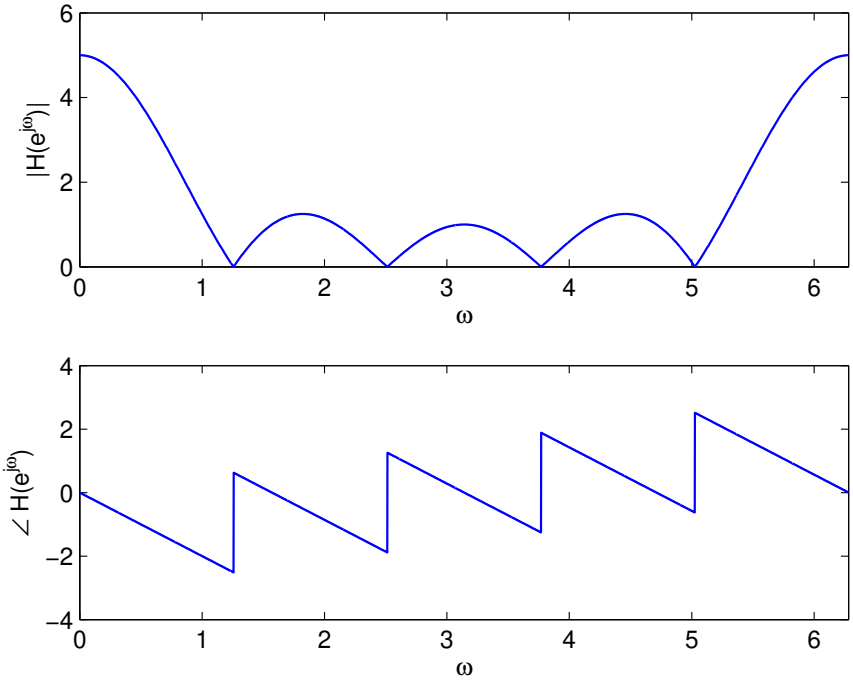


Figure 5.5: Magnitude and phase of the unit sample response.

5.4 Infinite Impulse Response (IIR) Filter

Infinite impulse response (IIR) system have $M > 0$ in Equation (5.1),

$$\sum_{k=0}^{M} a_k y(n-k) = \sum_{r=0}^{N-1} b_r x(n-r). \tag{5.3}$$

In the following, we restrict ourselves to the class of filters, which have a unit sample response $h(n)$ that is real, causal, and satisfies stability. Furthermore, the filter must possess a rational z-transform with $a_0 = 1$, i.e.,

$$H(z) = \frac{\sum_{r=0}^{N-1} b_r z^{-r}}{1 + \sum_{k=1}^{M} a_k z^{-k}}.$$

5.4.1 Stability

Stability of an IIR filter is given, if the roots of the transfer function

$$1 + \sum_{k=1}^{M} a_k z^{-k} = 0$$

lie strictly inside the unit circle, i.e. if $|z| < 1$ is fulfilled.

Example 5.4.1 *For $M = 1$, the denominator reduces to*

$$1 + a_1 z^{-1} = 0,$$

where stability is given if $|z| = |a_1| < 1$.

5.4.2 Implementation

In Figure 5.6, the implementation of IIR filters is schematically shown by a signal flow graph of direct form I structure.

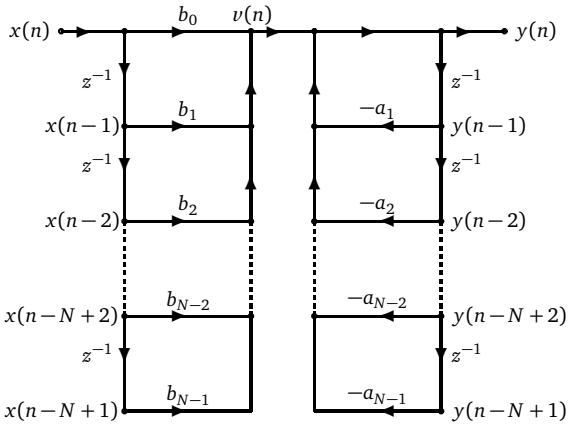


Figure 5.6: Signal flow graph of a direct form I structure for an Mth-order system ($N-1=M$).

Example 5.4.2 *The filter with the unit sample response*

$$h(n) = \left(\frac{1}{2}\right)^n \cdot u(n),$$

with $H(z)$ given by

$$H(z) = \frac{1}{1 - \frac{1}{2}z^{-1}}.$$

Can be realised by

$$y(n) = \frac{1}{2}y(n-1) + x(n).$$

5.4.3 Comparison to FIR Filters

When comparing IIR filter to FIR filter, three main properties can be distinguished:

- **Stability**: With an IIR filter, all the poles of $H(z)$ must lie inside the unit circle to ensure filter stability. FIR filters are stable by definition because $H(e^{j\omega})$ has no poles.
- **Linear phase**: It is very difficult to control the phase of an IIR filter. With FIR filter a perfect linear phase is easily achieved.
- **Complexity**: An IIR filter will meet a given amplitude response specification more efficiently than an FIR filter. As a consequence the complexity of IIR filters is lower.

5.5 Linear Filtering

Assume that we are given two stationary processes $X(n)$ and $Y(n)$ and a linear time-invariant (LTI) system with unit sample response $h(n)$, where the filter is stable as in Figure 5.7. A filter is stable in that $\sum|h(n)| < \infty$. Then for such a system

$$Y(n) = \sum_k h(k)X(n-k) = \sum_k h(n-k)X(k)$$

exists with probability one i.e. $\sum_{k=-N}^{M} h(k)X(n-k)$ converges with probability one to $Y(n)$ for $N, M \to \infty$. A sequence X_k converges to X with probability 1, if $\mathsf{P}(\lim_{k\to\infty}|\mathsf{X_k} - \mathsf{X}| = 0) = 1$.

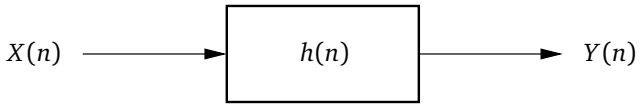


Figure 5.7: Filtering of $X(n)$ with an LTI system of unit sample response $h(n)$.

If $X(n)$ is stationary and $\mathsf{E}[|X(n)|] < \infty$, then $Y(n)$ is stationary. If $X(n)$ is white noise, i.e. $\mathsf{E}[X(n)] = 0$, $\mathsf{E}[X(n)X(k)] = \sigma_X^2\delta(n-k)$, with σ_X^2 the power of the noise and $\delta(n)$ the Kronecker's delta function

$$\delta(n) = \begin{cases} 1, & n = 0 \\ 0, & \text{else} \end{cases}$$

$Y(n)$ is called a linear process.

If we do not require the filter to be stable but $\int |H(e^{j\omega})|^2\, d\omega < \infty$, where

$$H(e^{j\omega}) = \sum_n h(n)e^{-j\omega n}$$

is the transfer function of the system, and if for $X(n)$, $\sum|c_{XX}(n)| < \infty$, then

$$Y(n) = \sum_k h(k)X(n-k) = \sum_k h(n-k)X(k)$$

exists in the MSS, i.e,

$$\sum_{k=-N}^{M} h(k)X(n-k)$$

converges in the MSS for $N, M \to \infty$ and $Y(n)$ is stationary in the wide sense with

$$\mu_Y \triangleq \mathsf{E}[Y(n)] = \sum_k h(k)\mathsf{E}[X(n-k)] = \mu_X H(e^{j\omega})|_{\omega=0}$$

Example 5.5.1 Hilbert Transforms

$$\begin{aligned} H(e^{j\omega}) &= \begin{cases} -j, & 0 < \omega < \pi \\ 0, & \omega = 0, -\pi \\ j, & -\pi < \omega < 0 \end{cases} \\ h(n) &= \begin{cases} 0, & n = 0 \\ (1-\cos(\pi n))/(\pi n), & \text{else} \end{cases} \end{aligned}$$

The filter is not stable, but $\int |H(e^{j\omega})|^2\, d\omega < \infty$. *$Y(n)$ is stationary in the wide sense. For example if* $X(n) = a\cos(\omega_o n + \phi)$, *then* $Y(n) = a\sin(\omega_o n + \phi)$.

Example 5.5.2 *Real-world application example: Before a sound reaches a person's eardrum, it is reflected and defracted by several objects, mainly: the head, the torso and the outer ear. These effects can be modelled by a linear time-invariant system which has a so-called Head-Related-Room-Impulse-Response or in the frequency domain a Head-Related-Transfer-Function (HRTF). Hence, an HRTF models the auditory space by characterizing how sound travels from a certain sound source to a listeners ear. HRTFs are used to process audio data in commercial products in order to produce a surround sound effect from only two speakers. For that and many other applications, HRTFs are, typically, measured from dummy-heads for both the left and right ears at a fixed radius from the head.*

We now determine the relationship between the covariance function and the spectrum, at the input and at the output. First, let us define

$$\begin{aligned}\tilde{Y}(n) \triangleq Y(n)-\mu_Y &= \sum_k h(k)X(n-k)-\sum_k h(k)\mathsf{E}[X(n-k)] \\ &= \sum_k h(k)\left(X(n-k)-\mu_X\right) \\ &= \sum_k h(k)\tilde{X}(n-k)\end{aligned}$$

where $\tilde{X}(n) \triangleq X(n)-\mu_X$. In the general case, the system unit sample response and the input process may be complex.

5.5.1 Output Covariance Function and Spectrum

The covariance function of the output is determined according to

$$\begin{aligned}c_{YY}(\kappa) &= \mathsf{E}\left[\tilde{Y}(n+\kappa)\tilde{Y}(n)^*\right] \\ &= \mathsf{E}\left[\left(\sum_k h(k)\tilde{X}(n+\kappa-k)\right)\left(\sum_l h(l)\tilde{X}(n-l)\right)^*\right] \\ &= \sum_k\sum_l h(k)h(l)^*\mathsf{E}\left[\tilde{X}(n+\kappa-k)\tilde{X}(n-l)^*\right]\end{aligned}$$

$$\boxed{c_{YY}(\kappa)=\sum_k\sum_l h(k)h(l)^*c_{XX}(\kappa-k+l)}$$

Taking the Fourier transform of both sides leads to

$$\boxed{C_{YY}(e^{j\omega})=|H(e^{j\omega})|^2C_{XX}(e^{j\omega})} \tag{5.4}$$

and applying the inverse Fourier transform formula yields[1]

$$c_{YY}(\kappa)=\int_{-\pi}^{\pi}|H(e^{j\omega})|^2C_{XX}(e^{j\omega})e^{j\omega\kappa}\,\frac{d\omega}{2\pi}$$

Example 5.5.3 FIR filter
Let $X(n)$ be white noise with power σ_X^2 and

$$Y(n)=X(n)-X(n-1).$$

From the above, one can see that

$$h(n)=\begin{cases}1, & n=0\\ -1, & n=1\\ 0, & \text{else}\end{cases}$$

[1] The integral exists because $C_{XX}(e^{j\omega})$ is bounded and $\int|H(e^{j\omega})|^2\,d\omega<\infty$.

Taking the Fourier transform, leads to

$$H(\mathrm{e}^{\mathrm{j}\omega}) = \sum_{n=-\infty}^{\infty} h(n)e^{-\mathrm{j}\omega n} = 1 - e^{-\mathrm{j}\omega} = e^{-\mathrm{j}\omega/2} \cdot 2\mathrm{j}\sin(\frac{\omega}{2})$$

and thus,

$$\begin{aligned} C_{YY}(\mathrm{e}^{\mathrm{j}\omega}) &= |H(\mathrm{e}^{\mathrm{j}\omega})|^2 C_{XX}(\mathrm{e}^{\mathrm{j}\omega}) = |1 - e^{-\mathrm{j}\omega}|^2 \sigma_X^2 \\ &= |2\mathrm{j}e^{-\mathrm{j}\omega/2}\sin(\omega/2)|^2 \sigma_X^2 \\ &= 4\sin(\omega/2)^2 \sigma_X^2 \end{aligned}$$

Note: $1 - e^{-j\omega} = e^{-j\omega/2} \cdot (e^{+j\omega/2} - e^{-j\omega/2}) = e^{-j\omega/2} \cdot 2j\sin(\omega/2)$

5.5.2 Cross-Covariance Function and Cross-Spectrum

The cross-covariance function of the output is determined according to

$$\begin{aligned} c_{YX}(\kappa) &= \mathsf{E}\left[\tilde{Y}(n+\kappa)\tilde{X}(n)^*\right] \\ &= \mathsf{E}\left[\left(\sum_k h(k)\tilde{X}(n+\kappa-k)\right)\tilde{X}(n)^*\right] \\ &= \sum_k h(k)\mathsf{E}\left[\tilde{X}(n+\kappa-k)\tilde{X}(n)^*\right] \end{aligned}$$

$$\boxed{c_{YX}(\kappa) = \sum_k h(k)c_{XX}(\kappa-k)}$$

Taking the Fourier transform of both sides, leads to

$$\boxed{C_{YX}(\mathrm{e}^{\mathrm{j}\omega}) = H(\mathrm{e}^{\mathrm{j}\omega})C_{XX}(\mathrm{e}^{\mathrm{j}\omega})}$$

and applying the inverse Fourier transform formula yields

$$c_{YX}(\kappa) = \int_{-\pi}^{\pi} H(\mathrm{e}^{\mathrm{j}\omega})C_{XX}(\mathrm{e}^{\mathrm{j}\omega})\mathrm{e}^{\mathrm{j}\omega\kappa}\,\frac{d\omega}{2\pi}.$$

Example 5.5.4 *Consider a system as depicted in Figure 5.7 where $X(n)$ is a stationary, zero-mean white noise process with variance σ_X^2. The unit sample response of the system is given by*

$$h(n) = \left[\frac{1}{4}\left(\frac{1}{6}\right)^n + \frac{3}{4}\left(-\frac{1}{2}\right)^n\right]u(n)$$

where $u(n)$ is the unit step function

$$u(n) = \begin{cases} 1, & n \geq 0 \\ 0, & n < 0 \end{cases}$$

and we wish to determine $C_{YY}(\mathrm{e}^{\mathrm{j}\omega})$ and $C_{YX}(\mathrm{e}^{\mathrm{j}\omega})$.

The system transfer function is given by

$$\begin{aligned} H(\mathrm{e}^{\mathrm{j}\omega}) &= \mathcal{F}\{h(n)\} = \frac{1/4}{1 - \frac{1}{6}\mathrm{e}^{-\mathrm{j}\omega}} + \frac{3/4}{1 + \frac{1}{2}\mathrm{e}^{-\mathrm{j}\omega}} \\ &= \frac{1}{1 + \frac{1}{3}\mathrm{e}^{\mathrm{j}\omega} - \frac{1}{12}\mathrm{e}^{\mathrm{j}2\omega}} \end{aligned}$$

The spectrum and cross-spectrum are given respectively by

$$\begin{aligned} C_{YY}(\mathrm{e}^{\mathrm{j}\omega}) &= |H(\mathrm{e}^{\mathrm{j}\omega})|^2 C_{XX}(\mathrm{e}^{\mathrm{j}\omega}) \\ &= \frac{\sigma_X^2}{\left|1 + \frac{1}{3}\mathrm{e}^{\mathrm{j}\omega} - \frac{1}{12}\mathrm{e}^{\mathrm{j}2\omega}\right|^2} \end{aligned}$$

$$\begin{aligned} C_{YX}(\mathrm{e}^{\mathrm{j}\omega}) &= H(\mathrm{e}^{\mathrm{j}\omega})C_{XX}(\mathrm{e}^{\mathrm{j}\omega}) \\ &= \frac{\sigma_X^2}{1 + \frac{1}{3}\mathrm{e}^{\mathrm{j}\omega} - \frac{1}{12}\mathrm{e}^{\mathrm{j}2\omega}} \end{aligned}$$

The spectrum of the input and output can be seen in Figure 5.8.

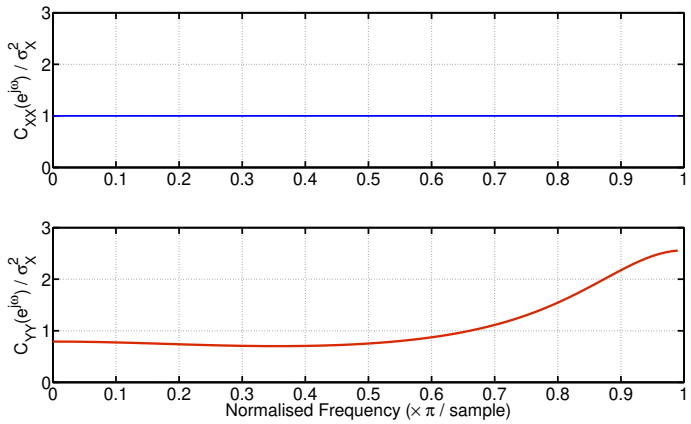


Figure 5.8: Spectrum of the input and output of the linear filter with frequency response $H(e^{j\omega}) = \frac{1}{1+\frac{1}{4}e^{j\omega}-\frac{1}{8}e^{j2\omega}}$.

5.5.3 Interaction of Two Linear Systems

The previous result may be generalized to obtain the cross-covariance or cross-spectrum of two linear systems. Let $X_1(n)$ and $Y_1(n)$ be respectively the input and output processes of the first system, with unit sample response $h_1(n)$, and $X_2(n)$ and $Y_2(n)$ be respectively the input and output processes of the second system with unit sample response $h_2(n)$ (see Figure 5.9). Suppose that $X_1(n)$ and $X_2(n)$ are wide-sense stationary and the systems are time-invariant and linear. Then the output cross-covariance function is:

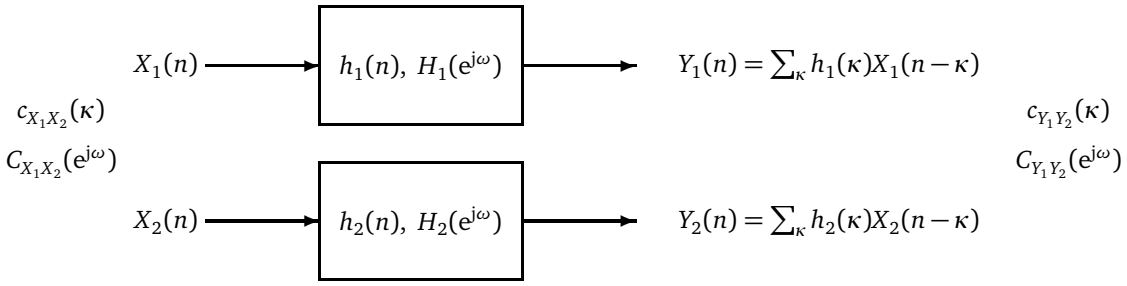


Figure 5.9: Interaction of two linear systems.

$$c_{Y_1Y_2}(\kappa) = \sum_k \sum_l h_1(k)h_2(l)^* c_{X_1X_2}(\kappa - k + l) \tag{5.5}$$

It can be seen from Equation (5.5) that

$$c_{Y_1Y_2}(\kappa) = h_1(\kappa) \star h_2(-\kappa)^* \star c_{X_1X_2}(\kappa)$$

where $\star$ denotes convolution. Taking the Fourier transform leads to

$$C_{Y_1Y_2}(e^{j\omega}) = H_1(e^{j\omega})H_2(e^{j\omega})^* C_{X_1X_2}(e^{j\omega}). \tag{5.6}$$

Example 5.5.5 *Consider the system shown in Figure 5.10 where $X(n)$ is stationary, zero-mean white noise with variance σ_X^2. Let $h_1(n)$ be the same system as discussed in Example 5.5.4 and $h_2(n)$ is given by*

$$h_2(n) = \delta(n) + \frac{1}{2}\delta(n-1) - \frac{1}{2}\delta(n-2)$$

where $\delta(n)$ is the Kronecker delta function. We wish to find the cross-spectrum $C_{Y_1Y_2}(e^{j\omega})$.

$$H_2(e^{j\omega}) = \sum_n h_2(n)e^{-j\omega n} = 1 + \frac{1}{2}e^{-j\omega} - \frac{1}{2}e^{-j2\omega}$$

$$C_{Y_1Y_2}(e^{j\omega}) = H_1(e^{j\omega})H_2(e^{j\omega})^* C_{XX}(e^{j\omega}) = \left[\frac{1 + \frac{1}{2}e^{j\omega} - \frac{1}{2}e^{j2\omega}}{1 + \frac{1}{3}e^{j\omega} - \frac{1}{12}e^{j2\omega}}\right]\sigma_X^2$$

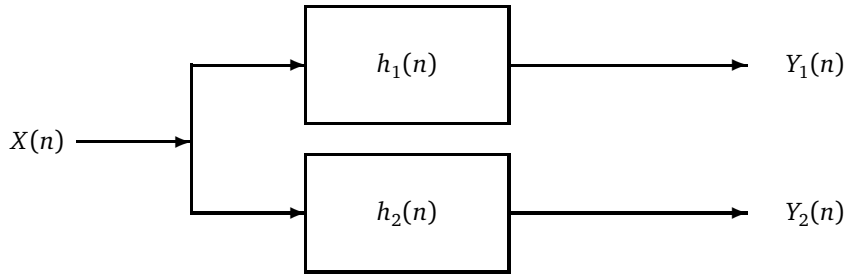


Figure 5.10: Interaction of two linear systems.

5.5.4 Cascade of Linear Systems

We consider a cascade of L linear systems in serial connection, as illustrated by Figure 5.11. From linear system theory, it is known that a cascade of linear systems can be equivalently represented by a system whose transfer function $H(e^{j\omega})$ is the product of the L individual system functions

$$H(e^{j\omega}) = H_1(e^{j\omega})H_2(e^{j\omega})\cdots H_L(e^{j\omega}).$$

When a random process is the input to such a system, the spectrum of the output and cross spectrum of the output with the input are given respectively by

$$\begin{aligned} C_{YY}(e^{j\omega}) &= C_{XX}(e^{j\omega})\left|\prod_{i=1}^{L} H_i(e^{j\omega})\right|^2 = C_{XX}(e^{j\omega})\prod_{i=1}^{L}\left|H_i(e^{j\omega})\right|^2, \\ C_{YX}(e^{j\omega}) &= C_{XX}(e^{j\omega})\prod_{i=1}^{L} H_i(e^{j\omega}). \end{aligned}$$

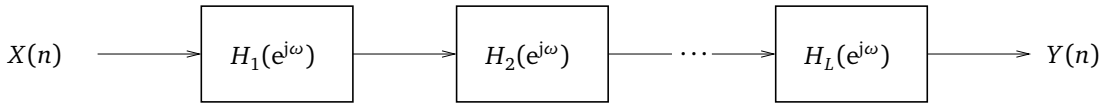


Figure 5.11: Cascade of L linear systems.

Example 5.5.6 *We consider again the system shown in Figure 5.10 where we wish to find the cross spectrum $C_{Y_1Y_2}(e^{j\omega})$. Instead of using Equation (5.6) with $X_1(n) = X_2(n) = X(n)$ as done in Example 5.5.5, we reformulate the problem as a cascade of two systems as shown in Figure 5.12.*

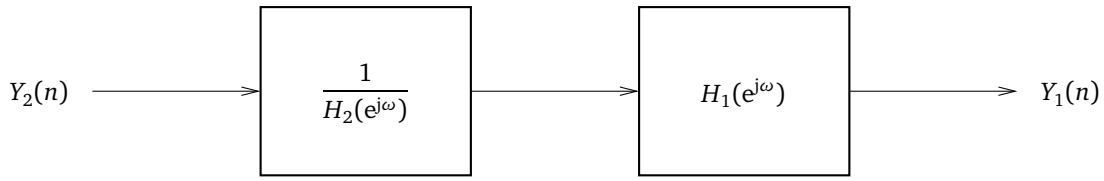


Figure 5.12: Cascade of two linear systems.

Remembering that $C_{Y_2Y_2}(e^{j\omega}) = |H_2(e^{j\omega})|^2 C_{XX}(e^{j\omega})$ (from Figure 5.10 and Equation (5.4)) we obtain

$$\begin{aligned} C_{Y_1Y_2}(e^{j\omega}) &= \frac{H_1(e^{j\omega})}{H_2(e^{j\omega})} C_{Y_2Y_2}(e^{j\omega}) \\ &= \frac{H_1(e^{j\omega})}{H_2(e^{j\omega})} |H_2(e^{j\omega})|^2 C_{XX}(e^{j\omega}) \\ &= H_1(e^{j\omega})H_2(e^{j\omega})^* C_{XX}(e^{j\omega}) \end{aligned}$$

Bibliography

J.F. Böhme. *Stochastische Signale*. Teubner Studienbücher, 1998.

A. Papoulis. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, Inc., third edition, 1991.

P. Z. Peebles, Jr. *Probability, Random Variables and Random Signal Principles*. McGraw-Hill, Inc., fourth edition, 2001.

6 Filtered Noise Processes: AR, MA, and ARMA Models

6.1 Linear Filtering of Random Processes Plus Noise

Consider the system shown in Figure 6.1. It is assumed that $X(n)$ is stationary with $\mathsf{E}[X(n)] = 0$, $V(n)$ stationary with $\mathsf{E}[V(n)] = 0$ and $c_{VX}(n) = 0$ ($X(n)$ and $V(n)$ uncorrelated). It is also assumed without loss of generality that $h(n)$, $X(n)$ and $V(n)$ are real-valued.

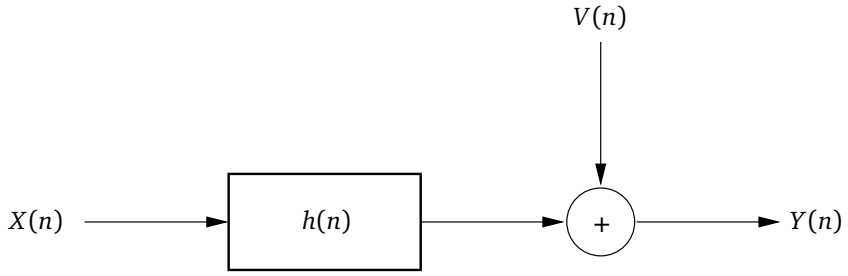


Figure 6.1: A linear filter plus noise.

The expected value of the output is derived below.

$$\begin{aligned} Y(n) &= \sum_{k=-\infty}^{\infty} h(k)X(n-k) + V(n) \\ \mathsf{E}[Y(n)] &= \sum_{k=-\infty}^{\infty} h(k)\mathsf{E}[X(n-k)] + \mathsf{E}[V(n)] = 0 \end{aligned}$$

6.1.1 Output Covariance Function and Spectrum

The covariance function of the output is determined according to

$$\begin{aligned} c_{YY}(\kappa) &= \mathsf{E}[Y(n+\kappa)Y(n)] \\ &= \mathsf{E}\left[\left(\sum_k h(k)X(n+\kappa-k) + V(n+\kappa)\right)\left(\sum_l h(l)X(n-l) + V(n)\right)\right] \\ &= \sum_k \sum_l h(k)h(l)\mathsf{E}[X(n+\kappa-k)X(n-l)] \\ &+ \sum_k h(k)\mathsf{E}[X(n+\kappa-k)V(n)] \\ &+ \sum_l h(l)\mathsf{E}[V(n+\kappa)X(n-l)] \\ &+ \mathsf{E}[V(n+\kappa)V(n)] \end{aligned}$$

$$\boxed{c_{YY}(\kappa) = \sum_k \sum_l h(k)h(l)c_{XX}(\kappa-k+l) + c_{VV}(\kappa)}$$

Applying the Fourier transform to both sides of the above expression yields

$$\boxed{C_{YY}(e^{j\omega}) = \mathcal{F}\{c_{YY}(\kappa)\} = |H(e^{j\omega})|^2 C_{XX}(e^{j\omega}) + C_{VV}(e^{j\omega})}$$

6.1.2 Cross-Covariance Function and Cross-Spectrum

The cross-covariance function of the output is determined according to

$$\begin{aligned} c_{YX}(\kappa) &= \mathsf{E}[Y(n+\kappa)X(n)] \\ &= \mathsf{E}\left[\left(\sum_k h(k)X(n+\kappa-k)+V(n+\kappa)\right)X(n)\right] \\ &= \sum_k h(k)\mathsf{E}[X(n+\kappa-k)X(n)] \end{aligned}$$

$$\boxed{c_{YX}(\kappa)=\sum_k h(k)c_{XX}(\kappa-k)}$$

Applying the Fourier transform to both sides of the above expression yields

$$\boxed{C_{YX}(\mathrm{e}^{\mathrm{j}\omega})=H(\mathrm{e}^{\mathrm{j}\omega})C_{XX}(\mathrm{e}^{\mathrm{j}\omega})}$$

Note that when the additive noise at the output, $V(n)$ is uncorrelated with the input process $X(n)$, it has no effect on the cross-spectrum of the input and output.

Example 6.1.1 *A complex-valued process $X(n)$ is comprised of a sum of K complex signals*

$$X(n)=\sum_{k=1}^{K} A_k \mathrm{e}^{\mathrm{j}(\omega_0 n+\phi_k)}$$

where ω_0 is a constant and A_k is a zero-mean random amplitude of the kth signal with variance σ_k^2, $k=1,2,\ldots,K$. The phase ϕ_k is uniformly distributed on $[-\pi,\pi)$. The random variables A_k and ϕ_k are statistically pairwise independent for all $k=1,\ldots,K$.

$X(n)$ is the input to a linear, stable time-invariant system with known impulse response $h(n)$. The output of the system is buried in zero-mean noise $V(n)$ of known covariance $c_{VV}(n)$, independent of $Z(n)$ (refer to Figure 6.1). We wish to determine the cross-spectrum $C_{YX}(\mathrm{e}^{\mathrm{j}\omega})$ and the auto-spectrum $C_{YY}(\mathrm{e}^{\mathrm{j}\omega})$.

We first calculate the covariance function of the input process $X(n)$

$$\mathsf{E}[X(n)] = \mathsf{E}\left[\textstyle\sum_{k=1}^{K} A_k \mathrm{e}^{\mathrm{j}(\omega_0 n+\phi_k)}\right]$$

and since A_k and ϕ_k are independent for all $k=1,\ldots,K$

$$\mathsf{E}[X(n)] = \sum_{k=1}^{K}\mathsf{E}[A_k]\mathsf{E}\left[\mathrm{e}^{\mathrm{j}(\omega_0 n+\phi_k)}\right]=0$$

$$c_{XX}(n+\kappa,n)=\mathsf{E}[X(n+\kappa)X(n)^*]$$

$$\begin{aligned} \mathsf{E}[X(n+\kappa)X(n)^*] &= \mathsf{E}\left[\sum_{k=1}^{K} A_k \mathrm{e}^{\mathrm{j}(\omega_0(n+\kappa)+\phi_k)}\sum_{l=1}^{K} A_l \mathrm{e}^{-\mathrm{j}(\omega_0 n+\phi_l)}\right] \\ &= \sum_{k=1}^{K}\sum_{l=1}^{K}\mathsf{E}[A_k A_l]\mathrm{e}^{\mathrm{j}\omega_0\kappa}\mathsf{E}\left[\mathrm{e}^{\mathrm{j}(\phi_k-\phi_l)}\right] \end{aligned}$$

$$\textit{Since}\quad \mathsf{E}\left[\mathrm{e}^{\mathrm{j}(\phi_k-\phi_l)}\right]=\begin{cases} 1, & k=l \\ 0, & k\neq l \end{cases}$$

$$c_{XX}(n+\kappa,n)=\sum_{k=1}^{K}\mathsf{E}\left[A_k^2\right]\mathrm{e}^{\mathrm{j}\omega_0\kappa}=\mathrm{e}^{\mathrm{j}\omega_0\kappa}\sum_{k=1}^{K}\sigma_k^2=c_{XX}(\kappa)$$

$$
\begin{aligned}
C_{YX}(e^{j\omega}) &= H(e^{j\omega})C_{XX}(e^{j\omega}) \\
&= H(e^{j\omega})\mathcal{F}\{c_{XX}(\kappa)\} \\
&= H(e^{j\omega})\left[\sum_{k=1}^{K}\sigma_k^2\right]2\pi\sum_{l}\delta(\omega+\omega_0+2\pi l) \\
&= \left[\sum_{k=1}^{K}\sigma_k^2\right]2\pi\sum_{l}H(e^{j(\omega_0-2\pi l)})\delta(\omega-\omega_0+2\pi l)
\end{aligned}
$$

$$
\begin{aligned}
C_{YY}(e^{j\omega}) &= |H(e^{j\omega})|^2 C_{XX}(e^{j\omega}) + C_{VV}(e^{j\omega}) \\
&= \left[\sum_{k=1}^{K}\sigma_k^2\right]2\pi\sum_{l}|H(e^{j(\omega_0-2\pi l)})|^2\delta(\omega-\omega_0+2\pi l) + C_{VV}(e^{j\omega})
\end{aligned}
$$

Example 6.1.2 *Multipath propagation is an important characteristic of mobile communications that results in the transmitted signal reaching the receiver by more than one path. This can be caused by reflections of the signal from surrounding objects, such as buildings or mountains or even water surfaces. Figure 6.2 shows the scenario of multi-path propagation where the transmitter transmits a signal that arrives at the receiver via different paths d_0, d_1 and d_2, leading to different arrival time delays k_0, k_1 and k_2. The signals on the different paths get attenuated with a_0, a_1 and a_2, leading to a received signal $Y(n) = a_0X(n-k_0) + a_1X(n-k_1) + a_2X(n-k2)$.*

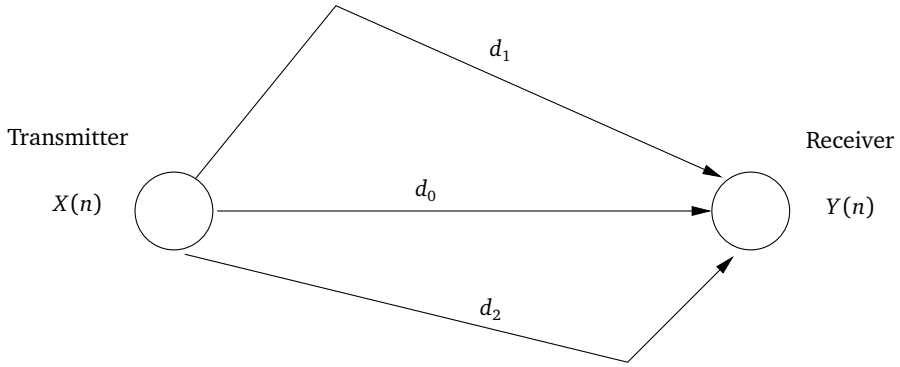


Figure 6.2: An example of multi-path propagation in mobile communications.

We will consider a two-path scenario where $a_0 = 1, k_0 = 0, a_2 = 0$. The transmitted signal $X(n)$ is assumed to be stochastic, stationary and zero-mean. The transmitter and the receiver produce stationary, pairwise uncorrelated, white noise, that can be modeled by $V(n)$ and $U(n)$ with respective variances σ_V^2 and σ_U^2. $X(n), V(n)$ and $U(n)$ are assumed to be pairwise uncorrelated. This scenario can be modeled by a linear, time-invariant system with unit sample response: $h(n) = \delta(n) + a_1\delta(n-k_1)$. The block diagram in Figure 6.3 depicts this scenario.

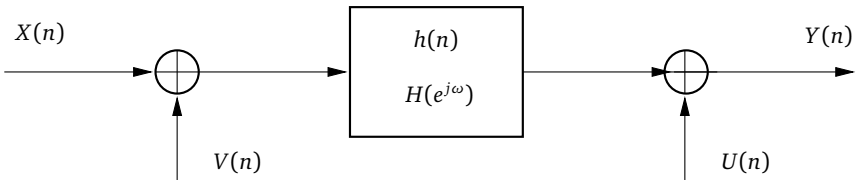


Figure 6.3: Model of a multi-path propagation scenario.

According to Figure 6.3, the received signal is given by $Y(n) = (X(n) + V(n)) \star h(n) + U(n)$. The transfer function of the system is: $H(e^{j\omega}) = 1 + a_1e^{-j\omega k_1}$. The spectrum at the output is given by:

$$
\begin{aligned}
C_{YY}(e^{j\omega}) &= |H(e^{j\omega})|^2(C_{XX}(e^{j\omega}) + C_{VV}(e^{j\omega})) + C_{UU}(e^{j\omega}) \\
&= |H(e^{j\omega})|^2 C_{XX}(e^{j\omega}) + |H(e^{j\omega})|^2 C_{VV}(e^{j\omega}) + C_{UU}(e^{j\omega}) \\
&= |1 + a_1e^{-j\omega k_1}|^2 C_{XX}(e^{j\omega}) + |1 + a_1e^{-j\omega k_1}|^2\sigma_V^2 + \sigma_U^2
\end{aligned}
$$

$$C_{YY}(e^{j\omega}) = (1+a_1^2)C_{XX}(e^{j\omega}) + a_1(C_{XX}(e^{j\omega})e^{-j\omega k_1} + C_{XX}(e^{j\omega})e^{j\omega k_1}) + (1+a_1^2)\sigma_V^2 + a_1(\sigma_V^2 e^{-j\omega k_1} + \sigma_V^2 e^{j\omega k_1}) + \sigma_U^2.$$

Let us now assume negligibly small noise variances, then the output spectrum becomes:

$$C_{YY}(e^{j\omega}) \approx (1+a_1^2)C_{XX}(e^{j\omega}) + a_1(C_{XX}(e^{j\omega})e^{-j\omega k_1} + C_{XX}(e^{j\omega})e^{j\omega k_1}).$$

and the inverse Fourier transform leads to:

$$c_{YY}(k) = (1+a_1^2)c_{XX}(k) + a_1(c_{XX}(k-k_1) + c_{XX}(k+k_1)).$$

Maxima of the covariance function $c_{YY}(k)$ will occur at $k_1 = 0$ and $k = \pm k_1$. By determining their positions we are able to determine the time delay parameter k_1.

6.2 Autoregressive (AR) Model

Consider a linearly filtered process $Y(n)$

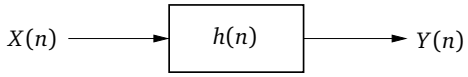


Figure 6.4: Linear filter.

where $X(n)$ is a stationary zero-mean white noise process with variance σ_X^2 and $h(n)$ and $X(n)$ are real-valued.

The main idea of the filter is to shape the white spectrum of the input signal as desired. Having observed $Y(n)$ on a time-limited interval, we are able to estimate the spectrum of $Y(n)$ via estimates of σ_X^2 and the coefficients of the filter.

6.2.1 Definition

The AR process $Y(n)$ of order p (AR(p)) is defined by the difference equation

$$Y(n) + \sum_{k=1}^{p} a_k Y(n-k) = X(n)$$

where $X(n)$ is a stationary zero-mean white noise process with variance σ_X^2 and the recursive filter is assumed to be stable.

The output process $Y(n)$ is stationary, has zero mean and spectrum

$$C_{YY}(e^{j\omega}) = \frac{\sigma_X^2}{|1 + \sum_{k=1}^{p} a_k e^{-j\omega k}|^2}.$$

6.2.2 Stability of the Filter

Stability is given, if the roots of the characteristic function

$$\Phi_p(z) = z^p + \sum_{k=1}^{p} a_k z^{p-k} = 0$$

lie strictly inside the unit circle, i.e. if $|z| < 1$ is fulfilled. Equally, the characteristic function can also be formulated as

$$\tilde{\Phi}_p(z) = z^p \Phi_p(1/z) = 1 + \left(\sum_{k=1}^{p} a_k z^k \right) = 0,$$

for which the roots have to be strictly outside the unit circle, i.e. $|z| > 1$. Stability of the filter ensures stationarity of $Y(n)$, provided $X(n)$ to be stationary as well. If the filter is unstable there exist two different interpretations:

1. The process is not an AR process

2. The process is a non-stationary AR process

In this manuscript the first interpretation is used.

Example 6.2.1 *Example of AR(1): For $p = 1$, the characteristic function*

$$\Phi_p(z) = z^p + \sum_{k=1}^{p} a_k z^{p-k} = 0$$

reduces to

$$z + a_1 = 0$$

and $|z| = |-a_1| = |a_1| < 1$ needs to be true for stability. Figure 6.5 shows three different realizations, where the coefficient a_1 is chosen to be stable (for $a_1 = -0.001$ and $a_1 = -0.999$) and unstable ($a_1 = -1.001$). For the unstable case, the filter output increases unbounded.

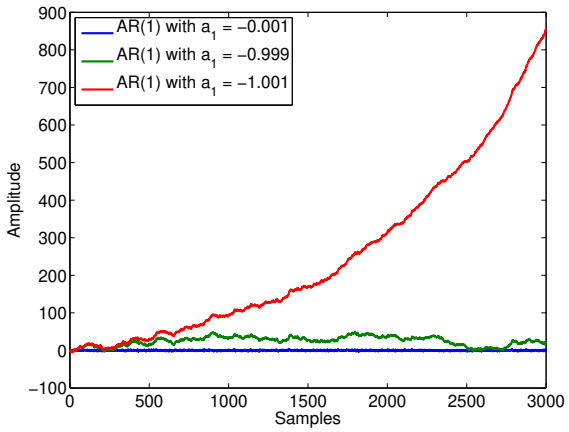


Figure 6.5: Filter outputs of two stable filters (AR(1) processes) and one unstable filter.

6.2.3 Output Spectrum

The output spectrum, $C_{YY}(e^{j\omega})$, of the AR(k) filter can be obtained by the difference equation

$$Y(n) + \sum_{k=1}^{p} a_k Y(n-k) = X(n).$$

From the difference equation, we can see that

$$H(e^{j\omega}) = \frac{1}{1 + \sum_{k=1}^{p} a_k e^{-j\omega k}}.$$

The output spectrum is finally given by

$$\begin{aligned} C_{YY}(e^{j\omega}) &= C_{XX}(e^{j\omega}) \cdot |H(e^{j\omega})|^2 \\ &= \sigma_X^2 \cdot \left| \frac{1}{1 + \sum_{k=1}^{p} a_k e^{-j\omega k}} \right|^2 \\ &= \frac{\sigma_X^2}{|1 + \sum_{k=1}^{p} a_k e^{-j\omega k}|^2}. \end{aligned}$$

6.3 Moving-Average (MA) Model

The MA process $Y(n)$ of order q (M(q)) is defined by the difference equation

$$Y(n) = X(n) + \sum_{l=1}^{q} b_l X(n-l)$$

where $X(n)$ is a stationary zero-mean white noise process with variance σ_X^2 and the filter that transforms $X(n)$ to $Y(n)$ is a transveral filter.

The output process $Y(n)$ is stationary, has zero mean and the spectrum

$$C_{YY}(e^{j\omega}) = |1 + \sum_{l=1}^{q} b_l e^{-j\omega l}|^2$$

6.3.1 Output Spectrum

The output spectrum, $C_{YY}(e^{j\omega})$, can be obtained by the difference equation

$$Y(n) = X(n) + \sum_{l=1}^{q} b_l X(n-l).$$

We can transpose it to the transfer function

$$H(e^{j\omega}) = 1 + \sum_{l=1}^{q} b_l e^{-j\omega l}.$$

The output spectrum is finally given by

$$\begin{aligned} C_{YY}(e^{j\omega}) &= C_{XX}(e^{j\omega}) \cdot |H(e^{j\omega})|^2 \\ &= \sigma_X^2 \cdot |1 + \sum_{l=1}^{q} b_l e^{-j\omega l}|^2. \end{aligned}$$

For $b_0 = 1$, the output spectrum is simplified to

$$C_{YY}(e^{j\omega}) = \sigma_X^2 \cdot |\sum_{l=0}^{q} b_l e^{-j\omega l}|^2.$$

6.4 Autoregressive Moving-Average (ARMA) Model

The ARMA process $Y(n)$ of order p and q (ARMA(p,q)) is defined by the difference equation

$$Y(n) + \sum_{k=1}^{p} a_k Y(n-k) = X(n) + \sum_{l=1}^{q} b_l X(n-l)$$

where $X(n)$ is a stationary zero-mean white noise process with variance σ_X^2 and the recursive part of the filter is assumed to be stable. Also the polynomials $1 + \sum_{k=1}^{p} a_k z^k$ and $1 + \sum_{l=1}^{q} b_l z^l$ have no common factors.

The output process $Y(n)$ is stationary, has zero mean and spectrum

$$C_{YY}(e^{j\omega}) = \sigma_X^2 \cdot \frac{|\sum_{l=0}^{q} b_l e^{-j\omega l}|^2}{|1 + \sum_{k=1}^{p} a_k e^{-j\omega k}|^2}, \quad \text{with} \quad b_0 = 1.$$

6.4.1 Output Spectrum

The output spectrum, $C_{YY}(e^{j\omega})$, can be obtained by the difference equation

$$Y(n)+\sum_{k=1}^{p} a_k Y(n-k) = X(n)+\sum_{l=1}^{q} b_l X(n-l).$$

In the frequency domain, the transfer function is given by

$$H(e^{j\omega}) = \frac{1+\sum_{l=1}^{q} b_l e^{-j\omega l}}{1+\sum_{k=1}^{p} a_k e^{-j\omega k}}.$$

The output spectrum is finally given by

$$\begin{aligned} C_{YY}(e^{j\omega}) &= C_{XX}(e^{j\omega})\cdot|H(e^{j\omega})|^2 \\ &= \sigma_X^2 \cdot \left|\frac{1+\sum_{l=1}^{q} b_l e^{-j\omega l}}{1+\sum_{k=1}^{p} a_k e^{-j\omega k}}\right|^2 \\ &= \sigma_X^2 \cdot \frac{|\sum_{l=0}^{q} b_l e^{-j\omega l}|^2}{|1+\sum_{k=1}^{p} a_k e^{-j\omega k}|^2}, \quad \text{with} \quad b_0 = 1. \end{aligned}$$

6.4.2 Concatenated Model

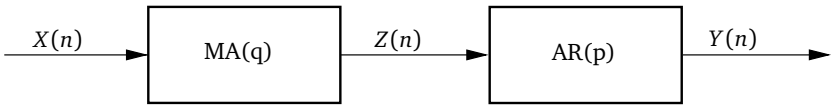


Figure 6.6: ARMA model as a concatenation of MA and AR model.

The ARMA model can be illustrated as two concatenated filters:

$$\begin{aligned} C_{ZZ}(e^{j\omega}) &= \sigma_X^2 \cdot |\sum_{l=0}^{q} b_l e^{-j\omega l}|^2 \\ C_{YY}(e^{j\omega}) &= \frac{C_{ZZ}(e^{j\omega})}{|1+\sum_{k=1}^{p} a_k e^{-j\omega k}|^2} \\ C_{YY}(e^{j\omega}) &= \sigma_X^2 \cdot \left|\frac{1+\sum_{l=1}^{q} b_l e^{-j\omega l}}{1+\sum_{k=1}^{p} a_k e^{-j\omega k}}\right|^2 \end{aligned}$$

6.4.3 Stability of the Filter

As this filter can be modelled as two concatenated filters, the AR model stability criterium still holds true. Therefore, stability is given, if the roots of the characteristic function

$$\Phi_p(z) = z^p + \sum_{k=1}^{p} a_k z^{p-k} = 0$$

lie strictly inside the unit circle, i.e. if $|z| < 1$ is fulfilled. Equally, the characteristic function can also be formulated as

$$\tilde{\Phi}_p(z) = z^p \Phi_p(1/z) = 1 + \left(\sum_{k=1}^{p} a_k z^k\right) = 0,$$

for which the roots have to be strictly outside the unit circle, i.e. $|z| > 1$. If the filter is unstable there exist two different interpretations:

1. The process is not an ARMA process
2. The process is a non-stationary ARMA process

In this manuscript the first interpretation is used.

Bibliography

J.F. Böhme. *Stochastische Signale*. Teubner Studienbücher, 1998.

7 Optimum Linear Systems

7.1 Introduction

A basic problem in the application of stochastic processes is the estimation of a signal in the presence of additive noise. The signal may be random or deterministic, and the noise may be colored or white. The problem consists of establishing the presence of the signal or of estimating its form[1]. The solution of this problem depends on the state of prior knowledge concerning the signal and the noise. For example, we may be able to specify signal and noise correlation functions, power spectra or probability densities.

System constraints define the form of the resulting system. For example we might allow the system to be linear, nonlinear, time-invariant, realizable, etc. We shall be exclusively concerned with linear time-invariant systems, but will not necessarily require that they be realizable.

7.2 The Matched Filter

We have previously developed techniques for describing random processes and analyzing the effect of linear systems on these processes. Here, we develop a technique for designing a linear filter to minimize the effect of additive noise.

$$X(n) = \begin{cases} s(n) + V(n) \\ V(n) \end{cases} \rightarrow \boxed{\begin{matrix} h(n) \\ H(e^{j\omega}) \end{matrix}} \rightarrow \begin{matrix} X_0(n) = s_0(n) + V_0(n) \\ X_0(n) = V_0(n) \end{matrix}$$

The known signal $X(n)$ could be either a signal in noise or noise only. The signal $s(n)$ is assumed to be real-valued, deterministic, and observed over the time-interval $[0, N)$. We assume that $\mathsf{E}[V(n)] = 0$ and the spectrum $C_{VV}(e^{j\omega})$ of the additive input noise $V(n)$ is known.

We wish to determine the filter characteristic such that the instantaneous output signal power is maximized at a sampling time n_0, when compared with the average output noise power. This problem is typical in radar where we wish to establish the presence and location of a signal $s(n)$ returning from a distant target. The effectiveness of the detection scheme depends on the ratio between the value $s(n_0)$ of $s(n)$ at certain n_0, and the standard deviation of the noise.

We want to find $h(n)$ or $H(e^{j\omega})$ so that

$$\left(\frac{S}{N}\right)_{\text{out}} = \frac{|s_0(n_0)|^2}{\mathsf{E}[V_0(n)^2]} \text{ is maximum}$$

Comment: The matched filter does not preserve the input signal wave shape. This is not a drawback because it is not the objective. The objective is to distort the input signal wave shape and filter the noise so that at the sampling time n_0 the output signal level will be as large as possible with respect to the (output) noise level.

Theorem 7.2.1 *The matched filter is the linear filter that maximizes*

$$\left(\frac{S}{N}\right)_{\text{out}} = \frac{|s_0(n_0)|^2}{\mathsf{E}[V_0(n)^2]}$$

and it has a transfer function given by:

$$H(e^{j\omega}) = k \cdot \frac{S(e^{j\omega})^*}{C_{VV}(e^{j\omega})} e^{-j\omega n_0}$$

where

[1] In particular we are concerned by designing a system that satisfies certain rules that make it optimum.

- $S(e^{j\omega}) = \mathcal{F}\{s(n)\}$ *is the Fourier transform of* $s(n)$
- $C_{VV}(e^{j\omega})$ *is the spectrum of the input noise*
- n_0 *is the sampling time when* (S/N) *is evaluated*
- k *is an arbitrary real constant*

Proof 7.2.1 *The output signal at time* n_0 *is:*

$$s_0(n_0) = \int_{-\pi}^{\pi} H(e^{j\omega}) S(e^{j\omega}) e^{j\omega n_0} \frac{d\omega}{2\pi}$$

The average power of the output noise is:

$$\begin{aligned} \mathsf{E}\left[V_0(n)^2\right] &= c_{V_0V_0}(0) \\ &= \int_{-\pi}^{\pi} \left|H(e^{j\omega})\right|^2 C_{VV}(e^{j\omega}) \frac{d\omega}{2\pi} \end{aligned}$$

Thus,

$$\left(\frac{S}{N}\right)_{\text{out}} = \frac{\left|\int_{-\pi}^{\pi} H(e^{j\omega}) S(e^{j\omega}) e^{j\omega n_0} \frac{d\omega}{2\pi}\right|^2}{\int_{-\pi}^{\pi} |H(e^{j\omega})|^2 C_{VV}(e^{j\omega}) \frac{d\omega}{2\pi}}$$

We want to find the particular $H(e^{j\omega})$ *that maximizes this ratio. For this, we use Schwarz's inequality:*

$$\left|\int_a^a \varphi_1(\omega)\varphi_2(\omega) d\omega\right|^2 \leq \left(\int_a^b |\varphi_1(\omega)|^2 d\omega\right) \cdot \left(\int_a^b |\varphi_2(\omega)|^2 d\omega\right)$$

where $\varphi_1(\omega)$ *and* $\varphi_2(\omega)$ *may be complex functions of the variable* ω*. Equality is obtained when*

$$\varphi_1(\omega) = k\varphi_2(\omega)^*$$

where k *is an arbitrary real-valued constant.*

Let $\varphi_1(\omega) = H(e^{j\omega})\sqrt{\frac{C_{VV}(e^{j\omega})}{2\pi}}$ *and* $\varphi_2(\omega) = \frac{S(e^{j\omega})e^{j\omega n_0}}{\sqrt{2\pi C_{VV}(e^{j\omega})}}$*. Then we have:*

$$\left(\frac{S}{N}\right)_{\text{out}} \leq \frac{\int_{-\pi}^{\pi} \left|H(e^{j\omega})\right|^2 C_{VV}(e^{j\omega}) \frac{d\omega}{2\pi} \cdot \int_{-\pi}^{\pi} \frac{\left|S(e^{j\omega})\right|^2}{C_{VV}(e^{j\omega})} \frac{d\omega}{2\pi}}{\int_{-\pi}^{\pi} \left|H(e^{j\omega})\right|^2 C_{VV}(e^{j\omega}) \frac{d\omega}{2\pi}}$$

with $C_{VV}(e^{j\omega})$ *being a non-negative function. Thus*

$$\left(\frac{S}{N}\right)_{\text{out}} \leq \int_{-\pi}^{\pi} \frac{\left|S(e^{j\omega})\right|^2}{C_{VV}(e^{j\omega})} \frac{d\omega}{2\pi}$$

The maximum $(S/N)_{\text{out}}$ *is obtained when* $H(e^{j\omega})$ *is chosen according to*

$$\begin{aligned} H(e^{j\omega})\sqrt{\frac{C_{VV}(e^{j\omega})}{2\pi}} &= k \cdot \frac{S(e^{j\omega})^* e^{-j\omega n_0}}{\sqrt{2\pi C_{VV}(e^{j\omega})}} \\ H(e^{j\omega}) &= k \cdot \frac{S(e^{j\omega})^* e^{-j\omega n_0}}{C_{VV}(e^{j\omega})}, \quad k \in \mathbb{R} \end{aligned}$$

■

Remarks:

- k is an arbitrary constant since the signal and the noise at the input are both multiplied by k. Both the output signal and noise levels depend on the value of the constant.
- The filter found may or may not be causal. If it is not causal, it has to be approximated by a causal filter in order to be realizable.

From the above, we find that the optimum filter's transfer function is proportional to the complex conjugate of the input signal's spectrum; we might say that the system is therefore matched to the specified signal since it depends so intimately on it. $H(e^{j\omega})$ depends also on k. This fact allows the optimal system to have arbitrary gain which intuitively is true because the gain cancels in the relation

$$\left(\frac{S}{N}\right)_{\text{out}} = \frac{|s_0(n_0)|^2}{\mathsf{E}[V_0(n)^2]}$$

since signal and noise are both affected by the gain.

7.2.1 The Matched Filter for White Noise

Theorem 7.2.2 *When the input noise is white, the unit sample response of the matched filter becomes $h(n) \equiv c \cdot s(n_0 - n)$, where c is an arbitrary real constant, n_0 is the time of the peak signal output, and $s(n)$ is the known input signal wave shape.*

Proof 7.2.2 *White noise has a spectrum $C_{VV}(e^{j\omega}) = \sigma_V^2$. Then $H(e^{j\omega}) = \frac{k}{\sigma_V^2} S(e^{j\omega})^* e^{-j\omega n_0}$ characterizes the matched filter. The unit sample response is then given by*

$$\begin{aligned} h(n) &= \mathcal{F}^{-1}\left\{H(e^{j\omega})\right\} = \frac{k}{\sigma_V^2}\int_{-\pi}^{\pi} S(e^{j\omega})^* e^{-j\omega n_0} e^{j\omega n}\frac{d\omega}{2\pi} \\ &= \frac{k}{\sigma_V^2}\left[\int_{-\pi}^{\pi} S(e^{j\omega}) e^{j\omega(n_0-n)}\frac{d\omega}{2\pi}\right]^* \\ &= \frac{k}{\sigma_V^2} s(n_0-n)^* \end{aligned}$$

If $s(n)$ is real-valued, then

$$h(n) = \frac{k}{\sigma_V^2} s(n_0 - n).$$

We let $\frac{k}{\sigma_V^2} = c$ and the theorem is proved.

■

In the white noise case, the impulse response of the matched filter is simply the known signal wave shape that is "played backward" and translated by an amount n_0. Thus, the filter is said to be "matched to the signal". The signal-to-noise power ratio at the output is given by

$$\left(\frac{S}{N}\right)_{\text{out}} = \int_{-\pi}^{\pi} \frac{\left|S(e^{j\omega})\right|^2}{\sigma_V^2}\frac{d\omega}{2\pi} = \frac{1}{\sigma_V^2}\sum_n |s(n)|^2 = \frac{E_s}{\sigma_V^2} \tag{7.1}$$

where $\sum_n |s(n)|^2 = E_s$ is the energy of the input signal of finite duration.

The above shows that :

- the signal-to-noise power ratio at the output of the filter $\left(\frac{S}{N}\right)_{\text{out}}$ depends on the signal energy and power level of the noise and not on the particular signal wave shape used.
- we can increase the signal amplitude or the signal duration, to improve the signal-to-noise power ratio.

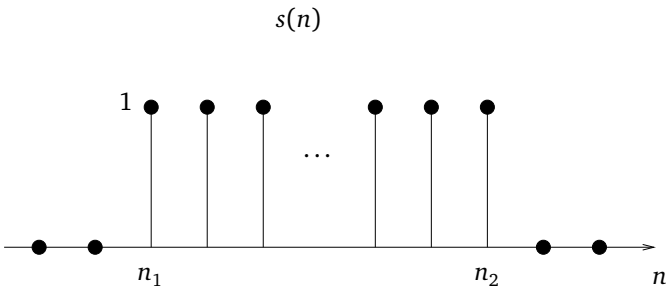


Figure 7.1: Input signal $s(n)$.

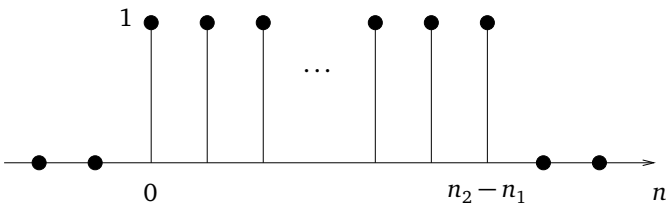


Figure 7.2: Matched filter unit sample response $h(n)$.

Example 7.2.1 *We shall find the matched filter for the known signal $s(n)$ given by*

$$s(n)=\begin{cases}1, & n_1\le n\le n_2\\ 0, & \text{otherwise}\end{cases}$$

of duration $N=n_2-n_1+1$, as depicted in Figure 7.1. We assume that the additive noise is white. The unit sample response of the matched filter is given by

$$h(n)=s(n_0-n)=s\left(-(n-n_0)\right).$$

In order to get a causal matched filter, we require $n_0\ge n_2$. Choosing $n_0=n_2$, the matched filter response is plotted in Figure 7.2 and the signal component of the matched filter output, $s_0(n)$, is plotted in Figure 7.3. The peak output level occurs at $n=n_0$. The input signal wave shape has been distorted by the filter in order to peak up the output signal at $n=n_0$.

Example 7.2.2 *Given a signal $s(n)=A\cdot e^{-bn}u(n),\ b>0$, in white noise, we wish to find the matched filter. The signal shown in Figure 7.4 is of finite energy but infinite time duration. We choose $n_0\ge 0$ and construct $h(n)=s(n_0-n)$. However, the unit sample response shown in Figure 7.5 is nonzero for $n<0$ and therefore the filter is non-causal.*

The unit sample response of the causal matched filter, $h_c(n)$ given below, is obtained by truncating the non-causal filter unit sample response as shown in Figure 7.5.

$$h_c(n)=\begin{cases}A\cdot e^{-b(n_0-n)}u(n_0-n), & n\ge 0\\ 0, & n<0\end{cases}$$

The signal-to-noise power ratio (SNR) at the output, at $n=n_0$ is given by:

$$\left(\frac{S}{N}\right)_{\text{out}}=\frac{s_0(n_0)^2}{\mathsf{E}[V_0(n)^2]}$$

For the causal filter

$$s_0(n_0) \quad = \quad \sum_k h_c(k)s(n_0-k)$$

$s_0(n), \quad n_0 = n_2$

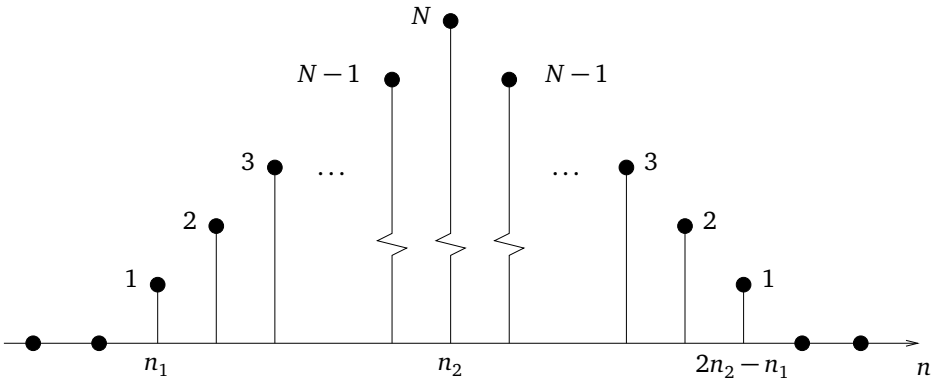


Figure 7.3: Signal out of the matched filter $s_0(n)$.

$s(n)$

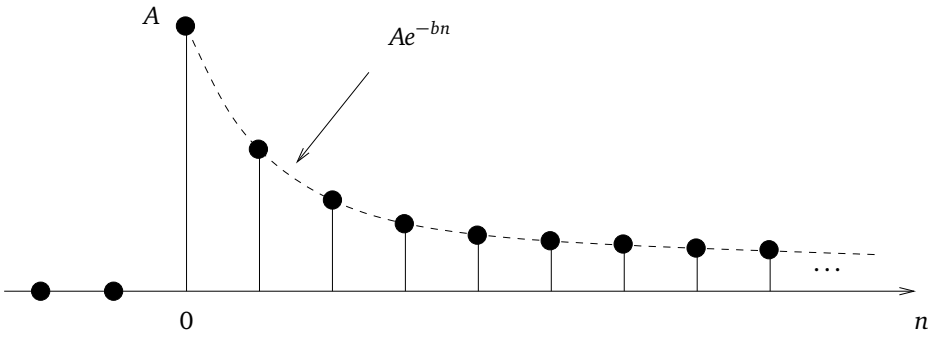


Figure 7.4: Input signal $s(n)$.

$$\begin{aligned} &= \sum_{k=0}^{n_0} A^2 e^{-2b(n_0-k)} \\ &= A^2\left(\frac{e^{2b}-e^{-2bn_0}}{e^{2b}-1}\right) \end{aligned}$$

where the closed form expression for the summation is obtained using the formula for a geometric series. The noise power at the output

$$\begin{aligned} \mathsf{E}[V_0(n)^2] = c_{V_0V_0}(0) &= \sum_k \sum_l h_c(k)h_c(l)c_{VV}(l-k) \\ &= \sum_k \sum_l h_c(k)h_c(l)\sigma_V^2\delta(l-k) \\ \sum_k \sum_l h_c(k)h_c(l)\sigma_V^2\delta(l-k) &= \sigma_V^2 \sum_{k=0}^{n_0} A^2 e^{-2b(n_0-k)} \\ &= \sigma_V^2 s_0(n_0) \end{aligned}$$

Therefore the maximum SNR at the output of the causal filter is given by

$$\left(\frac{S}{N}\right)_{\text{out}} = \frac{s_0(n_0)}{\sigma_V^2} = \frac{A^2}{\sigma_V^2}\left(\frac{e^{2b}-e^{-2bn_0}}{e^{2b}-1}\right) \tag{7.2}$$

For the non-causal filter, the maximum SNR at the output is given by Equation (7.1)

$$\left(\frac{S}{N}\right)_{\text{out}} = \frac{E_s}{\sigma_V^2}$$

$$h_c(n) = h(n)u(n)$$

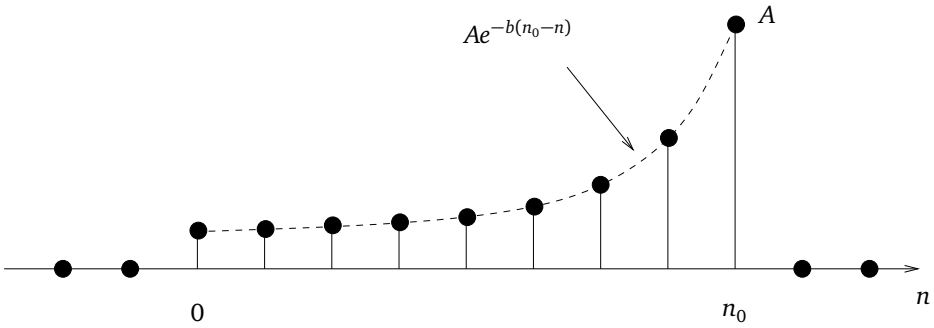


Figure 7.5: Causal matched filter $h_c(n)$.

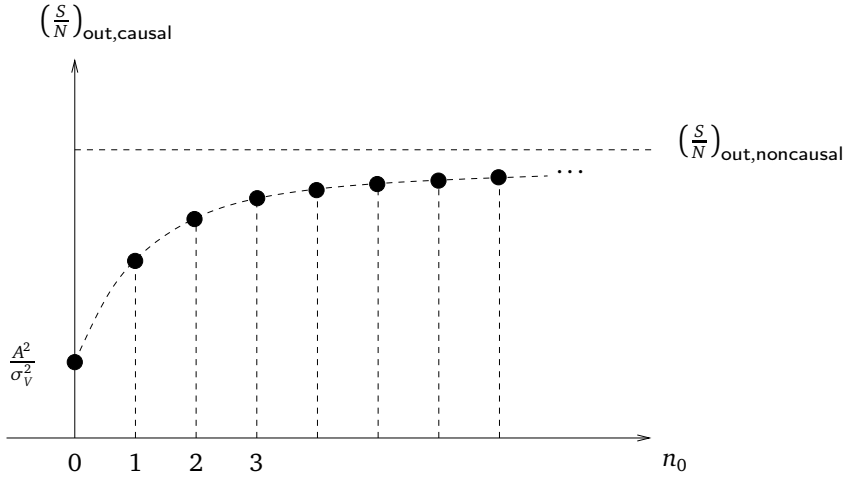


Figure 7.6: SNR of the optimal non-causal filter and its causal approximation with an increasing n_0.

The energy of the signal, E_s, is

$$E_s = \sum_n |s(n)|^2 = \sum_{n=0}^{\infty} A^2 e^{-2bn} = A^2 \left(\frac{e^{2b}}{e^{2b}-1} \right) \tag{7.3}$$

which results in a larger SNR than that obtained in Equation (7.2) for the causal filter response. However, comparing Equations (7.2) and (7.3), we see that

$$\lim_{n_0 \to \infty} \left(\frac{S}{N} \right)_{\text{out,causal}} = \left(\frac{S}{N} \right)_{\text{out,noncausal}}$$

To illustrate this concept, the SNR of the causal filter is plotted with the SNR of the non-causal filter, for increasing values of n_0, in Figure 7.6. From this figure, we can see that depending on the values of A, σ_V^2 and b, we may find a sufficiently large but finite value of n_0 so that the causal filter produces an SNR close to that of the optimal non-causal filter.

7.2.2 Matched Filter as Correlation Processing

Consider a known signal waveform $s(n)$ which is non-zero only for $n_1 \le n \le n_2$ (for example the rectangular signal of Example 7.2.1). The signal duration is $N = n_2 - n_1 + 1$ samples. For the case of additive white noise $V(n)$, the matched filter may be realized by correlating the filter input with $s(n)$,

$$X_0(n_0) = \sum_{n=n_0-N+1}^{n_0} X(n)s(n)$$

where $X(n) = s(n) + V(n)$ is the filter input.

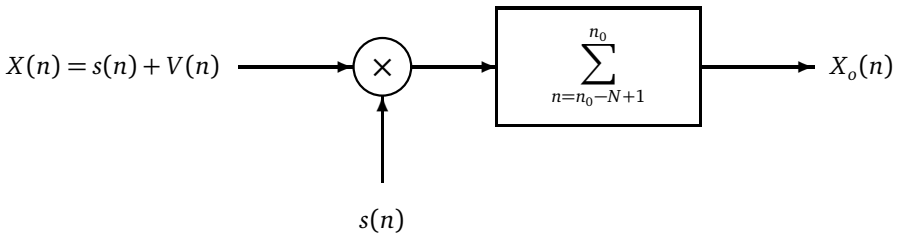


Figure 7.7: Matched filter realization by correlation processing.

Proof 7.2.3 *The output of the matched filter at time n_0 is:*

$$X_0(n_0) = [X(n) * h(n)]|_{n=n_0} = \sum_k X(k)h(n_0 - k) \tag{7.4}$$

We use the matched filter for white noise with $n_0 = n_2$

$$h(n) = \begin{cases} s(n_0 - n), & 0 \le n < N \\ 0, & \text{elsewhere} \end{cases}$$

so the summation (7.4) is non-zero for

$$\begin{aligned} 0 \le \quad & n_0 - k \quad < N \\ n_0 \ge \quad & k \quad \ge -N + n_0 + 1 \end{aligned}$$

Therefore

$$\begin{aligned} X_0(n_0) &= \sum_{k=n_0-N+1}^{n_0} X(k)h(n_0 - k) \\ &= \sum_{k=n_0-N+1}^{n_0} X(k)s(n_0 - (n_0 - k)) \end{aligned}$$

$$\boxed{X_0(n_0) = \sum_{k=n_0-N+1}^{n_0} X(k)s(k)}$$

■

The realization of a matched filter as a 'correlator' is shown in Figure 7.7.

Example 7.2.3 *Radar Example:*

A radar transmits the chirp signal $s(n)$ shown in Figure 7.8 a). A chirp signal is a signal in which the frequency decreases or increases with time. If the transmitted wave strikes an object it will reflect a delayed, noisy version of the transmitted signal back to the receiver. Based on the signal received by the radar, the aim is to decide whether a target is present or not. Thus, the problem at hand is that of having a signal of known shape that has been disturbed by noise. We wish to maximize the output signal-to-noise ratio at time n_0, to be able to assess whether a target is present or not. The corresponding matched filter at the receiver with the unit sample response $h(n) = s(n_0 - n)$ is shown in Figure 7.8 b). The transmitted signal arrives at the receiver with a delay of 1587 samples. In an ideal noise-free case, the received, delayed signal will correspond to the signal in the first image in Figure 7.9 a). However, usually the signal is corrupted by noise. We assume additive zero-mean white Gaussian noise with a variance σ^2, leading to a noisy received signal as in the second image of Figure 7.9 b). The matched filter output in Figure 7.9 c) shows its maximum at n_0, indicating the presence of a transmitted signal.

Now assume that there is no target. The radar sends the pulse and receives a noisy signal as in Figure 7.10 b). The output after the matched filter can be seen in Figure 7.10 c). It indicates the absence of a target. Comparing Figure 7.9 c) with Figure 7.10 c) one can easily see that the matched filter succeeds at detecting the absence or presence of a target. Figure 7.11 shows that even when the received signal is buried in noise the matched filter is still able to detect its presence.

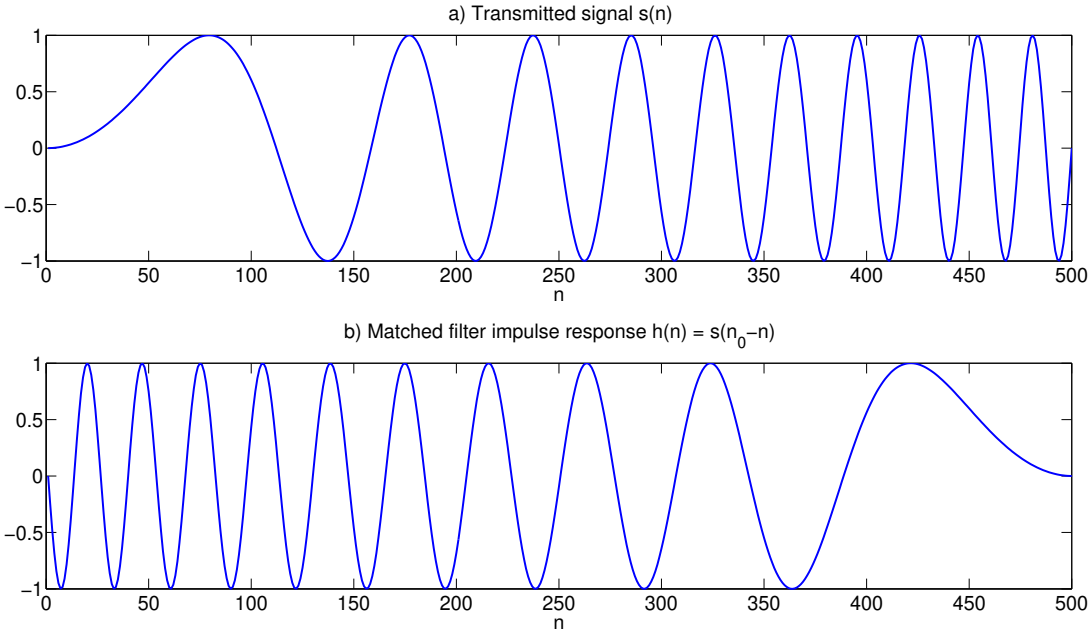


Figure 7.8: a) Transmitted signal $s(n)$. b) Matched filter unit sample response $h(n)$.

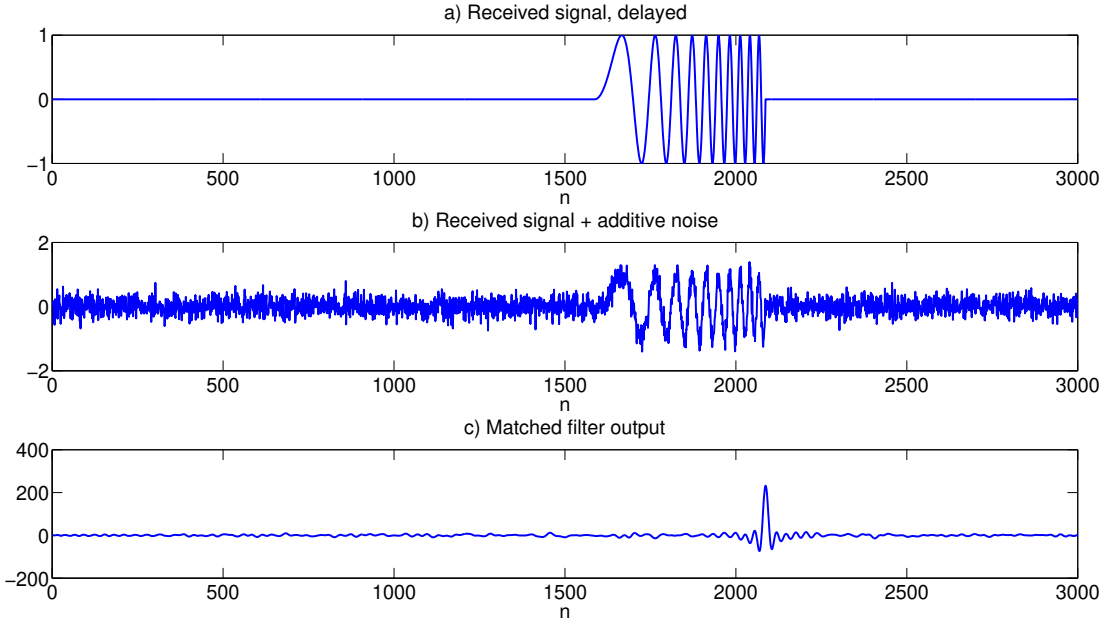


Figure 7.9: Radar signal transmitted from Figure 7.8, $SNR_{in} = 10dB$: a) Received signal delayed. b) Received signal plus additive noise. c) Matched filter output.

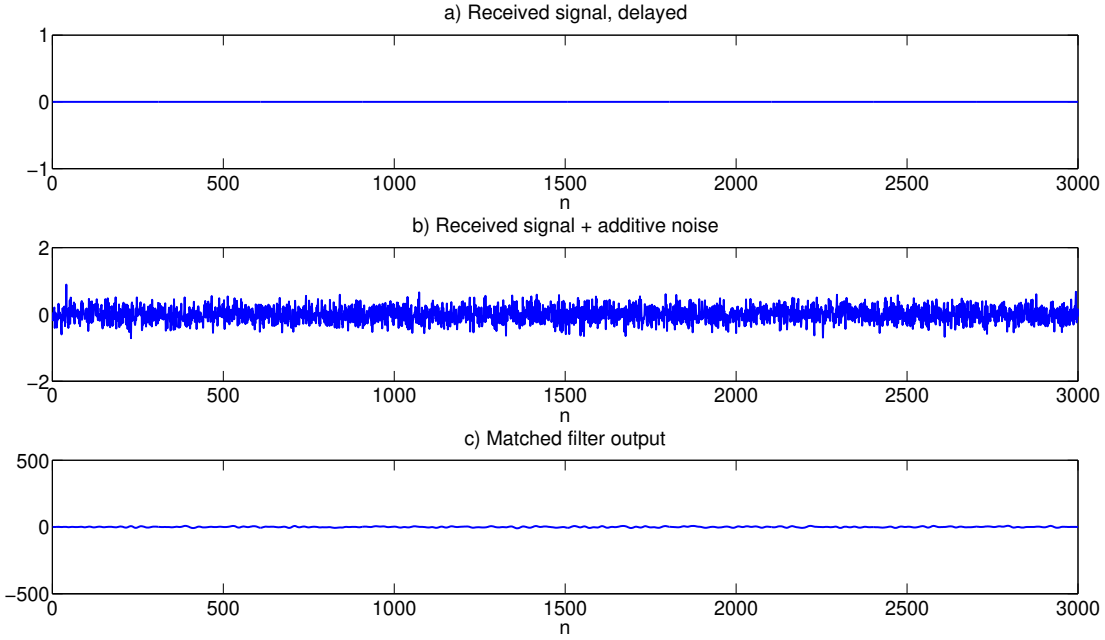


Figure 7.10: No radar signal transmitted: a) Received signal delayed. b) Received signal plus additive noise. c) Matched filter output.

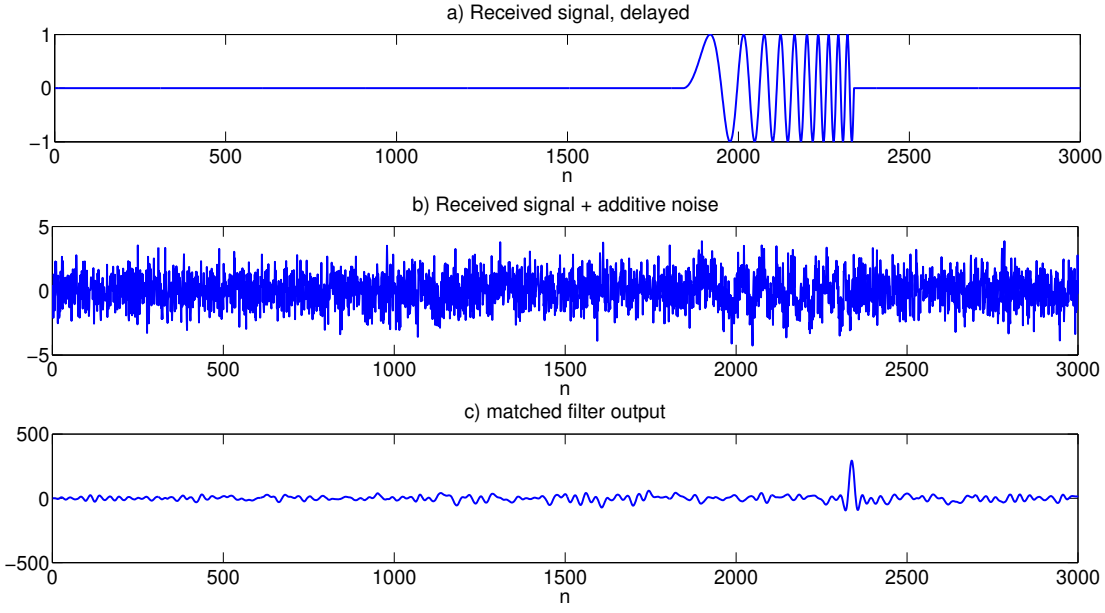


Figure 7.11: Radar signal transmitted from Figure 7.8, $SNR_{in} = -5dB$: a) Received signal delayed. b) Received signal plus additive noise. c) Matched filter output.

7.3 Introduction to Estimation Theory

Detection Theory:

- Detect a known signal in noise, e. g., in radar, if a target is present or not.
- The matched filter provides the highest SNR at the output for the detection of a known signal.

Estimation Theory:

- Estimate a non-observable random process from observations of another random process that contains innovations about the sought for process that contains information about the sought for process.
- In radar, e.g., estimate the range of a target from detected signals.
- The method of least squares minimizes the sum of squared residuals.
- The Wiener filter minimizes the mean squared error between the true process and the estimated one.

Example 7.3.1 *In radar systems, there are many parameters that can be estimated: e.g. the range to a target, the speed of a target, or acceleration of a target. Figure 7.12 shows a schematic radar system, in which the transmitted power by is much higher than the received power.*

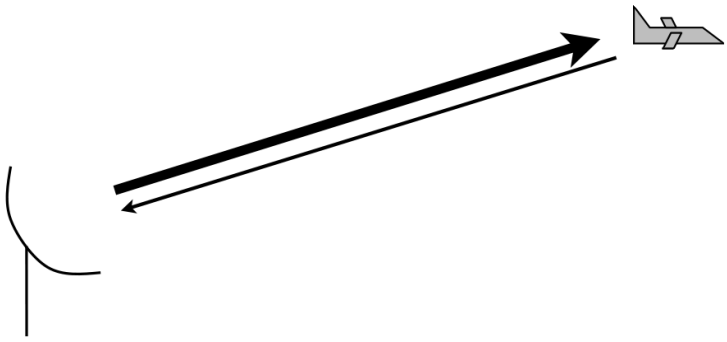


Figure 7.12: Radar system.

Example 7.3.2 *In biomedical signal processing, from physiological signals such as, e.g. electrocardiogram (ECG), photoplethysmogram (PPG), or electroencephalogram (EEG), several vital signs can be estimated, e.g. the heart rate, respiratory rate, blood pressure, etc. However, motion artifacts that are caused during movements when measuring the physiological signals make the estimation difficult, as shown in Figure 7.13.*

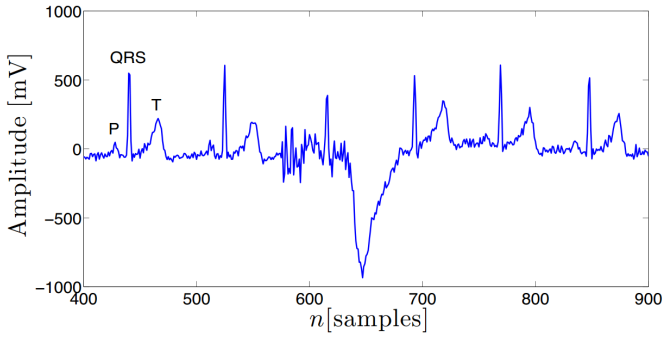


Figure 7.13: ECG with motion artifact.

Example 7.3.3 *There are many estimation tasks in speech signal processing, e.g. speech recognition, speaker recognition, pitch estimation, estimation of the direction of arrival, etc. How different words have distinguishable frequency characteristics is shown in the spectrogram in Figure 7.14.*

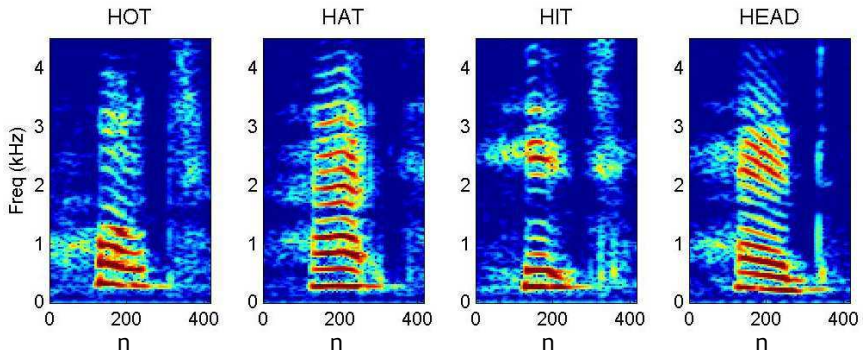


Figure 7.14: Spectrograms of different words (Source: Auditory Neuroscience).

Let $x_1,\dots,x_N$ be observations as realizations of the random variables (RVs) $X_1,\dots,X_N$. This multivariate distribution of $X_1,\dots,X_N$ depends on the unknown parameters $\theta_1,\dots,\theta_K$ which need to be estimated from the observations $x_1,\dots,x_N$.

Example 7.3.4 *Given an i.i.d. normally distributed RVs $X_1,\dots,X_N$ with parameters $\theta_1=\mu\in\mathbb{R}$, $\theta_2=\sigma^2>0$, and density*

$$f_{X_1,\dots,X_N}(x_1,\dots,x_N;\theta_1,\theta_2)=\frac{1}{(2\pi\theta_2)^{N/2}}\exp\left(-\frac{1}{2\theta_2}\sum_{n=1}^{N}(x_n-\theta_1)^2\right).$$

7.3.1 Goal of Estimation Theory

The **goal** of estimation theory is to determine the parameters $\theta_1,\dots,\theta_K$ under which the observations $x_1,\dots,x_N$ were generated. Or equally, find the function that maps the observations $x_1,\dots,x_N$ to the parameter $\theta_i, i=1,\dots,K$,

$$(x_1,\dots,x_N)\to\hat{\theta}_i(x_1,\dots,x_N),\qquad i=1,\dots,K.$$

The value $\hat{\theta}_i$ estimated by the observations is called the **estimate** of θ_i, $i=1,\dots,K$.

Two fundamental approaches to estimation exist:

1. Frequentist: θ_i is treated as an unknown deterministic parameter.
2. Bayesian: θ_i is treated as a RV with a known distribution.

We focus on the second, the Bayesian approach. However, the observations $x_1,\dots,x_N$ are only realizations of RV, which only approximate the real values of $\theta_1,\dots,\theta_K$.

7.3.2 Definitions

- We define the **estimator** $\hat{\Theta}_i$ as a function of RVs

$$\hat{\Theta}_i=\hat{\theta}_i(X_1,\dots,X_N),\qquad i=1,\dots,K.$$

- An **estimator** $\hat{\Theta}_i$ is a function of the RVs $X_1,\dots,X_N$
- An **estimate** $\hat{\theta}_i$ is a function of the observations $x_1,\dots,x_N$, and can be interpreted as a realization of $\hat{\Theta}_i$
- x_n is a realisation of the RV X_n
- $\hat{\theta}_i$ is a realisation of the RV $\hat{\Theta}_i$

In the following, we choose $K=1$ and denote $\hat{\Theta}_i$ by $\hat{\Theta}$ and $\hat{\theta}_i$ by $\hat{\theta}$.

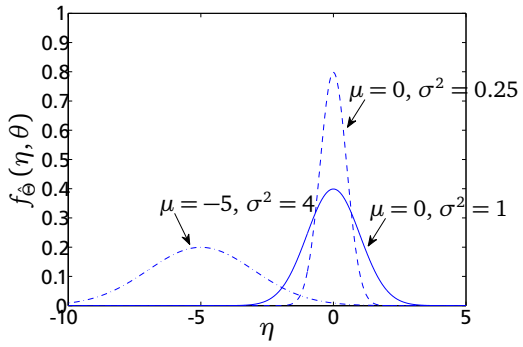


Figure 7.15: Three different normal probability density functions of estimator $\hat{\Theta}$.

7.3.3 Properties of an Estimator

To fully describe an estimator $\hat{\Theta}$, we need to know its density $f_{\hat{\Theta}}(\eta,\theta)$. Figure 7.15 shows three different normal probability density functions of estimator $\hat{\Theta}$.

However, densities are sometimes hard to determine, so we restrict ourselves to the **first two moments** of the estimator $\hat{\Theta}=\hat{\theta}(X_1,\ldots,X_N)$ of θ.

The **expectation** of an estimator (first moment) is given by

$$\mathsf{E}[\hat{\Theta}]=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\hat{\theta}(x_1,\ldots,x_N)\,f_{X_1,\ldots,X_N}(x_1,\ldots,x_N;\theta)\,dx_1\cdots dx_N.$$

The **bias** of an estimator is the difference of the expected value of the estimator and the real value

$$\text{bias}[\hat{\Theta}]=\mathsf{E}[\hat{\Theta}]-\theta$$

For $\text{bias}[\hat{\Theta}]=0$, the estimator is called unbiased.

The **variance** of an estimator (second moment) is given by

$$\mathsf{Var}[\hat{\Theta}]=\mathsf{E}[(\hat{\Theta}-\mathsf{E}[\hat{\Theta}])^2]=\mathsf{E}[\hat{\Theta}^2]-\mathsf{E}[\hat{\Theta}]^2.$$

The **mean square error** of an estimator can be calculated by

$$\begin{aligned}\text{MSE}[\hat{\Theta}]&=\mathsf{E}[(\hat{\Theta}-\theta)^2]\\&=\text{Var}[\hat{\Theta}]+\text{bias}[\hat{\Theta}]^2.\end{aligned}$$

An estimator based on N observations is *consistent* if

$$\lim_{N\to\infty}\Pr\left[|\hat{\Theta}-\theta|\geq\epsilon\right]=0,\quad\forall\,\epsilon>0$$

or equivalently

$$\lim_{N\to\infty}\Pr\left[|\hat{\Theta}-\theta|<\epsilon\right]=1,\quad\forall\,\epsilon>0.$$

For both equations, the estimator *converges in probability* to its true value θ,

$$\hat{\Theta}\xrightarrow[N\to\infty]{P}\theta.$$

Convergence is in the asymptotic sense, as $N\to\infty$.

7.4 Method of Least Squares

7.4.1 Introduction

One of the most intuitive procedures to estimate unknown parameters in a model is the **method of least squares**. It is commonly credited to Carl Friedrich Gauss (1795), but it was first published by Adrien–Marie Legendre (1805). When applying the method of least squares, no probabilistic assumptions about the distribution of the data are made, but a deterministic signal model is assumed. On the one hand, this makes the method of least squares widely applicable in practice. On the other hand, without an underlying probabilistic model, no prior knowledge about the data or noise distribution can be used to improve the estimation accuracy.

7.4.2 The Least Squares Criterion

The idea underlying the least squares approach is to minimize the sum of the squared differences between an observed, noisy signal and and an ideal, noise-free signal that is generated by some suitably chosen model. This model in turn depends on an unknown parameter θ that needs to be chosen such that it best explains the data. Given N observed input-output pairs, $(x_1, y_1), \ldots, (x_N, y_N)$, the signal model is

$$y_n = f(x_n; \theta) + z_n, \tag{7.5}$$

where y_n denotes the nth observed output, z_n denotes an unknown measurement error, and $f(x_n; \theta)$ denotes the noise-free signal, which is modeled as a function of the input x_n and the unknown parameter θ. In general, the input and output as well as the unknown parameter θ do not need to be scalars, but can be vector valued. In order to balance generality and simplicity, we assume y_n to be scalar, but allow vector valued inputs $\boldsymbol{x}_n = (x_{n1}, \ldots, x_{nM})^\top$ and vector valued parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)^\top$. Hence, the signal model in (7.5) becomes

$$y_n = f(\boldsymbol{x}_n; \boldsymbol{\theta}) + z_n. \tag{7.6}$$

The **least squares estimate** (LSE) for $\boldsymbol{\theta}$ is the parameter vector $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_K)^\top$ that minimizes the sum of the squared differences between the observed signal and the assumed signal model, i.e.,

$$\mathcal{S}(\boldsymbol{\theta}) = \sum_{n=1}^{N} z_n^2 = \sum_{n=1}^{N} \big(y_n - f(\boldsymbol{x}_n; \boldsymbol{\theta})\big)^2.$$

In general, no closed-form solution for the parameter vector that minimizes $\mathcal{S}(\boldsymbol{\theta})$ exists and often numerical optimization methods are used to find parameter vectors that are "good in the least squared sense", i.e., that lead to a model $f(\boldsymbol{x}; \hat{\boldsymbol{\theta}})$ with small squared residuals for the input-output pairs $(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_N, y_N)$.

A special case of the method of least squares, for which a closed-form solution exists, is obtained by restricting f in (7.6) to be a linear function of the parameters $\boldsymbol{\theta}$. This assumption simplifies the method of least squares significantly and is very common in practice.

7.4.3 Linear Least Squares

The linear least squares estimator is obtained by assuming f in (7.6) to be of the form

$$f(\boldsymbol{x}_n; \boldsymbol{\theta}) = \boldsymbol{x}_n^\top \boldsymbol{\theta} = \sum_{i=1}^{K} x_{ni} \theta_i,$$

where both $\boldsymbol{x}$ and $\boldsymbol{\theta}$ are of dimensions $1 \times K$. For N observed input-output pairs, $(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_N, y_N)$, the corresponding signal model is

$$y_n = \boldsymbol{x}_n^\top \boldsymbol{\theta} + z_n = \sum_{i=1}^{K} x_{ni} \theta_i + z_n \tag{7.7}$$

so that the sum of the squared errors calculates to

$$\mathcal{S}(\boldsymbol{\theta}) = \sum_{n=1}^{N} z_n^2 = \sum_{n=1}^{N} \big(y_n - \boldsymbol{x}_n^\top \boldsymbol{\theta}\big)^2 = \sum_{n=1}^{N} \left(y_n - \sum_{i=1}^{K} x_{ni} \theta_i \right)^2. \tag{7.8}$$

In order to find the minimum of $\mathcal{S}(\boldsymbol{\theta})$, we calculate the first derivative with respect to θ_j, $j = 1, \ldots, K$:

$$\left. \frac{\partial \mathcal{S}}{\partial \theta_j} \right|_{\hat{\boldsymbol{\theta}}} = -2 \sum_{n=1}^{N} x_{nj} \left(y_n - \sum_{i=1}^{K} x_{ni} \hat{\theta}_i \right) = 0, \qquad j = 1, \ldots, K.$$

These K equations can be rearranged into a system of linear equations

$$\sum_{i=1}^{K} \hat{\theta}_k \sum_{n=1}^{N} x_{ni} x_{nj} = \sum_{n=1}^{N} y_n x_{nj}, \qquad j = 1, \ldots, K.$$

This equation system admits a unique solution, $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_K)^\top$, if the coefficient matrix with elements

$$\left(\sum_{n=1}^{N} x_{ni} x_{nj} \right)_{i,j=1,\ldots,K}$$

is *non-singular*. This is guaranteed to be true if the vectors $(x_{1i}, \ldots, x_{Ni})^\top, i = 1, \ldots, K$ are *linearly independent*. In the one-dimensional case, i.e., $K = 1, \theta = \theta_1$ and $x_n = x_{ni} = x_{nj}$, the equation system in (7.4.3) reduces to a single linear equation that can be solved for

$$\hat{\theta} = \frac{\sum_{n=1}^{N} x_n y_n}{\sum_{n=1}^{N} x_n^2}.$$

Example 7.4.1 *For $K = 1$, the least squares problem is also called **linear regression**.*

With the regression line $\hat{y} = x\hat{\theta}$, we can predict y for each x. Figure 7.16 shows the data points in the x-y plane.

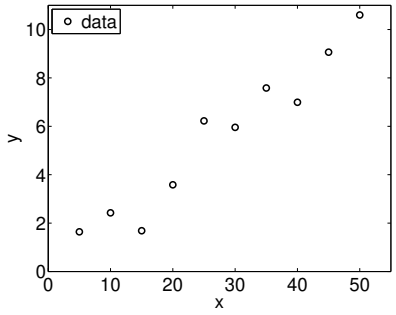


Figure 7.16: The data points in the x-y plane.

The estimation of θ yields

$$\hat{\theta} = \frac{\sum_{n=1}^{N} x_n y_n}{\sum_{n=1}^{N} x_n^2} = 0.2.$$

The regression line $\hat{y} = \hat{\theta} x$ is shown among the data points in Figure 7.17. The least squares estimator minimizes the

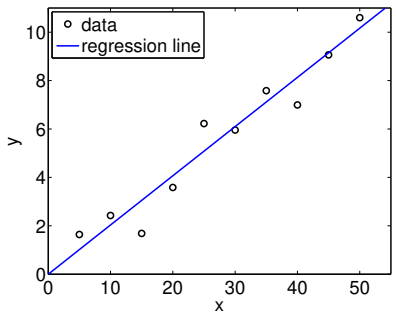


Figure 7.17: The regression line $\hat{y} = \hat{\theta} x$ with $\hat{\theta} = 0.2$ fitting the data points.

sum of squares of the ordinate distances of each point (x_i, y_i) to the regression line, as shown in Figure 7.18.

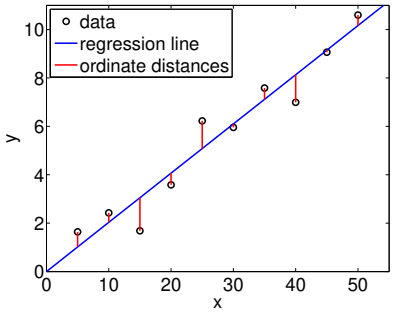


Figure 7.18: The ordinate distances of each point (x_i, y_i) to the regression line.

7.4.4 Linear Least Squares Using Matrix-Vector Notation

In the derivation of the linear LSE, we expanded the inner product of two vectors to sums over their individual elements. While this makes the derivation of the equation system (7.4.3) more transparent, the resulting expressions are rather unintuitive and prone to index errors. Therefore, the linear LSE is usually given in a more compact form, using **matrix-vector notation**. This is done by stacking the observations y_n into a vector $\boldsymbol{y} = (y_1, \ldots, y_N)^\top$ so that (7.7) can be written as

$$\mathbf{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{z}, \tag{7.9}$$

where

- $\boldsymbol{y} = (y_1, \ldots, y_N)^\top$ denotes the $N \times 1$ vector of observed signal,
- $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)^\top$ denotes the $N \times K$ matrix of known input signals,
- $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)^\top$ denotes the $K \times 1$ vector of parameters to be estimated,
- $\boldsymbol{z} = (z_1, \ldots, z_K)^\top$ denotes the $N \times 1$ vector of unknown measurement errors.

The sum of the squared errors in matrix-vector notation is given by

$$\mathcal{S}(\boldsymbol{\theta}) = \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|_2^2 = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta})^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}).$$

Following the same procedure as above, we can calculate the linear LSE in matrix-vector form. The gradient of $\mathcal{S}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ calculates to

$$\begin{aligned}\frac{\partial \mathcal{S}}{\partial \boldsymbol{\theta}} = \frac{\partial}{\partial \boldsymbol{\theta}} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta})^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}) &= \frac{\partial}{\partial \boldsymbol{\theta}} \left(\boldsymbol{y}^\top \boldsymbol{y} - \boldsymbol{y}^\top \boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{\theta}^\top \boldsymbol{X}^\top \boldsymbol{y} - \boldsymbol{\theta}^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\theta} \right) \\ &= -2\boldsymbol{X}^\top \boldsymbol{y} + 2\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\theta}.\end{aligned}$$

Setting the gradient equal to zero yields the linear equation system

$$\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\theta} = \boldsymbol{X}^\top \boldsymbol{y}, \tag{7.10}$$

which can be solved for the linear LSE

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y} = \boldsymbol{X}^\sharp \boldsymbol{y}. \tag{7.11}$$

The equations in (7.10) are called the **normal equations**. They appear in many problems that are related to a least squares criterion and have been studied extensively in the literature. The matrix $\boldsymbol{X}^\sharp = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top$, which defines the solution of the normal equations, is known as **generalized inverse**, **pseudoinverse**, or **Moore–Penrose inverse**. These names refer to the property of $\boldsymbol{X}^\sharp$ to satisfy $\boldsymbol{X}^\sharp \boldsymbol{X} = \boldsymbol{I}$, provided that $\boldsymbol{X}^\top \boldsymbol{X}$ is non-singular.

It is important to note that, although the signal model is linear in the parameters $\boldsymbol{\theta}$, the linear least squares approach can also be used to fit functions that are nonlinear in the input variables. Consider, for example, the following problem.

Example 7.4.2 *Let y_n be a sampled sinusoidal signal with additive noise z_n*

$$y_n = \eta \sin(\omega n + \phi) + z_n, \qquad n = 1, \dots, N,$$

where the normalized frequency ω is known. The amplitude η and the phase ϕ are unknown and need to be estimated. In the model, η is the linear parameter and ϕ is nonlinear one. We use the trigonometric equation for $\sin(\alpha + \beta)$ to reformulate y_n to

$$y_n = \eta \sin(\phi)\cos(\omega n) + \eta \cos(\phi)\sin(\omega n) + z_n, \qquad n = 1, \dots, N.$$

Given the reformulated signal y_n

$$y_n = \underbrace{\eta \sin(\phi)}_{\theta_1} \cdot \underbrace{\cos(\omega n)}_{x_{n1}} + \underbrace{\eta \cos(\phi)}_{\theta_2} \cdot \underbrace{\sin(\omega n)}_{x_{n2}} + z_n, \qquad n = 1, \dots, N,$$

we can express y_n as a linear model

$$y_n = \theta_1 x_{n1} + \theta_2 x_{n2} + z_n = \boldsymbol{x}_n^\top \boldsymbol{\theta}, \qquad n = 1, \dots, N,$$

with unknown parameters $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top$.

Another common example for using the method of least squares to estimate a nonlinear input-output relationship is the problem of fitting a polynomial to a given set of data points.

Example 7.4.3 *Assume that a scalar signal y_n is observed as a function of scalar input u_n. We would like to fit the polynomial signal model*

$$y_n = \theta_0 + u_n \theta_1 + u_n^2 \theta_2 + u_n^3 \theta_3 + z_n. \tag{7.12}$$

The linear least squares approach can be used to estimate $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^\top$ as follows. Instead of considering the scalar input output pairs $(u_1, y_1), \dots, (u_N, y_N)$, we introduce the vector-valued input

$$\boldsymbol{x}_n = (1, u_n, u_n^2, u_n^3)^\top$$

and consider the pairs $(\boldsymbol{x}_1, y_1), \dots, (\boldsymbol{x}_N, y_N)$. The polynomial in (7.12) then becomes a linear function of $\boldsymbol{x}_n$

$$y_n = \theta_0 + u_n \theta_1 + u_n^2 \theta_2 + u_n^3 \theta_3 = \boldsymbol{x}_n^\top \boldsymbol{\theta}$$

and the LSE of the polynomial coefficients can be obtained from (7.11).

Modeling a nonlinear system by a linear system and a nonlinear transformation of the input signal is a powerful method and commonly used in practice. The elements of the auxiliary input vector are often referred to as **features** and the corresponding transformation of the raw input as **feature extraction**. In the above example, additional features are introduced to account for nonlinearities in the signal model. However, feature extraction can also be used to reduce the size of the input vector, that is, to extract low dimensional, but highly informative representations of the input data.

Properties of the Linear LSE

As mentioned before, the LSE does not make any assumption about the probabilistic properties of the measurement noise z_n. The advantage of this approach is that the method of least squares is universally applicable. The disadvantage is that its performance can vary drastically, depending on the properties of the noise. Therefore, it is useful to study some statistical properties of the LSE as functions of the noise distribution. For the analysis it is assumed that the signal indeed follows a linear model, meaning that there exists a true parameter vector $\boldsymbol{\theta}$.

The expected value of the linear LSE calculates to

$$\mathsf{E}[\hat{\boldsymbol{\theta}}] = \mathsf{E}[\boldsymbol{X}^\sharp \boldsymbol{y}] = \boldsymbol{X}^\sharp \mathsf{E}[\boldsymbol{y}].$$

Using the signal model in (7.9), we hence obtain

$$\mathsf{E}[\hat{\boldsymbol{\theta}}] = \boldsymbol{X}^\sharp \mathsf{E}[\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{z}]$$

$$
\begin{aligned}
&= \boldsymbol{X}^{\sharp}\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{X}^{\sharp}\mathsf{E}[\boldsymbol{z}] \\
&= \boldsymbol{\theta} + \boldsymbol{X}^{\sharp}\mathsf{E}[\boldsymbol{z}],
\end{aligned}
$$

so that the **bias** of the linear LSE is given by

$$
\begin{aligned}
\mathrm{bias}[\hat{\boldsymbol{\theta}}] &= \mathsf{E}[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}] \\
&= \boldsymbol{X}^{\sharp}\mathsf{E}[\boldsymbol{z}].
\end{aligned}
$$

Consequently, the linear LSE is **unbiased** for $\mathsf{E}[\boldsymbol{z}] = \boldsymbol{0}$, i.e., zero mean noise.

For the **covariance** of the LSE, assuming zero mean noise, we obtain

$$
\begin{aligned}
\mathrm{Cov}[\hat{\boldsymbol{\theta}}] &= \mathsf{E}[(\boldsymbol{X}^{\sharp}\boldsymbol{y} - \boldsymbol{\theta})(\boldsymbol{X}^{\sharp}\boldsymbol{y} - \boldsymbol{\theta})^{\top}] \\
&= \mathsf{E}[\boldsymbol{X}^{\sharp}\boldsymbol{z}\boldsymbol{z}^{\top}\boldsymbol{X}^{\sharp\top}] \\
&= \boldsymbol{X}^{\sharp}\mathsf{E}[\boldsymbol{z}\boldsymbol{z}^{\top}]\boldsymbol{X}^{\sharp\top} \\
&= \boldsymbol{X}^{\sharp}\boldsymbol{\Sigma}\boldsymbol{X}^{\sharp\top}.
\end{aligned}
$$

If $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}$, i.e., the noise is white with variance σ^2, the covariance simplifies to

$$
\begin{aligned}
\mathrm{Cov}[\hat{\boldsymbol{\theta}}] &= \boldsymbol{X}^{\sharp}\sigma^2\boldsymbol{I}\boldsymbol{X}^{\sharp\top} \\
&= \sigma^2\boldsymbol{X}^{\sharp}\boldsymbol{X}^{\sharp\top} \\
&= \sigma^2(\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}\boldsymbol{X}^{\top}\left((\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}\boldsymbol{X}^{\top}\right)^{\top} \\
&= \sigma^2(\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}\boldsymbol{X}^{\top}\boldsymbol{X}\left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1} \\
&= \sigma^2(\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}.
\end{aligned}
$$

The variance of the ith element of $\hat{\boldsymbol{\theta}}$ is given by the ith diagonal element of $\mathrm{Cov}[\hat{\boldsymbol{\theta}}]$. The off-diagonal elements are the covariances between the corresponding elements of $\hat{\boldsymbol{\theta}}$.

As a final remark, it is important to keep in mind that the criterion of least squares is always applicable, but not always suitable. Especially in noise environments that occasionally generate very large values, so called heavy tailed or impulsive noise, the least squares criterion leads to highly biased estimators. In such cases, more advanced estimators need to be used that are robust with respect to outliers in the measurements. If you would like to learn more about estimation theory and statistical robustness in particular, we recommend our seminars on *Signal Detection and Parameter Estimation* (SDPE) and on *Advance Topics in Statistical Signal Processing* (ATISSP), which are held in the winter and summer semester, respectively.

7.5 Wiener Filter

The matched filter considered in the previous section is an optimal filter in the sense that it provides the highest SNR at the output of a linear time-invariant system for the *detection* of the presence of a known signal received in noise. The Wiener filter considered in this section aims at providing optimal *estimation* of the sample function of random process from observations of another random process. More specifically, we consider a system configuration as illustrated in Figure 7.19 where $X(n)$ is the random process to be estimated, $Y(n)$ is the observed random process and $\varepsilon(n)$ is the error process. The goal is to design a linear time-invariant filter with unit sample response $h(n)$ so that the ensemble average (expected value) of the squared-error process is minimized. This is known as the minimum mean-squared error (MSE) criterion.

The filter which minimizes the MSE criterion is known as a Wiener filter after Norbert Wiener who developed the continuous-time theory in the 1940's. The discrete-time theory which the following sections are based on, was actually developed in the same period by Andrei Kolmogorov yet it has become convention to refer to the discrete-time filter as Wiener filter. The assumptions made for the development in the following sections are that $X(n)$ and $Y(n)$ are real-valued, zero-mean, jointly wide-sense stationary (WSS) random processes.

Since the processes $X(n)$ and $Y(n)$ are jointly WSS and the filter with unit sample response $h(n)$ is assumed to be stable, the error process $\varepsilon(n)$ is also WSS and therefore the MSE, which is the second-order moment of $\varepsilon(n)$, does not depend on the time index n. The MSE is, however, expressed in terms of the filter response $h(n)$ as follows

$$
q(h) \triangleq \mathsf{E}\left[\varepsilon(n;h)^2\right]
$$

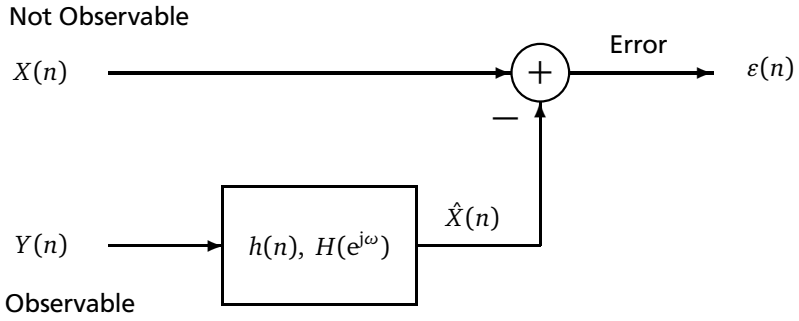


Figure 7.19: Wiener filter configuration.

$$
\begin{aligned}
&= \mathsf{E}\left[\left(X(n)-\sum_m h(m)Y(n-m)\right)^2\right] \\
&= \mathsf{E}\left[X(n)^2\right]+\sum_m\sum_l h(m)h(l)\mathsf{E}\left[Y(n-m)Y(n-l)\right] \qquad (7.13)\\
&\quad -2\sum_m h(m)\mathsf{E}\left[X(n)Y(n-m)\right] \\
&= c_{XX}(0)+\sum_m\sum_l h(m)h(l)c_{YY}(l-m)-2\sum_m h(m)c_{XY}(m) \qquad (7.14)
\end{aligned}
$$

The optimal (Wiener) filter is the filter with unit sample response $h_{\text{opt}}(n)$, defined such that

$$h_{\text{opt}}(n) = \arg\min_h q(h),\ n \in \mathbb{Z} \qquad (7.15)$$

7.5.1 Optimal Solution

To perform the minimization in Equation (7.15), we differentiate (7.14) with respect to h,

$$
\begin{aligned}
\frac{\partial q(h)}{\partial h} &= \frac{\partial \mathsf{E}\left[\varepsilon(n;h)^2\right]}{\partial h} \\
&= 2\mathsf{E}\left[\varepsilon(n;h)\cdot\frac{\partial\varepsilon(n;h)}{\partial h}\right] \\
&= -2\mathsf{E}\left[\varepsilon(n;h)\cdot\frac{\partial\sum_\kappa h(\kappa)Y(n-\kappa)}{\partial h}\right] \\
&= -2\sum_\kappa \mathsf{E}\left[\varepsilon(n)Y(n-\kappa)\right].
\end{aligned}
$$

The optimal filter $h_{\text{opt}}(n)$ is obtained by setting $\partial q(h)/\partial h = 0$ which requires

$$\boxed{c_{\varepsilon Y}(\kappa) = \mathsf{E}\left[\varepsilon(n)Y(n-\kappa)\right] = 0, \quad \forall\kappa\in\mathbb{Z},} \qquad (7.16)$$

Therefore the solution Equation (7.15) is a filter with unit unit sample response $h_{\text{opt}}(n)$, $n \in \mathbb{Z}$ for which the error process $\varepsilon(n)$ is orthogonal to the observed process $Y(n)$. This is known as the *orthogonality principle*. A geometric interpretation of this result is shown in Figure 7.20 where $X(n)$, $\varepsilon(n)$ and $Y(n-\kappa)$ are thought of as vectors. The optimal estimate $\hat{X}(n)$ is the projection of the desired signal $X(n)$ onto the observations $Y(n-\kappa)$, and the error vector is perpendicular to the plane on which the vectors $Y(n-\kappa)$, $\kappa\in\mathbb{Z}$ exist.

Expanding the cross-covariance function of the error and observations,

$$
\begin{aligned}
c_{\varepsilon Y}(\kappa) &= \mathsf{E}\left[\varepsilon(n)Y(n-\kappa)\right] \\
c_{\varepsilon Y}(\kappa) &= \mathsf{E}\left[\left(X(n)-\sum_m h(m)Y(n-m)\right)Y(n-\kappa)\right] \\
&= \mathsf{E}\left[X(n)Y(n-\kappa)\right]-\sum_m h(m)\mathsf{E}\left[Y(n-m)Y(n-\kappa)\right]
\end{aligned}
$$

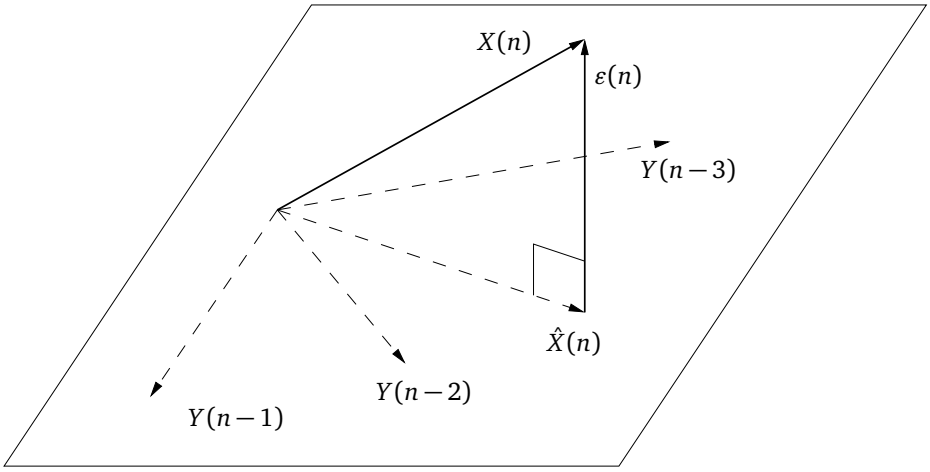


Figure 7.20: Geometric interpretation of the orthogonality principle, where the dashed (——) lines indicate co-planar vectors.

$$= \quad c_{XY}(\kappa) - \sum_m h(m) c_{YY}(\kappa - m). \tag{7.17}$$

If the orthogonality condition (7.16) is met, then from Equation (7.17) we obtain the following

$$\boxed{c_{XY}(\kappa) = h_{\text{opt}}(\kappa) \star c_{YY}(\kappa), \kappa \in \mathbb{Z}} \tag{7.18}$$

$$\boxed{C_{XY}(e^{j\omega}) = H_{\text{opt}}(e^{j\omega}) C_{YY}(e^{j\omega}), \omega \in \mathbb{R}} \tag{7.19}$$

which are known as the Wiener-Hopf equations. Clearly

$$\boxed{H_{\text{opt}}(e^{j\omega}) = \frac{C_{XY}(e^{j\omega})}{C_{YY}(e^{j\omega})}}$$

which in general provides a non-causal and therefore unrealizable solution.

The minimum MSE, q_{min}, obtained when using $h_{\text{opt}}(n)$ is found by substituting Equation (7.18) into (7.14)

$$\begin{aligned} q_{\text{min}} &= c_{XX}(0) - \sum_m \sum_l h_{\text{opt}}(m) h_{\text{opt}}(l) c_{YY}(l-m) \\ &= c_{XX}(0) - \sum_m h_{\text{opt}}(m) c_{XY}(m). \end{aligned} \tag{7.20}$$

An equivalent frequency-domain expression for q_{min} can be obtained by considering the following sequence

$$p(\kappa) = c_{XX}(\kappa) - h_{\text{opt}}(\kappa) \star c_{YX}(\kappa) = c_{XX}(\kappa) - \sum_m h_{\text{opt}}(m) c_{YX}(\kappa - m).$$

Clearly Equation (7.20) corresponds to $p(0)$, since $c_{YX}(-\kappa) = c_{XY}(\kappa)$ for real-valued $X(n)$ and $Y(n)$. Therefore we obtain

$$\begin{aligned} q_{\text{min}} = p(0) &= \int_{-\pi}^{\pi} P(e^{j\omega}) \frac{d\omega}{2\pi} \\ &= \int_{-\pi}^{\pi} \left[C_{XX}(e^{j\omega}) - H_{\text{opt}}(e^{j\omega}) C_{YX}(e^{j\omega}) \right] \frac{d\omega}{2\pi} \\ &= \int_{-\pi}^{\pi} \left[C_{XX}(e^{j\omega}) - \frac{|C_{XY}(e^{j\omega})|^2}{C_{YY}(e^{j\omega})} \right] \frac{d\omega}{2\pi} \end{aligned}$$

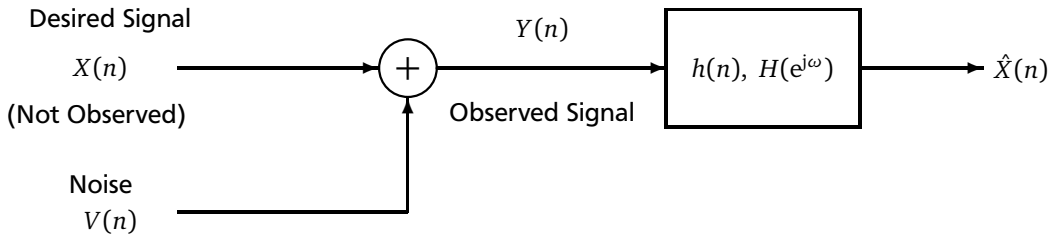


Figure 7.21: Wiener filter configuration for a random process observed in additive noise.

7.5.2 The Wiener Filter for Additive Noise

The particular case of additive noise is now considered, and the optimal filter is derived independently of the orthogonality principle. The observed process $Y(n)$ is assumed to be a sum of the desired random process $X(n)$ and a noise process $V(n)$ as illustrated in Figure 7.21.

The MSE for the case of additive noise is given by

$$\begin{aligned} q(h) &\triangleq \mathsf{E}\left[\varepsilon(n;h)^2\right] \\ &= \mathsf{E}\left[\left(X(n)-\sum_m h(m)[X(n-m)+V(n-m)]\right)^2\right] \\ &= c_{XX}(0)-\sum_m h(m)\left[c_{XX}(m)+c_{XV}(m)+c_{VX}(-m)+c_{XX}(-m)\right] \\ &\quad +\sum_m\sum_l h(m)h(l)\left[c_{XX}(l-m)+c_{XV}(l-m)+c_{VX}(l-m)+c_{VV}(l-m)\right] \end{aligned}$$

Assuming that the noise and the signal are uncorrelated, i.e. $c_{XV}(\kappa)=c_{VX}(\kappa)=0,\ \forall\kappa\in\mathbb{Z}$,

$$\begin{aligned} q(h) &= c_{XX}(0)-2\sum_m h(m)c_{XX}(m) \\ &\quad +\sum_m\sum_l h(m)h(l)\left[c_{XX}(l-m)+c_{VV}(l-m)\right] \end{aligned} \tag{7.21}$$

An equivalent frequency-domain expression for the MSE can be obtained by considering the following sequence

$$p(\kappa) = c_{XX}(\kappa)-2h(\kappa)\star c_{XX}(\kappa)+h(\kappa)\star\left[c_{XX}(\kappa)+c_{VV}(\kappa)\right]\star h(-\kappa)$$

where Equation (7.21) corresponds to $p(0)$. Assuming $h(n)$ is real-valued, the MSE can be expressed as

$$\begin{aligned} p(0) &= \int_{-\pi}^{\pi} P(e^{j\omega})\frac{d\omega}{2\pi} \\ &= \int_{-\pi}^{\pi}\left\{C_{XX}(e^{j\omega})-2H(e^{j\omega})C_{XX}(e^{j\omega})+H(e^{j\omega})H(e^{j\omega})^*\left[C_{XX}(e^{j\omega})+C_{VV}(e^{j\omega})\right]\right\}\frac{d\omega}{2\pi}. \end{aligned}$$

Since $h(n)=h^*(n)$, the property $H(e^{j\omega})=H(e^{-j\omega})^*$ holds, and therefore $2H(e^{j\omega})=H(e^{j\omega})+H(e^{-j\omega})^*$. Since $X(n)$ is real-valued, the spectrum $C_{XX}(e^{j\omega})$ is an even function of ω and

$$\int_{-\pi}^{\pi} H(e^{-j\omega})^* C_{XX}(e^{j\omega})\frac{d\omega}{2\pi} = \int_{-\pi}^{\pi} H(e^{j\omega})^* C_{XX}(e^{j\omega})\frac{d\omega}{2\pi},$$

which means that

$$p(0)=\int_{-\pi}^{\pi}\left\{C_{XX}(e^{j\omega})-\left[H(e^{j\omega})+H(e^{j\omega})^*\right]C_{XX}(e^{j\omega})+H(e^{j\omega})H(e^{j\omega})^*\left[C_{XX}(e^{j\omega})+C_{VV}(e^{j\omega})\right]\right\}\frac{d\omega}{2\pi}$$

It can be seen that the above integrand contains a quadratic expression plus a term independent of $H(e^{j\omega})$. Therefore, we can find the minimum with respect to $H(e^{j\omega})$ by finding the zeros (with respect to $H(e^{j\omega})$). Completing the square gives

$$p(0) = \int_{-\pi}^{\pi} \left\{ \left[H(e^{j\omega}) - \frac{C_{XX}(e^{j\omega})}{C_{XX}(e^{j\omega}) + C_{VV}(e^{j\omega})} \right] \left[H(e^{j\omega})^* - \frac{C_{XX}(e^{j\omega})}{C_{XX}(e^{j\omega}) + C_{VV}(e^{j\omega})} \right] \right.$$
$$\left. \left[C_{XX}(e^{j\omega}) + C_{VV}(e^{j\omega}) \right] \underbrace{- \frac{C_{XX}(e^{j\omega})^2}{C_{XX}(e^{j\omega}) + C_{VV}(e^{j\omega})} + C_{XX}(e^{j\omega})}_{\text{does not involve } H(e^{j\omega})} \right\} \frac{d\omega}{2\pi}$$

Clearly the minimum MSE can be obtained by choosing $H(e^{j\omega})$ as

$$H_{\text{opt}}(e^{j\omega}) = \frac{C_{XX}(e^{j\omega})}{C_{XX}(e^{j\omega}) + C_{VV}(e^{j\omega})} \tag{7.22}$$

where the minimum MSE is given by

$$q_{\min} = \int_{-\pi}^{\pi} \frac{C_{XX}(e^{j\omega}) C_{VV}(e^{j\omega})}{C_{XX}(e^{j\omega}) + C_{VV}(e^{j\omega})} \frac{d\omega}{2\pi}$$

It is noted that for the assumption of uncorrelated noise, $C_{XY}(e^{j\omega}) = C_{XX}(e^{j\omega})$ and $C_{YY}(e^{j\omega}) = C_{XX}(e^{j\omega}) + C_{VV}(e^{j\omega})$. Therefore the solution obtained in Equation (7.22) is consistent with that obtained from the Wiener-Hopf equations. The frequency response of the optimal filter given in Equation (7.22) can be expressed as

$$H_{\text{opt}}(e^{j\omega}) = \frac{1}{1 + \frac{C_{VV}(e^{j\omega})}{C_{XX}(e^{j\omega})}}$$

which may be interpreted as a frequency-domain weighting function. When the SNR at a particular frequency is high, then $H_{\text{opt}}(e^{j\omega})$ is close to one, i.e.

$$H_{\text{opt}}(e^{j\omega}) \simeq 1 \text{ when } \frac{C_{XX}(e^{j\omega})}{C_{VV}(e^{j\omega})} >> 1$$

and similarly, when the SNR is low then $H_{\text{opt}}(e^{j\omega})$ is close to zero.

Example 7.5.1 *In hearing aids noise cancellation is used to supress surrounding noise sources. Figure 7.22 illustrates the concept of noise cancellation in hearing aids.*

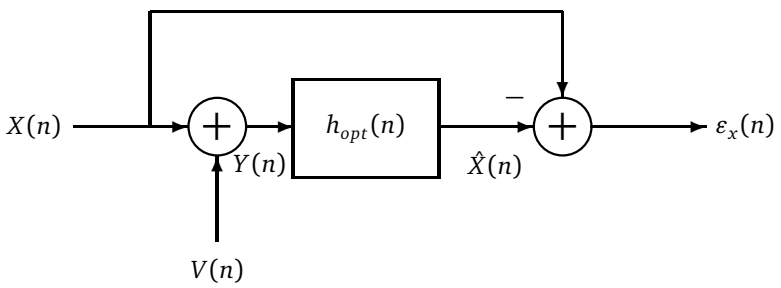


Figure 7.22: Noise cancellation in hearing aids.

The microphone of the hearing aid receives a zero-mean, noisy speech signal $Y(n)$. It consists of a zero-mean speech signal $X(n)$ with power σ_X^2 and an uncorrelated, zero-mean added noise $V(n)$ with power σ_V^2. The received microphone

signal $Y(n)$ is filtered with the unit sample response $h_{opt}(n)$. The aim of this filter is to estimate the clean speech signal $X(n)$. An error signal can, thus, be defined as $\epsilon_x(n)$, the difference between the clean signal $X(n)$ and the estimated signal $\hat{X}(n)$.
The problem at hand is that of estimating a random process we cannot observe, using an observed random process. Wiener filters provide an optimal solution to this problem, as they minimize the mean squared error $\mathsf{E}\left[\varepsilon_x^2(n)\right] = \mathsf{E}\left[(X(n)-\hat{X}(n))^2\right] = \mathsf{E}\left[(X(n)-Y(n)*h(n))^2\right]$.
Minimizing the MSE with respect to $h(n)$ leads to the Wiener-Hopf equations. Hence, the optimal frequency response $H_{\text{opt}}(\mathrm{e}^{\mathrm{j}\omega})$ of the filter is given by:

$$H_{\text{opt}}(\mathrm{e}^{\mathrm{j}\omega}) = \frac{C_{XY}(\mathrm{e}^{\mathrm{j}\omega})}{C_{YY}(\mathrm{e}^{\mathrm{j}\omega})}$$

The auto- and cross-covariance functions can be calculated:

$$\begin{aligned} c_{XY}(\kappa) &= \mathsf{E}[X(n+\kappa)Y(n)] \\ &= \mathsf{E}[X(n+\kappa)(X(n)+V(n))] \\ &= \mathsf{E}[X(n+\kappa)X(n)+X(n+\kappa)V(n)] \\ &= \mathsf{E}[X(n+\kappa)X(n)]+\mathsf{E}[X(n+\kappa)V(n)] \\ &= c_{XX}(\kappa)+c_{XV}(\kappa) \\ &= c_{XX}(\kappa). \end{aligned}$$

$$\begin{aligned} c_{YY}(\kappa) &= \mathsf{E}[Y(n+\kappa)Y(n)] \\ &= \mathsf{E}[(X(n+\kappa)+V(n+\kappa))(X(n)+V(n))] \\ &= \mathsf{E}[X(n+\kappa)X(n)]+\mathsf{E}[X(n+\kappa)V(n)] \\ &\quad +\mathsf{E}[V(n+\kappa)X(n)]+\mathsf{E}[V(n+\kappa)V(n)] \\ &= c_{XX}(\kappa)+c_{XV}(\kappa)+c_{VX}(\kappa)+c_{VV}(\kappa) \\ &= c_{XX}(\kappa)+c_{VV}(\kappa) \end{aligned}$$

Taking the Fourier transforms of the covariance functions leads to the auto- and cross-spectrum respectively:

$$C_{XY}(\mathrm{e}^{\mathrm{j}\omega}) = \sigma_X^2 \text{ and } C_{YY}(\mathrm{e}^{\mathrm{j}\omega}) = \sigma_X^2+\sigma_V^2.$$

The frequency response of the optimal filter can, thus, be given by:

$$H_{\text{opt}}(\mathrm{e}^{\mathrm{j}\omega}) = \frac{\sigma_X^2}{\sigma_X^2+\sigma_V^2} = \frac{1}{1+\frac{\sigma_V^2}{\sigma_X^2}}.$$

Let the SNR be defined as the quotient of the signal power and the noise power: $SNR = \frac{\sigma_X^2}{\sigma_V^2}$. If we assume a high SNR at a given frequency, we can see that the frequency response of the optimal filter approaches 1. This is an intuitive result, as we expect the filter to let everything pass if the input mainly consists of the desired signal. However, if the input signal consists mainly of noise, we expect the filter to block everything from passing. This is essentially the result we get when the SNR has very low values, the optimal frequency response at a given w approaches 0. This means that the hearing aid will adjust its behaviour depending on the incoming signal. The desired signal will be passed through while the noise will be blocked.
To be able to implement the Wiener filter in a real hearing aid, we need information about the auto-spectrum $C_{XX}(\mathrm{e}^{\mathrm{j}\omega})$ of the desired signal, which we obviously do not have. This is typically estimated using an estimate of $C_{VV}(\mathrm{e}^{\mathrm{j}\omega})$, which can be performed in speech pauses. Once this is given, an estimate $\hat{C}_{XX}(\mathrm{e}^{\mathrm{j}\omega})$ can be calculated as $\hat{C}_{XX}(\mathrm{e}^{\mathrm{j}\omega}) = \hat{C}_{YY}(\mathrm{e}^{\mathrm{j}\omega}) - \hat{C}_{VV}(\mathrm{e}^{\mathrm{j}\omega})$. Clearly, $\hat{C}_{YY}(\mathrm{e}^{\mathrm{j}\omega})$ is obtained from observations of $Y(n)$.

7.5.3 The Causal Wiener Filter

Causal Solution for White Processes

The optimal filter obtained from the Wiener-Hopf equations is, in general, non-causal. In order to implement the filter in practice, we seek a linear filter which provides the lowest MSE out of the class of *causal* filters. It will be shown that this solution can be obtained from the case when the observed random process $Y(n)$ is white. Considering again the system depicted in Figure 7.19; with white noise observations,

$$c_{YY}(\kappa) = \sigma_Y^2 \delta(\kappa).$$

Using Equation (7.14) the MSE is given by

$$\begin{aligned} q(h) &= c_{XX}(0) + \sum_m \sum_l h(m)h(l)\sigma_Y^2 \delta(l-m) - 2\sum_m h(m)c_{XY}(m) \\ &= c_{XX}(0) + \sum_m \left\{ [\sigma_Y h(m)]^2 - 2h(m)c_{XY}(m) \right\} \\ &= c_{XX}(0) + \sum_m \left[\sigma_Y h(m) - \frac{c_{XY}(m)}{\sigma_Y} \right]^2 - \frac{1}{\sigma_Y^2} \sum_m c_{XY}(m)^2 \end{aligned} \tag{7.23}$$

where only the middle term of (7.23) depends on $h(n)$ and is obviously minimized for

$$h(n) = \frac{c_{XY}(n)}{\sigma_Y^2}, \quad n \in \mathbb{Z}$$

which is in agreement with the solution provided by the Wiener-Hopf equations.

We now consider the restriction that the linear filter must be causal,

$$h(n) = 0, \quad n < 0.$$

Examining Equation (7.23), we see that this is obtained for

$$h(n) = \frac{1}{\sigma_Y^2} c_{XY}(n) u(n). \tag{7.24}$$

Therefore the optimal causal filter when $Y(n)$ is white noise is given by Equation (7.24) and will be denoted by $h_c(n)$. The overall MSE then becomes

$$\begin{aligned} q_{\min} &= c_{XX}(0) - \frac{1}{\sigma_Y^2} \sum_{m=0}^{\infty} c_{XY}(m)^2 \\ &= c_{XX}(0) - \sum_{m=0}^{\infty} h_c(m) c_{XY}(m) \end{aligned}$$

which is larger than the error for the non-causal solution. The error can also be expressed in the frequency-domain as

$$q_{\min} = \int_{-\pi}^{\pi} \left\{ C_{XX}(e^{j\omega}) - H_c(e^{j\omega}) C_{XY}(e^{j\omega})^* \right\} \frac{d\omega}{2\pi}$$

Causal Solution for Non-White Processes

In order to generalize this result to a non-white process, we model $Y(n)$ as shown in Figure 7.23 where $\tilde{Y}(n)$ is a white noise process of variance σ_Y^2 and $g(n)$ is a real, stable linear time-invariant filter. Assuming the system specified by $g(n)$ can be expressed as a linear constant coefficient difference equation, then $G(z) = B(z)/A(z)$ where $B(z)$ and $A(z)$ are polynomials in z whose roots denote the zeros and poles of $G(z)$ respectively.

The covariance function of $Y(n)$ is given by

$$c_{YY}(\kappa) = g(\kappa) \star c_{\tilde{Y}\tilde{Y}}(\kappa) \star g(-\kappa). \tag{7.25}$$

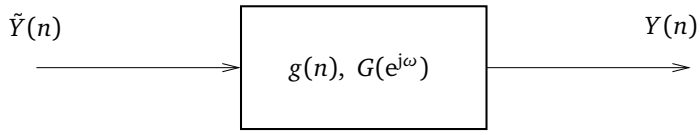


Figure 7.23: Modeling of $Y(n)$ as a filtered white noise process.

For a real linear time-invariant filter with unit sample response $g(n)$, the following relationship holds

$$g(-n) \leftrightarrow G(z^{-1})$$

where $G(z)$ is the $\mathcal{Z}$-transform of $g(n)$. Therefore from Equation (7.25) the spectrum of $Y(n)$ can be expressed as

$$C_{YY}(z) = \mathcal{Z}\{c_{YY}(\kappa)\} = G(z)G(z^{-1})\sigma_Y^2.$$

$C_{YY}(z)$ is then a rational function with zeros and poles in reciprocal pairs in the $\mathcal{Z}$-plane. By assigning all the poles and zeros of $C_{YY}(z)$ that are inside the unit circle to $G(z)$, and all those outside to $G(z^{-1})$ results in a causal, stable filter with transfer function $G(z)$. Provided no zeros lie exactly on the unit circle[2] then the inverse system with transfer function $G^{-1}(z)$ is also causal and stable. Applying $G^{-1}(z)$ to $Y(n)$ results in the white noise process $\tilde{Y}(n)$. Denoting $G^{-1}(z) \equiv H_1(z) \leftrightarrow h_1(n)$ we can represent the causal Wiener filter solution as shown in Figure 7.24. The application of the filter with unit sample response $h_1(n)$ to the signal $Y(n)$ is termed *pre-whitening* since the

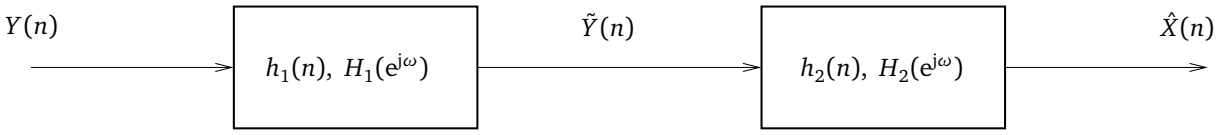


Figure 7.24: Causal Wiener filter system.

output $\tilde{Y}(n)$ is a white noise process. The filter with unit sample response $h_2(n)$, applied to $\tilde{Y}(n)$ is then determined according to Equation (7.24) as

$$\begin{aligned} h_2(n) = \frac{1}{\sigma_{\tilde{Y}}^2} c_{X\tilde{Y}}(n)u(n) &= \frac{1}{\sigma_{\tilde{Y}}^2} \mathsf{E}\left[X(m+n)\tilde{Y}(m)\right]u(n) \\ &= \frac{1}{\sigma_{\tilde{Y}}^2}\left[\sum_l h_1(l)\mathsf{E}\left[X(m+n)Y(m-l)\right]\right]u(n) \\ &= \frac{1}{\sigma_{\tilde{Y}}^2}\left[h_1(-n) \star c_{XY}(n)\right]u(n). \end{aligned}$$

The overall optimal causal filter system function is then described by

$$\begin{aligned} H_c(z) = H_1(z)H_2(z) &= H_1(z) \cdot \frac{1}{\sigma_{\tilde{Y}}^2}\left[H_1(z^{-1})C_{XY}(z)\right]^+ \\ &= \frac{1}{\sigma_{\tilde{Y}}^2 G(z)}\left[\frac{C_{XY}(z)}{G(z^{-1})}\right]^+ \end{aligned} \tag{7.26}$$

where $[\cdot]^+$ denotes the causal component, i.e.

$$[F(z)]^+ \triangleq \mathcal{Z}\{f(n)u(n)\}.$$

Example 7.5.2 *We wish to determine the Wiener filter for estimating a process $X(n)$ whose spectrum is given by*

$$C_{XX}(z) = \frac{0.19}{(1-0.9z^{-1})(1-0.9z)}, \quad 0.9 < |z| < (0.9)^{-1}.$$

[2]Equivalently we require $C_{YY}(e^{j\omega}) \neq 0,\ \omega \in [-\pi, \pi]$.

The observed random process $Y(n)$ consists of $X(n)$ embedded in white noise $V(n)$ which has variance σ_V^2 and is uncorrelated with $X(n)$.

The spectrum of the observed process is given by

$$\begin{aligned} C_{YY}(z) &= C_{XX}(z) + C_{VV}(z) \\ &= 1.436\left[\frac{1-0.627z^{-1}}{1-0.9z^{-1}}\right]\left[\frac{1-0.627z}{1-0.9z}\right] \\ &= \sigma_Y^2 G(z)G(z^{-1}) \end{aligned}$$

where the zero $z = 0.627$ and the pole $z = 0.9$ are assigned to $G(z)$. We find

$$\begin{aligned} \left[\frac{C_{XY}(z)}{G(z^{-1})}\right]^+ &= \left[\frac{0.19}{(1-0.9z^{-1})(1-0.627z)}\right]^+ \\ &= \left[\frac{0.436}{1-0.9z^{-1}} + \frac{0.273}{z^{-1}-0.627}\right]^+ \\ &= \frac{0.436}{1-0.9z^{-1}} \end{aligned}$$

where partial fraction expansion has been used to recover the causal component. From Equation (7.26) the optimal causal filter is described by

$$\begin{aligned} H_c(z) &= \frac{1}{\sigma_Y^2 G(z)} \frac{0.436}{1-0.9z^{-1}} \\ &= \frac{0.304}{1-0.627z^{-1}}. \end{aligned}$$

7.5.4 Finite Wiener Filters

We consider again the system depicted in Figure 7.19. However, it will be assumed that the random process $Y(n)$

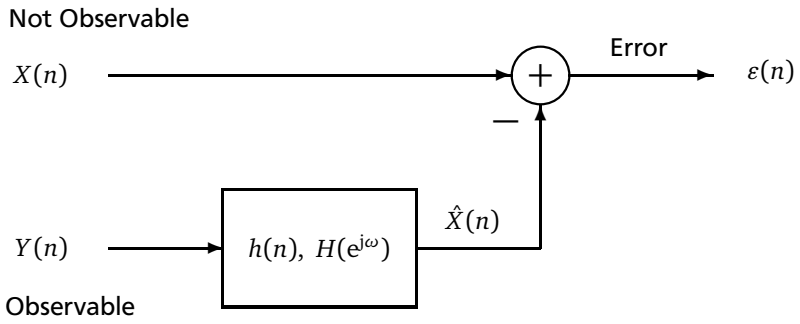


Figure 7.25: Wiener filter configuration.

is observed only over a finite discrete-time interval $\mathcal{I} = \{n-N_1, \ldots, n+N_2\}$ where $-N_2 \leq N_1$. It is desired to obtain an estimate $\hat{X}(n)$ for $X(n)$ by applying a linear filter with unit sample response $h(n)$,

$$\hat{X}(n) = \sum_{m=n-N_1}^{n+N_2} h(n-m)Y(m) = \sum_{m=-N_2}^{N_1} h(m)Y(n-m).$$

The filter $h(n)$ is therefore of finite length N_1+N_2+1 as shown in Figure 7.26. As discussed in the previous sections, we wish to determine the optimum solution, $h_{\text{opt}}(n)$, which minimizes the MSE of our estimate.

The relation between the time n at which $X(n)$ is to be estimated, and the observation interval $\mathcal{I}$, gives rise to three types of estimation problems:

1. *Filtering*
   For a given observation interval $\mathcal{I} = \{n-N_1, \ldots, n\}$, $N_1 > 0$, $X(n)$ is estimated from the N_1+1 most recent observations. The solution to this problem is a causal filter which can be implemented in 'real-time.'

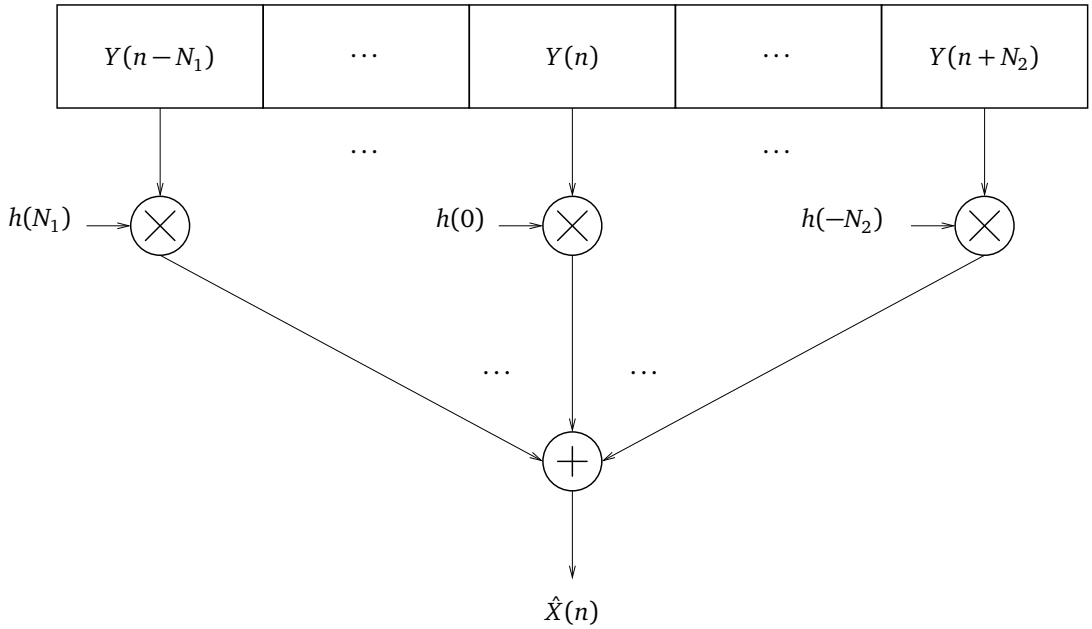


Figure 7.26: Finite length filter for estimating $X(n)$.

2. *Smoothing*
   The observation interval is $\mathcal{I} = \{n-N_1,\ldots,n+N_2\}$, $N_1,N_2 > 0$, so $X(n)$ is estimated from past and future observations. This is applicable in post-processing situations, when a realization of $Y(n)$ has been recorded and can be 'played back.'

3. *Prediction*
   The observation interval is given by $\mathcal{I} = \{n-N_1,\ldots,n-1\}$, where $N_1 > 0$. The problem is to predict $X(n)$ from past observations. Since $h(n)$ is specified to be the linear unit sample response of a system, then $\hat{X}(n)$ is a linear predictor of $X(n)$ in terms of the recent observations.

The optimum solution $h_{\mathsf{opt}}(n)$ satisfies the orthogonality condition,

$$\mathsf{E}\left[\varepsilon(n)Y(n-\kappa)\right] = 0, \quad \forall\, n-\kappa \in \mathcal{I} \;\equiv\; \forall\, \kappa \in [-N_2, N_1].$$

Therefore the Wiener-Hopf equations are given by

$$c_{XY}(\kappa) = \sum_{m=-N_2}^{N_1} h_{\mathsf{opt}}(m)c_{YY}(\kappa-m), \quad \forall\, \kappa \in [-N_2, N_1]. \tag{7.27}$$

The optimum filter must satisfy the N_1+N_2+1 equations given above, which can be represented in vector form as

$$\mathbf{c}_{XY} = \mathbf{C}_{YY}\mathbf{h} \tag{7.28}$$

where

$$\begin{aligned}
\mathbf{h} &\triangleq [h(-N_2),\ldots,h(N_1)]^T \\
\mathbf{c}_{XY} &\triangleq [c_{XY}(-N_2),\ldots,c_{XY}(N_1)]^T \\
\mathbf{C}_{YY} &\triangleq \begin{bmatrix} c_{YY}(0) & c_{YY}(1) & \cdots & c_{YY}(N_1+N_2) \\ c_{YY}(1) & c_{YY}(0) & \cdots & c_{YY}(N_1+N_2-1) \\ \vdots & \vdots & \ddots & \vdots \\ c_{YY}(N_1+N_2) & c_{YY}(N_1+N_2-1) & \cdots & c_{YY}(0) \end{bmatrix}
\end{aligned}$$

and $[\cdot]^T$ denotes the matrix transpose operation. The solution of (7.28) may be obtained as

$$\mathbf{h}_{\mathsf{opt}} = \left(\mathbf{C}_{YY}\right)^{-1}\mathbf{c}_{XY}$$

provided the matrix $\mathbf{C}_{YY}$ is invertible. The MSE of the optimum filter is given by

$$\begin{aligned} q_{\min} &= c_{XX}(0) - \sum_{m=-N_2}^{N_1} h_{\text{opt}}(m) c_{XY}(m) \\ &= c_{XX}(0) - \mathbf{h}_{\text{opt}}{}^T \mathbf{c}_{XY} \\ &= c_{XX}(0) - \mathbf{c}_{XY}{}^T \big(\mathbf{C}_{YY}\big)^{-1} \mathbf{c}_{XY} \end{aligned}$$

Example 7.5.3 *We consider the filtering problem for estimating $X(n)$ from the $N_1 + 1$ most recent noisy observations. The observed process is given by*

$$Y(n) = X(n) + V(n), \quad n \in \mathcal{I} = \{n - N_1, \ldots, n\}$$

where $V(n)$ is a noise process which is uncorrelated with $X(n)$.

Equation (7.27) becomes

$$c_{XY}(\kappa) = \sum_{m=0}^{N_1} h_{\text{opt}}(m) c_{YY}(\kappa - m), \quad \kappa \in [0, N_1].$$

The cross-covariance function is given by

$$c_{XY}(\kappa) = \mathsf{E}\big[X(n+\kappa)\big(X(n) + V(n)\big)\big] = c_{XX}(\kappa),$$

and the covariance function is given by

$$\begin{aligned} c_{YY}(\kappa) &= \mathsf{E}\big[\big(X(n+\kappa) + V(n+\kappa)\big)\big(X(n) + V(n)\big)\big] \\ &= c_{XX}(\kappa) + c_{VV}(\kappa) \end{aligned}$$

since $X(n)$ and $V(n)$ are uncorrelated. Therefore, the optimal filter is obtained from

$$c_{XX}(\kappa) = \sum_{m=0}^{N_1} h_{\text{opt}}(m)[c_{XX}(\kappa - m) + c_{VV}(\kappa - m)], \quad \kappa \in [0, N_1]$$

where the $N_1 + 1$ equations can be solved for the $N_1 + 1$ unknowns $h_{\text{opt}}(0), \ldots, h_{\text{opt}}(N_1)$.

Bibliography

J.F. Böhme. *Stochastische Signale*. Teubner Studienbücher, 1998.

Steven M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, Inc., 1993.

A. Papoulis. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, Inc., third edition, 1991.

P. Z. Peebles, Jr. *Probability, Random Variables and Random Signal Principles*. McGraw-Hill, Inc., fourth edition, 2001.

A Discrete-Time Fourier Transform

Sequence	Fourier Transform
$ax(n)+by(n)$	$aX(e^{j\omega})+bY(e^{j\omega})$
$x(n-n_d)$ (n_d integer)	$e^{-j\omega n_d}X(e^{j\omega})$
$e^{-j\omega_0 n}x(n)$	$X(e^{j(\omega+\omega_0)})$
$x(-n)$	$X(e^{-j\omega})$
	$X(e^{j\omega})^*$ if $x(n)$ real
$nx(n)$	$j\dfrac{dX(e^{j\omega})}{d\omega}$
$x(n)*y(n)$	$X(e^{j\omega})Y(e^{j\omega})$
$x(n)y(n)$	$\dfrac{1}{2\pi}\displaystyle\int_{-\pi}^{\pi} X(e^{j\theta})Y(e^{j(\omega-\theta)})d\theta$
$-jnx(n)$	$\dfrac{dX(e^{j\omega})}{d\omega}$
Symmetry properties:	
$x(n)^*$	$X(e^{-j\omega})^*$
$x(-n)^*$	$X(e^{j\omega})^*$
$x(n)$ real	$X(e^{j\omega})=X(e^{-j\omega})^*$
$x(n)$ real and even	$X(e^{j\omega})$ real and even
$x(n)$ real and odd	$X(e^{j\omega})$ pure imaginary and odd
Parseval's Theorem:	
$\displaystyle\sum_{n=-\infty}^{\infty}\lvert x(n)\rvert^2=\frac{1}{2\pi}\int_{-\pi}^{\pi}\lvert X(e^{j\omega})\rvert^2 d\omega$	
Special correspondences:	
$\delta(n)$	1
$\delta(n-n_0)$	$e^{-j\omega n_0}$
1, $(-\infty<n<\infty)$	$\displaystyle\sum_{k=-\infty}^{\infty}2\pi\delta(\omega+2\pi k)$
$a^n u(n)$, $(\lvert a\rvert<1)$	$\dfrac{1}{1-ae^{-j\omega}}$
$(n+1)a^n u(n)$ $(\lvert a\rvert<1)$	$\dfrac{1}{(1-ae^{-j\omega})^2}$
$u(n)$	$\dfrac{1}{1-e^{-j\omega}}+\displaystyle\sum_{k=-\infty}^{\infty}\pi\delta(\omega+2\pi k)$
$\dfrac{\sin(\omega_c n)}{\pi n}$	$X(e^{j\omega})=\begin{cases}1, & \lvert\omega\rvert<\omega_c\\ 0, & \omega_c<\lvert\omega\rvert\le\pi\end{cases}$
$x(n)=\begin{cases}1, & 0\le n\le M\\ 0, & \text{otherwise}\end{cases}$	$\dfrac{\sin[\omega(M+1)/2]}{\sin(\omega/2)}e^{-j\omega M/2}$
$e^{j\omega_0 m}$	$\displaystyle\sum_{k=-\infty}^{\infty}2\pi\delta(\omega-\omega_0+2\pi k)$
$\cos(\omega_0 n+\phi)$	$\displaystyle\sum_{k=-\infty}^{\infty}\pi\left[e^{j\phi}\delta(\omega-\omega_0+2\pi k)+e^{-j\phi}\delta(\omega+\omega_0+2\pi k)\right]$