Introduction to Random Processes
UDRC Summer School, 26th June 2017

Accompanying Notes

Dr James R. Hopgood

Room 2.05
Alexander Graham Bell Building
The King's Buildings
Mayfield Road
Edinburgh
EH9 3JL
Scotland, UK
James.Hopgood@ed.ac.uk
Telephone: +44 (0)131 650 5571
Fax: +44 (0)131 650 6554.

Major revision, Monday 19th June, 2017.
Last printed revision with minor corrections, 20 June, 2017.

Typeset by the author with the LaTeX 2_ε Documentation System, with $\mathcal{AMS}$-LaTeX Extensions, in 12/18 pt Times and Euler fonts.

INSTITUTE FOR DIGITAL COMMUNICATIONS,
School of Engineering,
College of Science and Engineering,
Kings's Buildings,
Edinburgh, EH9 3JL. U.K.

Copyright Statement

This document does not contain copyright material.

The author of this document

1. holds the copyright for all lecture and course materials in this module;
2. holds the copyright for students notes, summaries, or recordings that substantially reflect the lecture content or materials;
3. makes these materials available only for personal use by students studying this module;
4. reserves the right that no part of the notes, tutorials, solutions, or other course materials may be distributed or reproduced for commercial purposes without express written consent from the author; this does not prevent students from sharing notes on an individual basis for personal use.

These lecture notes consist of entirely original work, where all material has been written and typeset by the author. No figures or substantial pieces of text has been reproduced verbatim from other texts.

However, there is some material that has been based on work in a number of previous textbooks, and therefore some sections and paragraphs have strong similarities in structure and wording. These texts have been referenced and include, amongst a number of others, in order of contributions:

- Manolakis D. G., V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing*: *Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Processing*, McGraw Hill, Inc., 2000.

  IDENTIFIERS – *Paperback*, ISBN10: 0070400512, ISBN13: 9780070400511

- Therrien C. W., *Discrete Random Signals and Statistical Signal Processing*, Prentice-Hall, Inc., 1992.

  IDENTIFIERS – *Paperback*, ISBN10: 0130225452, ISBN13: 9780130225450
  Hardback, ISBN10: 0138521123, ISBN13: 9780138521127

- Kay S. M., *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Inc., 1993.

  IDENTIFIERS – *Hardback*, ISBN10: 0133457117, ISBN13: 9780133457117
  Paperback, ISBN10: 0130422681, ISBN13: 9780130422682

- Papoulis A. and S. Pillai, *Probability, Random Variables, and Stochastic Processes*, Fourth edition, McGraw Hill, Inc., 2002.

IDENTIFIERS – *Paperback*, ISBN10: 0071226613, ISBN13: 9780071226615
Hardback, ISBN10: 0072817259, ISBN13: 9780072817256

- Proakis J. G. and D. G. Manolakis, *Digital Signal Processing*: *Principles, Algorithms, and Applications*, Pearson New International Edition, Fourth edition, Pearson Education, 2013.

IDENTIFIERS – *Paperback*, ISBN10: 1292025735, ISBN13: 9781292025735

- Mulgew B., P. M. Grant, and J. S. Thompson, *Digital Signal Processing*: *Concepts and Applications*, Palgrave, Macmillan, 2003.

IDENTIFIERS – *Paperback*, ISBN10: 0333963563, ISBN13: 9780333963562

See `http://www.see.ed.ac.uk/~{}pmg/SIGPRO`

- Therrien C. W. and M. Tummala, *Probability and Random Processes for Electrical and Computer Engineers*, Second edition, CRC Press, 2011.

IDENTIFIERS – *Hardback*, ISBN10: 1439826986, ISBN13: 978-1439826980

- Press W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Receipes in C*: *The Art of Scientific Computing*, Second edition, Cambridge University Press, 1992.

IDENTIFIERS – *Paperback*, ISBN10: 0521437202, ISBN13: 9780521437202
Hardback, ISBN10: 0521431085, ISBN13: 9780521431088

The material in [Kay:1993] and [Papoulis:1991] is covered throughout the course; material in [Therrien:1992] and is covered primarily in the handouts on random processes. The following labelling convention is used for numbering equations that are taken from the various recommended texts. Equations labelled as:

M:v.w.xyz	are similar to those in [Manolakis:2001] with the corresponding label;
T:w.xyz	are similar to those in [Therrien:1992] with the corresponding label;
K:w.xyz	are similar to those in [Kay:1993] with the corresponding label;
P:v.w.xyz	are used in chapters referring to basic digital signal processing (DSP), and are references made to [Proakis:1996].

Contents

Acronyms

AC	autocorrelation
ACS	autocorrelation sequence
AIC	Akaike's information criterion
AIR	acoustic impulse response
AR	autoregressive
ARMA	autoregressive moving average
BIBO	bounded-input, bounded-output
BIC	B-Information criterion
BSS	blind source separation
CAT	Parzen's criterion autoregressive transfer function
CPSD	cross-power spectral density
DC	"direct current"
DFT	discrete Fourier transform
DSP	digital signal processing
DTFT	discrete-time Fourier transform
FIR	finite impulse response
FPE	final prediction error
FS	Fourier series
FT	Fourier transform
IDFT	inverse-DFT
KL	Karhunen-Loeve
LHS	left hand side
LS	least-squares
LSE	least-squares estimate
LTI	linear time-invariant

LTV	linear time-varying
MA	moving average
MDL	minimum description length
ML	maximum-likelihood
MLE	maximum-likelihood estimate
MS	mean-square
MSC	magnitude square coherence
MSE	mean-squared error
PSD	power spectral density
RHS	right hand side
ROC	region of convergence
SSS	strict-sense stationary
WGN	white Gaussian noise
WSP	wide-sense periodic
WSS	wide-sense stationary
cdf	cumulative distribution function
iff	if, and only if,
i. i. d.	independent and identically distributed
pdf	probability density function
RV	random variable
w. r. t.	with respect to

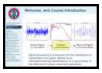
10;

1

Review of Fourier Transforms and Discrete-Time Systems

This handout will review complex Fourier series and Fourier transforms, followed by a review of discrete-time systems. It covers complex Fourier series, Fourier transforms, Discrete-time Fourier transforms, Discrete Fourier Transforms, Parseval's Theorem, the bilaterial Z-transform, frequency response, and rational transfer functions.

1.1 Obtaining the Latest Version of these Handouts



New slide

- This research tutorial is intended to cover a wide range of aspects which cover the fundamentals of statistical signal processing. It is written at a level which assumes knowledge of undergraduate mathematics and signal processing nomenclature, but otherwise should be accessible to most technical graduates.

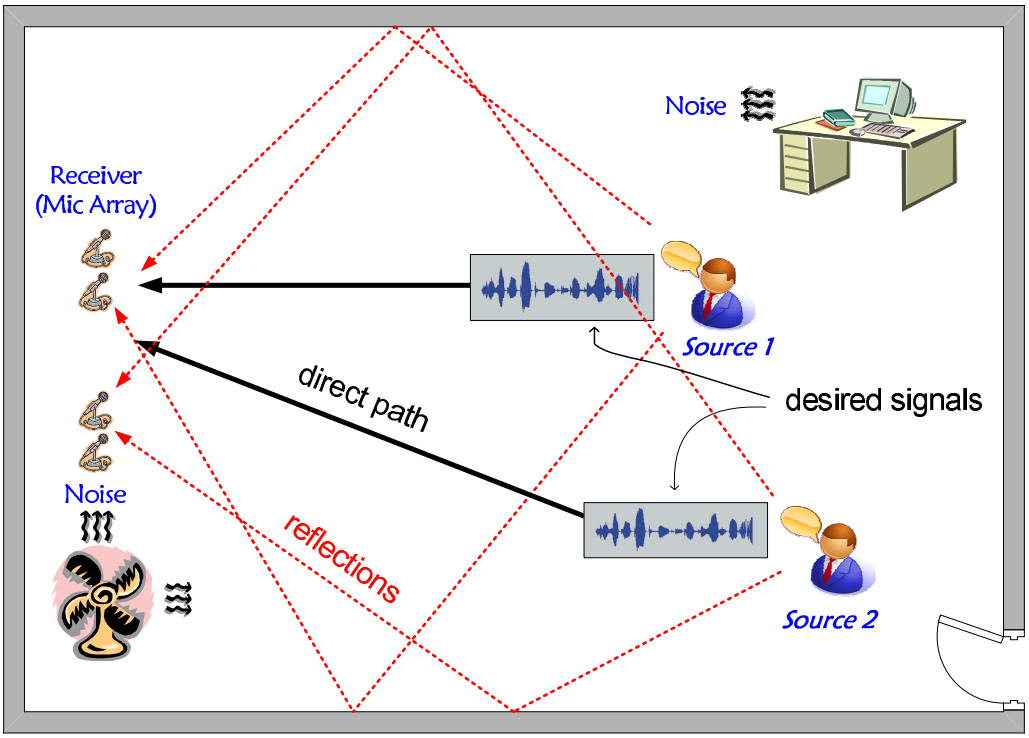


Figure 1.1: Source localisation and BSS. An example of topics using statistical signal processing.

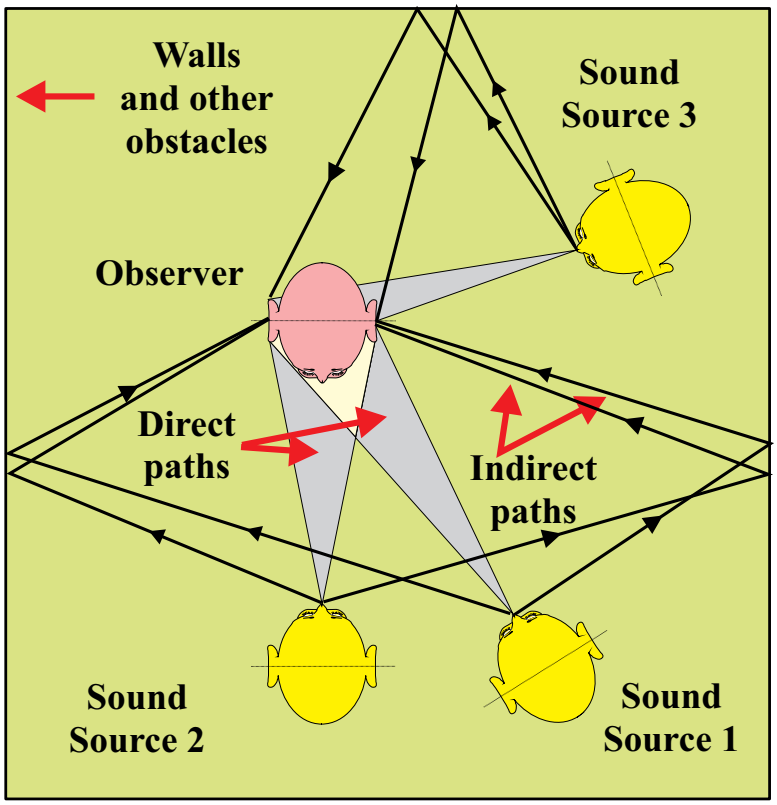


Figure 1.2: Humans turn their head in the direction of interest in order to reduce inteference from other directions; *joint detection, localisation, and enhancement*. An application of probability and estimation theory, and statistical signal processing.

KEYPOINT! (Latest Slides). Please note the following:

- This tutorial is being continually updated, and feedback is welcomed. The documents published on the USB stick may differ to the slides presented on the day. In particular, there are likely to be a few typos in the document, so if there is something that isn't clear, please feel free to email me so I can correct it (or make it clearer).
- The latest version of this document can be obtained from the author, Dr James R. Hopgood, by emailing him at: at:

  `mailto:james.hopgood@ed.ac.uk`

  (Update: The notes are no longer online due to the desire to maintain copyright control on the document.)
- Extended thanks are given to the many MSc students over the past 12 years who have helped proof-read and improve these documents.

1.2 Introduction

This handout will review complex **Fourier series** and **Fourier transform**s, followed by a review of **discrete-time system**s. The reader is expected to have previously covered most of the concepts in this handout, although it is likely that the reader might need to revise the material if it's been a while since it's been studied. Nevertheless, this revision material is included in the module as review material purely for completeness and reference. It is not intended as a full introduction, although some parts of the review cover the subject in detail.

As discussed in the first handout, if the reader wishes to revise these topics in more detail, the following book comes *highly* recommended:

> Proakis J. G. and D. G. Manolakis, *Digital Signal Processing*: *Principles, Algorithms, and Applications*, Pearson New International Edition, Fourth edition, Pearson Education, 2013.
>
> **IDENTIFIERS** – *Paperback*, ISBN10: 1292025735, ISBN13: 9781292025735

For undergraduate level text books covering signals and systems theory, which it is assumed you have covered, the following book is recommmended:

> Mulgew B., P. M. Grant, and J. S. Thompson, *Digital Signal Processing*: *Concepts and Applications*, Palgrave, Macmillan, 2003.
>
> **IDENTIFIERS** – *Paperback*, ISBN10: 0333963563, ISBN13: 9780333963562

See `http://www.see.ed.ac.uk/~{}pmg/SIGPRO`

The latest edition was printed in 2002, but any edition will do. An alternative presentation of roughly the same material is provided by the following book:

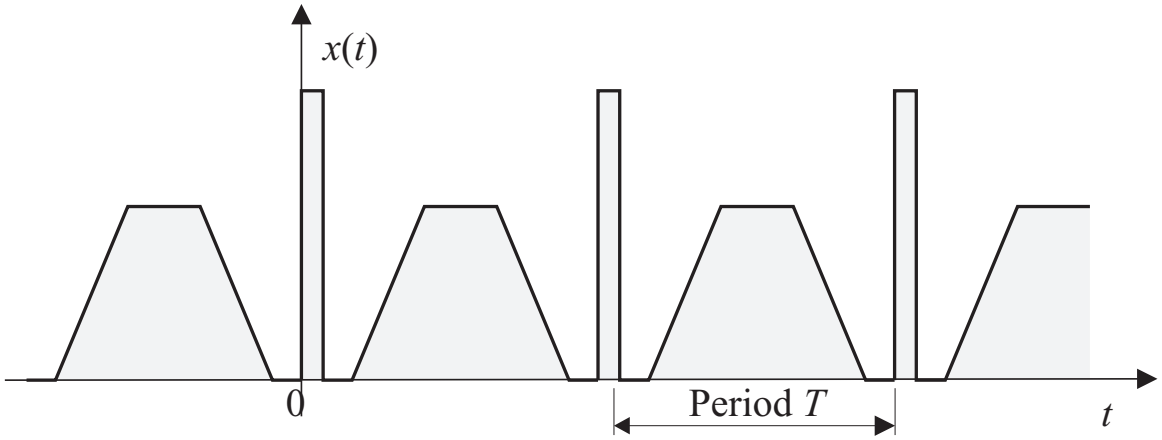


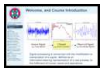
Figure 1.3: An example of a periodic signal with period T.

Balmer L., *Signals and Systems*: *An Introduction*, Second edition, Prentice-Hall, Inc., 1997.

IDENTIFIERS – *Paperback*, ISBN10: 0134954729, ISBN13: 9780134956725

In particular, the appendix on complex numbers may prove useful.

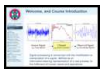
1.3 Signal Classification



New slide

Before considering the analysis of signals and systems, it is necessary to be aware of the general classifications to which signals can belong, and to be aware of the significance of some subtle characteristics that determine how a signal can be analysed. Not all signals can be analysed using a particular technique.

1.3.1 Types of signal



New slide

In general, there are four distinct types of signals that must be analysed:

Continuous-time periodic Such signals repeat themselves after a fixed length of time known as the period, usually denoted by T. This repetition continues ad-infinitum (i.e. forever).

Formally, a signal, $x(t)$, is periodic if

$$x(t) = x(t + m\,T)\,,\ \forall\, m \in \mathbb{Z} \tag{1.1}$$

where the notation $\forall\, m \in \mathbb{Z}$ means that m takes on *all* integer values: in other-words, $m = -\infty, \ldots, -2, -1, 0, 1, 2, \ldots, \infty$. The smallest positive value of T which satisfies this condition is the defined as the **fundamental period**.

These signals will be analysed using the **Fourier Series**, and are used to represent real-world waveforms that are near to being periodic over a sufficiently long period of time.

An example of a periodic signal is shown in Figure 1.3. This kind of signal vaguely represents a line signal in analogue television, where the rectangular pulses represent line synchronisation signals.

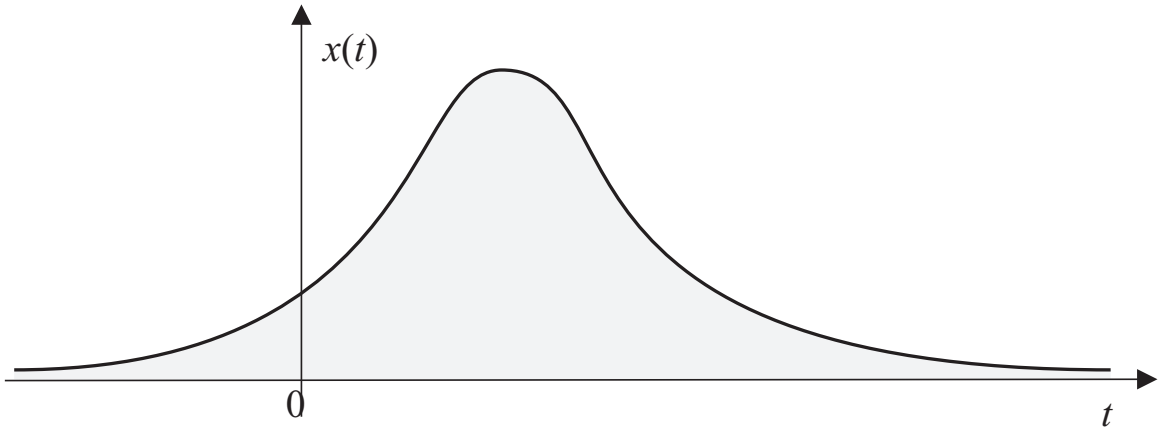


Figure 1.4: An example of an aperiodic signal.

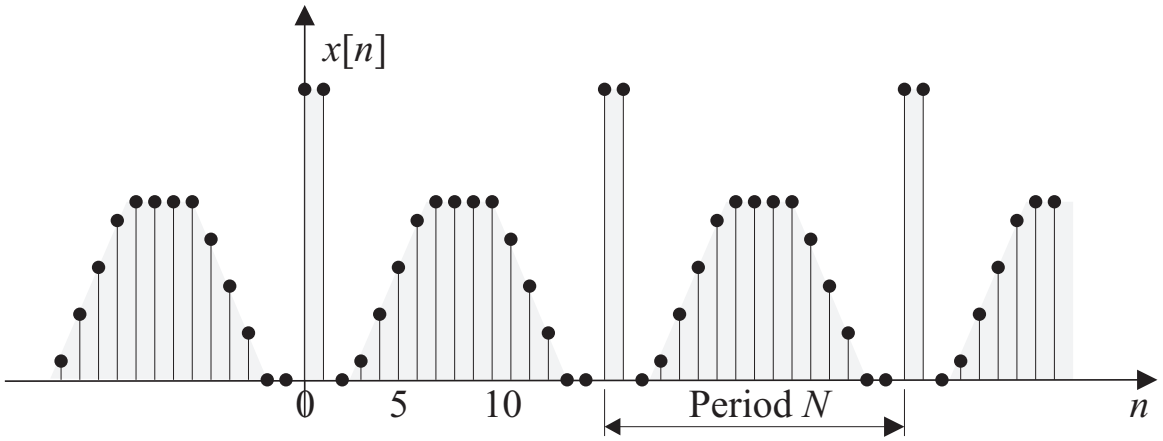


Figure 1.5: A discrete-time periodic signal.

Continuous-time aperiodic Continuous-time aperiodic signals are not periodic over all time, although they might be locally periodic over a short time-scale.

These signals can be analysed using the **Fourier transform** for most cases, and more often using the **Laplace transform**. Aperiodic signals are more representative of many real-world signals. Again, real signals don't last for all time, although can last for a considerably long time. An example of an aperiodic signal is shown in Figure 1.4.

Discrete-time periodic A discrete-time periodic signal is shown in Figure 1.5, which is essentially a *sampled* version of the signal shown in Figure 1.3. Note in this case, the period is often denoted by N, primarily to reflect the fact the time-index is now n.

A discrete-time signal, $x[n]$, is periodic if:

$$x[n] = x[n + m\,N]\,,\ \forall\, m \in \mathbb{Z} \tag{1.2}$$

This is, of course, similar to Equation 1.1.

Discrete-time aperiodic Analogous to the continuous-time aperiodic signal in Figure 1.4, a discrete-time aperiodic signal is shown in Figure 1.6.

Aperiodic discrete-time signals will be analysed using the z-transform and also the discrete-time Fourier transform (DTFT).

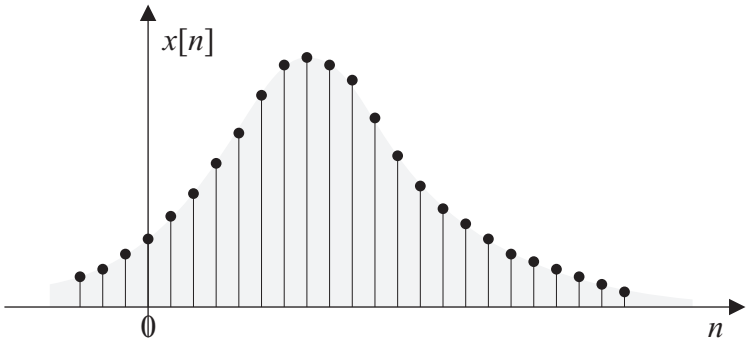


Figure 1.6: An example of a discrete-time aperiodic signal.

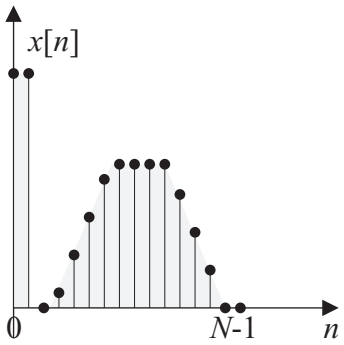
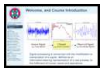


Figure 1.7: An example of a finite-duration signal.

Finite-length discrete-time signals Discrete-time signals can also be classified as being finite in length. In other words, they are not assumed to exist for all-time, and what happens outside the **window** of data is assumed unknown. These signals can be modelled using the so-called **discrete Fourier transform (DFT)**.

1.3.2 Energy and Power Signals



New slide

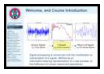
As stated in Section 1.3.1, signals can be analysed using a variety of frequency-domain transform methods, such as the **Fourier series**, **Fourier transform**, **Laplace transform**, and for discrete-time, the z**-transform**. For example, the Fourier transform is used to analyse aperiodic continuous-time signals.

However, not all aperiodic signals can be analysed using the Fourier transform, and the reason for this can be directly related to a fundamental property of a signal: a measure of *how much signal there is*.

Therefore it is relevant to consider the **energy** or **power** as a means for characterising a signal. The concepts of **power** and **energy** intuitively follow from their use in other aspects of the physical sciences. However, the concept of signals which exist for all time requires careful definitions, in order to determine when it has **energy** and when it has **power**.

Intuitively, energy and power measure *how big* a signal is. A motivating example of measuring the size of something is given in Sidebar 1.

1.3.2.1 Motivation for Energy and Power Expressions



New slide

Considering power from an electrical perspective, if a voltage $x(t)$ is connected across a resistance R, the dissipated power at time τ is given by:

Sidebar 1 Size of a Human Being

Suppose we wish to devise a signal number V as a measure of the size of a human being. Then clearly, the width (or girth) must also be taken into account as well as the height. One could make the simplifying assumption that the shape of a person is a cylinder of variable radius r (which varies with the height h). Then one possible measure of the size of a person of height H is the person's volume, given by:

$$V = \pi \int_0^H r^2(h)\ dh \tag{1.3}$$

This can be found by dividing the object into circular discs (which is an approximation), where each disc has a volume $\delta V \approx \pi r^2(h)\ \delta h$. Then the total volume is given by $V = \int dV$.

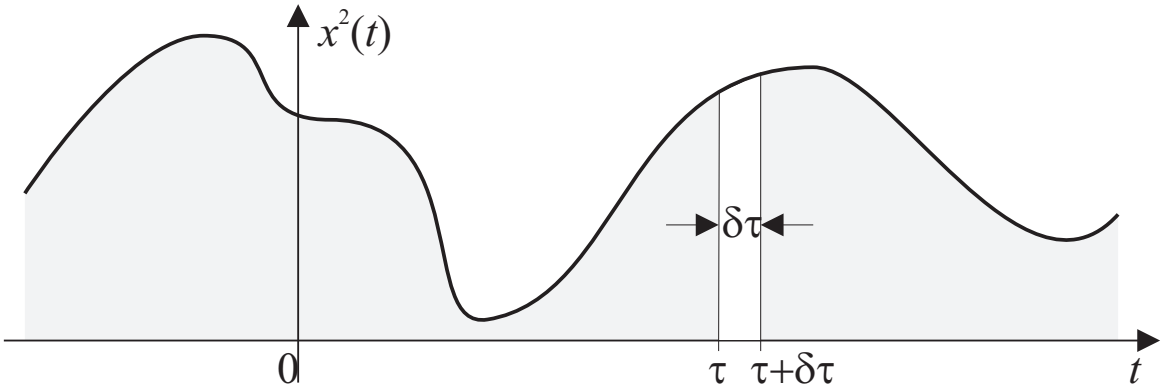


Figure 1.8: Energy Density.

$$P(\tau) = \frac{x^2(\tau)}{R} \propto x^2(\tau) \tag{1.4}$$

where $\propto$ denotes *proportional to*. Since energy and power are related through the expression

$$\text{Energy} = \text{Power} \times \text{Time}, \tag{1.5}$$

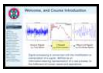
the energy dissipated between times τ and $\tau + \delta\tau$, as indicated in Figure 1.8, is:

$$\delta E(\tau) \propto P(\tau)\, \delta\tau \propto x^2(\tau)\delta\tau \tag{1.6}$$

The total energy over all time can thus be found by integrating over all time:

$$E \propto \int_{-\infty}^{\infty} x^2(\tau)\, d\tau \tag{1.7}$$

This leads to the formal definitions of energy and power.



New slide

1.3.2.2 Formal Definitions for Energy and Power

Based on the justification in Section 1.3.2.1, the formal abstract definitions for energy and power that are independent of how the energy or power is dissipated are defined below.

Energy Signals A continuous-time signal $x(t)$ is said to be an **energy signal** if the total energy, E, dissipated by the signal over all time is both *nonzero* and *finite*. Thus:

$$0 < E < \infty \quad \text{where} \quad E = \int_{-\infty}^{\infty} |x(t)|^2\ dt \tag{1.11}$$

Sidebar 2 Other signal measures

1. While the area under a signal $x(t)$ is a possible measure of its size, because it takes account not only of the amplitude but also of the duration, is defective since even for a very large signal, the positive and negative areas could cancel each other out, indicating a signal of a small size.

2. Using the sum of square values can potentially give undue weighting to any outliers in the signal, where an outlier is defined as an unusual signal variation that is not characteristic of the rest of the signal; an example might be a high-energy shot burst of interference.

3. Therefore, taking the absolute value, $|x(t)| \equiv \operatorname{abs} x(t)$ is a possible measure, and in some circumstances can be used. Unfortunately, dealing with the absolute value of a function can be difficult to manipulate mathematically. However, using the area under the square of the function is not only more mathematically tractable but is also more meaningful when compared with the electrical examples and the volume in Sidebar 1.

4. These notions lead onto the more general subject of **signal norms**. The L_p-norm is defined by:

$$L_p(x) \triangleq \left(\int_{-\infty}^{\infty} |x(t)|^p \, dt \right)^{\frac{1}{p}}, \qquad p \geq 1 \tag{1.8}$$

   In particular, the expression for energy is related to the L_2-norm, while using the absolute value of the signbal gives rise to the L_1-norm:

$$L_1(x) \triangleq \int_{-\infty}^{\infty} |x(t)| \, dt \tag{1.9}$$

   which is the integral of the absolute value as described above in part 3.

5. While Parseval's theorem exists between the time-domain and frequency-domain for relating the L_2-norms, in general no relation exists for other values of p.

6. Note that the L_p-norm generalises for discrete-time signals as follows:

$$L_p(x) \triangleq \left(\sum_{-\infty}^{\infty} |x[t]|^p \right)^{\frac{1}{p}}, \qquad p \geq 1 \tag{1.10}$$

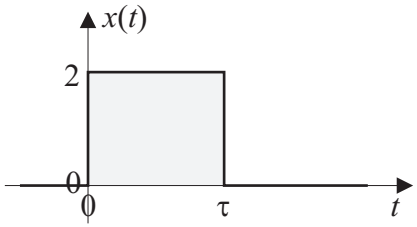


Figure 1.9: Rectangular pulse of length τ.

where $|x(t)\,|$ means the magnitude of the signal $x(t)$. If $x(t)$ is a real-signal, this is just its amplitude. If $x(t)$ is a complex-signal, then $|x(t)\,|^2 = x(t)\ x^*(t)$ where $*$ denotes complex-conjugate. In this course, however, only real signals will be encountered.

A necessary condition for the energy to be finite is that the signal amplitude $|x(t)\,| \to 0$ as $|t| \to \infty$, otherwise the integral in Equation 1.11 will not exist. When the amplitude of $x(t)$ does not tend to zero as $|t| \to \infty$, the signal energy is likely to be infinite. A more meaningful measure of the signal size in such a case would be the time average of the energy if it exists. This measure is called the **power** of the signal.

Power signals If the average power delivered by the signal over all time is both *nonzero* and *finite*, the signal is classified as a power signal:

$$0 < P < \infty \quad \text{where} \quad P = \lim_{W\to\infty} \frac{1}{2W} \int_{-W}^{W} |x(t)|^2\ dt \tag{1.12}$$

where the variable W can be considered as half of the width of a *window* that covers the signal and gets larger and larger.

Example 1.1. Name a type of signal which is not an example of an **energy signal**?

SOLUTION. A periodic signal has finite energy over one period, so consequently has infinite energy. However, as a result it has a finite average power and is therefore a power signal, and not an energy signal.

Example 1.2 (Rectangular Pulse). What is the energy of the rectangular pulse shown in Figure 1.9 as a function of τ, and for the particular case of $\tau = 4$?

SOLUTION. The signal can be represented by

$$x(t) = \begin{cases} 2 & 0 \le t < \tau \\ 0 & \text{otherwise} \end{cases} \tag{1.13}$$

so that the square of the signal is also rectangular and given by

$$x^2(t) = \begin{cases} 4 & 0 \le t < \tau \\ 0 & \text{otherwise} \end{cases} \tag{1.14}$$

Since an integral can be interpreted as the area under the curve (see Figure 1.10, the total energy is thus:

$$E = 4\tau \tag{1.15}$$

□

When $\tau = 4$, $E = 16$.

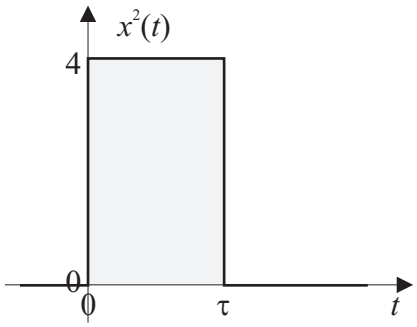
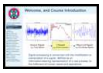


Figure 1.10: The total energy of the signal in Figure 1.9 can be found as the area under the curve representing the square of the rectangular pulse, as shown here.

1.3.2.3 Units of Energy and Power



New slide

It is important to consider the physical units associated with energy and power, and therefore to determine how the abstract definitions of E and P in Equation 1.11 and Equation 1.12 can be converted into real energy and power.

Consider again power from an electrical perspective. When considering "direct current" (DC) signals, power is given by

$$P_{DC} = \frac{V^2}{R} = \frac{\text{Volts}^2}{\text{Ohms}} = \text{Watts} \tag{1.16}$$

where V is the signal voltage, and R is the resistance through which the power is dissipated. Consider now the units of the abstract definition of power, P in Equation 1.12:

$$P = \frac{1}{\text{time}} \times \text{Volts}^2 \times \text{time} = \text{Volts}^2 = \text{Watts} \times \text{Ohms} \tag{1.17}$$

where the second unit of *time* comes from the integral term dt, and in which the integral may be considered as a summation. Therefore, by comparing Equation 1.16 and Equation 1.12, the abstract definition of power, P, can be converted to real power by **Ohms** for the case of electrical circuits.

Similarly, the units of energy in Equation 1.11 is given by

$$E = \text{volts}^2 \times \text{time} \tag{1.18}$$

Therefore, to convert the abstract energy to Joules, it is again necessary to divide by **Ohms** by noting that energy is power multiplied by time.

1.3.2.4 Power for Periodic Signals

The expression in Equation 1.12 can be simplified for periodic signals. Consider the periodic signal in Figure 1.3. Note here that there might be confusion with using the same symbol T to mean both the period of a periodic signal and the limit in Equation 1.12. To avoid ambiguity, rewrite Equation 1.12 with W instead of T where W denotes a *window length* over which the power is calculated, and define:

$$P_W = \frac{1}{2W} \int_{-W}^{W} |x(t)|^2 \, dt \tag{1.19}$$

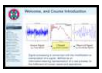
Thus, the average power over two periods is P_T, and the average power over $2N$ periods is P_{NT}. It should becomes clear that

$$P_T = P_{NT}, \forall N \in \mathbb{Z} \tag{1.20}$$

since the average in each period is the same. Consequently, **power** for a periodic signal with period T may be defined as:

$$P = \frac{1}{T}\int_0^T |x(t)|^2 \, dt \tag{1.21}$$

Note that the limits in Equation 1.21 may be over any period and thus can be replaced by $(\tau, \tau + T)$ for any value of τ.

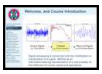


New slide

1.4 Fourier Series and Fourier Transforms

In this review of Fourier series and transforms, the topics covered are:

- Complex Fourier series
- **Fourier transform**
- The discrete-time Fourier transform
- Discrete Fourier transform



New slide

1.4.1 Complex Fourier series

The complex Fourier series is a frequency analysis tool for continuous time periodic signals. Examples of periodic signals encountered in practice include square waves, triangular waves, sawtooth waves, pulse waves and, of course, sinusoids and complex exponentials, as well as half-wave recitifed, full-wave rectified and p-phased rectified sinusoids. The basic mathematical representation of periodic signals is the Fourier series, which is a linear weighted sum of harmonically related sinusoids or complex exponentials.

A **periodic continuous-time** deterministic signal, $x_c(t)$, with fundamental period T_p can be expressed as a linear combination of harmonically related complex exponentials:

$$x_c(t) = \sum_{k=-\infty}^{\infty} \check{X}_c(k)\, e^{jk\omega_0 t}, \quad t \in \mathbb{R}, \tag{M:2.2.1}$$

where $\omega_0 = 2\pi F_0 = \frac{2\pi}{T_p}$ is the **fundamental frequency**. Here, ω_0 is in radians per second, and the fundamental frequency, in Hertz, is given by $F_0 = \frac{1}{T_p}$. Moreover,

$$\check{X}_c(k) = \frac{1}{T_p}\int_0^{T_p} x_c(t)\, e^{-jk\omega_0 t}\, dt, \quad k \in \mathbb{Z} \tag{M:2.2.2}$$

are termed the **Fourier coefficients**, or **spectrum** of $x_c(t)$. Note that although the region of integration in Equation M:2.2.2 is from 0 to T_p, it can actually be over any period of the waveform, since the signal, $x_c(t)$, is periodic with period T_p.

The kth frequency component corresponds to frequency $\omega_k = k\omega_0 = k\frac{2\pi}{T_p}$. The set of exponential functions

$$\mathcal{F}(t) = \{e^{j\omega_0 kt},\, k \in \mathbb{Z}\} \tag{1.22}$$

can be thought of as the basic *building blocks* from which periodic signals of various types can be constructed with the proper choice of fundamental frequency and Fourier coefficients.

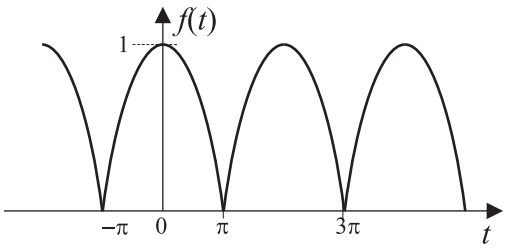


Figure 1.11: Function $f(t)$ of Example 1.3

Example 1.3 (Complex Fourier Series). Find the complex form of the Fourier series expansion of the periodic function $f(t)$ defined by:

$$\begin{aligned} f(t) &= \cos\frac{1}{2}t \quad (-\pi < t < \pi) \\ f(t+2\pi) &= f(t) \end{aligned} \tag{1.23}$$

SOLUTION. A graph of the function $f(t)$ over the interval $-\pi \le t \le 3\pi$ is shown in Figure 1.11. The period $T_p = 2\pi$, so therefore the complex coefficients, denoted by F_n, are given by Equation M:2.2.2 as:

$$F_n = \frac{1}{T_p}\int_0^{T_p} f(t)\, e^{-jn\omega_0 t}\, dt, \quad n \in \mathbb{Z} \tag{1.24}$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos\frac{t}{2}\, e^{-jnt}\, dt = \frac{1}{4\pi}\int_{-\pi}^{\pi} \left(e^{j\frac{t}{2}} + e^{-j\frac{t}{2}}\right) e^{-jnt}\, dt \tag{1.25}$$

$$= \frac{1}{4\pi}\int_{-\pi}^{\pi} \left(e^{-j\left(n-\frac{1}{2}\right)t} + e^{-j\left(n+\frac{1}{2}\right)t}\right) dt \tag{1.26}$$

which, after some trivial integration, gives:

$$F_n = \frac{1}{4\pi}\left[\frac{-2e^{-j(2n-1)\frac{t}{2}}}{j(2n-1)} - \frac{2e^{-j(2n+1)\frac{t}{2}}}{j(2n+1)}\right]_{-\pi}^{\pi} \tag{1.27}$$

$$= \frac{j}{2\pi}\left[\left(\frac{e^{-jn\pi}\, e^{j\frac{\pi}{2}}}{2n-1} + \frac{e^{-jn\pi}\, e^{-j\frac{\pi}{2}}}{2n+1}\right) - \left(\frac{e^{jn\pi}\, e^{-j\frac{\pi}{2}}}{2n-1} + \frac{e^{jn\pi}\, e^{j\frac{\pi}{2}}}{2n+1}\right)\right] \tag{1.28}$$

Noting that $e^{\pm j\frac{\pi}{2}} = \pm j$, and $e^{\pm jn\pi} = \cos n\pi = (-1)^n$, then it follows that:

$$F_n = \frac{j}{2\pi}\left(\frac{j}{2n-1} - \frac{j}{2n+1} + \frac{j}{2n-1} - \frac{j}{2n+1}\right)(-1)^n \tag{1.29}$$

$$= \frac{(-1)^n}{\pi}\left(\frac{1}{2n+1} - \frac{1}{2n-1}\right) = \frac{2(-1)^{n+1}}{(4n^2-1)\pi} \tag{1.30}$$

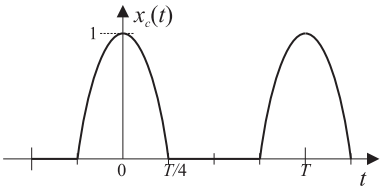
Note that in this case, the coefficients F_n are real. This is expected, since the function $f(t)$ is an even function of t. The complex Fourier series expansion for $f(t)$ is therefore:

$$f(t) = \sum_{n=-\infty}^{\infty} \frac{2(-1)^{n+1}}{(4n^2-1)\pi}\, e^{jnt} \tag{1.31}$$

□

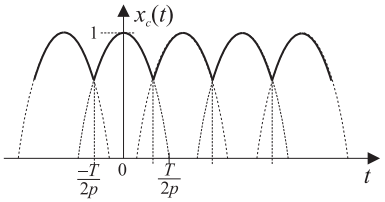
1.4.1.1 Common Fourier Series Expansions

In the following Fourier series expansions, $\omega_0 = \frac{2\pi}{T}$ is the fundamental frequency.



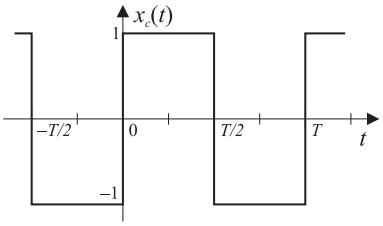
Half-wave rectified cosine-wave:

$$x_c(t) = \frac{1}{\pi} + \frac{1}{2}\cos\omega_0 t + \frac{2}{\pi}\sum_{n=1}^{\infty}(-1)^{n+1}\frac{\cos(2n\omega_0 t)}{4n^2-1}$$



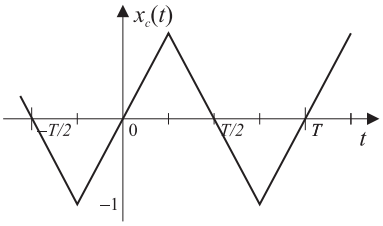
p-phase rectified cosine-wave ($p \geq 2$):

$$x_c(t) = \frac{p}{\pi}\sin\frac{\pi}{p}\left[1 + 2\sum_{n=1}^{\infty}(-1)^{n+1}\frac{\cos(pn\omega_0 t)}{p^2n^2-1}\right]$$



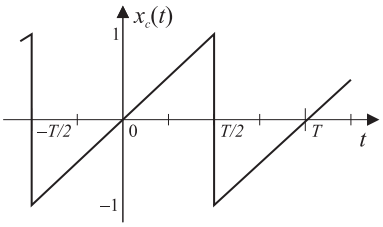
Square wave:

$$x_c(t) = \frac{4}{\pi}\sum_{n=1}^{\infty}\frac{\sin(2n-1)\omega_0 t}{2n-1}$$



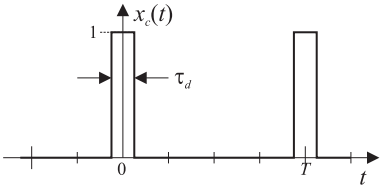
Triangular wave:

$$x_c(t) = \frac{8}{\pi^2}\sum_{n=1}^{\infty}(-1)^{n+1}\frac{\sin(2n-1)\omega_0 t}{(2n-1)^2}$$



Sawtooth wave:

$$x_c(t) = \frac{2}{\pi}\sum_{n=1}^{\infty}(-1)^{n+1}\frac{\sin n\omega_0 t}{n}$$



Pulse wave:

$$x_c(t) = \frac{\tau_d}{T}\left[1 + 2\sum_{n=1}^{\infty}\frac{\sin(n\pi\frac{\tau_d}{T})}{(n\pi\frac{t_d}{T})}\cos(n\omega_0 t)\right]$$

1.4.1.2 Dirichlet Conditions

An important issue that arises in the representation of the continuous time periodic signal $x_c(t)$ by the Fourier series representation,

$$\bar{x}_c(t) = \sum_{k=-\infty}^{\infty}\check{X}_c(k)\,e^{jk\omega_0 t}, \qquad \text{(P:4.1.5)}$$

is whether or not the series converges for every value of $t \in \mathbb{R}$; i.e., is it true that

$$\bar{x}_c(t) \stackrel{?}{=} x_c(t), \quad \forall t \in \mathbb{R} \tag{1.32}$$

The so-called **Dirichlet conditions** guarantee that the Fourier series converges everywhere except at points of discontinuity. At these points, the Fourier series representation $\bar{x}_c(t)$ converges to the midpoint, or average value, of the discontinuity.

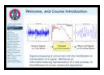
The **Dirichlet conditions** require that the signal $x_c(t)$:

1. has a finite number of discontinuities in any period;
2. contains a finite number of maxima and minima during any period;
3. is **absolutely integrable** in any period; that is:

$$\int_{T_p} |x_c(t)|\, dt < \infty \tag{P:4.1.6}$$

where the integral is over one period. Many periodic signals of practical interest easily satisfy these conditions, and it is reasonable to assume that all practical periodic signals do. However, it is important to beware that pathological cases can make certain proofs or algorithms collapse.

1.4.1.3 Parseval's Theorem (for Fourier series)



New slide

It is sometimes relevant to consider the **energy** or **power** as a means for characterising a signal. These concepts of **power** and **energy** intuitively follow from their use in other aspects of the physical sciences. However, the concept of signals which exist for all time requires careful definitions for when it has **energy** and when it has **power**. Consider the following signal classifications:

Energy Signals A signal $x_c(t)$ is said to be an **energy signal** if the total energy, E, dissipated by the signal over all time is both *nonzero* and *finite*. Thus:

$$0 < E < \infty \quad \text{where} \quad E = \int_{-\infty}^{\infty} |x_c(t)|^2 \, dt \tag{1.33}$$

Power signals If the average power delivered by the signal over all time is both *nonzero* and *finite*, the signal is classified as a power signal:

$$0 < P < \infty \quad \text{where} \quad P = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |x_c(t)|^2 \, dt \tag{1.34}$$

A periodic signal has infinite energy, but finite average power. The average power of $x_c(t)$ is given by **Parseval's theorem**:

$$P_x = \frac{1}{T_p} \int_0^{T_p} |x_c(t)|^2 \, dt = \sum_{k=-\infty}^{\infty} |\check{X}_c(k)|^2 \tag{M:2.2.3}$$

The term $|\check{X}_c(k)|^2$ represents the power in the kth frequency component, at frequency $\omega_k = k\frac{2\pi}{T_p}$. Hence,

$$\check{P}_x(k) = |\check{X}_c(k)|^2, \quad -\infty < k < \infty,\ k \in \mathbb{Z} \tag{1.35}$$

is called the **power spectrum** of $x_c(t)$. Consequently, it follows P_x may also be written as:

$$P_x = \sum_{k=-\infty}^{\infty} \check{P}_x(k) \tag{1.36}$$

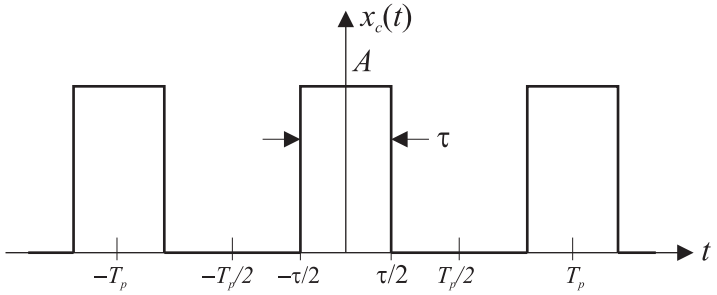


Figure 1.12: Continuous-time periodic train of rectangular pulses.

PROOF. Starting with

$$P_x = \frac{1}{T_p}\int_0^{T_p} x_c(t)\, x_c^*(t)\, dt \tag{1.37}$$

then substituting for the Fourier series expansion of $x_c(t)$ gives:

$$P_x = \frac{1}{T_p}\int_0^{T_p} x_c(t) \left\{ \sum_{k=-\infty}^{\infty} \check{X}_c(k)\, e^{jk\omega_0 t} \right\}^* dt \tag{1.38}$$

Noting that the conjugate of a summation (multiplication) is the summation (multiplication) of the conjugates, then:

$$P_x = \frac{1}{T_p}\int_0^{T_p} x_c(t) \sum_{k=-\infty}^{\infty} \check{X}_c^*(k)\, e^{-jk\omega_0 t}\, dt \tag{1.39}$$

Rearranging the order of the integration and the summation gives:

$$P_x = \sum_{k=-\infty}^{\infty} \check{X}_c^*(k) \underbrace{\left\{ \frac{1}{T_p}\int_0^{T_p} x_c(t)\; e^{-jk\omega_0 t}(t)\, dt \right\}}_{X_c(k)} \tag{1.40}$$

□

which is the desired result and completes the proof.

Later in this course, the notion of a **power spectrum** will be extended to *stochastic* signals.

Example 1.4 ([Proakis:1996, Example 4.1.1, Page 237]). Determine the Fourier series and the power density spectrum of a rectangular pulse train that is defined over *one* period as follows:

$$x_c(t) = \begin{cases} 0 & \text{if } -\frac{T_p}{2} \le t < -\frac{\tau}{2} \\ A & \text{if } -\frac{\tau}{2} \le t < \frac{\tau}{2} \\ 0 & \text{if } \frac{\tau}{2} \le t < \frac{T_p}{2} \end{cases} \tag{1.41}$$

where $\tau < T_p$.

SOLUTION. The signal is periodic with fundamental period T_p and, clearly, satisfies the Dirichlet conditions. Consequently, this signal can be represented by the Fourier series. Hence, it follows that

$$\check{X}_c(k) = \frac{1}{T_p}\int_{-\frac{T_p}{2}}^{\frac{T_p}{2}} x_c(t)\, e^{-jk\omega_0 t}\, dt = \frac{A}{T_p}\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} e^{-jk\omega_0 t}\, dt \tag{1.42}$$

Two different integrals need to be performed depending on whether $k = 0$ or not. Considering the case when $k = 0$, then the average value of the signal is obtained and given by:

$$\check{X}_c(0) = \frac{1}{T_p}\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} x_c(t)\,dt = \frac{1}{T_p}\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} A\,dt = \frac{A\tau}{T_p} \tag{1.43}$$

For $k \neq 0$, then

$$\check{X}_c(k) = \frac{A}{T_p}\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} e^{-jk\omega_0 t}\,dt = \frac{A}{T_p}\left[\frac{e^{-jk\omega_0 t}}{-jk\omega_0}\right]_{-\frac{\tau}{2}}^{\frac{\tau}{2}} \tag{1.44}$$

$$= \frac{A}{jk\omega_0 T_p}\left(e^{jk\omega_0\frac{\tau}{2}} - e^{-jk\omega_0\frac{\tau}{2}}\right) = \frac{A\tau}{T_p}\frac{\sin\frac{\tau\omega_0 k}{2}}{k\omega_0\frac{\tau}{2}} \tag{1.45}$$

$$= \frac{A\tau}{T_p}\text{sinc}\,\frac{\tau\omega_0 k}{2} \quad \text{where sinc}\,x \triangleq \frac{\sin x}{x} \tag{1.46}$$

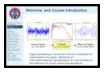
Hence, the power density spectrum for the rectangular pulse is:

$$\left|\check{X}_c(k)\right|^2 = \left(\frac{A\tau}{T_p}\right)^2 \text{sinc}^2\,\frac{\tau\omega_0 k}{2}, \quad k \in \mathbb{Z} \tag{P:4.1.19}$$

□

where it is noted that sinc $(0) = 1$.

1.4.2 Fourier transform



New slide

An **aperiodic continuous-time** deterministic signal, $x_c(t)$, can be expressed in the frequency domain using the **Fourier transform** pairs:

$$x_c(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X_c(\omega)\,e^{j\omega t}\,d\omega \tag{M:2.2.5}$$

and

$$X_c(\omega) = \int_{-\infty}^{\infty} x_c(t)\,e^{-j\omega t}\,dt \tag{M:2.2.4}$$

$X_c(\omega)$ is called the **spectrum** of $x_c(t)$. Again, note that [Manolakis:2000] uses the defintion $F = \frac{\omega}{2\pi}$. Continuous-time aperiodic signals have continuous aperiodic spectra.

There are a few mathematical requirements that $x_c(t)$ must satisfy for $X_c(\omega)$ to exist; these can be summarised by the phrase that *the signal must be well-behaved*. More specifically, the set of conditions that guarantee the existence of the Fourier transform are the Dirichlet conditions which are the same as for Fourier series.

Example 1.5 (Fourier Transforms). Find the Fourier transform of the one-sided exponential function

$$f(t) = H(t)\,e^{-at} \quad \text{where } a > 0 \tag{1.47}$$

and where $H(t)$ is the Heaviside unit step function show in Figure 1.13 and given by:

$$H(t) = \begin{cases} 1 & \text{if } t \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{1.48}$$

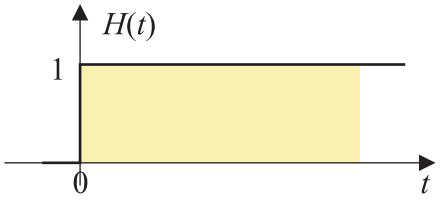


Figure 1.13: The Heaviside step function $H(t)$.

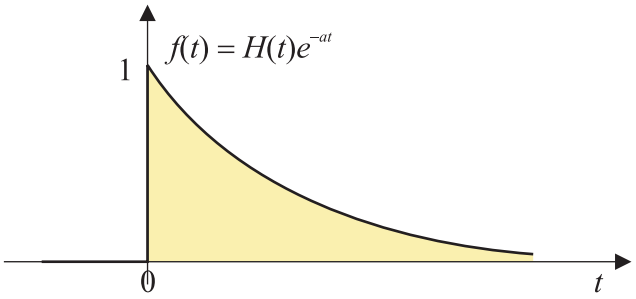


Figure 1.14: Exponential decaying function, $f(t) = H(t)\, e^{-at}$ for $a > 0$.

SOLUTION. Since $f(t) \to 0$ as $t \to \infty$, then the signal energy is bounded, as indicated by plotting the graph of $f(t)$ as shown in Figure 1.14.

A Fourier transform therefore exists, and is given by:

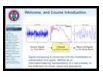
$$X_c(\omega) = \int_{-\infty}^{\infty} H(t)\, e^{-at}\, e^{-j\omega t}\, dt \tag{1.49}$$

$$= \int_{0}^{\infty} e^{-(a+j\omega)t}\, dt = \left[-\frac{e^{-(a+j\omega)t}}{a+j\omega} \right]_0^{\infty} \tag{1.50}$$

giving

$$X_c(\omega) = \frac{1}{a+j\omega}, \qquad \text{for } -\infty < \omega < \infty \tag{1.51}$$

□



New slide

1.4.2.1 Parseval's theorem (for Fourier transforms)

The *energy* of $x_c(t)$ is, as for **Fourier series**, computed in either the time or frequency domain by **Parseval's theorem**:

$$E_x = \int_{-\infty}^{\infty} |x_c(t)|^2\, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |X_c(\omega)|^2\, d\omega \tag{M:2.2.7}$$

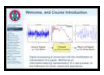
The function $|X_c(\omega)|^2 \geq 0$ shows the distribution of energy of $x_c(t)$ as a function of frequency, ω, and is called the **energy spectrum** of $x_c(t)$.

PROOF. The derivation of Parseval's theorem for Fourier transforms follows a similar line to the derivation of Parseval's theorem for Fourier series; it proceeds as follows:

$$\begin{aligned} E_x &= \int_{-\infty}^{\infty} x_c(t)\, x_c^{\star}(t)\, dt = \int_{-\infty}^{\infty} x_c(t)\, \frac{1}{2\pi} \int_{-\infty}^{\infty} X_c^{\star}(\omega)\, e^{-j\omega t}\, d\omega\; dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} X_c^{\star}(\omega) \int_{-\infty}^{\infty} x_c(t)\; e^{-j\omega t}\; dt\, d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} X_c^{\star}(\omega)\, X_c(\omega)\, d\omega \end{aligned} \tag{1.52}$$

□

1.4.3 The discrete-time Fourier transform



New slide

Turning to discrete-time deterministic signals, the natural starting point is to consider aperiodic signals that exist for all discrete-time; i.e. $\{x[n]\}_{-\infty}^{\infty}$. It is interesting to note that there are fewer convergence issues with transforms for discrete-time signals than there are in the continuous-time case.

An **aperiodic discrete-time** deterministic signal, $\{x[n]\}_{-\infty}^{\infty}$, can be synthesised from its **spectrum** using the inverse-discrete-time Fourier transform, given by:

$$x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X\left(e^{j\omega T}\right)\, e^{j\omega n}\, d\omega, \quad n \in \mathbb{Z} \tag{M:2.2.13}$$

and the discrete-time Fourier transform (DTFT):

$$X\left(e^{j\omega T}\right) = \sum_{\text{all } n} x[n]\; e^{-j\omega n}, \quad \omega \in \mathbb{R} \tag{M:2.2.12}$$

$X\left(e^{j\omega T}\right)$ is the **spectrum** of $x[n]$.

Since $X\left(e^{j\omega T}\right) = X\left(e^{j(\omega+2\pi k)}\right)$, discrete-time aperiodic signals have continuous periodic spectra with **fundamental period** 2π. However, this property is just a consequence of the fact that the frequency range of any discrete-time signal is limited to $[-\pi, \pi)$ or $[0, 2\pi)$; any frequency outside this interval is equivalent to some frequency within this interval.

There are two basic differences between the Fourier transform of a discrete-time finite-energy aperiodic signal, as represented by the discrete-time Fourier transform, and the Fourier transform of a finite-energy continuous-time aperiodic signal:

1. For continuous-time signals, the Fourier transform, and hence the spectrum of the signal, have a frequency range of $(-\infty, \infty)$. In contrast, the frequency range for a discrete-time signal is unique over the frequency range $[-\pi, \pi)$ or, equivalently, $[0, 2\pi)$.
2. Since $X\left(e^{j\omega T}\right)$ in the DTFT is a periodic function of frequency, it has a Fourier series expansion, provided that the conditions for the existence of the Fourier series are satisfied. In fact, from the fact that $X\left(e^{j\omega T}\right)$ is given by the summation of exponentially weighted versions of $x[n]$ is is clear that the DTFT already has the form of a Fourier series. This is not true for the Fourier transform.

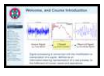
In order for $X\left(e^{j\omega T}\right)$ to exist, $x[n]$ must be absolutely summable:

$$\sum_{\text{all } n} |x[n]\,| < \infty \tag{M:2.2.11}$$

Finally, as for the Fourier series, and the Fourier transform, discrete-time aperiodic signals have energy which satisfies Parseval's theorem:

$$E_x = \sum_{n=-\infty}^{\infty} |x[n]\,|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |X\left(e^{j\omega T}\right)\,|^2\, d\omega \tag{P:4.2.41}$$

1.4.4 Discrete Fourier transform



New slide

Any finite-length or **periodic discrete-time** deterministic signal, $\{x[n]\}_0^{N-1}$, can be written by the Fourier series, or inverse-DFT (IDFT):

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X_k\, e^{j\frac{2\pi}{N}nk}, \quad n \in \mathcal{N} \tag{M:2.2.8}$$

where $\mathcal{N} = \{0, 1, \ldots, N-1\} \subset \mathbb{Z}^+$, and where the DFT:

$$X_k = \sum_{n=0}^{N-1} x[n]\ e^{-j\frac{2\pi}{N}nk}, \quad k \in \mathcal{N} \tag{M:2.2.9}$$

are the corresponding Fourier coefficients. The sequence X_k, $k \in \mathbb{R}$ is the **spectrum** of $x[n]$. X_k is discrete and periodic with the same period as $x[n]$.

Note that a finite-length discrete-time signal of length N has the same Fourier transform, through the DFT, as an infinite-length discrete-time periodic signal of period N. Hence, these equivalent perspectives will be considered synonymous.

PROOF (DERIVATION OF DISCRETE FOURIER TRANSFORM). If the **discrete-time** signal $x[n]$ is **periodic** over N samples, then it can be written over one period in continuous time as:

$$x_c(t) = T_p \sum_{n\in\mathcal{N}} x[n]\ \delta(t - n\,T_s)\,, \quad 0 \le t < T_p \tag{1.53}$$

where $\mathcal{N} = \{0, \ldots, N-1\}$, T_s is the sampling period, and $T_p = N\,T_s$ is the period of the process. Substituting into the expression for the **Fourier series**, using the **sifting property** and noting that $\omega_0 = \frac{2\pi}{T_p} = \frac{2\pi}{NT_s}$, gives:

$$\begin{aligned}
X_k &= \frac{1}{T_p}\int_0^{T_p} x_c(t)\ e^{-jk\omega_0 t}\,dt && (1.54)\\
&= \frac{1}{T_p}\int_0^{T_p} \left\{ T_p \sum_{n\in\mathcal{N}} x[n]\ \delta(t - n\,T_s) \right\} e^{-jk\omega_0 t}\,dt && (1.55)\\
&= \sum_{n\in\mathcal{N}} x[n] \int_0^{T_p} \delta(t - n\,T_s)\, e^{-jk\omega_0 t}\,dt && (1.56)\\
&= \sum_{n\in\mathcal{N}} x[n]\ e^{-j\frac{2\pi}{N}nk} && (1.57)
\end{aligned}$$

□

The IDFT can be obtained using the appropriate identities to ensure a unique inverse.

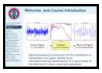
1.4.4.1 Parseval's Theorem for Finite Length Discrete-Time Signals

The average power of a finite length or periodic discrete-time signal with period N is given by

$$P_x = \sum_{n=0}^{N-1} |x[n]\,|^2 \tag{P:4.2.10}$$

Through the same manipulations as for Parseval's theorems in the cases presented above, which are left as an exercise for the reader, it is straightforward to show that:

$$P_x = \sum_{n=0}^{N-1} |x[n]\,|^2 = \frac{1}{N}\sum_{k=0}^{N-1} |X_k|^2 \tag{P:4.2.11}$$



New slide

1.4.4.2 The DFT as a Linear Transformation

The formulas for the DFT and IDFT may be expressed as:

$$X_k = \sum_{n=0}^{N-1} x[n]\ W_N^{nk}, \quad k \in \mathcal{N} \tag{P:5.1.20}$$

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X_k\, W_N^{-nk}, \quad n \in \mathcal{N} \tag{P:5.1.21}$$

where, by definition:

$$W_N = e^{-j\frac{2\pi}{N}} \tag{P:5.1.22}$$

which is the Nth root of unity. Note here that, if W_N has been pre-calculated, then the computation of each point of the DFT can be accomplished by N complex multiplications and $N-1$ complex additions. Hence, the N-point DFT can be computed in a total of N^2 complex multiplications and $N(N-1)$ complex additions.

It is instructive to view the DFT and IDFT as linear transformations on the sequences $\{x[n]\}_0^{N-1}$ and $\{X_k\}_0^{N-1}$. Defining the following vectors and matrices:

$$\mathbf{x}_N = \begin{bmatrix} x[0] \\ \vdots \\ x[N-1] \end{bmatrix}, \quad \mathbf{X}_N = \begin{bmatrix} X_0 \\ \vdots \\ X_{N-1} \end{bmatrix} \tag{1.58}$$

$$\mathbf{W}_N = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & W_N & W_N^2 & \cdots & W_N^{N-1} \\ 1 & W_N^2 & W_N^4 & \cdots & W_N^{2(N-1)} \\ \vdots & \vdots & \vdots & \cdot & \vdots \\ 1 & W_N^{N-1} & W_N^{2(N-1)} & \cdots & W_N^{(N-1)(N-1)} \end{bmatrix} \tag{1.59}$$

Observe that X_k can be obtained by the inner-product of the $(k-1)$ th-order row with the column $\mathbf{x}_N$:

$$X_k = \begin{bmatrix} 1 & W_N^k & W_N^{2k} & \cdots & W_N^{(N-1)k} \end{bmatrix} \begin{bmatrix} x[0] \\ \vdots \\ x[N-1] \end{bmatrix} \tag{1.60}$$

Then the N-point DFT may be expressed in vector-matrix form as:

$$\mathbf{X}_N = \mathbf{W}_N \mathbf{x}_N \tag{P:5.1.24}$$

where $\mathbf{W}_N$ is the matrix of the linear transformation. Observe that $\mathbf{W}_N$ is a symmetric matrix. Assuming that the inverse of $\mathbf{W}_N$ exists, then Equation P:5.1.24 can be inverted by pre-multiplying both sides by $\mathbf{W}_N^{-1}$, to obtain:

$$\mathbf{x}_N = \mathbf{W}_N^{-1} \mathbf{X}_N \tag{P:5.1.25}$$

This is the expression for the IDFT, which can also be expressed in matrix form as:

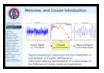
$$\mathbf{x}_N = \frac{1}{N} \mathbf{W}_N^* \mathbf{X}_N \tag{P:5.1.26}$$

where $\mathbf{W}_N^*$ denotes the complex conjugate of the matrix $\mathbf{W}_N$. Hence, it follows that:

$$\mathbf{W}_N^{-1} = \frac{1}{N} \mathbf{W}_N^* \quad \text{or} \quad \mathbf{W}_N \mathbf{W}_N^* = N\mathbf{I}_N \tag{P:5.1.27}$$

where $\mathbf{I}_N$ is the $N \times N$ identity matrix. Hence, $\mathbf{W}_N$ is an orthogonal or unity matrix.

1.4.4.3 Properties of the discrete Fourier transforms



New slide

There are some important basic properties of the DFT that should be noted. The notation used to denote the N-point DFT pair $x[n]$ and X_k is

$$x[n] \overset{\text{DFT}}{\rightleftharpoons} X_k \tag{1.61}$$

Periodicity If $x[n] \overset{\text{DFT}}{\rightleftharpoons} X_k$, then:

$$x[n+N] = x[n] \quad \text{for all } n \tag{P:5.2.4}$$
$$X_{k+N} = X_k \quad \text{for all } k \tag{P:5.2.5}$$

These periodicities in $x[n]$ and X_k follow immediately from the definitions of the DFT and IDFT.

Linearity If $x[n] \overset{\text{DFT}}{\rightleftharpoons} X_k$ and $y[n] \overset{\text{DFT}}{\rightleftharpoons} Y_k$, then

$$\alpha_1 x[n] + \alpha_2 y[n] \overset{\text{DFT}}{\rightleftharpoons} \alpha_1 X_k + \alpha_2 Y_k \tag{P:5.2.6}$$

for any real or complex-valued constants α_1 and α_2.

Symmetry of real-valued sequences If the sequence $x[n] \overset{\text{DFT}}{\rightleftharpoons} X_k$ is real, then

$$X_{N-k} = X_k^* = X_{-k} \tag{P:5.2.24}$$

Complex-conjugate properties If $x[n] \overset{\text{DFT}}{\rightleftharpoons} X_k$ then

$$x^*[n] \overset{\text{DFT}}{\rightleftharpoons} X_{N-k}^* \tag{P:5.2.45}$$

PROOF. The DFT of the sequence $x[n]$ is given by:

$$X_k = \sum_{n=0}^{N-1} x[n]\ e^{-j\frac{2\pi}{N}nk}, \quad k \in \mathcal{N} \tag{M:2.2.9}$$

and the DFT of $y[n] = x^*[n]$ is given by:

$$Y_k = \sum_{n=0}^{N-1} x^*[n]\ e^{-j\frac{2\pi}{N}nk} \tag{1.62}$$

Taking complex conjugates, and noting that $e^{j\frac{2\pi}{N}mk} = e^{-j\frac{2\pi}{N}m(N-k)}$, then:

$$Y_k^* = \sum_{n=0}^{N-1} x[n]\ e^{-j\frac{2\pi}{N}m(N-k)} = X_{N-k} \tag{1.63}$$

□

Hence, giving $x^*[n] \overset{\text{DFT}}{\rightleftharpoons} X_{N-k}^*$ as required.

Time reversal of a sequence If $x[n] \overset{\text{DFT}}{\rightleftharpoons} X_k$ then

$$x[N-n] \overset{\text{DFT}}{\rightleftharpoons} X_{N-k} \tag{P:5.2.42}$$

Hence, reversing the N-point sequence in time is equivalent to reversing the DFT values in frequency.

PROOF. From the definition of the DFT, if $y[n] = x[N-n]$, then:

$$Y_k = \sum_{n=0}^{N-1} x[N-n]\ e^{-j\frac{2\pi}{N}nk} = \sum_{m=1}^{N} x[m]\ e^{-j\frac{2\pi}{N}(N-m)k} \tag{1.64}$$

where the second summation comes from changing the index from n to $m = N - n$. Noting then, that $x[N] = x[0]$,then this may be written as

$$Y_k = \sum_{m=0}^{N-1} x[m]\ e^{-j\frac{2\pi}{N}(N-m)k} = \sum_{m=0}^{N-1} x[m]\ e^{j\frac{2\pi}{N}mk} \tag{1.65}$$

$$= \sum_{m=0}^{N-1} x[m]\ e^{-j\frac{2\pi}{N}m(N-k)} = X_{N-k} \tag{1.66}$$

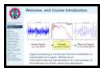
□

as required.

Circular Convolution As with many linear transforms, convolution in the time-domain becomes multiplication in the frequency domain, and vice-versa. Since the signals are periodic, it is necessary to introduce the idea of circular convolution. Details of this are discussed in depth in [Proakis:1996, Section 5.2.2, Page 415] and are currently ommitted here. However, assuming that convolution is interpreted in the circular sense (i.e. taking advantage of the periodicity of the time-domain signals), then if $x[n] \overset{\text{DFT}}{\rightleftharpoons} X_k$ and $y[n] \overset{\text{DFT}}{\rightleftharpoons} Y_k$, then:

$$x[n] * y[n] \overset{\text{DFT}}{\rightleftharpoons} X_k\, Y_k \tag{P:5.2.41}$$

1.5 Review of discrete-time systems

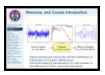


New slide

The following aspects of **discrete-time system**s are reviewed:

- Basic discrete-time signals
- The z-transform
- Review of **linear time-invariant** systems
- Rational **transfer functions**

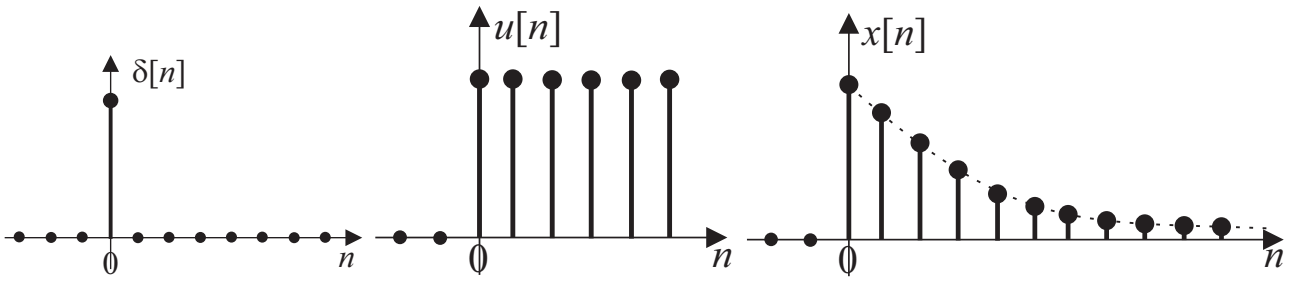
1.5.1 Basic discrete-time signals



New slide

In general, the notation $x[n]$ is used to denote a sequence of numbers that represent a discrete-time signal. The nth sample refers to the value of this sequence for a specific value of n. In a strict sense, this terminology is only correct if the discrete-time signal has been obtained by sampling a continuous-time signal $x_c(t)$. In the case of periodic sampling with sampling period T, then $x[n] = x_c(nT)$, $n \in \mathbb{Z}$; that is, $x[n]$ is the nth sample of $x_c(t)$.

There are some basic discrete-time signals that will be used repeatedly throughout the course, and these are shown in Figure 1.15:



(a) The unit sample or unit impulse sequence $\delta[n]$.

(b) The unit step sequence $u[n]$.

(c) The exponential decay sequence.

Figure 1.15: Basic discrete-time signals.

1. The **unit sample** or **unit impulse** sequence $\delta[n]$ is defined as:

$$\delta[n] = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases} \tag{M:2.1.1}$$

2. The **unit step** sequence, $u[n]$ is defined as:

$$u[n] = \begin{cases} 1 & n \geq 0 \\ 0 & n < 0 \end{cases} \tag{M:2.1.2}$$

3. The **exponential sequence** is of the form

$$x[n] = a^n, \quad -\infty < n < \infty,\ n \in \mathbb{Z} \tag{M:2.1.3}$$

If a is a complex number, such that $a = r\,e^{j\omega_0}$ for $r > 0$, $\omega_0 \neq 0, \pi$, then $x[n]$ is complex valued and given by:

$$\begin{aligned} x[n] &= r^n\, e^{j\omega_0 n} = x_R[n] + j x_I[n] && \text{(M:2.1.4)} \\ &= r^n \cos\omega_0 n + j r^n \sin\omega_0 n && (1.67) \end{aligned}$$

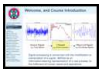
where $x_R[n]$ and $x_I[n]$ are real sequences given by:

$$x_R[n] = r^n \cos\omega_0 n \quad \text{and} \quad x_I[n] = r^n \sin\omega_0 n \tag{M:2.1.5}$$

4. The **critical decay sequence** is of the form

$$x[n] = a\, n\, r^n\, u[n]\,, \quad n \in \mathbb{Z} \tag{1.68}$$

which is discussed further in Sidebar 3.



New slide

1.5.2 The z-transform

The z-transform of a sequence is a very powerful tool for the analysis of discrete linear and time-invariant systems; it plays the same role in the analysis of discrete-time signals and linear time-invariant (LTI) systems as the Laplace transform does in the analysis of continuous-time signals

Sidebar 3 The signal $n\,r^n$

The discrete-time signal

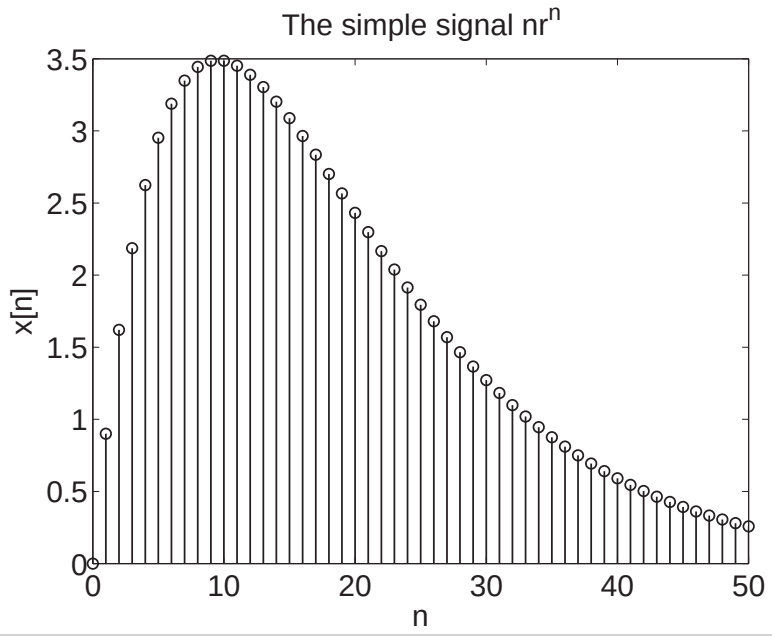
$$x[n] = a\,n\,r^n \tag{1.69}$$

is equivalent to the continuous-time signal $x[t] = t\,e^{-\alpha t}$, and both are important, as they represent the response of a **critically damped system**, as will be seen later. Note in both cases that:

$$\lim_{n\to\infty} n\,r^n \to 0 \tag{1.70}$$

The shape of $x[n]$ is shown below for $r = 0.9$, and note the relationship derived in Sidebar 4 that:

$$n\,r^n \overset{z^+}{\rightleftharpoons} \frac{r}{(1-r)^2} \quad \text{if } |r| < 1 \tag{1.71}$$



and LTI systems. For example, as will be seen, in the z-domain, also known as the complex z-plane, the convolution of two time-domain signals is equivalent to multiplication of their corresponding z-transforms. This property greatly simplifies the analysis of the response of an LTI system to various inputs.

Although the Fourier transform also satisfies the property that convolution in the time domain becomes multiplication in the frequency domain, it is not always possible to calculate the Fourier transform of a signal, $x[n]$, even for some elementary signals that are important for the analysis of systems. For example, if $x[n]$ is a **power signal** (having finite power), rather than an **energy signal**, the discrete-time Fourier transform (DTFT) does not exist.

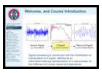
One such signal, of practical importance, is the unit step function, $u[t]$, as can be illustrated by attempting to evaluate the DTFT:

$$U\left(e^{j\omega T}\right) = \sum_{n=-\infty}^{\infty} u[n]\ e^{-j\omega n} = \sum_{n=0}^{\infty} e^{-j\omega n} \tag{1.72}$$

This is a geometric series, of the form $\sum_{n=0}^{\infty} r^n$ where $r = e^{-j\omega}$; however, this series *diverges* since $|r| = 1$. Therefore, the DTFT does not exist; this could also have been deduced from the fact that $u[n]$ is not absolutely summable, which a necessary condition for a Fourier transform to exist:

$$\sum_{\text{all } n} |u[n]| = \sum_{n=0}^{\infty} 1 \not< \infty \tag{1.73}$$

The solution is to multiply the signal by a convergence factor, which leads to the z-transform. Details of the derivation can be found in some text books.



New slide

1.5.2.1 Bilateral z-transform

The bilateral z-transform is defined by the following pairs of equations:

$$X(z) \triangleq \mathcal{Z}[x[n]] = \sum_{n=-\infty}^{\infty} x[n]\ z^{-n} \tag{M:2.2.29}$$

$$x[n] = \frac{1}{2\pi j} \oint_C X(z)\ z^{n-1}\, dz \tag{M:2.2.30}$$

where z is a complex variable. This is usually denoted as:

$$x[n] \overset{z}{\rightleftharpoons} X(z) \quad \text{or} \quad X(z) = \mathcal{Z}[x[n]] \tag{1.74}$$

The set of values of z for which the power series in the (direct) z-transform converges is called the region of convergence (ROC) of $X(z)$. A sufficient condition for convergence is:

$$\sum_{n=-\infty}^{\infty} |x[n]\,||z^{-n}| < \infty \tag{M:2.2.31}$$

The unilateral or one-sided z-transform, which is more commonly encountered in undergraduate Engineering courses, is discussed below in Section 1.5.2.3. For the moment, it suffices to mention that the difference between them usually comes down to the initial conditions, and therefore a discussion of the bilateral transform is not too restrictive at this point.

Sidebar 4 The Ubiquitous Geometric Progression

The **geometric progresson** occurs frequently in discrete-time analysis due to the existance of the summation operator and the commonality of exponential decay functions. It is essentially the discrete-time equivalent of integrating an exponential function. The geometric progression is given by

$$\sum_{n=0}^{N} a\, r^n = a\frac{1-r^{N+1}}{1-r} \tag{1.75}$$

$$\sum_{n=0}^{\infty} a\, r^n = a\frac{1}{1-r} \quad \text{if } |r|<1 \tag{1.76}$$

More interesting are variants of the geometric progression that can be obtained by simple manipulations, such as differentiating both sides of Equation 1.76 with respect to (w. r. t.) r:

$$\frac{d}{dr}\left[\sum_{n=0}^{\infty} a\, r^n\right] = \frac{d}{dr}\left[a\frac{1}{1-r}\right] \tag{1.77}$$

$$\sum_{n=0}^{\infty} a\, n\, r^{n-1} = a\frac{1}{(1-r)^2} \tag{1.78}$$

or, multiplying both sides by r, gives:

$$\sum_{n=0}^{\infty} a\, n\, r^{n} = a\frac{r}{(1-r)^2} \quad \text{if } |r|<1 \tag{1.79}$$

which is also a useful identity. The signal $x[n] = n\, r^n$ is an important one and discussed further in Sidebar 3. Differentiating repeated times gives a general expresion for $\sum n^p\, r^n$ which can often be useful.

By evaluating the z-transform on the unit circle of the z-plane, such that $z = e^{j\omega}$, then:

$$X(z)|_{z=e^{j\omega}} = X\left(e^{j\omega T}\right) = \sum_{n=-\infty}^{\infty} x[n]\, e^{-j\omega n} \tag{M:2.2.32}$$

$$x[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} X\left(e^{j\omega T}\right)\, e^{j\omega n}\, d\omega \tag{M:2.2.33}$$

which are the DTFT and inverse-DTFT relating the signals $x[n]$ and $X\left(e^{j\omega T}\right)$. This relation holds only if the unit circle is inside the ROC.

Example 1.6 ([Proakis:1996, Example 3.1.3, Page 154]). Determine the z-transform of the signal:

$$x[n] = \alpha^n\, u[n] \equiv \begin{cases} \alpha^n & n \geq 0 \\ 0 & n < 0 \end{cases} \tag{1.80}$$

SOLUTION. From the definition of the z-transform, it follows that:

$$X(z) = \sum_{k=0}^{\infty} \alpha^n z^{-n} = \sum_{n=0}^{\infty} \left(\alpha\, z^{-1}\right)^n \tag{1.81}$$

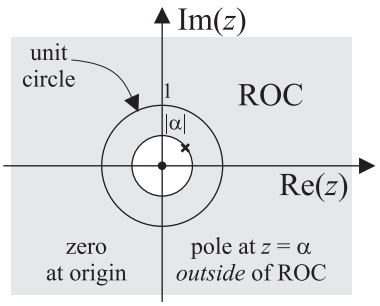


Figure 1.16: The region of convergence (ROC) for the transfer function in Equation P:3.1.7.

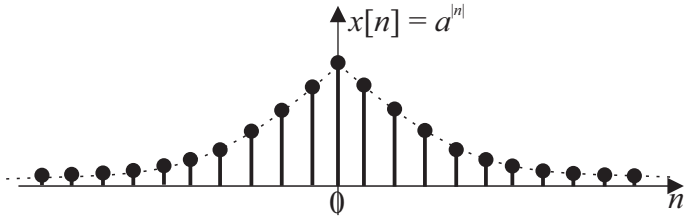


Figure 1.17: The sequence $x[n] = a^{|n|}$.

The summation on the right is a geometric progression, and converges to $\frac{1}{1-\alpha\, z^{-1}}$if, and only if, (iff) $|\alpha z^{-1}| < 1$ or, equivalently, $|z| > |\alpha|$. Further details on the geometric progression are given in Sidebar 4. Thus, this gives the z-transform pair:

$$x[n] = \alpha^n\, u[n] \overset{z}{\rightleftharpoons} X\,(z) = \frac{1}{1-\alpha z^{-1}} \quad \text{ROC: } |z| > |\alpha| \tag{P:3.1.7}$$

Note that, in general, α need not be real. The ROC is the exterior of a circle having radius $|\alpha|$. The ROC is shown in Figure 1.16. The z-transform in Equation P:3.1.7 may be written as:

$$X\,(z) = \frac{z}{z-\alpha} \quad \text{ROC: } |z| > |\alpha| \tag{1.82}$$

□

and therefore it has a pole at $z = \alpha$ and a zero at $z = 0$. The position of the pole is outside the ROC, which is as expected since the z-transform does not converge at a pole; it tends to infinity instead. However, simply because there is a zero at the origin does not mean the z-transform converges at that point – it doesn't, since it is outside of the ROC. However, the concept of the zero is still important and is thus still drawn on the pole-zero diagram.

Example 1.7 (Two-sided exponential (Laplacian exponential)). What is the bilateral z-transform of the sequence $x[n] = a^{|n|}$ for all n and some real constant a, where $|a| < 1$?

SOLUTION. The bilateral z-transform of a sequence $x[n] = a^{|n|}$, shown in Figure 1.17, is given by:

$$X\,(z) = \sum_{n=-\infty}^{\infty} x[n]\; z^{-n} = \sum_{n=-\infty}^{\infty} a^{|n|}\, z^{-n} \tag{1.83}$$

$$= \sum_{n=-\infty}^{-1} a^{-n}\, z^{-n} + \sum_{n=0}^{\infty} a^n\, z^{-n} \tag{1.84}$$

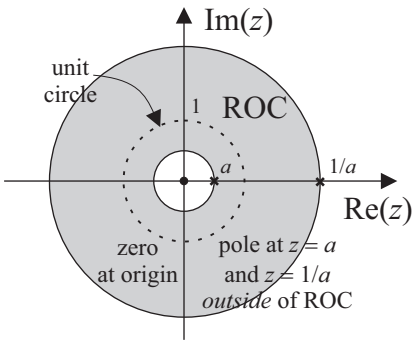


Figure 1.18: The region of convergence (ROC) for the transfer function in Equation 1.88.

Setting $m = -n$ in the first summation, noting that when $n = -\infty$ then $m = \infty$, and when $n = 0$ then $m = 0$, gives:

$$X(z) = \sum_{n=1}^{\infty} (az)^n + \sum_{n=0}^{\infty} \left(\frac{a}{z}\right)^n \tag{1.85}$$

$$= \sum_{n=0}^{\infty} (az)^n - 1 + \sum_{n=0}^{\infty} \left(\frac{a}{z}\right)^n \tag{1.86}$$

$$= \frac{1}{1-az} - 1 + \frac{1}{1-\frac{a}{z}} \tag{1.87}$$

where the expression for an infinite geometric progression has been used. Note, however, that each summation has different convergence constraints. Thus, note that the first summation only exists for $|az| < 1$, while the second summation only exists for $\left|\frac{a}{z}\right| < 1$. This means that the ROC for this transform is the ring $|a| < z < \frac{1}{|a|}$. The ROC is thus shown in Figure 1.18.

Combining the various terms and a slight rearrangement gives the expression

$$X(z) = \frac{1-a^2}{(1-az)(1-az^{-1})} \tag{1.88}$$

□

which has a zero at $z = 0$ and poles at $z = a$ and $z = \frac{1}{a}$.

1.5.2.2 Properties of the z-transform

The power of the z-transform is a consequence of some very important properties that the transform possesses. Some of these properties are listed below, as a re-cap. Note that the proof of many of these properties follows immediately from the definition of the property itself and the z-transform, and is left as an exercise for the reader. Alternatively, cheat and look in, for example, [Proakis:1996].

Linearity If $x_1[n] \overset{z}{\rightleftharpoons} X_1(z)$ and $x_2[n] \overset{z}{\rightleftharpoons} X_2(z)$, then by linearity

$$x[n] = \alpha_1 x_1[n] + \alpha_2 x_2[n] \overset{z}{\rightleftharpoons} X(z) = \alpha_1 X_1(z) + \alpha_2 X_2(z) \tag{P:3.2.1}$$

for any constants α_1 and α_2. Obviously, this property can be generalised for an arbitrary number of signals, and therefore if $x_m[n] \overset{z}{\rightleftharpoons} X_m(z)$ for $m = \{1, \ldots, M\}$

$$x[n] = \sum_{m=1}^{M} \alpha_m x_m[n] \overset{z}{\rightleftharpoons} X(z) = \sum_{m=1}^{M} \alpha_m X_m(z) \tag{1.89}$$

for any constants $\{\alpha_m\}_1^M$.

Time shifting If $x[n] \overset{z}{\rightleftharpoons} X(z)$ then:

$$x[n-k] \overset{z}{\rightleftharpoons} z^{-k} X(z) \tag{1.90}$$

The ROC of $z^{-k}X(z)$ is the same as that of $X(z)$ except for $z = 0$ if $k > 0$ and $z = \infty$ if $k < 0$.

Scaling in the z-domain If $x[n] \overset{z}{\rightleftharpoons} X(z)$ with ROC $r_1 < |z| < r_2$, then

$$a^n x[n] \overset{z}{\rightleftharpoons} X(a^{-1} z) \quad \text{ROC: } |a|r_1 < |z| < |a|r_2 \tag{P:3.2.9}$$

for any constant a.

Time reversal If $x[n] \overset{z}{\rightleftharpoons} X(z)$ with ROC $r_1 < |z| < r_2$, then

$$x[-n] \overset{z}{\rightleftharpoons} X(z^{-1}) \quad \text{ROC: } \frac{1}{r_1} < |z| < \frac{1}{r_2} \tag{P:3.2.12}$$

Differentiation in the z-domain If $x[n] \overset{z}{\rightleftharpoons} X(z)$ then

$$nx[n] \overset{z}{\rightleftharpoons} -z\frac{dX(z)}{dz} \tag{P:3.2.14}$$

PROOF. Since

$$X(z) = \sum_{n=-\infty}^{\infty} x[n]\, z^{-n} \tag{1.91}$$

then differentiating both sides gives:

$$\frac{dX(z)}{dz} = -z^{-1} \sum_{n=-\infty}^{\infty} [nx[n]]\, z^{-n} = -z^{-1}\mathcal{Z}[nx[n]] \tag{1.92}$$

□

Both transforms have the same ROC.

Convolution If $x_1[n] \overset{z}{\rightleftharpoons} X_1(z)$ and $x_2[n] \overset{z}{\rightleftharpoons} X_2(z)$, then

$$x[n] = x_1[n] * x_2[n] \overset{z}{\rightleftharpoons} X(z) = X_1(z)X_2(z) \tag{3.2.17}$$

The ROC of $X(z)$ is, at least, the intersection of that for $X_1(z)$ and $X_2(z)$.

PROOF. The convolution of $x_1[n]$ and $x_2[n]$ is defined as:

$$x[n] = \sum_{k=-\infty}^{\infty} x_1[k]\, x_2[n-k] \tag{1.93}$$

The z-transform of $x[n]$ is:

$$X(z) = \sum_{n=-\infty}^{\infty} x[n]\, z^{-n} = \sum_{n=-\infty}^{\infty} \left[\sum_{k=-\infty}^{\infty} x_1[k]\, x_2[n-k] \right] z^{-n} \tag{1.94}$$

Upon changing the order of the summations, then:

$$X(z) = \sum_{k=-\infty}^{\infty} x_1[k] \underbrace{\left[\sum_{n=-\infty}^{\infty} x_2[n-k]\, z^{-n} \right]}_{X_2(z)\, z^{-k}} = X_2(z) \underbrace{\sum_{k=-\infty}^{\infty} x_1[k]\, z^{-k}}_{X_1(z)} \tag{1.95}$$

□

giving the desired result.

The Initial Value Theorem If $x[n] = 0,\ n < 0$ is a **causal** sequence, then

$$x[0] = \lim_{z\to\infty} X\,(z) \tag{P:3.2.23}$$

PROOF. Since $x[n]$ is causal, then:

$$X\,(z) = x[0] + x[1]\ z^{-1} + x[2]\ z^{-2} + \cdots \tag{1.96}$$

□

Hence, as $z \to \infty$, $z^{-n} \to 0$ since $n > 0$, and thus the desired result is obtained.

1.5.2.3 The Unilateral z-transform

The two-sided z-transform requires that the corresponding signals be specified for the entire time range $n \in \mathbb{Z}$. This requirement prevents its used for systems that are described by difference equations with nonzero initial conditions. Since the input is applied at a finite time, say n_0, both input and output signals are specified for $n \geq n_0$, but are not necessarily zero for $n < 0$. Thus the two sided z-transform cannot be used

The one-sided **unilateral z-transform** of a signal $x[n]$ is defined by:

$$X^+(z) \equiv \sum_{n=0}^{\infty} x[n]\ z^{-n} \tag{P:3.5.1}$$

This is usually denoted as:

$$x[n] \overset{z^+}{\rightleftharpoons} X^+(z) \quad \text{or} \quad X^+(z) = \mathcal{Z}^+[x[n]] \tag{1.97}$$

The unilateral z-transform differs from the bilateral transform in the lower limit of the summation, which is always zero, whether or not the signal $x[n]$ is zero for $n < 0$ (i.e., causal). Therefore, the unilateral z-transform contains no information about the signal $x[n]$ for negative values of time, and is therefore *unique* only for causal signals. The unilateral and bilateral z-transforms are, consequentially, identical for the signal $x[n]\,u[n]$ where $u[n]$ is the step function. Since $x[n]\,u[n]$ is causal, the ROC of its transform, and hence the ROC of $X^+(z)$, is always the exterior of a circle. Thus, when discussing the unilateral z-transform, it is not necessary to refer to their ROC - which perhaps explains why this is the more commonly discussed transform in undergraduate courses.

Almost all the properties for the bilateral z-transform carry over to the unilateral transform with the exception of the shifting property.

Shifting property: Time Delay If $x[n] \overset{z^+}{\rightleftharpoons} X^+(z)$ then:

$$x[n-k] \overset{z^+}{\rightleftharpoons} z^{-k}X\,(z) + \underbrace{\sum_{n=-k}^{-1} x[n]\ z^{-(n+k)}}_{\text{initial conditions}}, \quad k > 0 \tag{1.98}$$

PROOF. Since

$$X^+(z) \equiv \sum_{n=0}^{\infty} x[n]\ z^{-n} \tag{P:3.5.1}$$

then it follows that

$$\mathcal{Z}^+[x[n-k]] = \sum_{n=0}^{\infty} x[n-k]\ z^{-n} = \sum_{m=-k}^{\infty} x[m]\ z^{-(m+k)} \tag{1.99}$$

by the change of index $m = n - k$,

$$= z^{-k} \sum_{m=-k}^{-1} x[m]\ z^{-m} + z^{-k} \underbrace{\sum_{m=0}^{\infty} x[m]\ z^{-m}}_{X^+(z)} \tag{1.100}$$

This is the desired result. □

Shifting property: Time Advance If $x[n] \overset{z^+}{\rightleftharpoons} X^+(z)$ then:

$$x[n+k] \overset{z^+}{\rightleftharpoons} z^k X\,(z) - \sum_{n=0}^{k-1} x[n]\ z^{k-n}, \quad k > 0 \tag{1.101}$$

PROOF. From the definition of the unilateral transform, it follows

$$\mathcal{Z}^+[x[n+k]] = \sum_{n=0}^{\infty} x[n+k]\ z^{-n} = \sum_{m=k}^{\infty} x[m]\ z^{-(m-k)} \tag{1.102}$$

by the change of index $m = n + k$. Thus,

$$= z^k \underbrace{\sum_{0}^{\infty} x[m]\ z^{-m}}_{X^+(z)} - z^k \sum_{m=1}^{k-1} x[m]\ z^{-m} \tag{1.103}$$

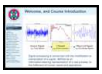
This is the desired result. □

Final Value Theorem If $x[n] \overset{z^+}{\rightleftharpoons} X^+(z)$ then:

$$\lim_{n\to\infty} x[n] = \lim_{z\to 1}(z-1)X^+(z) \tag{P:3.5.6}$$

The limit on the right hand side (RHS) exists if the ROC of $(z-1)X^+(z)$ includes the unit circle.

Further information can be found in books on discrete-time systems, for example [Proakis:1996, Section 3.5, Page 197].



New slide

1.5.3 Review of linear time-invariant systems

- Systems which are **LTI** can be elegantly analysed in both the time and frequency domain: **convolution** in time, multiplication in frequency.
- For signals and sequences, it is common to write $\{y[n]\}_{n=-\infty}^{\infty}$, or even $\{y[n]\}_{n\in\mathbb{Z}}$ rather than simply $y[n]$: the latter is sufficient for these notes.

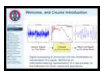
- Output, $y[n]$, of a **LTI** system is the **convolution** of the input, $x[n]$, and the **impulse response** of the system, $h[n]$:

$$y[n] = x[n] * h[n] \triangleq \sum_{k\in\mathbb{Z}} x[k]\ h[n-k] \tag{M:2.3.2}$$

- By making the substitution $\hat{k} = n - k$, it follows:

$$y[n] = \sum_{k\in\mathbb{Z}} h[k]\ x[n-k] = h[n] * x[n] \tag{M:2.3.3}$$

1.5.3.1 Matrix-vector formulation for convolution



New slide

If $x[n]$ and $h[n]$ are sequences of finite duration, the **convolution** operation can be written in matrix-vector form. Let $x[n]\,,\ 0 \le n \le N-1$ and $h[n]\,,\ 0 \le n \le M-1$ be finite-duration sequences, then $y[n]\,,\ 0 \le n \le L-1$, where $L = N + M - 1$, can be written as:

$$\begin{bmatrix} y[0] \\ y[1] \\ \vdots \\ y[M-1] \\ \vdots \\ y[N-1] \\ \vdots \\ y[L-2] \\ y[L-1] \end{bmatrix} = \begin{bmatrix} x[0] & 0 & \cdots & 0 \\ x[1] & x[0] & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ x[M-1] & \cdots & \cdots & x[0] \\ \vdots & \ddots & \ddots & \vdots \\ x[N-1] & \cdots & \cdots & x[N-M] \\ 0 & \ddots & & \vdots \\ \vdots & \ddots & x[N-1] & x[N-2] \\ 0 & \cdots & 0 & x[N-1] \end{bmatrix} \begin{bmatrix} h[0] \\ h[1] \\ \vdots \\ h[M-1] \end{bmatrix} \tag{M:2.3.4}$$

or

$$\mathbf{y} = \mathbf{X}\,\mathbf{h} \tag{M:2.3.5}$$

- Here, $\mathbf{y} \in \mathbb{R}^L$, $\mathbf{X} \in \mathbb{R}^{L\times M}$, and $\mathbf{h} \in \mathbb{R}^M$.
- The matrix $\mathbf{X}$ is termed an **input data matrix**, and has the property that it is **toeplitz**. [1]
- The observation or output vector $\mathbf{y}$ can also be written in a similar way as:

$$\mathbf{y} = \mathbf{H}\,\mathbf{x} \tag{M:2.3.6}$$

  in which $\mathbf{H}$ is also **toeplitz**.
- A system is **causal** if the present output sample depends only on past and/or present input samples.
- Assume system is asymptotically stable.

1.5.3.2 Transform-domain analysis of LTI systems

New slide

Time-domain **convolution**:

$$y[n] = \sum_{k\in\mathbb{Z}} x[k]\ h[n-k] \tag{M:2.3.2}$$

[1] A **Toeplitz** matrix is one in which the elements along each diagonal, parallel to the main diagonal each descending from left to right, are constant. Note that the anti-diagonals are not necessarily equal.

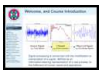
or

$$y[n] = \sum_{k\in\mathbb{Z}} h[k]\ x[n-k] \tag{M:2.3.3}$$

Taking z-**transforms** gives:

$$Y(z) = H(z)\ X(z) \tag{M:2.3.8}$$

where $X(z)$, $Y(z)$ and $H(z)$ are the z-**transforms** of the input, output, and impulse response sequences respectively. $H(z) = \mathcal{Z}[h[n]]$ is the **system function** or **transfer function**.



New slide

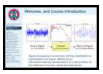
1.5.3.3 Frequency response of LTI systems

The **frequency response** of the system is found by evaluating the z-transform on the unit circle, so $z = e^{j\omega}$:

$$Y\left(e^{j\omega T}\right) = H\left(e^{j\omega T}\right)\ X\left(e^{j\omega T}\right) \tag{M:2.3.9}$$

- $|H(e^{j\omega})|$ is the **magnitude response** of the system, and $\arg H(e^{j\omega})$ is the **phase response**.
- The **group delay** of the system is a measure of the average delay of the system as a function of frequency:

$$\tau(e^{j\omega}) = -\frac{d}{d\omega}\arg H(e^{j\omega}) \tag{M:2.3.11}$$



New slide

1.5.3.4 Frequency response to Periodic Inputs

Although the convolution summation formula can be used to compute the response of a stable system to any input, the frequency-domain input-output relationship for a **LTI** cannot be used with periodic inputs, since periodic signals do not strictly possess a z-transform. However, it is possible to develop an expression for the frequency response of LTI from first principles. Let $x[n]$ be a periodic signal with fundamental period N. This signal can be expanded using an IDFT as:

$$x[n] = \frac{1}{N}\sum_{k=0}^{N-1} X_k\, e^{j\frac{2\pi}{N}kn}, \quad n \in \{0,\ldots,N-1\} \tag{M:2.3.19}$$

where X_k are the Fourier components.

Hence, it follows that on substitution into the convolution equation:

$$y[n] = \sum_{m=-\infty}^{\infty} h[m]\ x[n-m] = \frac{1}{N}\sum_{m=-\infty}^{\infty} h[m] \sum_{k=0}^{N-1} X_k\, e^{j\frac{2\pi}{N}k(n-m)} \tag{M:2.3.20}$$

which, by interchanging the order of summation (noting that the limits are over a rectangular region of summation), gives;

$$y[n] = \frac{1}{N}\sum_{k=0}^{N-1} X_k\, e^{j\frac{2\pi}{N}kn} \underbrace{\sum_{m=-\infty}^{\infty} h[m]\ e^{-j\frac{2\pi}{N}km}}_{H(e^{j\frac{2\pi}{N}k})} \tag{1.104}$$

where $H(e^{j\frac{2\pi}{N}k})$ are samples of $H(e^{j\omega})$. Hence,

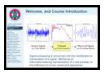
$$y[n] = \frac{1}{N}\sum_{k=0}^{N-1} \left\{ H(e^{j\frac{2\pi}{N}k})\, X_k \right\} e^{j\frac{2\pi}{N}kn} \tag{1.105}$$

However, this is just the inverse-DFT expansion of $y[n]$, and therefore:

$$Y_k = H(e^{j\frac{2\pi}{N}k})\, X_k \quad k \in \{0, \ldots, N-1\} \tag{M:2.3.21}$$

Thus, the response of a LTI system to a periodic input is also periodic with the same period. The magnitude of the input components is modified by $|H(e^{j\frac{2\pi}{N}k})|$, and the phase is modified by $\arg H(e^{j\frac{2\pi}{N}k})$.

1.5.4 Rational transfer functions



New slide

Many systems can be expressed in the z-domain by a **rational transfer function**. They are described in the time domain by:

$$y[n] = -\sum_{k=1}^{P} a_k\, y[n-k] + \sum_{k=0}^{Q} d_k\, x[n-k] \tag{M:2.3.12}$$

Taking z-transforms gives:

$$H\,(z) = \frac{Y\,(z)}{X\,(z)} = \frac{\sum_{k=0}^{Q} d_k\, z^{-k}}{1 + \sum_{k=1}^{P} a_k\, z^{-k}} \triangleq \frac{D\,(z)}{A\,(z)} \tag{M:2.3.13}$$

This can be described in the complex z-plane as:

$$H\,(z) = \frac{D\,(z)}{A\,(z)} = G\frac{\prod_{k=1}^{Q}(1 - z_k\, z^{-1})}{\prod_{k=1}^{P}(1 - p_k\, z^{-1})} \tag{M:2.3.14}$$

where p_k are the poles of the system, and z_k are the zeros.

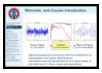
2

Discrete-Time Stochastic Processes

Introduces the notion of time-series or random processes. Gives an interpretation using ensembles, and covers second-order statistics including correlation sequences. Discusses types of stationary processes, ergodicity, joint-signal statistics, and correlation matrices.

2.1 A Note on Notation

Note that, unfortunately, for this module, a slightly different (and abusive use of) notation for random quantities is used than what was presented in the first four handouts of the *Probability, Random Variables, and Estimation Theory (PET)* module. In the literature, most time series are described using lower-case letters, primarily since once the notation for the representation of a random process in the frequency domain is discussed, upper-case letters are exclusively reserved to denote spectral representations. Moreover, lower-case letters for time-series are generally more recognisable and readable, and helps with the clarity of the presentation. Hence, random variables and vectors in this handout will not always be denoted using upper-case letters.



New slide

2.2 Definition of a Stochastic Process

After studying random variables and vectors, these concepts can now (easily) be extended to discrete-time signals or sequences.

- Natural discrete-time signals can be characterised as random signals, since their values cannot be determined precisely; that is, they are **unpredictable**. A natural mathematical framework for the description of these discrete-time random signals is provided by discrete-time stochastic processes.
- To obtain a formal definition, consider an experiment with a finite or infinite number of unpredictable outcomes from a sample space $\mathcal{S} = \{\zeta_k, k \in \mathbb{Z}^+\}$, each occurring with probability $\Pr(\zeta_k)$. Assign by some rule to each $\zeta_k \in \mathcal{S}$ a deterministic sequence $x[n, \zeta_k]$, $n \in \mathbb{Z}$.

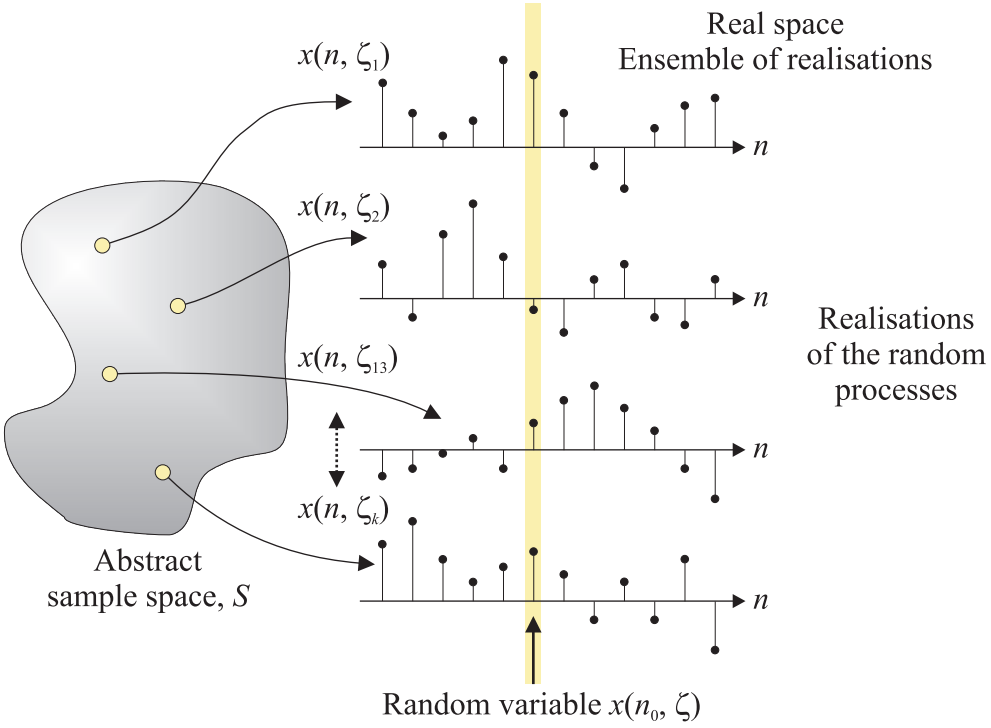
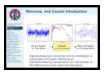


Figure 2.1: A graphical representation of a random process.

- The sample space $\mathcal{S}$, probabilities $\Pr(\zeta_k)$, and the sequences $x[n, \zeta_k]$, $n \in \mathbb{Z}$ constitute a **discrete-time stochastic process**, or **random sequence**.
- Formally, $x[n, \zeta_k]$, $n \in \mathbb{Z}$ is a random sequence or **stochastic process** if, for a fixed value $n_0 \in \mathbb{Z}^+$ of n, $x[n_0, \zeta]$, $n \in \mathbb{Z}$ is a random variable.
- A random or stochastic process is also known as a **time series** in the statistics literature.
- It is an infinite sequence of random variables, so could be thought of as an infinite-dimensional random vector. Indeed, finite-length random signals and sequences can specifically be represented by the concept of a random vector.

2.2.1 Interpretation of Sequences



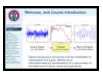
New slide

The set of all possible sequences $\{x[n, \zeta]\}$ is called an **ensemble**, and each individual sequence $x[n, \zeta_k]$, corresponding to a specific value of $\zeta = \zeta_k$, is called a **realisation** or a **sample sequence** of the ensemble. Hence, when a random process is observed through the outcome of a single experiment, one member of the ensemble is selected randomly and presented. A graphical representation of a random process is shown in Figure 2.1.

There are four possible interpretations of $x[n, \zeta]$:

	ζ Fixed	ζ Variable
n Fixed	Number	Random variable
n Variable	Sample sequence	Stochastic process

Use simplified notation $x[n] \equiv x[n, \zeta]$ to denote both a stochastic process, and a single realisation. The word *stochastic* is derived from the Greek word *stochasticos*, which means skillful in aiming or guessing. Use the terms **random process** and **stochastic process** interchangeably throughout this course.



New slide

2.2.2 Predictable Processes

A deterministic signal is by definition exactly predictable; it assumes there exists a certain functional relationship that completely describes the signal, even if that functional relationship is not available or is extremely difficult to describe. The unpredictability of a random process is, in general, the combined result of the following two characteristics:

1. The selection of a single realisation of a stochastic process is based on the outcome of a random experiment; in other-words, it depends on ζ.
2. No functional description is available for *all* realisations of the *ensemble*. In other-words, even if a functional relationship is available for a subset of the ensemble, it might not be available for all members of the ensemble.

In some special cases, however, a functional relationship is available. This means that after the occurrence of all samples of a particular realisation up to a particular point, n, all future values can be predicted exactly from the past ones.

If this is the case for a random process, then it is called **predictable**, otherwise it is said to be **unpredictable** or a **regular process**.

KEYPOINT! (Predictable Process). As an example of a predictable process, consider the signal:

$$x[n,\zeta] = A\,\sin\left(\omega\, n + \phi\right) \tag{2.1}$$

□

where A is a known amplitude, ω is a known normalised angular frequency, and ϕ is a random phase, where $\phi \sim f_\Phi\left(\phi\right)$ is its probability density function (pdf).

As an outline of this idea, suppose that all the samples of a stochastic process $x[n,\zeta]$ upto sample $n-1$ are known; thus, $\{x[k,\zeta]\}_{k=-\infty}^{n-1}$ are known. Then the predicted value of $x[n]$ might, for example, be expressed as:

$$\hat{x}[n] = -\sum_{k=1}^{\infty} a_k^*\, x[n-k] \tag{T:7.189}$$

The error in this prediction is given by

$$\epsilon[n] = x[n] - \hat{x}[n] = \sum_{k=0}^{\infty} a_k^*\, x[n-k] \tag{T:7.190}$$

where $a_0 = 1$. The process is said to be **predictable** if the $\{a_k\}$'s can be chosen such that:

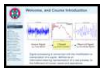
$$\sigma_\epsilon^2 = \mathbb{E}\left[|\epsilon[n]\,|^2\right] = 0 \tag{T:7.191}$$

Otherwise the process is not predictable. The phrase *not predictable* is somewhat misleading, since the **linear prediction** in Equation T:7.189 can be applied to any process, whether predictable or not, with satisfactory results. If a process is not predictable, it just means that the prediction error variance is not zero.

An example of **predictable process** is the process $x[n,\zeta] = c$, where c is a random variable, since every realisation of the discrete-time signal has a constant amplitude, and once $x[n_0,\zeta_k]$ is known for a particular realisation, all other samples of that process have also been determined.

The notion of predictable and regular processes is formally presented through the **Wold decomposition**, and further details of this very important theorem can be found in [Therrien:1992, Section 7.6, Page 390] and [Papoulis:1991, Page 420].

2.2.3 Description using pdfs



New slide

For fixed $n = n_0$, it is clear from Figure 2.1 that $x[n_0, \zeta]$ is a random variable. Moreover, the random vector formed from the k random variables $\{x[n_j]\,,\, j \in \{1, \ldots\, k\}\}$ is characterised by the joint-cumulative distribution function (cdf) and pdfs:

$$F_X\left(x_1 \ldots x_k \mid n_1 \ldots n_k\right) = \Pr\left(x[n_1] \leq x_1, \ldots, x[n_k] \leq x_k\right) \tag{2.2}$$

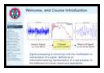
$$f_X\left(x_1 \ldots x_k \mid n_1 \ldots n_k\right) = \frac{\partial^k F_X\left(x_1 \ldots x_k \mid n_1 \ldots n_k\right)}{\partial x_1 \cdots \partial x_k} \tag{2.3}$$

In exactly the same way as with random variables and random vectors, it is:

- difficult to estimate these probability functions without considerable additional information or assumptions;
- possible to frequently characterise stochastic processes usefully with much less information.

Thus, the density and distribution functions are characterised using moments and, in particular, second-order moments.

2.3 Second-order Statistical Description



New slide

Random variables can be characterised, upto second-order statistics, using the mean and variance; random vectors are characterised by the mean vector, auto-correlation and auto-covariance matrices. Random processes, however, are characterised by sequences, where a particular sample, n_0, of this sequence characterises the random variable $x[n_0, \zeta]$. These sequences are the mean and variance sequence, the autocorrelation and autocovariance sequences, as outlined below.

Mean and Variance Sequence At time n, the **ensemble** mean and variance are given by:

$$\mu_x[n] = \mathbb{E}\left[x[n]\right] \tag{M:3.3.3}$$

$$\sigma_x^2[n] = \mathbb{E}\left[\left|x[n] - \mu_x[n]\right|^2\right] = \mathbb{E}\left[\left|x[n]\right|^2\right] - \left|\mu_x[n]\right|^2 \tag{M:3.3.4}$$

Both $\mu_x[n]$ and $\sigma_x^2[n]$ are deterministic sequences.

Autocorrelation sequence The second-order statistic $r_{xx}[n_1, n_2]$ provides a measure of the dependence between values of the process at two different times; it can provide information about the time variation of the process:

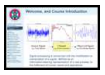
$$r_{xx}[n_1, n_2] = \mathbb{E}\left[x[n_1]\; x^*[n_2]\right] \tag{M:3.3.5}$$

Autocovariance sequence The autocovariance sequence provides a measure of how similar the deviation from the mean of a process is at two different time instances:

$$\begin{aligned}\gamma_{xx}[n_1, n_2] &= \mathbb{E}\left[(x[n_1] - \mu_x[n_1])(x[n_2] - \mu_x[n_2])^*\right] \\ &= r_{xx}[n_1, n_2] - \mu_x[n_1]\; \mu_x^*[n_2]\end{aligned} \tag{M:3.3.6}$$

To show how these deterministic sequences of a stochastic process can be calculated, several examples are considered in detail below.

2.3.1 Example of calculating autocorrelations



New slide

These examples assume that the notion of stationarity has been met; this, in fact, is not discussed until Section 2.5. Either the reader can skip these examples and return to read them after reading Section 2.5, or for the moment the reader can proceed by using the simple definition that a "stationary" process is one for which the autocorrelation function $r_{xx}[n,m] = r_{xx}[n-m] = r_{xx}[\ell]$ is simply a function of the time (or sample index) differences, also called the lag: $\ell = n - m$.

Example 2.1 (Straightforward example). A random variable $y[n]$ is defined to be:

$$y[n] = x[n] + x[n+m] \tag{2.4}$$

where m is some integer, and $x[n]$ is a stationary stochastic process whose autocorrelation function is given by:

$$r_{xx}[\ell] = e^{-\ell^2} \tag{2.5}$$

Derive an expression for the autocorrelation of the stochastic process $y[n]$.

SOLUTION. In this example, it is simplest to form the product:

$$\begin{aligned} y(n)\,y^*(n-l) &= [x(n) + x(n+m)]\;[x^*(n-l) + x^*(n+m-l)] && (2.6)\\ &= x(n)\,x^*(n-l) + x(n+m)\,x^*(n-l) \\ &\quad + x(n)\,x^*(n+m-l) + x(n+m)\,x^*(n+m-l) && (2.7) \end{aligned}$$

then, taking expectations, noting $x(n)$ is a stationary signal, it follows:

$$r_{yy}(l) = r_{xx}(l) + r_{xx}(m+l) + r_{xx}(l-m) + r_{xx}(l) \tag{2.8}$$

giving, in this particular case,

$$r_{yy}(l) = 2\,e^{-l^2} + e^{-(l+m)^2} + e^{-(l-m)^2} \tag{2.9}$$

□

Example 2.2 ([Manolakis:2000, Ex 3.9, page 144]). The harmonic process $x[n]$ is defined by:

$$x[n] = \sum_{k=1}^{M} A_k\,\cos(\omega_k n + \phi_k), \quad \omega_k \neq 0 \tag{M:3.3.50}$$

where M, $\{A_k\}_1^M$ and $\{\omega_k\}_1^M$ are constants, and $\{\phi_k\}_1^M$ are pairwise independent random variables uniformly distributed in the interval $[0, 2\pi]$.

1. Determine the mean of $x[n]$.

2. Show the autocorrelation sequence is given by

$$r_{xx}[\ell] = \frac{1}{2}\sum_{k=1}^{M} |A_k|^2\,\cos\omega_k\ell, \quad -\infty < \ell < \infty \tag{2.10}$$

SOLUTION. 1. The expected value of the process is straightforwardly given by:

$$\mathbb{E}\left[x[n]\right] = \mathbb{E}\left[\sum_{k=1}^{M} A_k \cos(\omega_k n + \phi_k)\right] = \sum_{k=1}^{M} A_k \,\mathbb{E}\left[\cos(\omega_k n + \phi_k)\right] \tag{2.11}$$

Recall from results derived earlier in the course that if $x[n,\zeta] = g(n,\phi(\zeta))$ is a random variable obtained by transforming $\phi(\zeta)$ through a known function, g, the expectation of $x[n] = x[n,\zeta]$ is:

$$\mathbb{E}\left[x(n)\right] = \int_{-\infty}^{\infty} g(n,\phi)\, p_\Phi(\phi)\, d\phi \tag{2.12}$$

It is important to consider n as a constant.

Since a co-sinusoid is zero-mean, then:

$$\mathbb{E}\left[\cos(\omega_k n + \phi_k)\right] = \int_0^{2\pi} \cos(\omega_k n + \phi_k) \times \frac{1}{2\pi} \times d\phi_k = 0 \tag{2.13}$$

Hence, it follows:

$$\mathbb{E}\left[x(n)\right] = 0, \quad \forall n \tag{2.14}$$

2. The autocorrelation $r_{xx}(n_1,n_2) = \mathbb{E}\left[x(n_1)\, x^*(n_2)\right]$ follows similarly:

$$r_{xx}(n_1,n_2) = \mathbb{E}\left[\sum_{k=1}^{M} A_k \cos(\omega_k n_1 + \phi_k) \sum_{j=1}^{M} A_j^* \cos(\omega_j n_2 + \phi_j)\right] \tag{2.15}$$

$$= \sum_{k=1}^{M}\sum_{j=1}^{M} A_k\, A_j^* \mathbb{E}\left[\cos(\omega_k n_1 + \phi_k)\, \cos(\omega_j n_2 + \phi_j)\right] \tag{2.16}$$

After some algebra, it can be shown that:

$$\mathbb{E}\left[\cos(\omega_k n_1 + \phi_k)\, \cos(\omega_j n_2 + \phi_j)\right] = \begin{cases} \frac{1}{2}\cos\omega_k(n_1 - n_2) & k = j \\ 0 & \text{otherwise} \end{cases} \tag{2.17}$$

The proof of this statement is obtained by considering the term

$$r(\phi_k,\phi_j) = \mathbb{E}\left[\cos(\omega_k n_1 + \phi_k)\, \cos(\omega_j n_2 + \phi_j)\right] \tag{2.18}$$

for the cases when $k \neq j$, and when $k = j$. Considering the former case first, $k \neq j$, then

$$r(\phi_k,\phi_j) = \frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi} \cos(\omega_k n_1 + \phi_k)\, \cos(\omega_j n_2 + \phi_j)\, d\phi_j\, d\phi_k \tag{2.19}$$

$$= \frac{1}{4\pi^2}\int_0^{2\pi} \cos(\omega_k n_1 + \phi_k)\, d\phi_k \int_0^{2\pi} \cos(\omega_j n_2 + \phi_j)\, d\phi_j \tag{2.20}$$

$$= 0 \tag{2.21}$$

An alternative derivation which might be considered more straightforward is to observe that Equation 2.18 might also be written as:

$$r(\phi_k,\phi_j) = \mathbb{E}\left[g(\phi_k)\, h(\phi_j)\right] = \mathbb{E}\left[g(\phi_k)\right]\mathbb{E}\left[h(\phi_j)\right] \tag{2.22}$$

where $g(\phi_k) = \cos(\omega_k n_1 + \phi_k)$ and $h(\phi_k) = \cos(\omega_j n_2 + \phi_j)$, and the fact that ϕ_k and ϕ_j are independent implies the expectation function may be factorised.

For the case when $k = j$ such that $\phi = \phi_k = \phi_j$ and $\omega = \omega_k = \omega_j$, then:

$$r(\phi, \phi) = \frac{1}{2\pi} \int_0^{2\pi} \cos(\omega n_1 + \phi)\, \cos(\omega n_2 + \phi)\, d\phi \tag{2.23}$$

Using the trigonometric identity $\cos A\, \cos B = \frac{1}{2}\left(\cos(A+B) + \cos(A-B)\right)$, then:

$$r(\phi_k, \phi_j) = \frac{1}{4\pi} \int_0^{2\pi} \left\{\cos\omega(n_1 - n_2) + \cos(\omega(n_1 + n_2) + 2\phi)\right\} d\phi \tag{2.24}$$

$$= \frac{1}{2} \cos\omega(n_1 - n_2) \tag{2.25}$$

giving the result above; namely:

$$\mathbb{E}\left[\cos(\omega_k n_1 + \phi_k)\, \cos(\omega_j n_2 + \phi_j)\right] = \frac{1}{2} \cos\omega_k(n_1 - n_2)\, \delta(k - j) \tag{2.26}$$

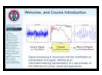
Substituting this expression into

$$r_{xx}(n_1, n_2) = \frac{1}{2} \sum_{k=1}^{M} \sum_{j=1}^{M} A_k\, A_j^* \mathbb{E}\left[\cos(\omega_k n_1 + \phi_k)\, \cos(\omega_j n_2 + \phi_j)\right] \tag{2.27}$$

thus leads to the desired result. It can be seen that the process $x(n)$ must be a stationary process, as it is only a function of the lag l:

$$r_{xx}(l) = \frac{1}{2} \sum_{k=1}^{M} |A_k|^2\, \cos\omega_k l, \quad -\infty < l < \infty \tag{2.28}$$

□



New slide

2.4 Types of Stochastic Processes

Some useful types of stochastic properties, based on their statistical properties, are now introduced:

Independence A stochastic process is independent if, and only if, (iff)

$$f_X\left(x_1, \ldots, x_N \mid n_1, \ldots, n_N\right) = \prod_{k=1}^{N} f_{X_k}\left(x_k \mid n_k\right) \tag{M:3.3.10}$$

$\forall N$, n_k, $k \in \{1, \ldots, N\}$. Here, therefore, $x[n]$ is a sequence of independent random variables.

An independent and identically distributed (i. i. d.) process is one where all the random variables $\{x[n_k, \zeta]\,,\, n_k \in \mathbb{Z}\}$ have the same pdf, and $x(n)$ will be called an **i. i. d.** random process.

Example 2.3 (Independence: i. i. d. processes). I am selling my house, and have decided to accept the first offer exceeding K pounds. Assuming that the offers are i. i. d. random variables, with common cumulative distribution function $F_X(x)$, where x is the offer price, find the expected number of offers received before I sell the house.

SOLUTION. Suppose that I sell the house after N offers. Then there are $N-1$ offers that are less than K, which occur with probability $F_X(K)$. Thus, the probability of selling the house after N offers is:

$$\Pr(N=n) = F_X(K)^{n-1}\left[1 - F_X(K)\right] \quad n \geq 1 \tag{2.29}$$

This is a **geometric distribution**, and its mean can either be looked up in tables, or calculated:

$$\mu_N = \sum_{n=1}^{\infty} n \Pr(N=n) = \sum_{n=1}^{\infty} n\, F_X(K)^{n-1}\left[1 - F_X(K)\right] \tag{2.30}$$

$$= \left[\frac{1-r}{r}\right] \sum_{n=0}^{\infty} n\, r^n \tag{2.31}$$

where $r = F_X(K)$. There is a general result which can be found in mathematical tables that [Gradshteyn:1994]:

$$\sum_{n=0}^{N-1}(a+nb)r^n = \frac{a-[a+(N-1)b]r^N}{1-r} + \frac{br(1-r^{N-1})}{(1-r)^2}, \quad r \neq 0,\ N > 1 \tag{2.32}$$

Therefore, in the case when $a = 0$, $r = 1$, and $N \to \infty$, and $0 < r < 1$ then:

$$\sum_{n=0}^{\infty} n\, r^n = \frac{r}{(1-r)^2}, \quad 0 < r < 1 \tag{2.33}$$

Hence, this gives the mean of the geometric distribution as:

$$\mu_N = \left[\frac{1-r}{r}\right] \frac{r}{(1-r)^2} = \frac{1}{1-r} = \left[1 - F_X(K)\right]^{-1} \tag{2.34}$$

□

An uncorrelated processes is a sequence of uncorrelated random variables:

$$\gamma_{xx}[n_1, n_2] = \sigma_x^2[n_1]\ \delta[n_1 - n_2] \tag{M:3.3.11}$$

Alternatively, the autocorrelation function can be written as:

$$r_{xx}[n_1, n_2] = \begin{cases} \sigma_x^2[n_1] + |\mu_x[n_1]|^2 & n_1 = n_2 \\ \mu_x[n_1]\,\mu_x^*[n_2] & n_1 \neq n_2 \end{cases} \tag{M:3.3.12}$$

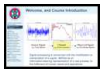
An orthogonal process is a sequence of **orthogonal random variables**, and is given by:

$$r_{xx}[n_1, n_2] = \mathbb{E}\left[|x[n_1]|^2\right]\ \delta[n_1 - n_2] \tag{M:3.3.13}$$

If a process is zero-mean, then it is both **orthogonal** and **uncorrelated** since $\gamma_{xx}[n_1, n_2] = r_{xx}[n_1, n_2]$. More often than not, in this course, we shall consider zero-mean processes.

A stationary process is a random process where its statistical properties do not vary with time. Put another way, it would be impossible to distinguish the statistical characteristics of a process at time t from those at some other time, t'. Processes whose statistical properties **do** change with time are referred to as **nonstationary**.

2.5 Stationary Processes

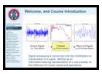


New slide

A random process $x(n)$ has been called **stationary** if its statistics determined for $x[n]$ are equal to those for $x[n+k]$, for every k. There are various formal definitions of **stationarity**, along with **quasi-stationary** processes, which are discussed below.

- **Order-N and strict-sense stationarity**
- **Wide-sense stationarity**
- **Wide-sense periodicity** and **cyclo-stationarity**
- Local- or **quasi-stationary** processes

After this, some examples of various stationary processes will be given.



2.5.1 Order-N and strict-sense stationarity

New slide

Definition 2.1 (Stationary of order-N). A stochastic process $x[n]$ is called **stationary of order-N** if:

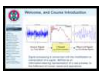
$$f_X\left(x_1,\ldots,x_N \mid n_1,\ldots,n_N\right) = f_X\left(x_1,\ldots,x_N \mid n_1+k,\ldots,n_N+k\right) \tag{M:3.3.21}$$

♢

for any value of k. If $x[n]$ is stationary for all orders $N \in \mathbb{Z}^+$, it is said to be **strict-sense stationary (SSS)**.

Clearly, any stochastic process that is stationary of order-N is also stationary of order-M, where $M \leq N$.

An independent and identically distributed process is SSS since, in this case, $f_{X_k}\left(x_k \mid n_k\right) = f_X\left(x_k\right)$ is independent of n, and therefore also of $n+k$. However, SSS is more restrictive than necessary in practical applications, and is a rarely required property.



2.5.2 Wide-sense stationarity

New slide

A more relaxed form of stationarity, which is sufficient for practical problems, occurs when a random process is stationary order-2; such a process is **wide-sense stationary (WSS)**.

Definition 2.2 (Wide-sense stationarity). A random signal $x[n]$ is called wide-sense stationary if:

- the mean and variance is constant and independent of n:

$$\mathbb{E}\left[x[n]\right] = \mu_x \tag{M:3.3.22}$$

$$\text{var}\left[x[n]\right] = \sigma_x^2 \tag{M:3.3.23}$$

- the autocorrelation depends only on the time difference $\ell = n_1 - n_2$, called the lag:

$$\begin{aligned} r_{xx}[n_1,n_2] &= r_{xx}^*[n_2,n_1] = \mathbb{E}\left[x[n_1]\ x^*[n_2]\right] \\ &= r_{xx}[\ell] = r_{xx}[n_1-n_2] = \mathbb{E}\left[x[n_1]\ x^*[n_1-\ell]\right] \\ &= \mathbb{E}\left[x[n_2+\ell]\ x^*[n_2]\right] \end{aligned} \tag{M:3.3.24}$$

♢

Additionally:

- The autocovariance sequence is given by:

$$\gamma_{xx}[\ell] = r_{xx}[\ell] - |\mu_x|^2 \tag{2.35}$$

- Since 2nd-order moments are defined in terms of 2nd-order pdf, then strict-sense stationary are always WSS, but not necessarily *vice-versa*, except if the signal is Gaussian.
- In practice, however, it is very rare to encounter a signal that is stationary in the wide-sense, but not stationary in the strict sense.

Example 2.4 ([Manolakis:2000, Example 3.3.1, Page 102]). Let $w(n)$ be a zero-mean, uncorrelated Gaussian random sequence with variance $\sigma_w^2(n) = 1$.

1. Characterise the random sequence $w(n)$.
2. Define $x(n) = w(n) + w(n-1)$, $n \in \mathbb{Z}$. Determine the mean and autocorrelation of $x(n)$. Also, characterise $x(n)$.

SOLUTION. Note that the variance of $w(n)$ is a constant.

1. Since uncorrelatedness implies independence for Gaussian random variables, then $w(n)$ is an independent random sequence. Since its mean and variance are constants, it is at least stationary of first-order. Furthermore, from Equation M:3.3.12 or from Equation M:3.3.13, then:

$$r_{ww}(n_1, n_2) = \sigma_w^2 \delta(n_1 - n_2) = \delta(n_1 - n_2) \tag{2.36}$$

   Since the autocorrelation sequence depends only on the lag $n_1 - n_2$, then by definition it is WSS process.

2. The mean of $x(n)$ is zero for all n since $w(n)$ is a zero-mean process. Next, consider:

$$\begin{aligned} r_{xx}(n_1, n_2) &= \mathbb{E}\left[x(n_1)\, x^*(n_2)\right] && (2.37)\\ &= \mathbb{E}\left[[w(n_1) + w(n_1 - 1)][w^*(n_2) + w^*(n_2 - 1)]\right] && (2.38)\\ &= r_{ww}(n_1, n_2) + r_{ww}(n_1, n_2 - 1) + r_{ww}(n_1 - 1, n_2) + r_{ww}(n_1 - 1, n_2 - 1) && (2.39)\\ &= 2\delta(n_1 - n_2) + \delta(n_1 - n_2 + 1) + \delta(n_1 - n_2 - 1) && (2.40)\\ &= 2\delta(l) + \delta(l + 1) + \delta(l - 1), \quad l = n_1 - n_2 && (2.41) \end{aligned}$$

   □

   Hence, since $r_{xx}(n_1, n_2) \equiv r_{xx}(l)$ is a function of the difference between n_1 and n_2 only, then $x(n)$ is a WSS sequence. However, it is not an independent process since both $x(n)$ and $x(n+1)$ both depend on $w(n)$.

Example 2.5 ([Manolakis:2000, Example 3.3.2, Page 103]: Wiener Process). A coin is tossed at each $n \in \mathbb{Z}$. Let:

$$w[n] = \begin{cases} +S & \text{if heads is the outcome, with probability } \Pr(H) = p \\ -S & \text{if tails is the outcome, with probability } \Pr(T) = 1 - p \end{cases} \tag{2.42}$$

where S is some arbitrary increment or step size in the process $w[n]$. Since $w[n]$, for a given n, is a discrete-random variable taking on two possible values (either S or $-S$), then $w[n]$ is an independent random process with mean:

$$\mathbb{E}\left[w[n]\right] = S\Pr\left(H\right) + (-S)\Pr\left(T\right) \tag{2.43}$$
$$\mu_w = Sp + (-S)(1-p) = S\left(2p-1\right) \tag{2.44}$$

and second moment:

$$\mathbb{E}\left[w^2[n]\right] = \sigma_w^2 + \mu_w^2 \tag{2.45}$$
$$= S^2\Pr\left(H\right) + (-S)^2\Pr\left(T\right) \tag{2.46}$$
$$= S^2 p + S^2\left(1-p\right) = S^2 \tag{2.47}$$

This in turn means that the autocorrelation function for $w[n]$ is given by:

$$r_{ww}[n,\, m] = \begin{cases} S^2 & \text{if } n = m \\ \mu_w^2 = S^2\left(2p-1\right)^2 & \text{if } n \neq n \end{cases} \tag{2.48}$$

Not only is the process $w[n]$ an i. i. d. process, it is also SSS, and therefore, it is also WSS.

Now, define a new random process $x[n]\,,\ n \geq 1$, as:

$$x[1] = w[1] \tag{2.49}$$
$$x[2] = x[1] + w[2] = w[1] + w[2] \tag{2.50}$$
$$\vdots \tag{2.51}$$
$$x[n] = x[n-1] + w[n] \tag{2.52}$$
$$= \sum_{k=1}^{n} w[n] \tag{2.53}$$

Note that $x[n]$ is a running or cummulative sum of independent increments; this is known as an **independent increment process**. Such a sequence is called a **discrete Wiener process** or **random walk**. It can easily be seen that the mean is given by:

$$\mu_x[n] = \mathbb{E}\left[x[n]\right] = \mathbb{E}\left[\sum_{k=1}^{n} w[n]\right] \tag{2.54}$$
$$= n\,S\left(2p-1\right) \tag{2.55}$$

The variance of $x[n]$ is given by:

$$\sigma_x^2[n] = \mathbb{E}\left[x^2[n]\right] - \mu_x^2[n] = \mathbb{E}\left[\sum_{k=1}^{n} w[k] \sum_{\ell=1}^{n} w[\ell]\right] - \mu_x^2[n] \tag{2.56}$$
$$= \mathbb{E}\left[\sum_{k=1}^{n}\sum_{\ell=1}^{n} w[k]\, w[\ell]\right] - \mu_x^2[n] = \sum_{k=1}^{n}\sum_{\ell=1}^{n} r_{ww}[k-\ell] - \mu_x^2[n] \tag{2.57}$$
$$= \sum_{k=1}^{n}\left[S^2 + (n-1)\,S^2(2p-1)^2\right] - \left(n\,S\left(2p-1\right)\right)^2 \tag{2.58}$$
$$= n\,S^2 + \left(n(n-1) - n^2\right) S^2(2p-1)^2 = \left[1 - (2p-1)^2\right] n\,S^2 \tag{2.59}$$
$$= 4p\left(1-p\right)\, nS^2 \tag{2.60}$$

Therefore, the random walk is a nonstationary (or evolutionary) process with a mean and variance that grows linearly with n, the number of steps taken.

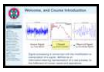
It is worth noting that finding the autocorrelation the process $x[n]$ is somewhat more involved, as it involves a calculation involving different limits in each summation:

$$\mathbb{E}\left[x[n]\ x[m]\right] = \sum_{k=1}^{n}\sum_{\ell=1}^{m}\mathbb{E}\left[w[k]\ w[\ell]\right] \tag{2.61}$$

$\bowtie$

Substituting the expression for $r_{ww}[k,\ \ell]$, and rearranging will give the desired answer. This is left as an exercise to the reader, but note that you will need to consider the cases when $m < n$ and $n \geq m$.

2.5.3 Wide-sense periodicity and cyclo-stationarity



New slide

A signal whose statistical properties vary *cyclically* with time is called a cyclostationary process. A cyclostationary process can be viewed as several interleaved stationary processes. For example, the maximum daily temperature in Edinburgh can be modeled as a cyclostationary process: the maximum temperature on July 21 is statistically different from the temperature on December 18; however, the temperature on December 18 of different years has (arguably) identical statistics.

Two classes of **cyclostationary signals** that are actually **nonstationary process** which, in part, have properties resembling stationary signals are:

1. A wide-sense periodic (WSP) process is classified as signals whose mean is periodic, and whose autocorrelation sequence (ACS) is periodic in both dimensions:

$$\mu_x(n) = \mu_x(n+N) \tag{M:3.3.14}$$

$$\begin{aligned} r_{xx}(n_1,n_2) &= r_{xx}(n_1+N,n_2) = r_{xx}(n_1,n_2+N) \\ &= r_{xx}(n_1+N,n_2+N) \end{aligned} \tag{M:3.3.15}$$

for all n, n_1 and n_2. These are quite tight constraints for real signals.

2. A wide-sense cyclo-stationary process has similar but less restrictive properties than a WSP process, in that the mean is periodic, but the autocorrelation function is now just invariant to a shift by N in both of its arguments:

$$\mu_x(n) = \mu_x(n+N) \tag{M:3.3.16}$$

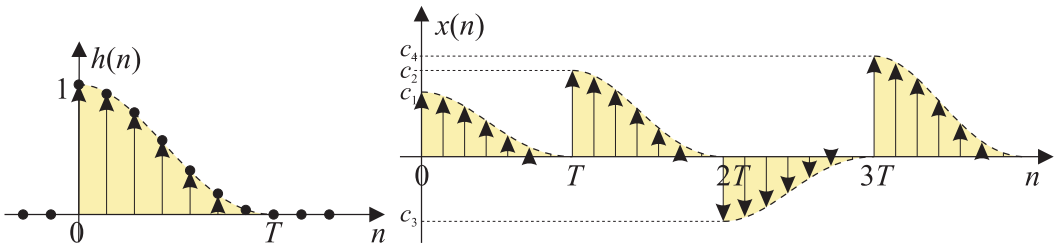
$$r_{xx}(n_1,n_2) = r_{xx}(n_1+N,n_2+N) \tag{M:3.3.17}$$

for all n, n_1 and n_2. This type of nonstationary process has more practical applications, as the following example will show.

Example 2.6 (Pulse-Amplitude Modulation). An important example of a cyclo-stationary process is the random signal

$$x[n] = \sum_{m=-\infty}^{\infty} c_m\, h[n-mT] \tag{2.62}$$

for some period T, and where c_m is a stationary sequence of random variables (RVs) with autocorrelation function $r_{cc}[n_1,n_2] = \mathbb{E}\left[c_{n_1}\, c^*_{n_2}\right] = r_{cc}[n_1-n_2]$, and $h[n]$ is a given deterministic sequence. An example of a particular pulse shape for $h[n]$ and a typical sequence $x[n]$ is shown in Figure 2.2.



(a) An example of a pulse shape.

(b) Typical transmitted signal.

Figure 2.2: An example pulse shape and typical transmit signal in a communications system.

The stochastic process $x[n]$ represents the signal for several different types of linear modulation techniques used in digital communication systems. The sequence $\{c_m\}$ represents the digital information (of symbols) that is transmitted over the communication channel, and $\frac{1}{T}$ represents the rate of transmission of the information symbols.

Note that this example demonstrates why notation can become an issue: how is it possible to determine that c_n is a RV, while $h[n]$ is not?

To see that this is a wide-sense cyclo-stationary process, first begin by writing:

$$\mu_x[n] = \mathbb{E}\left[x[n]\right] = \sum_{m=-\infty}^{\infty} \mathbb{E}\left[c_m\right]\, h[n - mT] = \mu_c \sum_{m=-\infty}^{\infty} h[n - mT] \tag{2.63}$$

where $\mu_c[n] = \mu_c$ since it is a stationary process. Thus, observe that

$$\mu_x[n + kT] = \mu_c \sum_{m=-\infty}^{\infty} h[n + kT - Tm] = \mu_c \sum_{r=-\infty}^{\infty} h[n - Tr] = \mu_x[n] \tag{2.64}$$

by a change of variables $r = m - k$.

Next consider the autocorrelation function given by:

$$r_{xx}[n_1, n_2] = \mathbb{E}\left[x(n_1)x^*(n_2)\right] = \sum_{m=-\infty}^{\infty} \sum_{\ell=-\infty}^{\infty} h[n_1 - Tm]\; h[n_2 - T\ell]\; r_{cc}[m - \ell] \tag{2.65}$$

where it has been noted that $r_{cc}[m, \ell] = r_{cc}[m - \ell]$ since it is a stationary process. Similar to the approach with the mean above, observe that

$$r_{xx}[n_1 + pT,\, n_2 + qT] = \sum_{m=-\infty}^{\infty} \sum_{\ell=-\infty}^{\infty} h[n_1 - T(m - p)]\; h[n_2 - T(l - q)]\; r_{cc}[m - \ell] \tag{2.66}$$

Again, by the change of variables $r = m - p$ and $s = \ell - q$, it can be seen that

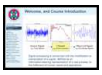
$$r_{xx}[n_1 + pT, n_2 + qT] = \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} h[n_1 - Tr]\; h[n_2 - Ts]\; r_{cc}[r - s + p - q] \tag{2.67}$$

In the case that $p = q$, then comparing Equation 2.65 and Equation 2.67, it finally follows that:

$$r_{xx}[n_1 + pT, n_2 + pT] = r_{xx}[n_1, n_2] \tag{2.68}$$

⋈

By definition, $x[n]$ is therefore a cyclo-stationary process.



2.5.4 Local- or quasi-stationary processes

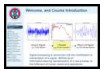
New slide

At the introduction of this lecture course, it was noted that in the analysis of speech signals, the speech waveform is broken up into short segments whose duration is typically 10 to 20 milliseconds.

This is because speech can be modelled as a **locally stationary** or **quasi-stationary** process. Such processes possess statistical properties that change *slowly* over short periods of time. They are *globally* nonstationary, but are approximately *locally* stationary, and are modelled as if the statistics *actually are* stationary over a short segment of time.

Quasi-stationary models are, in fact, just a special case of nonstationary processes, but are distinguished since their characterisation closely resemble stationary processes.

2.6 Autocorrelation properties for WSS processes



New slide

The average power of a WSS process $x(n)$ satisfies:

$$r_{xx}(0) = \sigma_x^2 + |\mu_x|^2 \qquad \text{(M:3.3.27)}$$

$$r_{xx}(0) \geq r_{xx}(l), \quad \text{for all } l \qquad \text{(M:3.3.28)}$$

The expression for power can be broken down as follows:

Average DC Power: $|\mu_x|^2$

Average AC Power: σ_x^2

Total average power: $r_{xx}(0)$

In otherwords,

$$\text{Total average power} = \text{Average DC power} + \text{Average AC power} \qquad \text{(M:3.3.27)}$$

To prove $r_{xx}[0] \geq r_{xx}[\ell]$, observe that $\mathbb{E}\left[|x[n+\ell] \pm x[n]\,|^2\right] \geq 0$. On expansion, this yields the desired result; this is left as an exercise to the reader, see [Manolakis:2000, Exercise 3.21, Page 145].

Moreover, it follows that $\gamma_{xx}(0) \geq \gamma_{xx}(l)$.

It is also the intuitively obvious, since the autocorrelation of a function should be maximum when it is "self-aligned" with itself. This property also it useful for **template-matching** time-series; i.e. to find which of a particular set of realisations is *most like* a given separate realisation.

It is left as an exercise to show that the autocorrelation sequence $r_{xx}(l)$ is:

- a conjugate symmetric function of the lag l:

$$r_{xx}^*(-l) = r_{xx}(l) \qquad \text{(M:3.3.29)}$$

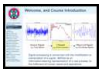
- a **nonnegative-definite** or **positive semi-definite** function, such that for any sequence $\alpha(n)$:

$$\sum_{n=1}^{M}\sum_{m=1}^{M} \alpha^*(n)\, r_{xx}(n-m)\alpha(m) \geq 0 \qquad \text{(M:3.3.30)}$$

Note that, more generally, even a correlation function for a nonstationary random process is **positive semi-definite**:

$$\sum_{n=1}^{M}\sum_{m=1}^{M} \alpha^*(n)\, r_{xx}(n,m)\alpha(m) \geq 0 \quad \text{for any sequence } \alpha(n) \qquad (2.69)$$

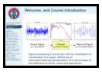
When dealing with stationary processes, this course will exclusively consider wide-sense stationary (WSS) rather than strict-sense stationary (SSS) processes. Therefore, the term *stationary* will be used to mean WSS form here onwards.



2.7 Estimating statistical properties

New slide

- A stochastic process consists of the ensemble, $x(n, \zeta)$, and a probability law, $f_X(\{x\} \mid \{n\})$. If this information is available $\forall n$, the statistical properties are easily determined.
- In practice, only a limited number of realisations of a process is available, and often only one: i.e. $\{x(n, \zeta_k),\ k \in \{1, \ldots, K\}\}$ is known for some K, but $f_X(x \mid n)$ is unknown.
- Is is possible to infer the statistical characteristics of a process from a single realisation? Yes, for the following class of signals:
  - **ergodic** processes;
  - nonstationary processes where additional structure about the autocorrelation function is known (beyond the scope of this course).



2.7.1 Ensemble and Time-Averages

New slide

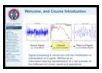
Ensemble averaging, as considered so far in the course, is not frequently used in practice since it is impractical to obtain the number of realisations needed for an accurate estimate.

A statistical average that can be obtained from a **single** realisation of a process is a **time-average**, defined by:

$$\langle g(x(n)) \rangle \triangleq \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{N} g(x(n)) \qquad \text{(M:3.3.32)}$$

For every ensemble average, a corresponding time-average can be defined; the time-average above corresponds to: $\mathbb{E}[g(x(n))]$.

Time-averages are random variables since they implicitly depend on the particular realisation, given by ζ. Averages of deterministic signals are fixed numbers or sequences, even though they are given by the same expression.



2.7.2 Ergodicity

New slide

A stochastic process, $x(n)$, is **ergodic** if its ensemble averages can be estimated from a single realisation of a process using time averages.

The two most important degrees of ergodicity are:

Mean-Ergodic (or ergodic in the mean) processes have identical expected values and sample-means:

$$\langle x(n) \rangle = \mathbb{E}[x(n)] \qquad \text{(M:3.3.34)}$$

Covariance-Ergodic Processes (or ergodic in correlation) have the property that:

$$\langle x(n)\, x^*(n-l) \rangle = \mathbb{E}[x(n)\, x^*(n-l)] \qquad \text{(M:3.3.35)}$$

Another form of ergodicity is a **distribution-ergodic** process, but this will not be discussed here.

- It should be intuitiveness obvious that ergodic processes must be stationary and, moreover, that a process which is ergodic both in the mean and correlation is WSS.

- WSS processes are not necessarily ergodic.
- Ergodic is often used to mean both ergodic in the mean and correlation.
- In practice, only finite records of data are available, and therefore an estimate of the time-average will be given by

$$\langle g(x(n))\rangle = \frac{1}{N}\sum_{n\in\mathcal{N}} g(x(n)) \tag{M:3.3.37}$$

where N is the number of data-points available. The performance of this estimator will be discussed elsewhere in this course.

2.7.3 More Details on Mean-Ergodicity

Returning to the definition of mean-ergodicity, a little more detail of conditions on the random process is given. Consider the **time-average** over $2N+1$ samples, $\{x(n)\}_{-N}^{N}$ is given by:

$$\mu_x|_N = \langle x(n)\rangle = \frac{1}{2N+1}\sum_{n=-N}^{N} x(n) \tag{2.70}$$

Clearly, $\mu_X|_N$ is a RV with mean:

$$\mathbb{E}\left[\mu_x|_N\right] = \frac{1}{2N+1}\sum_{n=-N}^{N}\mathbb{E}\left[x(n)\right] = \mu_x \tag{2.71}$$

since $x(n)$ is a stationary stochastic process. As will be seen in later lectures, this is known as an **unbiased estimate** since the **sample mean** is equal to the **ensemble mean**. Since $\mu_x|_N$ is a RV, then it must have a variance as well:

$$\text{var}\left[\mu_x|_N\right] = \text{var}\left[\frac{1}{2N+1}\sum_{n=-N}^{N} x(n)\right] \tag{2.72}$$

Noting the mean of the expression in the square brackets on the right hand side (RHS) is equal to μ_x, then:

$$\text{var}\left[\mu_x|_N\right] = \frac{1}{(2N+1)^2}\mathbb{E}\left[\sum_{n=-N}^{N}\sum_{m=-N}^{N} x(n)x^*(m)\right] - \mu_x^2 \tag{2.73}$$

$$= \frac{1}{(2N+1)^2}\left\{\sum_{n=-N}^{N}\sum_{m=-N}^{N} r_{xx}(n-m)\right\} - \mu_x^2 \tag{2.74}$$

since $x(n)$ is a stationary process, and therefore its autocorrelation function only depends on the time difference. With a little manipulation, then noting that the autocovariance is given by $\gamma_{xx}(l) = r_{xx}(l) - \mu_x^2$, it follows that:

$$\text{var}\left[\mu_x|_N\right] = \frac{1}{(2N+1)^2}\sum_{n=-N}^{N}\sum_{m=-N}^{N}\gamma_{xx}(n-m) \tag{2.75}$$

A change of variable can now be performed by setting $l = n - m$. Hence:

$$\text{var}\left[\mu_x|_N\right] = \frac{1}{(2N+1)^2}\sum_{n=-N}^{N}\sum_{l=n-N}^{n+N}\gamma_{xx}(l) \tag{2.76}$$

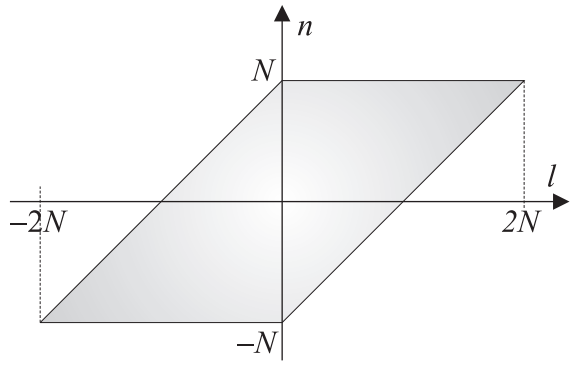


Figure 2.3: Region of summation for deriving the variance of the time-average.

The region of summation is shown in Figure 2.3.

Thus, the next step is to change the order of summation (as this is the usual trick), and so considering the region of summation, then summing l first:

$$\text{var}\left[\mu_x|_N\right] = \frac{1}{(2N+1)^2}\sum_{l=-2N}^{2N}\sum_{n=\max\{-N,l-N\}}^{\min\{N,l+N\}}\gamma_{xx}(l) \tag{2.77}$$

$$= \frac{1}{(2N+1)^2}\sum_{l=-2N}^{2N}(2N+1-|l|)\gamma_{xx}(l) \tag{2.78}$$

$$= \frac{1}{2N+1}\sum_{l=-2N}^{2N}\left(1-\frac{|l|}{2N+1}\right)\gamma_{xx}(l) \tag{2.79}$$

If the variance $\lim_{N\to\infty}\text{var}\left[\mu_x|_N\right] = 0$, then $\mu_x|_N \to \mu_x$ in the mean-square sense. In this case, it is said that the time average $\mu_x|_N$ computed from a single realisation of $x(n)$ is close to μ_x with probability close to 1. If this is true, then the technical definition is that the process $x(n)$ is **mean-ergodic**. The result presented above leads to the following conclusion:

Theorem 2.1 (Mean-ergodic processes). A discrete-random process $x(n)$ with autocovariance $\gamma_{xx}(l)$ is mean-ergodic iff:

$$\lim_{N\to\infty}\frac{1}{2N+1}\sum_{l=-2N}^{2N}\left(1-\frac{|l|}{2N+1}\right)\gamma_{xx}(l) = 0 \tag{2.80}$$

PROOF. See discussion above.

Example 2.7 ([Papoulis:1991, Example 13.3, Page 429]). A stationary stochastic process $x(n)$ has an autocovariance function given by $\gamma_{xx}(l) = q\,e^{-c\,|l|}$ for some constants q and l. Is the process $x(n)$ ergodic in the mean?

SOLUTION. Writing:

$$\text{var}\left[\mu_x|_N\right] = \frac{1}{2N+1}\sum_{l=-2N}^{2N}\left(1-\frac{|l|}{2N+1}\right)\gamma_{xx}(l) = \frac{q}{2N+1}\sum_{l=-2N}^{2N}\left(1-\frac{|l|}{2N+1}\right)e^{-c|l|} \tag{2.81}$$

which can be written as:

$$\text{var}\left[\mu_x|_N\right] = \frac{q}{2N+1}\left\{2\sum_{l=0}^{2N}\left(1-\frac{l}{2N+1}\right)e^{-cl}-1\right\} \tag{2.82}$$

Now, again noting the general result which can be found in mathematical tables [Gradshteyn:1994]:

$$\sum_{n=0}^{N-1}(a+nb)r^n = \frac{a-[a+(N-1)b]r^N}{1-r} + \frac{br(1-r^{N-1})}{(1-r)^2}, \quad r\neq 0,\ N>1 \tag{2.83}$$

then by setting $a=1$, $b=-\frac{1}{2N+1}$ and $r=e^{-c}$, with $n=l$ and $N\to 2N+1$:

$$\text{var}\left[\mu_x|_N\right] = 2q\left[\frac{\frac{1}{M}-\frac{1}{M^2}e^{-Mc}}{1-e^{-c}} + \frac{\frac{1}{M^2}e^{-c}-\frac{1}{M^2}e^{-Mc}}{(1-e^{-c})^2} - \frac{1}{2M}\right] \tag{2.84}$$

where $M=2N+1$. Now, by setting $N\to\infty$, which is equivalent to $M\to\infty$, and noting the relationship that:

$$\lim_{n\to\infty} n^s\, x^n \to 0 \quad \text{if } |x|<1 \text{ for any real value of } s \tag{2.85}$$

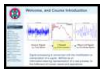
it can easily be seen that

$$\lim_{N\to\infty}\text{var}\left[\mu_x|_N\right] = 0 \tag{2.86}$$

□

and therefore $x(n)$ is mean-ergodic.

2.8 Joint Signal Statistics



New slide

Next, it is important to consider the dependence between two different random processes, and these follow similar definitions to those introduced for random vectors. In this section, consider the interaction between two random processes $x(n)$ and $y(n)$.

Cross-correlation and cross-covariance A measure of the dependence between values of two *different* stochastic processes is given by the **cross-correlation** and **cross-covariance** functions:

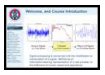
$$r_{xy}(n_1,n_2) = \mathbb{E}\left[x(n_1)\,y^*(n_2)\right] \tag{M:3.3.7}$$

$$\gamma_{xy}(n_1,n_2) = r_{xy}(n_1,n_2) - \mu_x(n_1)\,\mu_y^*(n_2) \tag{M:3.3.8}$$

Normalised cross-correlation (or cross-covariance) The cross-covariance provides a measure of similarity of the deviation from the respective means of two processes. It makes sense to consider this deviation relative to their **standard deviation**s; thus, **normalised cross-correlation**s:

$$\rho_{xy}(n_1,n_2) = \frac{\gamma_{xy}(n_1,n_2)}{\sigma_x(n_1)\,\sigma_y(n_2)} \tag{M:3.3.9}$$

2.8.1 Types of Joint Stochastic Processes



New slide

The definitions introduced earlier for a single stochastic process can be extended to the case of two joint stochastic processes:

Statistically independence of two stochastic processes occurs when, for every n_x and n_y,

$$f_{XY}\left(x, y \mid n_x, n_y\right) = f_X\left(x \mid n_x\right) f_Y\left(y \mid n_y\right) \tag{M:3.3.18}$$

Uncorrelated stochastic processes have, for all n_x & $n_y \neq n_x$:

$$\begin{aligned} \gamma_{xy}(n_x, n_y) &= 0 \\ r_{xy}(n_x, n_y) &= \mu_x(n_x)\,\mu_y(n_y) \end{aligned} \tag{M:3.3.19}$$

Joint stochastic processes that are statistically independent are uncorrelated, but not necessarily vice-versa, except for Gaussian processes. Nevertheless, a measure of uncorrelatedness is often used as a measure of independence. More on this later.

Further definitions include:

Orthogonal joint processes have, for every n_1 and $n_2 \neq n_1$:

$$r_{xy}(n_1, n_2) = 0 \tag{M:3.3.20}$$

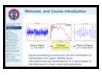
Joint WSS is a similar to WSS for a single stochastic process, and is useful since it facilitates a spectral description, as discussed later in this course:

$$r_{xy}(l) = r_{xy}(n_1 - n_2) = r^*_{yx}(-l) = \mathbb{E}\left[x(n)\, y^*(n-l)\right] \tag{2.87}$$

$$\gamma_{xy}(l) = \gamma_{xy}(n_1 - n_2) = \gamma^*_{yx}(-l) = r_{xy}(l) - \mu_x\,\mu^*_y \tag{2.88}$$

Joint-Ergodicity applies to two ergodic processes, $x(n)$ and $y(n)$, whose ensemble cross-correlation can be estimated from a time-average:

$$\langle x(n)\, y^*(n-l)\rangle = \mathbb{E}\left[x(n)\, y^*(n-l)\right] \tag{M:3.3.36}$$



New slide

2.9 Correlation Matrices for Random Processes

A stochastic process can also be represented as a random vector, and its second-order statistics given by the mean vector and the correlation matrix. Obviously these quantities are functions of the index n.

Let an M-dimensional random vector $\mathbf{X}(n, \zeta) \equiv \mathbf{X}(n)$ be derived from the random process $x(n)$ as follows:

$$\mathbf{X}(n) \triangleq \begin{bmatrix} x(n) & x(n-1) & \cdots & x(n-M+1) \end{bmatrix}^T \tag{M:3.4.56}$$

Then its mean is given by an M-vector

$$\boldsymbol{\mu}_{\mathbf{X}}(n) \triangleq \begin{bmatrix} \mu_x(n) & \mu_x(n-1) & \cdots & \mu_x(n-M+1) \end{bmatrix}^T \tag{M:3.4.57}$$

and the $M \times M$ correlation matrix is given by:

$$\mathbf{R}_{\mathbf{X}}(n) = \mathbb{E}\left[\mathbf{X}(n)\, \mathbf{X}^H(n)\right] \tag{T:4.23}$$

which can explicitly be written as:

$$\mathbf{R}_{\mathbf{X}}(n) \triangleq \begin{bmatrix} r_{xx}(n,n) & \cdots & r_{xx}(n,n-M+1) \\ \vdots & \ddots & \vdots \\ r_{xx}(n-M+1,n) & \cdots & r_{xx}(n-M+1,n-M+1) \end{bmatrix} \tag{M:3.4.58}$$

Clearly $\mathbf{R}_{\mathbf{X}}(n)$ is Hermitian, since $r_{xx}(n-i,n-j) = \mathbb{E}\left[x(n-i)x^*(n-j)\right] = r^*_{xx}(n-j,n-i)$, $0 \leq i,j \leq M-1$. This vector representation can be useful in discussion of optimum filters.

For stationary processes, the correlation matrix has an interesting additional structure. Note that:

1. $\mathbf{R}_{\mathbf{X}}(n)$ is a constant matrix $\mathbf{R}_{\mathbf{X}}$;
2. $r_{xx}(n-i,n-j) = r_{xx}(j-i) = r_{xx}(l)$, $l = j-i$;
3. conjugate symmetry gives $r_{xx}(l) = r^*_{xx}(-l)$.

Hence, the matrix $\mathbf{R}_{xx}$ is given by:

$$\mathbf{R}_{\mathbf{X}} \triangleq \begin{bmatrix} r_{xx}(0) & r_{xx}(1) & r_{xx}(2) & \cdots & r_{xx}(M-1) \\ r^*_{xx}(1) & r_{xx}(0) & r_{xx}(1) & \cdots & r_{xx}(M-2) \\ r^*_{xx}(2) & r^*_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(M-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r^*_{xx}(M-1) & r^*_{xx}(M-2) & r^*_{xx}(M-3) & \cdots & r_{xx}(0) \end{bmatrix} \tag{M:3.4.60}$$

It can easily be seen that $\mathbf{R}_{\mathbf{X}}$ is Hermitian and **Toeplitz**; a **Toeplitz** matrix is one in which the elements along each diagonal, parallel to the main diagonal, are equal. Note that the anti-diagonals are not necessarily equal. Thus, the autocorrelation matrix of a stationary process is Hermitian, nonnegative definite, and Toeplitz.

Example 2.8 (Correlation matrices). The correlation function for a certain random process $x[n]$ has the exponential form:

$$r_{xx}(l) = 4(-0.5)^{|l|} \tag{2.89}$$

Hence, the correlation matrix for $N = 3$ is given by:

$$\mathbf{R}_{\mathbf{X}} = \begin{bmatrix} r_{xx}(0) & r_{xx}(1) & r_{xx}(2) \\ r^*_{xx}(1) & r_{xx}(0) & r_{xx}(1) \\ r^*_{xx}(2) & r^*_{xx}(1) & r^*_{xx}(0) \end{bmatrix} \tag{2.90}$$

$$= \begin{bmatrix} 4(-0.5)^0 & 4(-0.5)^1 & 4(-0.5)^2 \\ 4(-0.5)^1 & 4(-0.5)^0 & 4(-0.5)^1 \\ 4(-0.5)^2 & 4(-0.5)^1 & 4(-0.5)^0 \end{bmatrix} = \begin{bmatrix} 4 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 4 \end{bmatrix} \tag{2.91}$$

⋈

which is clearly Toeplitz.

Note that the definition of a covariance matrix for a random process follows an almost identical form, except with the elements of the autocorrelation functions replaced by the autocovariance functions. Finally, note that is is possible to define a correlation or covariance matrix for a random vector that consists of non-consecutive samples from a random process. Hence, if

$$\mathbf{X}(\{n\}) \triangleq \begin{bmatrix} x(n_1) & x(n_2) & \cdots & x(n_M) \end{bmatrix}^T \tag{2.92}$$

where $\{n_k\}_1^M$ are unique arbitrary indices to samples from the random process, then the correlation matrix is still defined as:

$$\mathbf{R}_{\mathbf{X}}(\{n\}) = \mathbb{E}\left[\mathbf{X}(\{n\})\,\mathbf{X}^H(\{n\})\right] \tag{T:4.23}$$

2.10 Markov Processes

New slide

Finally, in this handout, a powerful model for a stochastic process known as a **Markov model** is introduced; such a process that satisfies this model is known as a **Markov process**. Quite simply, a Markov process is one in which the probability of any particular value in a sequence is dependent upon the preceding sample values. The simplest kind of dependence arises when the probability of any sample depends only upon the value of the *immediately preceding* sample, and this is known as a **first-order Markov process**. This simple process is a surprisingly good model for a number of practical signal processing, communications and control problems.

As an example of a Markov process, consider the process generated by the difference equation

$$x(n) = -a\,x(n-1) + w(n) \tag{T:3.17}$$

where a is a known constant, and $w(n)$ is a sequence of zero-mean i. i. d. Gaussian random variables with variance σ_W^2 density:

$$f_W\left(w(n)\right) = \frac{1}{\sqrt{2\pi\sigma_W^2}} \exp\left\{-\frac{w(n)^2}{2\sigma_W^2}\right\} \tag{T:3.18}$$

The conditional density of $x(n)$ given $x(n-1)$ is also Gaussian, and using the probability transformation rule for which the Jacobian evaluates to one, it can be shown that

$$f_X\left(x(n) \mid x(n-1)\right) = \frac{1}{\sqrt{2\pi\sigma_W^2}} \exp\left\{-\frac{(x(n) + ax(n-1))^2}{2\sigma_W^2}\right\} \tag{T:3.19}$$

In fact, if $w(n)$ is independent with *any* density $f_W\left(w(n)\right)$, the conditional density of $x(n)$ given $x(n-1)$ is $f_W\left(x(n) + a\,x(n-1)\right)$. Note that $x(n-1)$ completely determines the distribution for $x(n)$, and $x(n)$ completely determines the distribution for $x(n+1)$ and so forth. Thus, the value of the sequence at any time n_0 completely determines the distribution of $x(n)$ for any $n > n_0$. The following serves as a formal definition of a Markov process.

Definition 2.3 (Markov Process). A random process is a Pth-order Markov process if the distribution of $x(n)$, given the infinite past, depends only on the previous P samples $\{x(n-1), \ldots, x(n-P)\}$; that is, if:

$$f_X\left(x(n) \mid x(n-1),\, x(n-2),\, \ldots\right) = f_X\left(x(n) \mid x(n-1),\, \ldots,\, x(n-P)\right) \tag{T:3.20}$$

♢

Finally, it is noted that if $x(n)$ takes on a countable (discrete) set of values, a Markov random process is called a **Markov chain**. This will always be the case in digital signal processing since the values of the random sequence are represented with a finite number of bits. There is a tremendous volume of results on Markov chains, but they will not presently be covered in this course.

3

Frequency-Domain Description of Stationary Processes

Introduces the notion of a frequency-domain description of stationary random processes, defining the power spectral density (PSD) as the Fourier transform of the autocorrelation function. Considers the properties of the PSD including the PSD of harmonic processes. Defines the cross-PSD and the complex spectral density.

3.1 Introduction to the power spectral density

New slide

Frequency- and transform-domain methods including the Fourier-transform and z-transform are very powerful tools for the analysis of deterministic sequences. It seems natural to extend these techniques to analysis stationary **random processes**.

So far in this course, **stationary stochastic processes** have been considered in the time-domain through the use of the **autocorrelation sequence (ACS)**. Since the ACS for a stationary process is a function of a single-discrete time process, then the question arises as to what the discrete-time Fourier transform (DTFT) corresponds to. It turns out to be known as the **power spectral density (PSD)** of a stationary random process, and the PSD is an extremely powerful and conceptually appealing tool in statistical signal processing. This handout will study the PSD in some detail.

In signal theory for deterministic signals, spectra are used to represent a function as a superposition of exponential functions. For random signals, the notion of a spectrum has two interpretations:

Transform of averages The first involves transform of averages (or moments). As will be seen, this will be the Fourier transform of the autocorrelation function.

Stochastic decomposition The second interpretation represents a stochastic process as a superposition of exponentials, where the coefficients are themselves random variables. Hence, a stochastic process $x[n]$ can be represented as:

$$x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega})\, e^{j\omega n}\, d\omega, \quad n \in \mathbb{R} \tag{3.1}$$

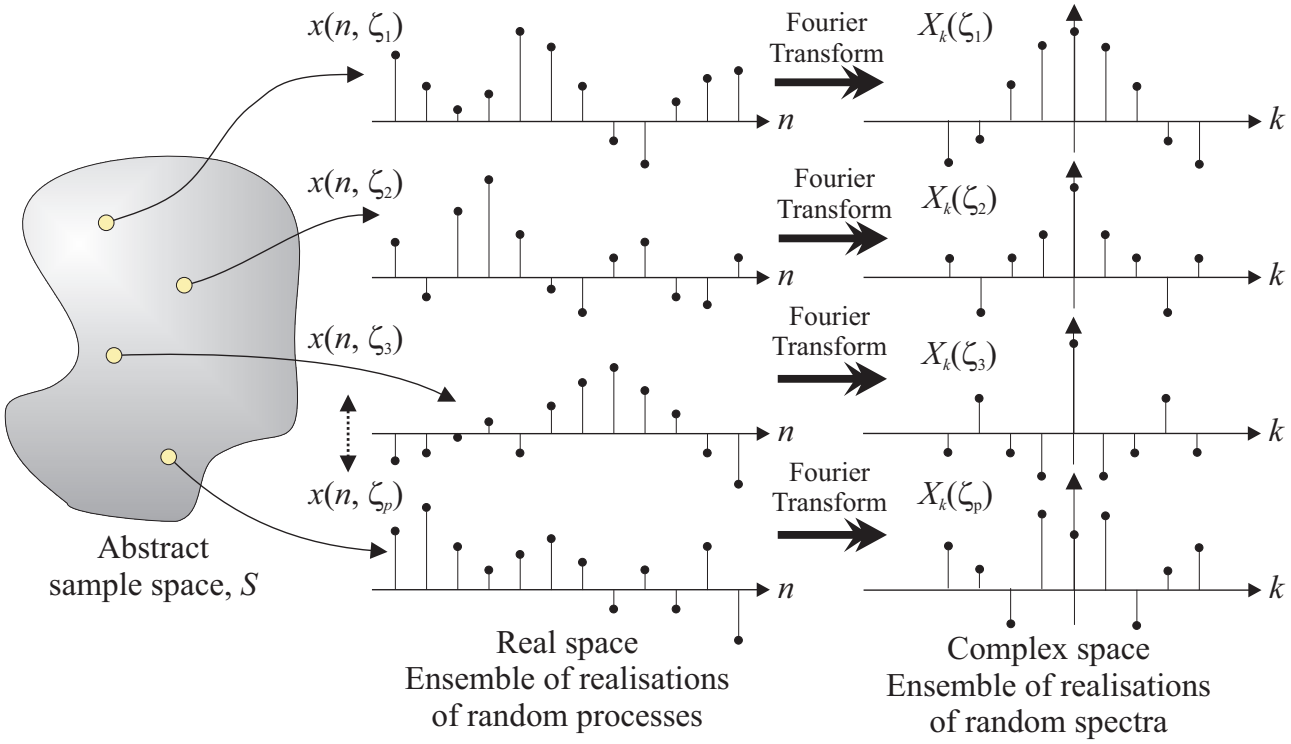


Figure 3.1: A graphical respresentation of random spectra.

where $X(e^{j\omega})$ is a random variable for a given value of ω. Alternatively, $X(e^{j\omega})$ can be considered as a continuous random-process, as a function of ω. This interpretation is extremely powerful, and can in fact be extended to the superposition of any set of **basis functions**; the Karhunen-Loeve (KL) transform is an example of such a decomposition. Unfortunately, there is not time in this course to consider this spectral representation, extremely interesting as it is.

3.2 Motivating the power spectral density

It is important to appreciate that most realisations of random signals, $x[n,\, \zeta]$, do not have finite energy, as they usually don't decay away as $n \to \pm\infty$. Therefore, technically, they do not possess a corresponding DTFT, and hence it is not possible simply to take the DTFT of a random signal. Noting that a random signal is actually an ensemble of realisations, each realisation occuring with a different probability, it is, in any case, somewhat meaningless to take the DTFT of a random process. It should also be remembered that the DTFT of a particular observed realisation, even if it existed, is itself a **random process**, albeit as a function of frequency rather than time. Therefore, it is necessary to take an alternative perspective, as discussed in this section.

Motivated by the stochastic decomposition in Equation 3.1, and restricting the analsysis to wide-sense stationary (WSS) processes, consider the random variable (RV), $X(e^{j\omega})$, resulting from the DTFT of a random signal, $x[n]$:

$$X\left(e^{j\omega}\right) = \sum_{n=-\infty}^{\infty} x[n]\, e^{-j\omega n} \tag{3.2}$$

It is of interest to consider the **total power** in the rv, $X\left(e^{j\omega}\right)$, which is given by the second moment:

$$P_{XX}\left(e^{j\omega}\right) = \mathbb{E}\left[\left|X\left(e^{j\omega}\right)\right|^2\right] \tag{3.3}$$

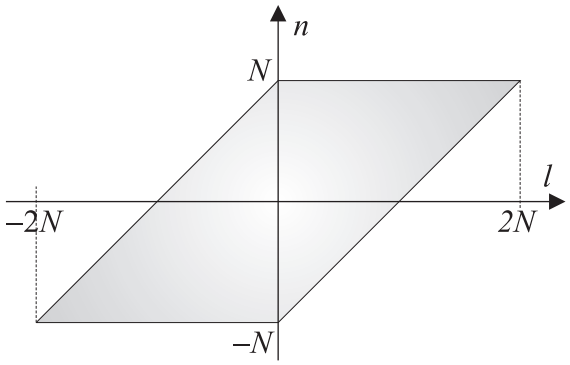


Figure 3.2: Region of summation for deriving the variance of the time-average.

Since random signals are not finite energy, then this expression will diverge, so consider instead the definition:

$$P_{XX}\left(e^{j\omega}\right) = \lim_{N\to\infty} \frac{1}{2N+1}\mathbb{E}\left[\left|X_N\left(e^{j\omega}\right)\right|^2\right] \tag{3.4}$$

where $X_N\left(e^{j\omega}\right)$ is the truncated Fourier transform of $x[n]$, or basically a **windowed** version of the sequence between $-N$ and N, as given by:

$$X_N\left(e^{j\omega}\right) \triangleq \sum_{n=-N}^{N} x[n]\, e^{-j\omega n} = \sum_{n=-\infty}^{\infty} w[n]\, x[n]\, e^{-j\omega n} \tag{3.5}$$

where $w[n]$ is the window function:

$$w[n] = \begin{cases} 1 & -N \le n \le N \\ 0 & \text{otherwise} \end{cases} \tag{3.6}$$

Then, substituting Equation 3.5 into Equation 3.4 and rearranging gives:

$$P_{XX}\left(e^{j\omega}\right) = \lim_{N\to\infty} \frac{1}{2N+1}\mathbb{E}\left[\sum_{n=-N}^{N} x[n]\, e^{-j\omega n} \sum_{m=-N}^{N} x^*[m]\, e^{j\omega m}\right] \tag{3.7}$$

$$= \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N}\sum_{m=-N}^{N} \mathbb{E}\left[x[n]\, x^*[m]\right] e^{-j\omega(n-m)} \tag{3.8}$$

Substitute the variable $\ell = n - m$, such that when $m = \pm N$, then $\ell = n \mp N$. Since the summation is over integers, which means that $\sum_a^b(\cdot) = \sum_b^a(\cdot)$, and noting that for WSS processes, $\mathbb{E}\left[x[n]\, x^*[n-\ell]\right] = r_{xx}[\ell]$ this means Equation 3.8 becomes:

$$P_{XX}\left(e^{j\omega}\right) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N}\sum_{\ell=n-N}^{n+N} r_{xx}[\ell]\, e^{-j\omega\ell} \tag{3.9}$$

The region of summation is shown in Figure 3.2. Changing the order of summation (as this is the usual trick), to sum over ℓ first, then it can be seen that ℓ varies from $-2N$ to $2N$, while n will vary from $\max\{-N, \ell - N\}$ to $\min\{N, \ell + N\}$. Hence, Equation 3.9 becomes:

$$P_{XX}\left(e^{j\omega}\right) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{\ell=-2N}^{2N} \sum_{n=\max\{-N,\,\ell-N\}}^{\min\{N,\,\ell+N\}} r_{xx}[\ell]\, e^{-j\omega\ell} \tag{3.10}$$

$$P_{XX}\left(e^{j\omega}\right) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{\ell=-2N}^{2N} r_{xx}[\ell]\, e^{-j\omega\ell} \left[\sum_{n=\max\{-N,\,\ell-N\}}^{\min\{N,\,\ell+N\}} 1 \right] \tag{3.11}$$

The second summation in the square brackets can be shown by, simple counting, to simplify to $2N + 1 - |\ell|$, and therefore:

$$P_{XX}\left(e^{j\omega}\right) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{\ell=-2N}^{2N} (2N+1-|\ell|)\, r_{xx}[\ell]\, e^{-j\omega\ell} \tag{3.12}$$

$$= \sum_{\ell=-\infty}^{\infty} r_{xx}[\ell]\, e^{-j\omega\ell} - \lim_{N\to\infty} \sum_{\ell=-2N}^{2N} \frac{|\ell|}{2N+1} r_{xx}[\ell]\, e^{-j\omega\ell} \tag{3.13}$$

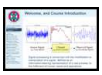
Assuming the mild assumption that the autocorrelation sequence $r_{xx}[\ell]$ decays sufficiently rapidly such that:

$$\lim_{N\to\infty} \sum_{\ell=-2N}^{2N} |\ell|\, |r_{xx}[\ell] = 0 \tag{3.14}$$

then Equation 3.13 simplifies to:

$$P_{XX}\left(e^{j\omega}\right) = \sum_{\ell=-\infty}^{\infty} r_{xx}[\ell]\, e^{-j\omega\ell} \tag{3.15}$$

Hence, $P_{XX}\left(e^{j\omega}\right)$ can be viewed as the average power, or energy, of the Fourier transform of a random process at frequency ω. Clearly, this gives an indication of whether, *on average*, there are dominant frequencies present in the realisations of $x[n]$.



New slide

3.3 The power spectral density

The discrete-time Fourier transform of the autocorrelation sequence of a stationary stochastic process $x[n, \zeta]$ is known as the **power spectral density (PSD)**, is denoted by $P_{xx}(e^{j\omega})$, and is given by:

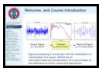
$$P_{xx}(e^{j\omega}) = \sum_{\ell\in\mathbb{Z}} r_{xx}[\ell]\; e^{-j\omega\ell} \tag{M:3.3.39}$$

where ω is frequency in radians per sample.

The autocorrelation sequence, $r_{xx}[\ell]$, can be recovered from the **PSD** by using the inverse-**DTFT**:

$$r_{xx}[\ell] = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_{xx}(e^{j\omega})\, e^{j\omega\ell}\, d\omega, \quad \ell \in \mathbb{Z} \tag{M:3.3.41}$$

Sometimes the PSD is called the auto-PSD to distinguish it from the cross-PSD introduced in Section 3.4. In the case that $r_{xx}[\ell]$ is periodic, corresponding to a wide-sense periodic stochastic process, then the power spectral density is defined as the discrete Fourier transform of the autocorrelation sequence. This natural extension is easily obtained once the aperiodic-case is considered in depth.



New slide

3.3.1 Properties of the power spectral density

There are a number of properties of the power spectral density that follow from the corresponding properties of the autocorrelation sequence, and the discrete-time Fourier transform.

- $P_{xx}(e^{j\omega}) : \omega \to \mathbb{R}^+$; in otherwords, the PSD is real valued, and nonnegative definite. i.e.

$$P_{xx}(e^{j\omega}) \geq 0 \tag{M:3.3.44}$$

  This property follows from the positive semi-definiteness of the autocorrelation sequence.

- $P_{xx}(e^{j\omega}) = P_{xx}(e^{j(\omega+2n\pi)})$; in otherwords, the PSD is periodic with period 2π.

- If $x[n]$ is real-valued, then:

  - $r_{xx}[\ell]$ is real and even;
  - $P_{xx}(e^{j\omega}) = P_{xx}(e^{-j\omega})$ is an even function of ω.

- The area under $P_{xx}(e^{j\omega})$ is nonnegative and is equal to the average power of $x[n]$. Hence:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} P_{xx}(e^{j\omega})\, d\omega = r_{xx}[0] = \mathbb{E}\left[|x[n]\,|^2\right] \geq 0 \tag{M:3.3.45}$$

Example 3.1 ([Manolakis:2001, Example 3.3.4, Page 109]). Determine the PSD of a zero-mean WSS process $x(n)$ with $r_{xx}(l) = a^{|l|}$, $-1 < a < 1$.

SOLUTION. Using the definition of the PSD directly, then:

$$\begin{aligned} P_{xx}(e^{j\omega}) &= \sum_{l\in\mathbb{Z}} r_{xx}(l)\, e^{-j\omega l} && (3.16)\\ &= \sum_{l\in\mathbb{Z}} a^{|l|}\, e^{-j\omega l} && (3.17)\\ &= \sum_{l=0}^{\infty} \left(a\, e^{-j\omega}\right)^l + \sum_{l=0}^{\infty} \left(a\, e^{j\omega}\right)^l - 1 && (3.18) \end{aligned}$$

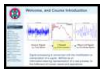
which, by using the expressions for geometric series, gives:

$$\begin{aligned} P_{xx}(e^{j\omega}) &= \frac{1}{1-a\, e^{-j\omega}} + \frac{1}{1-a\, e^{j\omega}} - 1 && \text{(M:3.3.42)}\\ &= \frac{1-a^2}{1-2a\cos\omega + a^2} && (3.19) \end{aligned}$$

□

which is a real-valued, even, and nonnegative function of ω.

3.3.2 General form of the PSD



New slide

A process, $x[n]$, and therefore its corresponding autocorrelation sequence (ACS), $r_{xx}[\ell]$, can always be decomposed into a zero-mean aperiodic component, $r_{xx}^{(a)}[\ell]$, and a non-zero-mean periodic component, $r_{xx}^{(p)}[\ell]$:

$$r_{xx}[\ell] = r_{xx}^{(a)}[\ell] + r_{xx}^{(p)}[\ell] \tag{3.20}$$

Theorem 3.1 (PSD of a non-zero-mean process with periodic component). The most general definition of the PSD for a non-zero-mean stochastic process with a periodic component is given by:

$$P_{xx}(e^{j\omega}) = P_{xx}^{(a)}(e^{j\omega}) + \frac{2\pi}{K}\sum_{k\in\mathcal{K}} P_{xx}^{(p)}(k)\,\delta\left(\omega-\omega_k\right) \tag{T:4.41}$$

The term $P_{xx}^{(a)}(e^{j\omega})$ is the DTFT of the aperiodic component $r_{xx}^{(a)}[\ell]$, while $P_{xx}^{(p)}(k)$ are the discrete Fourier transform (DFT) coefficients for the periodic component $r_{xx}^{(p)}[\ell]$ assuming a periodicity of length K, and where $\omega_k = \frac{2\pi k}{K}$.

Moreover, it can be seen that $P_{xx}^{(a)}(e^{j\omega})$ represents the continuous part of the spectrum, while the sum of weighted impulses represent the discrete part or *lines* of the spectrum.

PROOF. The non-zero-mean periodic component, $r_{xx}^{(p)}(l)$ can itself be decomposed using a **discrete Fourier transform**:

$$r_{xx}^{(p)}(l) = \frac{1}{K}\sum_{k\in\mathcal{K}} P_{xx}^{(p)}(k)\,e^{j\omega_k l} \tag{3.21}$$

where $\mathcal{K} = \{0,\ldots,K-1\}$, and $\omega_k = \frac{2\pi}{K}k$. Thus, the **PSD** of $X(\zeta)$, becomes:

$$P_{xx}(e^{j\omega}) = P_{xx}^{(a)}(e^{j\omega}) + \frac{1}{K}\sum_{\ell\in\mathbb{Z}}\sum_{k\in\mathcal{K}} P_{xx}^{(p)}(k)\,e^{j\omega_k\ell}\,e^{-j\omega\ell} \tag{3.22}$$

As usual, change the order of summation:

$$= P_{xx}^{(a)}(e^{j\omega}) + \frac{1}{K}\sum_{k\in\mathcal{K}} P_{xx}^{(p)}(k)\sum_{\ell\in\mathbb{Z}} e^{-j\ell(\omega-\omega_k)} \tag{3.23}$$

$$= P_{xx}^{(a)}(e^{j\omega}) + \frac{2\pi}{K}\sum_{k\in\mathcal{K}} P_{xx}^{(p)}(k)\,\delta\left(\omega-\omega_k\right) \tag{3.24}$$

where **Poisson's formula**, which can be derived by writing down the Fourier series for an impulse train, is used:

$$\sum_{n=-\infty}^{\infty}\delta\left(t-nT\right) = \frac{1}{T}\sum_{\ell=-\infty}^{\infty} e^{-j\ell\omega_0 t} \tag{3.25}$$

where $\omega_0 = \frac{2\pi}{T}$. Thus, by letting $T = 2\pi$, and $t = \omega - \omega_k$, then:

$$2\pi\sum_{n=-\infty}^{\infty}\delta\left(\omega-\omega_k-2\pi n\right) = \sum_{\ell=-\infty}^{\infty} e^{-j\ell(\omega-\omega_k)} \tag{3.26}$$

Since $-2\pi < \omega_k \leq 2\pi$, and $P_{xx}(e^{j\omega})$ is periodic in ω with period 2π, then it is sufficient to write for $|\omega| \leq 2\pi$, that:

$$2\pi\delta\left(\omega-\omega_k\right) = \sum_{l=-\infty}^{\infty} e^{-jl(\omega-\omega_k)} \tag{3.27}$$

which can be substituted to give the desired result. □

Example 3.2 ([Manolakis:2001, Harmonic Processes, Page 110-111]). Determine the PSD of the **harmonic process** introduced in the previous handout and defined by:

$$x[n] = \sum_{k=1}^{M} A_k \cos(\omega_k n + \phi_k), \quad \omega_k \neq 0 \tag{M:3.3.50}$$

where M, $\{A_k\}_1^M$ and $\{\omega_k\}_1^M$ are constants, and $\{\phi_k\}_1^M$ are pairwise independent and identically distributed (i. i. d.) RVs uniformly distributed in the interval $[0, 2\pi]$.

SOLUTION. As shown in the previous handout, $x[n]$ is a stationary process with zero-mean, and autocorrelation sequence (ACS):

$$r_{xx}[\ell] = \frac{1}{2} \sum_{k=1}^{M} |A_k|^2 \cos \omega_k \ell, \quad -\infty < \ell < \infty \tag{M:3.3.52}$$

Note that $r_{xx}[\ell]$ consists of a sum of *in-phase* cosines with the same frequencies as in $x[n]$. By writing

$$\cos \omega_k \ell = \frac{e^{j\omega_k \ell} + e^{-j\omega_k \ell}}{2} \tag{3.28}$$

then Equation M:3.3.52 may be written as:

$$\begin{aligned} r_{xx}[\ell] &= \frac{1}{4} \sum_{k=1}^{M} |A_k|^2 \left(e^{j\omega_k \ell} + e^{-j\omega_k \ell} \right) \\ &= \sum_{k=1}^{M} \frac{|A_k|^2}{4} e^{j\omega_k \ell} + \sum_{k=1}^{M} \frac{|A_k|^2}{4} e^{-j\omega_k \ell} \\ &= \sum_{k=1}^{M} \frac{|A_k|^2}{4} e^{j\omega_k \ell} + \sum_{\hat{k}=-1}^{-M} \frac{|A_{-\hat{k}}|^2}{4} e^{-j\omega_{-\hat{k}} \ell} \end{aligned} \tag{3.29}$$

Hence, the ACS can be written as:

$$r_{xx}[\ell] = \sum_{k=-M}^{M} \frac{|A_k|^2}{4} e^{j\omega_k \ell}, \quad -\infty < \ell < \infty \tag{3.30}$$

where the following are defined: $A_0 = 0$, $A_k = A_{-k}$, and $\omega_{-k} = -\omega_k$.

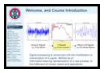
Hence, it directly follows using the results above that:

$$P_{xx}(e^{j\omega}) = 2\pi \sum_{k=-M}^{M} \frac{|A_k|^2}{4} \delta(\omega - \omega_k) = \frac{\pi}{2} \sum_{k=-M}^{M} |A_k|^2 \delta(\omega - \omega_k) \tag{3.31}$$

□

The harmonic process is predictable because any given realisation is a sinusoidal sequence with fixed amplitude, frequency and phase. The independence and uniform distribution of the phase, however, is strictly required to ensure the stationarity of the process $x[n]$.

3.4 The cross-power spectral density



New slide

The cross-power spectral density (CPSD) of two jointly stationary stochastic processes, $x[n]$ and $y[n]$, provides a description of their statistical relations in the frequency domain. It is defined, naturally, as the DTFT of the cross-correlation, $r_{xy}(\ell) \triangleq \mathbb{E}\left[x(n)\, y^*(n-\ell)\right]$:

$$P_{xy}(e^{j\omega}) = \mathcal{F}\{r_{xy}(\ell)\} = \sum_{l\in\mathbb{Z}} r_{xy}(\ell)\, e^{-j\omega\ell} \tag{M:3.3.56}$$

The cross-correlation $r_{xy}(l)$ can be recovered by using the inverse-DTFT:

$$r_{xy}(l) = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_{xy}(e^{j\omega})\, e^{j\omega l}\, d\omega, \quad l \in \mathbb{R} \tag{M:3.3.57}$$

Since this integral is essentially a summation, then an interpretation that can be given to the cross-spectrum is that $P_{xy}(e^{j\omega_0})$ measures the correlation between two RVs at a given frequency ω_0.

The cross-spectrum $P_{xy}(e^{j\omega})$ is, in general, a complex function of ω.

Some properties of the CPSD and related definitions include:

1. $P_{xy}(e^{j\omega})$ is periodic in ω with period 2π.

2. Since $r_{xy}(l) = r^*_{yx}(-l)$, then it follows:

$$P_{xy}(e^{j\omega}) = P^*_{yx}(e^{j\omega}) \tag{M:3.3.58}$$

   Thus, $P_{xy}(e^{j\omega})$ and $P_{yx}(e^{j\omega})$ have the same magnitude, but opposite phase.

3. If the process $x(n)$ is real, then $r_{xy}(l)$ is real, and:

$$P_{xy}(e^{j\omega}) = P^*_{xy}(e^{-j\omega}) \tag{3.32}$$

4. The normalised cross-correlation, or **coherence function**, is given by:

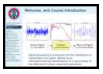
$$\Gamma_{xy}(e^{j\omega}) \triangleq \frac{P_{xy}(e^{j\omega})}{\sqrt{P_{xx}(e^{j\omega})}\sqrt{P_{yy}(e^{j\omega})}} \tag{M:3.3.59}$$

   Its squared magnitude is known as the magnitude square coherence (MSC) function.

$$\left|\Gamma_{xy}(e^{j\omega})\right|^2 = \frac{\left|P_{xy}(e^{j\omega})\right|^2}{P_{xx}(e^{j\omega})P_{yy}(e^{j\omega})} \tag{3.33}$$

   If $y[n] = x[n]$, then $\Gamma_{xy}(e^{j\omega})$, corresponding to maximum correlation, whereas if $x[n]$ and $y[n]$ are uncorrelated, then $r_{xy}[\ell] = 0$, and therefore $\Gamma_{xy}(e^{j\omega}) = 0$. Hence:

$$0 \leq |\Gamma_{xy}(e^{j\omega})|^2 \leq 1 \tag{M:3.3.60}$$



3.5 Complex Spectral Density Functions

New slide

The analysis of discrete-deterministic signals is also performed through the the z-transform and, therefore, in addition to using the Fourier transform, it is also very important to analyse stationary random processes using this transform; it is a perfectly natural extension.

The second moment quantities that described a random process in the z-transform domain are known as the **complex spectral density** and **complex cross-spectral density** functions. The PSD and CPSD functions discussed previously can be considered as special cases of the complex spectral density functions when the latter are evaluated on the unit circle.

If the sequences $r_{xx}(l)$ and $r_{xy}(l)$ are absolutely summable within a certain ring of the complex z-plane, then their z-transforms exist. Hence, $r_{xx}(l) \overset{z}{\rightleftharpoons} P_{xx}(z)$ and $r_{xy}(l) \overset{z}{\rightleftharpoons} P_{xy}(z)$, where:

$$P_{xx}(z) = \sum_{l\in\mathbb{Z}} r_{xx}(l)\, z^{-l} \tag{M:3.3.61}$$

$$P_{xy}(z) = \sum_{l\in\mathbb{Z}} r_{xy}(l)\, z^{-l} \tag{M:3.3.62}$$

Note that these are bilateral z-transforms. If the unit circle, defined by $z = e^{j\omega}$ is within the region of convergence of these summations, then:

$$P_{xx}(e^{j\omega}) = P_{xx}(z)\big|_{z=e^{j\omega}} \tag{M:3.3.63}$$

$$P_{xy}(e^{j\omega}) = P_{xy}(z)\big|_{z=e^{j\omega}} \tag{M:3.3.64}$$

The inverse of the complex spectral and cross-spectral densities are given by the contour integral:

$$r_{xx}(l) = \frac{1}{2\pi j}\oint_C P_{xx}(z)\, z^{l-1}\, dz \tag{3.34}$$

$$r_{xy}(l) = \frac{1}{2\pi j}\oint_C P_{xy}(z)\, z^{l-1}\, dz \tag{3.35}$$

where the contour of integration C is to be taken counterclockwise and in the region of convergence. In practice, these integrals are usually never performed, and tables, instead, are used.

Some properties of the complex spectral densities include:

1. Conjugate-symmetry:

$$P_{xx}(z) = P^*_{xx}(1/z^*) \quad \text{and} \quad P_{xy}(z) = P^*_{xy}(1/z^*) \tag{3.36}$$

2. For the case when $x(n)$ is real, then:

$$P_{xx}(z) = P_{xx}(z^{-1}) \tag{3.37}$$

The possible existence of lines in the PSD function due to a periodic component of the random process, as discussed in Section 3.3.2, poses some mathematical problems in defining the complex spectral density function since the z-transform does not exist. Try, for example, finding the complex spectral density of a function with a non-zero mean. In the case of the PSD, these functions were included formally as impulses in frequency. A similar approach to that in Equation T:4.41 is used here, and the complex spectral density function is written as:

$$P_{xx}(z) = P^{(a)}_{xx}(z) + 2\pi \sum_{k\in\mathcal{K}} P^{(p)}_{xx}(k)\, \delta\left(z - e^{j\omega_k}\right) \tag{3.38}$$

where $P^{(a)}_{xx}(z)$ corresponds to the aperiodic component of the autocorrelation function, and the second summation term denotes the line spectra.

Sidebar 5 Equivalent Expressions

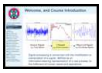
A variety of equivalent expressions can result from some simple manipulations; thus, other tables of z-transforms may appear to list different results, but are actually equivalent. Some examples include:

$$x(n) = \begin{cases} a^{\left|\frac{n}{2}\right|} & n \in \{0, \text{ even}\} \\ 0 & \text{for } n \text{ odd} \end{cases}$$
$$\overset{z}{\rightleftharpoons} \frac{1}{1-az^2} + \frac{az^{-2}}{1-az^{-2}} = \left\{\frac{az^2}{1-az^2} + 1\right\} + \left\{\frac{1}{1-az^{-2}} - 1\right\}$$
$$= \frac{az^2}{1-az^2} + \frac{1}{1-az^{-2}}$$

and

$$x(n) = a^{|n|} \overset{z}{\rightleftharpoons} \frac{1}{1-az^{-1}} + \frac{az}{1-az} = \left\{\frac{1}{1-az^{-1}} - 1\right\} + \left\{\frac{az}{1-az} + 1\right\}$$
$$= \frac{az^{-1}}{1-az^{-1}} + \frac{1}{1-az}$$

The fact that there are so many equivalent expressions means that sometimes it can be difficult to find the exact transform relation in tables. The particular form of the z-transform that needs to be inverted can vary depending on how it is calculated.



New slide

3.6 Table of bilateral z-transforms

The **bilateral z-transform** is defined by the following pairs of equations:

$$X(z) \triangleq \mathcal{Z}[x(n)] = \sum_{n=-\infty}^{\infty} x(n)\, z^{-n} \tag{M:2.2.29}$$

$$x(n) = \frac{1}{2\pi j} \oint_C X(z)\, z^{n-1}\, dz \tag{M:2.2.30}$$

In the following table, it is assumed that $|a| \leq 1$. It is important to note that this is a crucial condition, as it will distinguish signals that exist only for $n \geq 0$ and those for $x < 0$. To use these tables, it is crucial to match an expression with an identity exactly, otherwise the incorrect inverse transform might accidentally be used.

For the purposes of the table, recall that $u(n)$ is the **discrete-time step function** given by:

$$u(n) = \begin{cases} 1 & n \geq 0 \\ 0 & n < 0 \end{cases} \tag{3.39}$$

The region of convergence (ROC) is also shown for completeness, although it is usual to assume that z is only considered within the ROC. Note that if the signal $x(n) = 0$ for $n < 0$, it is known as a **causal sequence**, and if $x(n) = 0$ for $n > 0$, it is known as an **anticausal sequence**.

Notes	$x[n]$	$X(z)$	ROC
$x[n]=0, n<0$	$u[n]$	$\dfrac{1}{1-z^{-1}} \equiv \dfrac{z}{z-1}$	$\lvert z\rvert > 1$
$x[n]=0, n>0$	$u[-n]$	$\dfrac{1}{1-z}$	$\lvert z\rvert < 1$
$x[n]=0, n<0$	$a^n\,u[n]$	$\dfrac{z}{z-a} \equiv \dfrac{1}{1-az^{-1}}$	$\lvert z\rvert > \lvert a\rvert$
$x[n]=0, n\leq 0$	$a^n\,u[n-1]$	$\dfrac{a}{z-a} \equiv \dfrac{az^{-1}}{1-az^{-1}}$	$\lvert z\rvert > \lvert a\rvert$
$x[n]=0, n>0$	$a^{-n}\,u[-n]$	$\dfrac{1}{1-az} \equiv \dfrac{z^{-1}}{z^{-1}-a}$	$\lvert z\rvert < \dfrac{1}{\lvert a\rvert}$
$x[n]=0, n\geq 0$	$a^{-n}\,u[-n-1]$	$\dfrac{az}{1-az} \equiv \dfrac{a}{z^{-1}-a}$	$\lvert z\rvert < \dfrac{1}{\lvert a\rvert}$
$x[n]=0, n<0$	$na^n\,u[n]$	$\dfrac{az^{-1}}{\left(1-az^{-1}\right)^2}$	$\lvert z\rvert > \lvert a\rvert$
$x[n]=0, n\geq 0$	$-na^{-n}\,u[-n-1]$	$\dfrac{az}{\left(1-az\right)^2}$	$\lvert z\rvert < \dfrac{1}{\lvert a\rvert}$
See note 3	$\begin{cases} a^{\left\lvert\frac{n}{2}\right\rvert} & n\in\{0,\text{ even}\} \\ 0 & \text{for } n \text{ odd}\end{cases}$	$\dfrac{1}{1-az^2}+\dfrac{az^{-2}}{1-az^{-2}}$ or $\dfrac{1-a^2}{\left(1-az^2\right)\left(1-az^{-2}\right)}$	$\lvert a\rvert^{\frac{1}{2}} < \lvert z\rvert < \dfrac{1}{\lvert a\rvert^{\frac{1}{2}}}$
	$\begin{cases} a^{\left\lvert\frac{n}{2}\right\rvert+\frac{1}{2}} & \text{for } n \text{ odd} \\ 0 & \text{otherwise}\end{cases}$	$\dfrac{az}{1-az^2}+\dfrac{az^{-1}}{1-az^{-2}}$	$\lvert a\rvert^{\frac{1}{2}} < \lvert z\rvert < \dfrac{1}{\lvert a\rvert^{\frac{1}{2}}}$
See notes 1 , 3	$a^{\lvert n\rvert}$	$\dfrac{1}{1-az^{-1}}+\dfrac{az}{1-az}$ or $\dfrac{1-a^2}{\left(1-az\right)\left(1-az^{-1}\right)}$	$\lvert a\rvert < \lvert z\rvert < \dfrac{1}{\lvert a\rvert}$
See note 2	$\lvert n\rvert a^{\lvert n\rvert}$	$\dfrac{az^{-1}}{\left(1-az^{-1}\right)^2}+\dfrac{az}{\left(1-az\right)^2}$	$\lvert a\rvert < \lvert z\rvert < \dfrac{1}{\lvert a\rvert}$

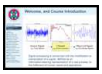
Notes:

1. This identity follows since $a^{|n|} \equiv a^n\,u[n] + a^{-n}\,u[-n-1]$.
2. Similarly, note that $|n|a^{|n|} = na^n\,u[n] - na^{-n}\,u[-n-1]$.
3. Note other similar expressions result, as shown in Sidebar 5.

4

Linear Systems with Stationary Random Inputs

Considers the concept of applying a stochastic signal to the input of a system and determining the resulting output. Looks at the special case of linear time-invariant (LTI) systems with stationary inputs. Analysis by looking at the input and output statistics, as well as the input-output joint-statistics. Discusses system identification using cross-correlation. Provides examples for systems with rationale transfer functions (using time domain analysis by solving difference equations and frequency domain analysis).



New slide

4.1 Systems with Stochastic Inputs

Signal processing involves the transformation of signals to enhance certain characteristics; for example, to suppress noise, or to extract meaningful information. This handout considers the processing of random processes by systems, and in particular linear systems.

What does it mean to apply a stochastic signal to the input of a system? This question is an interesting one since a stochastic process is not just a single sequence but an ensemble of sequences.

If the system is a general nonlinear possibly time-varying system, then one approach of expressing the relationship is as follows: Given a stochastic process $x(n, \zeta)$, assign according to some rule to each of its realisations $x[n, \zeta_k]$ a function $y[n, \zeta_k]$. Thus, another process has been created in which:

$$y[n] = T\left[x[n]\right] \tag{4.1}$$

whose realisations are the functions $\{y[n, \zeta_k]\}$. This process $y[n]$ so formed can be considered as the output of a **system** or transformation with, as its input, the process $x[n]$. The system is completely specified in terms of the transformation function (or operator) T; that is, the rule of correspondence between the samples of the input $x[n]$ and the output $y[n]$. This relationship is indicated in Figure 4.1.

In principle, the statistics of the output of a system can be expressed in terms of the statistics of the input. However, in general this is a complicated problem except in special cases. A special case is that of *linear systems*, and this is considered in the next section. In particular, if the input is a stationary stochastic process, and the system linear time-invariant (LTI), the statistics are even simpler. Moreover, it leads to a slightly simpler and intuitive explanation for the response of the

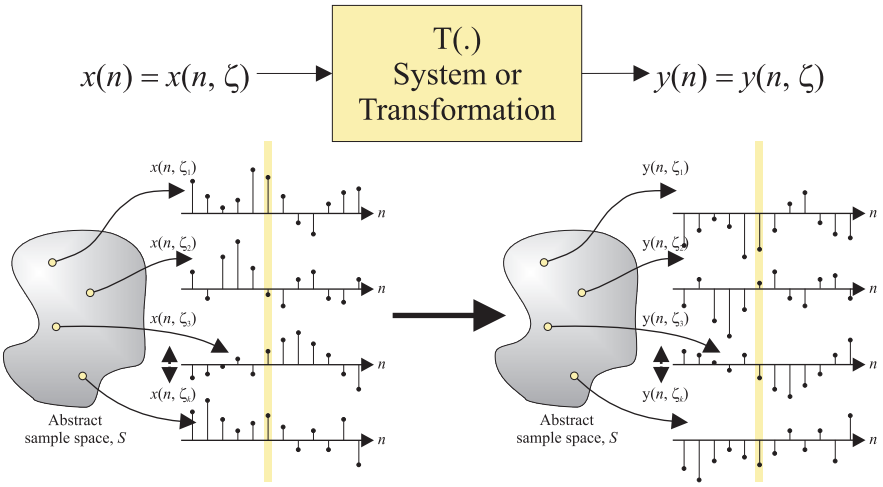
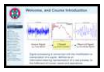


Figure 4.1: A graphical representation of a random process at the output of a system in relation to a random process at the input of the system.

system to the input. There are other systems that can be analysed, but due to time constraints, they are not considered in this course. For more information see, for example, [Papoulis:1991, Chapter 10].

4.2 LTI Systems with Stationary Inputs



New slide

The notation:

$$y[n] = L[x[n]] \tag{P:10-76}$$

will indicate that $y[n]$ is the output of a **linear system** with input $x[n]$. This means that for K random processes $\{x_k[n]\}_{k=1}^{K}$ and K scalar values $\{\alpha_k\}_{k=1}^{K}$, then

$$y[n] = L\left[\sum_{k=1}^{K} \alpha_k\, x_k[n]\right] = \sum_{k=1}^{K} \alpha_k\, L[x_k[n]] \tag{P:10-77}$$

Since each sequence (realisation) of a stochastic process is a deterministic signal, there is a well-defined input signal producing a well-defined output signal corresponding to a single realisation of the output stochastic process:

$$y(n,\zeta) = \sum_{k=-\infty}^{\infty} h(k)\, x(n-k,\zeta) \tag{M:3.4.1}$$

This is the familiar convolution integral for LTI systems, and the impulse response of this system is given by:

$$h(n) = L[\delta(n)] \tag{P:10-78}$$

If the sum in the right hand side (RHS) of Equation M:3.4.1 exists for all ζ such that $\Pr(\zeta) = 1$, then it is said that this sum has *almost-everywhere convergence* with probability of 1.

Theorem 4.1 (Input-output realisations for a LTI). If the process $x(n,\zeta)$ is stationary with $\mathbb{E}[|x(n,\zeta)|] < \infty$ and if the system is bounded-input, bounded-output (BIBO) stable, such that $\sum_{-\infty}^{\infty} |h(k)| < \infty$, then the output $y(n,\zeta)$ of the system in Equation M:3.4.1 converges absolutely with probability 1, or:

$$y(n,\zeta) = \sum_{k=-\infty}^{\infty} h(k)\, x(n-k,\zeta) \quad \text{for all } \zeta \in \mathcal{A},\ \Pr(\mathcal{A}) = 1 \tag{M:3.4.2}$$

♢

- A complete description of $y[n, \zeta]$ requires the computation of an infinite number of convolutions, corresponding to each value of ζ.

- Thus, a better description would be to consider the statistical properties of $y[n, \zeta]$ in terms of the statistical properties of the input and the characteristics of the system. For Gaussian signals, which are used very often in practice, first- and second- order statistics are sufficient, since higher-order statistics are completely specified by these first two moments.

To investigate the statistical input-output properties of a linear system, note the following fundamental theorem:

Theorem 4.2 (Expectation in Linear Systems). For any linear system,

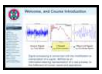
$$\mathbb{E}\left[L[x[n]]\right] = L[\mathbb{E}\left[x[n]\right]] \tag{4.2}$$

In other words, the mean $\mu_y(n)$ of the output $y(n)$ equals the response of the system to the mean $\mu_x(n)$ of the input:

$$\mu_y(n) = L[\mu_x(n)] \tag{4.3}$$

PROOF. This is a simple extension of the linearity of expected values to arbitrary linear operators.

This result will be used throughout the next section.



New slide

4.2.1 Input-output Statistics of a LTI System

If a stationary stochastic process $x[n]$ with mean value μ_x and correlation $r_{xx}[\ell]$ is applied to the input of a LTI system with impulse response $h[n]$ and transfer function $H(e^{j\omega})$, then the:

Output mean value is given by:

$$\mu_y = \mu_x \sum_{k=-\infty}^{\infty} h[k] = \mu_x\, H(e^{j0}) \tag{M:3.4.4}$$

This is easily shown by using the linearity property of the expectation operator:

$$\mu_y = \mathbb{E}\left[\sum_{k=-\infty}^{\infty} h[k]\, x[n-k]\right] = \sum_{k=-\infty}^{\infty} h[k]\, \mathbb{E}\left[x[n-k]\right] \tag{M:3.4.4}$$

and since the process is stationary, then $\mathbb{E}\left[x[n-k]\right] = \mu_x$, giving the desired result. Since μ_x and $H(e^{j0})$ are constant, μ_y is also constant. Note that $H(e^{j0})$ is the "direct current" (DC) gain of the spectrum.

Input-output cross-correlation is given by:

$$r_{xy}[\ell] = h^*[-\ell] * r_{xx}[\ell] = \sum_{k=-\infty}^{\infty} h^*[-k]\; r_{xx}[\ell - k] \tag{M:3.4.5}$$

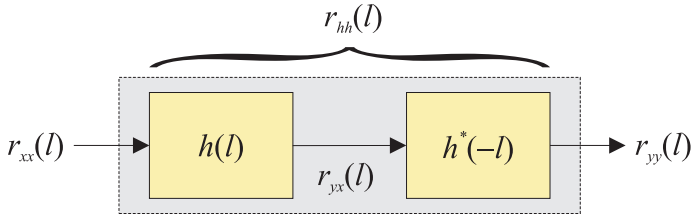


Figure 4.2: An equivalent LTI system for autocorrelation filtration.

This can be shown by writing:

$$r_{xy}(l) = \mathbb{E}\left[x(n)\, y^*(n-l)\right] = \mathbb{E}\left[x(n+l)\, y^*(n)\right] \tag{4.4}$$

$$= \mathbb{E}\left[x(n+l) \sum_{k=-\infty}^{\infty} h^*(k)\, x^*(n-k)\right] \tag{4.5}$$

$$= \sum_{k=-\infty}^{\infty} h^*(k)\mathbb{E}\left[x(n+l)\, x^*(n-k)\right] \tag{4.6}$$

$$= \sum_{k=-\infty}^{\infty} h^*(k)\, r_{xx}(l+k) \tag{4.7}$$

which by making the substitution $m = -k$, gives:

$$r_{xy}(l) = \sum_{m=-\infty}^{\infty} h^*(-m)\, r_{xx}(l-m) = h^*(-l) * r_{xx}(l) \tag{4.8}$$

as required.

Similarly, it follows that $r_{yx}(l) = h(l) * r_{xx}(l)$. Since $r_{xy}(l)$ depends only on the lag l, then the input and output processes of a BIBO stable linear time-invariant system, when driven by a wide-sense stationary (WSS) input, are jointly WSS.

Output autocorrelation is obtained by pre-multiplying the system-output by $y^*(n-l)$ and taking expectations:

$$r_{yy}(l) = \sum_{k=-\infty}^{\infty} h(k)\, \mathbb{E}\left[x(n-k)\, y^*(n-l)\right] = h(l) * r_{xy}(l) \tag{M:3.4.8}$$

Substituting the expression for $r_{xy}(l)$ gives:

$$r_{yy}(l) = h(l) * h^*(-l) * r_{xx}(l) = r_{hh}(l) * r_{xx}(l) \tag{M:3.4.10}$$

where $r_{hh}(l) = r^*_{hh}(-l)$ is the *autocorrelation*, for want of a better phrase, of the system impulse response:

$$r_{hh}(l) \triangleq h(l) * h^*(-l) = \sum_{n=-\infty}^{\infty} h(n)\, h^*(n-l) \tag{M:3.4.11}$$

where $\triangleq$ means *defined as*. If the relationship in Equation M:3.4.11 is not apparent, it can be proven by writing $g(l) = h^*(-l)$, such that the standard convolution formula gives:

$$r_{hh}(l) \triangleq h(l) * g(l) = \sum_{n=-\infty}^{\infty} h(n)\, g(l-n) \tag{4.9}$$

and, since $g(l-n) = h^*(-[l-n]) = h^*(n-l)$, Equation M:3.4.11 follows. However, this equation can also be written in an alternative form by making the substitution $m = n - l$ such that when $n \to \pm\infty$, $m \to \pm\infty$, and Equation M:3.4.11 becomes:

$$r_{hh}(l) \triangleq h(l) * h^*(-l) = \sum_{m=-\infty}^{\infty} h(m+l)\, h^*(m) \tag{M:3.4.11}$$

Both of these forms of the convolution $r_{hh}[\ell] \triangleq h[\ell] * h^*[-\ell]$ are equally valid. It is straightforward to show that $r_{hh}[\ell] = r^*_{hh}[-\ell]$ by writing:

$$r^*_{hh}[-\ell] = (h[-\ell] * h^*[+\ell])^* = h[-\ell]^* * h[+\ell] = r_{hh}[\ell] \tag{4.10}$$

Since μ_y, as given by Equation M:3.4.4 is constant, and $r_{yy}(l)$ depends only on the lag l, the response of a BIBO stable linear time-invariant to a stationary input is also a stationary process. A careful examination of Equation M:3.4.10 shows that when a signal $x[n]$ is filtered by a LTI system with impulse response $h[n]$, its autocorrelation sequence is *filtered* by a system with impulse response equal to the *autocorrelation* of its impulse response. This idea is illustrated in Figure 4.2.

Output-power of the process $y(n)$ is given by $r_{yy}(0) = \mathbb{E}\left[|y(n)|^2\right]$, and therefore since $r_{yy}(l) = r_{hh}(l) * r_{xx}(l)$,

$$P_{yy} = r_{hh}(l) * r_{xx}(l)|_{l=0} = \sum_{k=-\infty}^{\infty} r_{hh}(k)\, r_{xx}(-k) \tag{4.11}$$

Noting power, P_{yy}, is real, then taking complex-conjugates using $r^*_{xx}(-l) = r_{xx}(l)$:

$$P_{yy} = \sum_{k=-\infty}^{\infty} r^*_{hh}(k)\, r_{xx}(k) = \sum_{n=-\infty}^{\infty} h^*(n) \sum_{k=-\infty}^{\infty} r_{xx}(n+k)\, h(k) \tag{4.12}$$

This last step can be shown as follows:

$$P_{yy} = \sum_{k=-\infty}^{\infty} r^*_{hh}(k)\, r_{xx}(k) = \sum_{k=-\infty}^{\infty} \left\{ \sum_{n=-\infty}^{\infty} h^*(n)\, h(n-k) \right\} r_{xx}(n) \tag{4.13}$$

Hence, by rearranging the order of summation, and bringing the $h^*[n]$ forward, this gives:

$$= \sum_{n=-\infty}^{\infty} h^*(n) \sum_{k=-\infty}^{\infty} h(n-k)\, r_{xx}(n) \tag{4.14}$$

Then, by letting $m = n - k$, the desired result is obtained.

Output probability density function (pdf) In general, it is very difficult to calculate the pdf of the output of a LTI system, except in special cases, namely Gaussian processes.

If $x[n]$ is a Gaussian process, then the output is also a Gaussian process with mean and autocorrelation sequence given by Equation M:3.4.4 and Equation M:3.4.10 above. Also, if $x[n]$ is independent and identically distributed (i. i. d.), the pdf of the output is obtained by noting that $y[n]$ is a weighted sum of independent random variables (RVs). Indeed, as shown in earlier handouts, the pdf of the sum of independent RVs is the convolution of their pdfs or the product of their characteristic functions.

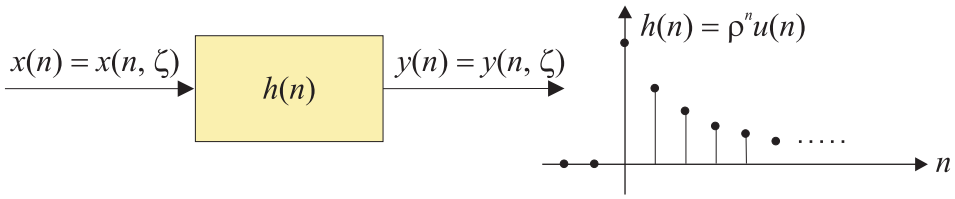


Figure 4.3: A LTI system for [Therrien:1991, Example 5.1, Page 229].

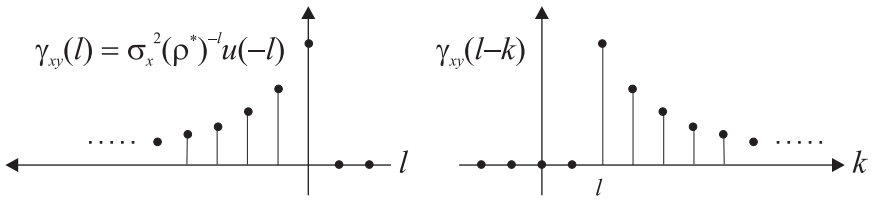


Figure 4.4: The input-output cross-covariance sequences for [Therrien:1991, Example 5.1, Page 229].

Finally, before concluding this section, note that the covariance sequences (or functions) is just the correlation sequences for the random process with the mean removed. As a result, the covariance functions satisfy a set of equations analogous to those derived above. For completeness, they are listed below:

$$\gamma_{yx}(l) = h(l) * \gamma_{xx}(l) \tag{T:5.18}$$
$$\gamma_{xy}(l) = h^*(-l) * \gamma_{xx}(l) \tag{T:5.19}$$
$$\gamma_{yy}(l) = h(l) * \gamma_{xy}(l) \tag{T:5.20}$$
$$= h(l) * h^*(-l) * \gamma_{xx}(l) \tag{T:5.21}$$

The following example illustrates the application of these results.

Example 4.1 ([Therrien:1991, Example 5.1, Page 229]). The LTI system shown in Figure 4.3 is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}(l) = \sigma_x^2\delta(l)$; note that this input process is white noise with an added nonzero mean. Calculate the mean, autocorrelation and autocovariance sequences of the output, $y(n)$, as well as the cross-correlation and cross-covariance functions between the input and the output.

SOLUTION. Each of these functions may be calculated using the equations listed in this section. Hence:

Output mean value Using Equation M:3.4.4, then:

$$\mu_y = \mu_x \sum_{k=-\infty}^{\infty} h(k) = \mu_x \sum_{k=0}^{\infty} \rho^k = \frac{\mu_x}{1-\rho} \tag{4.15}$$

Input-output cross-covariance Since the input and the output both have nonzero mean, then it is easiest to first calculate the auto- and cross-*covariance* functions, and then use these to find the auto- and cross-*correlation* functions. Thus, the output-input cross-covariance is given by Equation T:5.18:

$$\gamma_{yx}(l) = h(l) * \gamma_{xx}(l) = \big(\rho^l u(l)\big) * \big(\sigma_x^2\delta(l)\big) = \sigma_x^2\rho^l u(l) \tag{4.16}$$

and therefore the input-output cross-covariance is

$$\gamma_{xy}(l) = \gamma_{yx}^*(-l) = \sigma_x^2(\rho^*)^{-l}u(-l) \tag{4.17}$$

Output autocovariance Next, using Equation T:5.20, then:

$$\gamma_{yy}(l) = h(l) * \gamma_{xy}(l) = \sum_{k=-\infty}^{\infty} h(k)\,\gamma_{xy}(l-k) \tag{4.18}$$

The input-output cross-covariance sequence, $\gamma_{xy}(l)$, is plotted in Figure 4.4, along with $\gamma_{xy}(l-k)$ as a function of k. Hence, if $l > 0$ it follows

$$\gamma_{yy}(l) = \sum_{k=l}^{\infty} h(k)\,\gamma_{xy}(l-k) = \sum_{k=l}^{\infty} \rho^k\,\sigma_x^2(\rho^*)^{-(l-k)} \tag{4.19}$$

Substituting $m = k - l$, such that when $k = l,\ \infty$, then $m = 0,\ \infty$, and so:

$$\gamma_{yy}(l) = \sigma_x^2 \sum_{m=0}^{\infty} \rho^l\,\rho^m\,(\rho^*)^m = \sigma_x^2\rho^l \sum_{m=0}^{\infty} \left(|\rho|^2\right)^m = \frac{\sigma_x^2\rho^l}{1-|\rho|^2},\ l > 0 \tag{4.20}$$

If $l \leq 0$, then the summation is slightly different:

$$\gamma_{yy}(l) = \sum_{k=0}^{\infty} \rho^k\,\sigma_x^2(\rho^*)^{-(l-k)} = \sigma_x^2(\rho^*)^{-l} \sum_{k=0}^{\infty} \left(|\rho|^2\right)^k = \frac{\sigma_x^2(\rho^*)^{-l}}{1-|\rho|^2},\ l \leq 0 \tag{4.21}$$

Input-output cross-correlation This can now be calculated using the relationship:

$$r_{xy}(l) = \gamma_{xy}(l) + \mu_x\,\mu_y^* \tag{4.22}$$

$$= \sigma_x^2(\rho^*)^{-l}u(-l) + \mu_x\,\frac{\mu_x^*}{1-\rho^*} \tag{4.23}$$

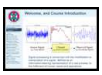
$$= \sigma_x^2(\rho^*)^{-l}u(-l) + \frac{|\mu_x|^2}{1-\rho^*} \tag{4.24}$$

Output autocorrelation In a similar manner, the autocorrelation of the output is given by:

$$r_{yy}(l) = \gamma_{yy}(l) + |\mu_y|^2 = \begin{cases} \frac{\sigma_x^2\rho^l}{1-|\rho|^2} + \left|\frac{\mu_x}{1-\rho}\right|^2 & l > 0 \\ \frac{\sigma_x^2(\rho^*)^{-l}}{1-|\rho|^2} + \left|\frac{\mu_x}{1-\rho}\right|^2 & l \leq 0 \end{cases} \tag{4.25}$$

□

Note that these results show that a process with the exponential correlation function can always be generated by applying white noise to a stable first-order system.



New slide

4.2.2 System identification by cross-correlation

The input-output cross-correlation of a LTI system is the basis for a classical method of identification of an unknown linear system.

The system is excited with a white Gaussian noise (WGN) input with autocorrelation function:

$$r_{xx}(l) = \delta(l) \tag{4.26}$$

Since the output-input cross-correlation can be written as:

$$r_{yx}(l) = h(l) * r_{xx}(l) \tag{M:3.4.6}$$

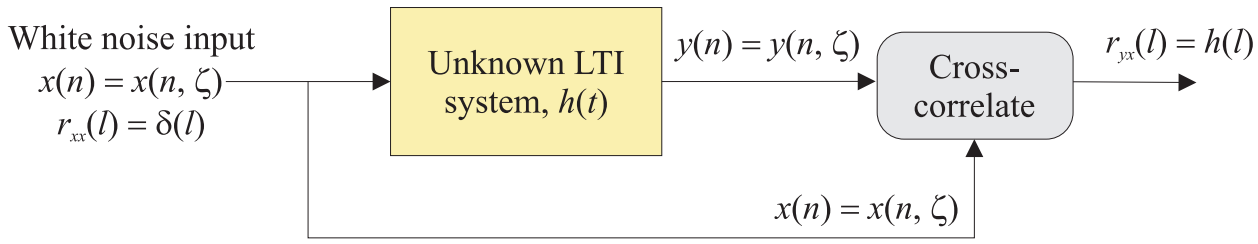


Figure 4.5: System identification by cross-correlation.

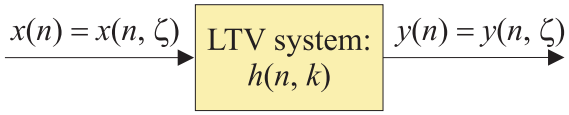


Figure 4.6: General linear time-varying (LTV) system with nonstationary input; the impulse response $h(n, k)$ is the response at index n to an impulse occurring at time index k.

then, with $r_{xx}(l) = \delta(l)$, it follows:

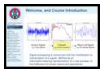
$$r_{yx}(l) = h(l) * \delta(l) = h(l) \tag{4.27}$$

Hence, the impulse response of an unknown LTI system can be estimated by exciting the system with WGN and evaluating the input-output cross-correlation.

If the discrete system represents a sampled continuous system, this method of estimating the impulse response out-performs an estimation based on simply driving the system by an impulse since:

1. it is easier to generate an approximation to white noise than to generate an approximation to an impulse, since the latter must have finite energy in an almost zero-width pulse;
2. application of an impulse to a physical system requires driving it *very* hard, albeit for a very short time, and may cause damage. Driving a system with white noise is less traumatic. As an example, consider estimating the acoustic impulse response (AIR) of a concert hall or office; one method of generating an impulse is to fire a gun and, obviously, this will damage the concert hall, which is less than desirable.

4.3 LTV Systems with Nonstationary Inputs



New slide

It is also possible to analyse a general linear system that is not necessarily time-invariant, as shown in Figure 4.6; such a system is called linear time-varying (LTV).

The input and output are related by the generalised convolution:

$$y(n) = \sum_{k=-\infty}^{\infty} h(n, k)\, x(k) \tag{T:5.1}$$

where $h(n, k)$ is the response at time-index n to an impulse occurring at the system input at time-index k. The mean, autocorrelation and autocovariance sequences of the output, $y(n)$, as well as the cross-correlation and cross-covariance functions between the input and the output, can be calculated in a similar way as for LTI systems with stationary inputs. It is left as an exercise to the reader to derive these, but the results are summarised in the next section.

4.3.1 Input-output Statistics of a LTV System

It is important to note that the input-output statistics of a LTI system with a stationary input are simply special cases of the following results. Thus, it is perhaps preferable to remember these more general results and simplify them as necessary.

Output mean value is given by

$$\mu_y(n) = \sum_{k=-\infty}^{\infty} h(n,k)\mu_x(k) \tag{T:5.2}$$

This can be written as:

$$\mu_y(n) = L[\mu_x(n)] \tag{P:10-80}$$

Output-input cross-correlation is given by

$$r_{yx}(n,m) = \sum_{k=-\infty}^{\infty} h(n,k)\, r_{xx}(k,m) \tag{T:5.5}$$

and the input-output cross-correlation is:

$$r_{xy}(n,m) = r^*_{yx}(m,n) \tag{T:5.4}$$

Output autocorrelation is a similar form, given by:

$$r_{yy}(n,m) = \sum_{k=-\infty}^{\infty} h(n,k)\, r_{xy}(k,m) \tag{T:5.3}$$

Output-input cross-covariance has an identical form to that for the input-output cross-correlation functions:

$$\gamma_{yx}(n,m) = r_{yx}(n,m) - \mu_y(n)\,\mu^*_x(m) \tag{4.28}$$

$$= \sum_{k=-\infty}^{\infty} h(n,k)\,\gamma_{xx}(k,m) \tag{T:5.9}$$

and

$$\gamma_{yx}(n,m) = \gamma^*_{xy}(m,n) \tag{T:5.8}$$

Output autocovariance is given by:

$$\gamma_{yy}(n,m) = r_{yy}(n,m) - \mu_y(n)\,\mu^*_y(m) \tag{T:5.6}$$

$$= \sum_{k=-\infty}^{\infty} h(n,k)\,\gamma_{xy}(k,m) \tag{T:5.7}$$

Note that if the impulse response of the system has *finite support*, in the sense the region over which it has non-zero values is a well-defined finite region, then it is possible to represent the correlation functions and the impulse response function in matrix form:

$$\mathbf{R}_{yy} = \mathbf{H}\mathbf{R}_{xx}\mathbf{H}^H \tag{4.29}$$

Correlation matrices were introduced in an earlier handout.

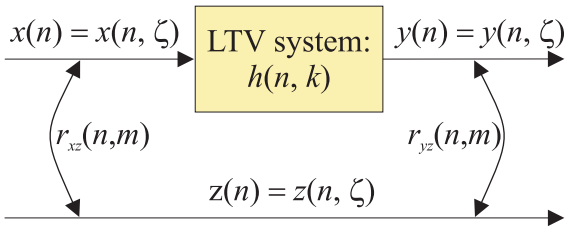


Figure 4.7: Cross-correlation with respect to a third random process.

4.3.2 Effect of Linear Transformations on Cross-correlation

Another situation worth considering is the cross-correlation with respect to a third random process, as shown in Figure 4.7.

A random process $x(n)$ is transformed by a LTV system to produce another signal $y(n)$. The process $x(n)$ is related to a third process $z(n)$, and $r_{xz}(n_1,\, n_2)$ is known. It is desirable to find $r_{yz}(n_1,\, n_2)$. The response of the LTV system to $x(n)$ is:

$$y(n) = \sum_{k\in\mathbb{Z}} h(n,k)\, x(k) \tag{T:5.22}$$

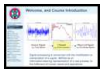
Hence, multiplying both sides by $z^*(m)$ and taking expectations:

$$r_{yz}(n,m) = \sum_{k\in\mathbb{Z}} h(n,k)\, r_{xz}(k,m) = h(n,k) * r_{xz}(k,m) \tag{T:5.24}$$

If the system is LTI, then this simplifies to:

$$r_{yz}(l) = \sum_{k\in\mathbb{Z}} h(k)\, r_{xz}(l-k) = h(l) * r_{xz}(l) \tag{4.30}$$

4.4 Difference Equation



New slide

A mathematically elegant analysis of stochastic systems comes about when a LTI system can be represented by difference equations. This will be particularly useful in the next handout on linear signal models. Although the results of the preceding sections apply to these systems, the difference equation offers an alternative representation of the results that can sometimes be quite useful and important.

Consider a LTI system that can be represented by a difference equation:

$$\sum_{p=0}^{P} a_p\, y[n-p] = \sum_{q=0}^{Q} b_q\, x[n-q] \tag{4.31}$$

where $a_0 \triangleq 1$. Assuming that both $x(n)$ and $y(n)$ are stationary processes, such that $\mathbb{E}\left[x(n-p)\right] = \mu_x$ and $\mathbb{E}\left[y(n-q)\right] = \mu_y$, then taking expectations of both sides gives, after a little rearrangement:

$$\mu_y = \frac{\sum_{q=0}^{Q} b_q}{1 + \sum_{p=1}^{P} a_p}\mu_x \tag{4.32}$$

Next, multiplying the system equation throughout by $y^*(m)$ and taking expectations gives:

$$\sum_{p=0}^{P} a_p \, r_{yy}(n-p, m) = \sum_{q=0}^{Q} b_q \, r_{xy}(n-q, m) \tag{4.33}$$

Similarly, rather than multiplying throughout the system equation by $y^*(m)$, instead multiply though by $x^*(m)$ to give:

$$\sum_{p=0}^{P} a_p \, r_{yx}(n-p, m) = \sum_{q=0}^{Q} b_q \, r_{xx}(n-q, m) \tag{4.34}$$

These two difference equations may be used to solve for $r_{yy}(n_1, n_2)$ and $r_{xy}(n_1, n_2)$. Similar expressions can be obtained for the covariance functions. They are given by:

$$\sum_{p=0}^{P} a_p \, \gamma_{yy}(n-p, m) = \sum_{q=0}^{Q} b_q \, \gamma_{xy}(n-p, m) \tag{4.35}$$

and

$$\sum_{p=0}^{P} a_p \, \gamma_{yx}(n-p, m) = \sum_{q=0}^{Q} b_q \, \gamma_{xx}(n-p, m) \tag{4.36}$$

Example 4.2 ([Manolakis:2000, Example 3.6.2, Page 141]). Let $x(n)$ be a random process generated by the first order difference equation given by:

$$x(n) = \alpha \, x(n-1) + w(n), \quad |\alpha| \leq 1, n \in \mathbb{Z} \tag{4.37}$$

where $w(n) \sim \mathcal{N}(\mu_w, \sigma_w^2)$ is an i. i. d. WGN process.

- Demonstrate that the process $x(n)$ is stationary, and determine the mean μ_x.
- Determine the autocovariance and autocorrelation function, $\gamma_{xx}(l)$ and $r_{xx}(l)$.

SOLUTION. Note that this is a first-order autoregressive (AR) process, which will be discussed in more detail later in the lecture course. The case written above is, in fact, the stationary case, and [Manolakis, Exercise 3.23, Page 145] poses the case where there is an initial transient, resulting in a nonstationary autocorrelation function. This exercise is left for those interested, although be forewarned that this is not an easy exercise. This example uses the theory described above.

- The output of a LTI system with a stationary input is always stationary. It follows directly from the results above that:

$$\mu_x = \frac{\mu_w}{1-\alpha} \tag{4.38}$$

- Using the results for the input-output covariance of a LTI system represented by difference equation:

$$\gamma_{xx}(n, m) - \alpha \, \gamma_{xx}(n-1, m) = \gamma_{wx}(n, m) \tag{4.39}$$

$$\gamma_{xw}(n, m) - \alpha \, \gamma_{xw}(n-1, m) = \gamma_{ww}(n, m) \tag{4.40}$$

  which, since the system is stationary, can be written as:

$$\gamma_{xx}(l) - \alpha \, \gamma_{xx}(l-1) = \gamma_{wx}(l) \tag{4.41}$$

$$\gamma_{xw}(l) - \alpha \, \gamma_{xw}(l-1) = \gamma_{ww}(l) \tag{4.42}$$

Noting $x(n)$ cannot depend on future values of $w(n)$, then $\gamma_{xw}(n+l,n) = \gamma_{xw}(l) = 0$, $l < 0$. This can also be demonstrated by explicitly evaluating $\gamma_{xw}(n,m)$, $m < n$, and noting that $x(n)$ and $w(n)$ are independent and zero mean. Since $\gamma_{ww}(l) = \sigma_w^2\,\delta(l)$, the second of the equations above becomes:

$$\gamma_{xw}(l) = \begin{cases} \alpha\,\gamma_{xw}(l-1) & l > 0 \\ \sigma_w^2 & l = 0 \\ 0 & l < 0 \end{cases} \tag{4.43}$$

Solving for $l \geq 0$ gives by repeated substitution, $\gamma_{xw}(l) = \alpha^l\,\sigma_w^2$, and zero for $l < 0$. Since $\gamma_{wx}(l) = \gamma_{xw}^*(-l)$, then the difference equation for the autocovariance function of $x(n)$ simplifies to:

$$\gamma_{xx}(l) - \alpha\,\gamma_{xx}(l-1) = \begin{cases} 0 & l > 0 \\ \alpha^{-l}\,\sigma_w^2 & l \leq 0 \end{cases} \tag{4.44}$$

Note the solution for $l > 0$ is the solution of the homogeneous equation. Hence, since $\gamma_{xx}(l) = \gamma_{xx}(-l)$ for a real process, then this equation is solved by assuming the solution:

$$\gamma_{xx}(l) = a\,\alpha^{|l|} + b \tag{4.45}$$

The values of a and b can be found by directly substituting the proposed solution for $l \leq 0$ into the difference equation:

$$a\,\alpha^{-l} + b - \alpha\left(a\,\alpha^{-(l-1)} + b\right) = \alpha^{-l}\,\sigma_w^2 \tag{4.46}$$

$$\alpha^{-l}\left(1-\alpha^2\right)a + (1-\alpha)\,b = \alpha^{-l}\,\sigma_w^2 \tag{4.47}$$

from which it directly follows that $b = 0$ and $a = \sigma_x^2 = \frac{\sigma_w^2}{1-\alpha^2}$, corresponding to the case when $l = 0$. Hence, in conclusion

$$\gamma_{xx}(l) = \frac{\sigma_w^2}{1-\alpha^2}\,\alpha^{|l|} \tag{4.48}$$

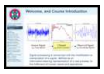
Using the relationship that $r_{xx}(l) = \gamma_{xx}(l) + \mu_x^2$, it follows that the output auto-correlation is given by

$$r_{xx}(l) = \frac{\sigma_w^2}{1-\alpha^2}\,\alpha^{|l|} + \frac{\mu_w^2}{(1-\alpha)^2} \tag{4.49}$$

□

As usual, if $\mu_w = 0$, then $r_{xx}(l) = \gamma_{xx}(l)$.

4.5 Frequency-Domain Analysis of LTI systems



New slide

Now consider how a LTI transformation affects the power spectra and complex power density spectra of a stationary random process. Recall that the power spectral density (PSD) is the Fourier transform of the **autocorrelation** functions. Alternatively, it is possible to note that the frequency response of a system is the z-transform evaluated on the unit circle.

This therefore leads to the following spectral densities:

$$P_{xy}(e^{j\omega}) = H^*(e^{j\omega})\,P_{xx}(e^{j\omega}) \tag{M:3.4.19}$$

$$P_{yx}(e^{j\omega}) = H(e^{j\omega})\,P_{xx}(e^{j\omega}) \tag{M:3.4.20}$$

$$P_{yy}(e^{j\omega}) = |H(e^{j\omega})|^2\,P_{xx}(e^{j\omega}) \tag{M:3.4.21}$$

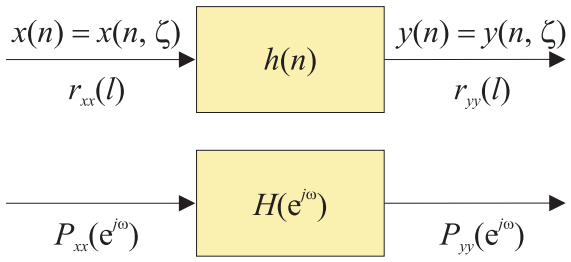


Figure 4.8: The PSD at the input and output of a LTI system with stationary input.

These results are derived very easily from the results in Section 4.2.1 and the properties of the Fourier transform, especially that convolution becomes multiplication. It is important to stress the similarity of these results with those for the frequency analysis of linear time-invariant systems with deterministic signal inputs. The system is depicted in Figure 4.8.

- If the input and output autocorrelations or autospectral densities are known, the magnitude response of a system $|H(e^{j\omega})|$ can be determined, but not the phase response.
- Only cross-correlation or cross-spectral information can help determine the phase response.

A set of similar relations to Equation M:3.4.19, Equation M:3.4.20 and Equation M:3.4.21 can also be derived for the complex spectral density function. Specifically, if $h(l) \overset{z}{\rightleftharpoons} H(z)$, then $h^*(-l) \overset{z}{\rightleftharpoons} H^*\left(\frac{1}{z^*}\right)$ and therefore:

$$P_{xy}(z) = H^*\left(\frac{1}{z^*}\right) P_{xx}(z) \tag{T:5.41}$$

$$P_{yx}(z) = H(z)\, P_{xx}(z) \tag{T:5.40}$$

$$P_{yy}(z) = H(z)\, P_{xy}(z) \tag{T:5.42}$$

$$P_{yy}(z) = H(z)\, H^*\left(\frac{1}{z^*}\right) P_{xx}(z) \tag{T:5.44}$$

Note that $P_{yy}(z)$ satisfies the required property for a complex spectral density function, namely that $P_{yy}(z) = P_{yy}^*\left(\frac{1}{z^*}\right)$. Also, note the following result for real filters that make the above equations simplify accordingly.

Theorem 4.3 (Transfer function for a real filter). For a real filter:

$$h(-l) \overset{z}{\rightleftharpoons} H^*\left(\frac{1}{z^*}\right) = H(z^{-1}) \tag{4.50}$$

PROOF. Writing:

$$H(z) = \sum_{n=-\infty}^{\infty} h(n)\, z^{-n} \tag{4.51}$$

then setting $z \to \frac{1}{z^*}$ gives:

$$H\left(\frac{1}{z^*}\right) = \sum_{n=-\infty}^{\infty} h(n) \left[\frac{1}{z^*}\right]^{-n} \tag{4.52}$$

Now, taking complex-conjugates, using the following facts:

- the conjugate of a sum/product of complex numbers is the sum/product of the conjugates of the complex numbers, or in otherwords $(a+b)^* = a^* + b^*$ and $(ab)^* = a^*b^*$,
- the filter coefficients are real, such that $h(n)^* = h(n)$,

then

$$H^*\left(\frac{1}{z^*}\right) = \sum_{n=-\infty}^{\infty} h(n)\, z^n \equiv \sum_{m=-\infty}^{\infty} h(-m)\, z^{-m} \tag{4.53}$$

□

where in the last step, the substitution $m = -n$ has been made. Hence, this gives the desired result. It is straightfoward to adapt the final stage of this proof to show that $h^*(-l) \overset{z}{\rightleftharpoons} H^*\left(\frac{1}{z^*}\right)$ in general.

Consider again the earlier example based on [Therrien:1991, Example 5.1, Page 229].

Example 4.3 ([Therrien:1991, Example 5.3, Page 237]). Again, the LTI system shown in Figure 4.3 is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}(l) = \sigma_x^2\,\delta(l)$. Calculate the power spectral density (PSD), cross-power spectral density (CPSD) and the complex cross-spectral densities.

SOLUTION. The first-order system with impulse response $h(n) = \rho^n u(n)$ has system transfer function

$$H(z) = \frac{1}{1-\rho\, z^{-1}} \tag{4.54}$$

The complex spectral density function for the white noise with added mean is given by the z-transform of the autocorrelation sequence. Since $\gamma_{xx}(l) = \sigma_x^2\delta(l)$, then $r_{xx}(l) = \gamma_{xx}(l) + \mu_x^2 = \sigma_x^2\delta(l) + |\mu_x|^2$. Taking z-transforms gives:

$$P_{xx}(z) = \sigma_x^2 + 2\pi|\mu_x|^2\delta(z - e^{j0}) \tag{4.55}$$

$$= \sigma_x^2 + 2\pi|\mu_x|^2\delta(z-1) \tag{4.56}$$

where the complex spectral density result in Equation (T:4.59) at the end of the previous handout has been used. Hence, the complex cross-spectral density is given by:

$$P_{xy}(z) = H^*\left(\frac{1}{z^*}\right)\, P_{xx}(z) = \left(\frac{1}{1-\rho\left[\frac{1}{z^*}\right]^{-1}}\right)^* \left[\sigma_x^2 + 2\pi|\mu_x|^2\delta(z-1)\right] \tag{4.57}$$

$$= \frac{\sigma_x^2}{1-\rho^* z} + \frac{2\pi|\mu_x|^2}{1-\rho^* z}\delta(z-1) \tag{4.58}$$

Moreover, the complex spectral density is given by:

$$P_{yy}(z) = H(z)\, P_{xy}(z) = \left(\frac{1}{1-\rho z^{-1}}\right)\left(\frac{1}{1-\rho^* z}\right)\left[\sigma_x^2 + 2\pi|\mu_x|^2\delta(z-1)\right] \tag{4.59}$$

$$= \frac{\sigma_x^2}{1+|\rho|^2-\rho^* z-\rho z^{-1}} + \frac{2\pi|\mu_x|^2}{|1-\rho|^2}\delta(z-1) \tag{4.60}$$

Thus, the CPSD and the PSD can be obtained by setting $z = e^{j\omega}$ to obtain:

$$P_{xy}(e^{j\omega}) = \frac{\sigma_x^2}{1-\rho^* e^{j\omega}} + \frac{2\pi|\mu_x|^2}{1-\rho^* e^{j\omega}}\delta(e^{j\omega}-1) \tag{4.61}$$

Moreover, the complex spectral density is given by:

$$P_{yy}(e^{j\omega}) = \frac{\sigma_x^2}{1 + |\rho|^2 - 2|\rho|\cos(\omega - \arg\rho)} + \frac{2\pi|\mu_x|^2}{|1-\rho|^2}\delta(e^{j\omega} - 1) \tag{4.62}$$

where the simplification that:

$$\begin{aligned} \rho^* e^{j\omega} + \rho e^{-j\omega} &= |\rho| \left[e^{-j\arg\rho} e^{j\omega} + e^{j\arg\rho} e^{-j\omega} \right] = |\rho| \left[e^{j(\omega - \arg\rho)} + e^{-j(\omega - \arg\rho)} \right] && (4.63) \\ &= 2|\rho|\cos(\omega - \arg\rho) && (4.64) \end{aligned}$$

□

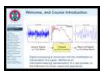
has been used.

5

Linear Signal Models

This handout looks at the special class of stationary signals that are obtained by driving a linear time-invariant (LTI) system with white noise. A particular focus is placed on system functions that are rational; that is, they can be expressed at the ratio of two polynomials. Thus, the time-domain and frequency domain characteristics of pole-zero, all-pole, and all-zero models are investigated, including their time-series equivalents.

5.1 Abstract



New slide

- In the last lecture, the response of a linear-system when a stochastic process is applied at the input was considered. General linear systems were considered, and no focus on their interpretation or their practical applications was discussed.

- This lecture looks at the special class of stationary signals that are obtained by driving a linear time-invariant (LTI) system with white noise. A particular focus is placed on **system function**s that are rational; that is, they can be expressed at the ratio of two polynomials. The power spectral density (PSD) of the resulting process is also rational, and its shape is completely determined by the filter coefficients. As a result, linear signal models provide a method for modelling the PSD of a process, and thus leads to **parametric PSD estimation**, also known as **modern spectral estimation**.

- The following models are considered in detail:

  - **All-pole** systems and autoregressive (AR) processes;
  - **All-zero** systems and moving average (MA) processes;
  - and **pole-zero** systems and autoregressive moving average (ARMA) processes.

- **Pole-zero** models are widely used for modelling stationary signals with short memory; the concepts will be extended, in overview at least, to nonstationary processes.

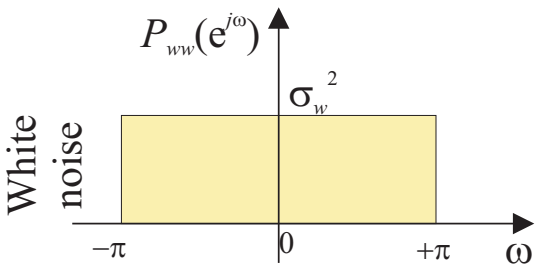
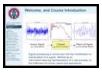


Figure 5.1: White noise PSD.

Linear signal models are developed first by assuming that the second order moments of the random process are known, and equations are developed whose solution provides the model parameters. In most practical applications of the theory, however, the fixed quantities in the equations, namely the correlation functions and the model orders, are not known *a priori* but need to be estimated from the data. This, as a result, introduces the issue of estimation of the model parameters and leads to the notion of, for example, maximum likelihood estimation and least squares estimates as discussed in the next handout.



New slide

5.2 The Ubiquitous WGN Sequence

The simplest random signal model is the wide-sense stationary (WSS) white Gaussian noise (WGN) sequence:

$$w(n) \sim \mathcal{N}\left(0,\, \sigma_w^2\right) \tag{5.1}$$

The sequence is independent and identically distributed (i. i. d.), and has a flat PSD: $P_{ww}(e^{j\omega}) = \sigma_w^2,\ -\pi < \omega \leq \pi$. The PSD is shown below in Figure 5.1.

5.2.1 Generating WGN samples

Recall that the **probability transformation rule** takes random variables from one distribution as inputs and outputs random variables in a new distribution function:

Theorem 5.1 (Probability transformation rule (revised)). If $\{x_1,\, \ldots\, x_n\}$ are random variables with a joint-probability density function (pdf) $f_X\left(x_1, \ldots, x_n\right)$, and if $\{y_1,\, \ldots\, y_n\}$ are random variables obtained from functions of $\{x_k\}$, such that $y_k = g_k(x_1,\, x_2\, \ldots\, x_n)$, then the joint-pdf, $f_Y\left(y_1, \ldots, y_n\right)$, is given by:

$$f_Y\left(y_1, \ldots, y_n\right) = \frac{1}{|J(x_1, \ldots, x_n)|} f_X\left(x_1, \ldots, x_n\right) \tag{5.2}$$

where $J(x_1,\, \ldots,\, x_n)$ is the **Jacobian** of the transformation given by:

$$J(x_1, \ldots, x_n) = \frac{\partial(y_1, \ldots\, y_n)}{\partial(x_1, \ldots\, x_n)} \tag{5.3}$$

$\diamondsuit$

One particular well-known example is the *Box-Muller* (1958) transformation that takes two uniformly distributed random variables, and transforms them to a bivariate Gaussian distribution. Consider the transformation between two uniform random variables given by,

$$f_{X_k}\left(x_k\right) = \mathbb{I}_{0,1}\left(x_k\right), \quad k = 1, 2 \tag{5.4}$$

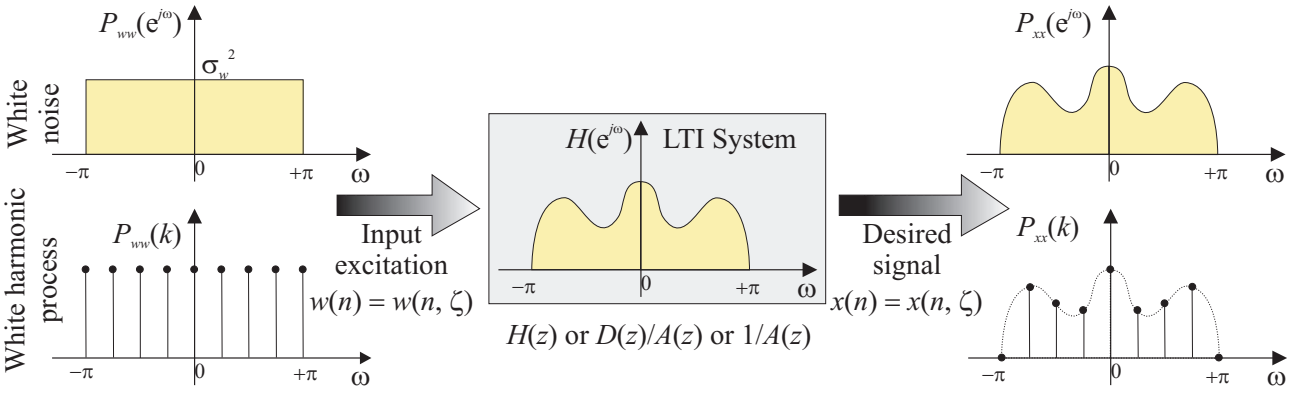


Figure 5.2: Signal models with continuous and discrete (line) power spectrum densities.

where $\mathbb{I}_{\mathcal{A}}(x) = 1$ if $x \in \mathcal{A}$, and zero otherwise, and the two random variables $y_1,\ y_2$ given by:

$$y_1 = \sqrt{-2\ln x_1}\,\cos 2\pi x_2 \tag{5.5}$$

$$y_2 = \sqrt{-2\ln x_1}\,\sin 2\pi x_2 \tag{5.6}$$

It follows, by rearranging these equations, that:

$$x_1 = \exp\left[-\frac{1}{2}(y_1^2 + y_2^2)\right] \tag{5.7}$$

$$x_2 = \frac{1}{2\pi}\arctan\frac{y_2}{y_1} \tag{5.8}$$

The Jacobian determinant can be calculated as:

$$J(x_1, x_2) = \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{vmatrix} = \begin{vmatrix} \frac{-1}{x_1\sqrt{-2\ln x_1}}\cos 2\pi x_2 & -2\pi\sqrt{-2\ln x_1}\,\sin 2\pi x_2 \\ \frac{-1}{x_1\sqrt{-2\ln x_1}}\sin 2\pi x_2 & 2\pi\sqrt{-2\ln x_1}\,\cos 2\pi x_2 \end{vmatrix} = \frac{2\pi}{x_1} \tag{5.9}$$

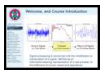
Hence, it follows:

$$f_Y(y_1, y_2) = \frac{x_1}{2\pi} = \left[\frac{1}{\sqrt{2\pi}}e^{-y_1^2/2}\right]\left[\frac{1}{\sqrt{2\pi}}e^{-y_2^2/2}\right] \tag{5.10}$$

since the domain $[0, 1]^2$ is mapped to the range $(-\infty, \infty)^2$, thus covering the range of real numbers. This is the product of y_1 alone and y_2 alone, and therefore each y is i. i. d. according to the normal distribution, as required.

Consequently, this transformation allows one to sample from a uniform distribution in order to obtain samples that have the same pdf as a Gaussian random variable.

5.2.2 Filtration of WGN



New slide

By filtering a WGN through a stable LTI system, it is possible to obtain a stochastic signal at the output with almost any arbitrary aperiodic correlation function or continuous PSD. The PSD of the output is given by:

$$P_{xx}(e^{j\omega}) = \sigma_w^2|H(e^{j\omega})|^2 = G^2\frac{\prod_{k=1}^{Q}|1 - z_k\,e^{-j\omega}|^2}{\prod_{k=1}^{P}|1 - p_k\,e^{-j\omega}|^2} \tag{5.11}$$

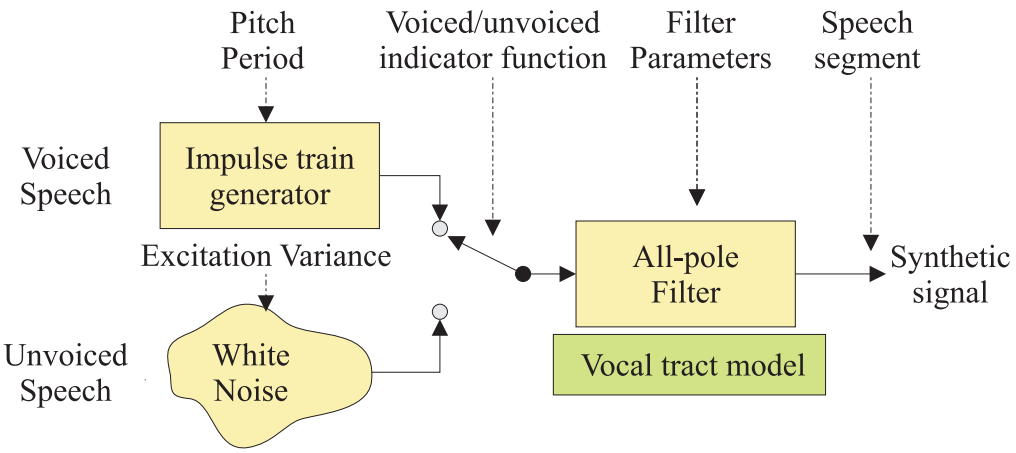
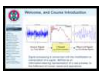


Figure 5.3: The speech synthesis model.

Note that the shape of the power spectrum depends only upon the magnitude of the filter's frequency response.

Random signals with line PSDs can be generated by using the **harmonic process** model, which is a linear combination of sinusoidal sequences with statistically independent random phases. Signal models with mixed PSDs can be obtained by combining these two models; a process justified by the **Wold decomposition**. This is highlighted in Figure 5.2; contrast this with the speech synthesis model shown in Figure 5.3, which was also shown in the introductory handout.



New slide

5.3 Nonparametric and parametric signal models

Nonparametric models have no restriction on its form, or the number of parameters characterising the model. For example, specifying a LTI filter by its impulse response is a nonparametric model.

If the input $w(n)$ is a zero-mean white noise process with variance σ_w^2, autocorrelation $r_{ww}(l) = \sigma_w^2 \delta(l)$ and $P_{ww}(e^{j\omega}) = \sigma_w^2$, $-\pi < \omega \leq \pi$, then the autocorrelation, complex spectral density, and PSD of the output $x(n)$ are given by, respectively:

$$r_{xx}(l) = \sigma_w^2 \sum_{k=-\infty}^{\infty} h(k)\, h^*(k-l) = \sigma_w^2 r_{hh}(l) \tag{M:4.1.2}$$

$$P_{xx}(z) = \sigma_w^2 H(z) H^*\left(\frac{1}{z^*}\right) \tag{M:4.1.3}$$

$$P_{xx}(e^{j\omega}) = \sigma_w^2 |H(e^{j\omega})|^2 \tag{M:4.1.4}$$

Notice that the shape of the autocorrelation and the power spectrum of the output signal are completely characterised by the system. This is known as a **system based signal model**, and in the case of linear systems, is also known as the **linear random signal model**, or the **general linear process model**.

Parametric models, on the other hand, describe a system with a finite number of parameters. For example, if a LTI filter is specified by a finite-order rational **system function**, it is a parametric model.

Two important analysis tools present themselves for parametric modelling:

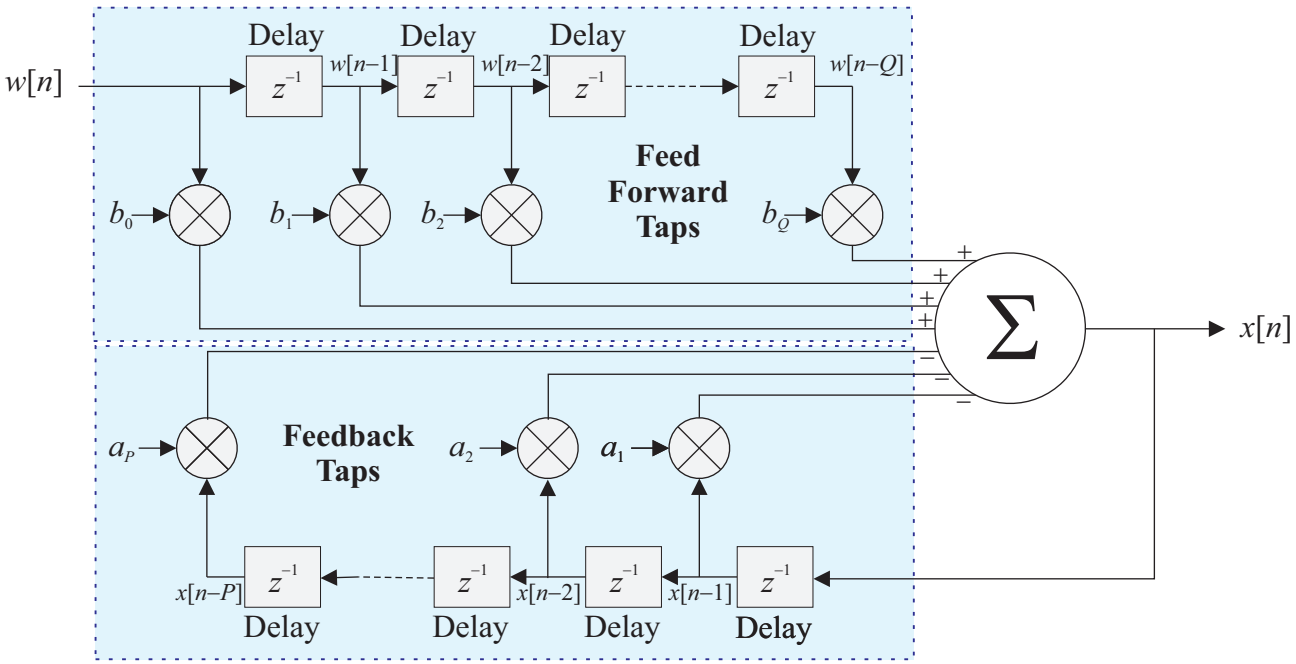
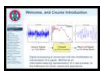


Figure 5.4: Filter block diagram for ARMA model.

1. given the parameters of the model, analyse the characteristics of that model (in terms of moments etc.);
2. design of a parametric system model to produce a random signal with a specified **autocorrelation** function or PSD. This problem is known as **signal modelling**.

5.4 Parametric Pole-Zero Signal Models



New slide

Parametric models describe a system with a finite number of parameters. Consider a system described by the following linear constant-coefficient difference equation:

$$x(n) = -\sum_{k=1}^{P} a_k\, x(n-k) + \sum_{k=0}^{Q} d_k\, w(n-k) \qquad \text{(M:4.1.21)}$$

This rational transfer function was introduced in the first lecture, and the filter block diagram is shown in Figure 5.4. Taking z-transforms gives the system function:

$$H(z) = \frac{X(z)}{W(z)} = \frac{\sum_{k=0}^{Q} d_k\, z^{-k}}{1 + \sum_{k=1}^{P} a_k\, z^{-k}} \qquad \text{(M:4.1.22)}$$

$$\triangleq \frac{D(z)}{A(z)} = G\,\frac{\prod_{k=1}^{Q}(1 - z_k\, z^{-1})}{\prod_{k=1}^{P}(1 - p_k\, z^{-1})} \qquad \text{(M:4.1.23)}$$

This system has Q zeros, $\{z_k,\, k \in \mathcal{Q}\}$ where $\mathcal{Q} = \{1, \ldots, Q\}$, and P poles, $\{p_k,\, k \in \mathcal{P}\}$. Note that poles and zeros at $z = 0$ are not considered here. The term G is the system gain. It is assumed that the polynomials $A(z)$ and $D(z)$ do not have any common roots; that is, common poles and zeros have already been cancelled.

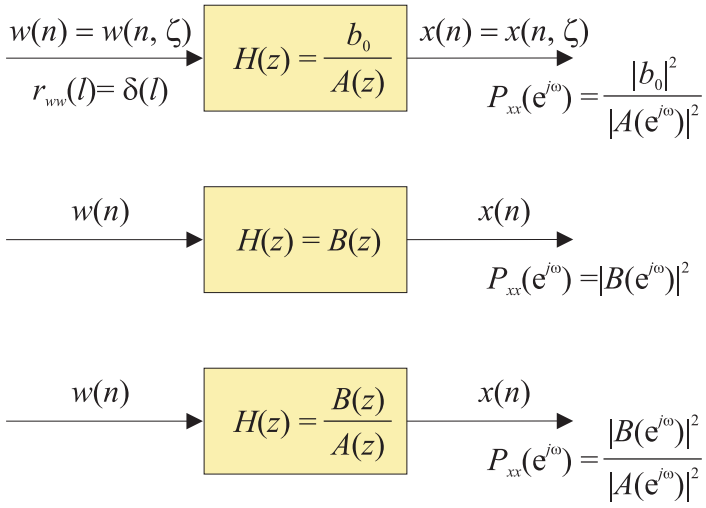
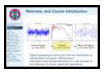


Figure 5.5: Types of linear model; top to bottom, these are the AR, MA and ARMA models.

5.4.1 Types of pole-zero models



New slide

There are three cases of interest as shown in Figure 5.5:

All-pole model when $Q = 0$. The input-output difference equation is given by:

$$x(n) = -\sum_{k=1}^{P} a_k\, x(n-k) + d_0\, w(n) \qquad \text{(M:4.1.26)}$$

This is commonly denoted as the $AP(P)$ model.

All-zero model when $P = 0$. The input-output relation is given by:

$$x(n) = \sum_{k=0}^{Q} d_k\, w(n-k) \qquad \text{(M:4.1.25)}$$

This is commonly denoted as the $AZ(Q)$ model.

Pole-zero model when $P > 0$ and $Q > 0$.

This is commonly denoted as the $PZ(P, Q)$ model, and if it is assumed to be causal, is given by Equation M:4.1.21. This is, of course, the most general of the linear signal models.

If a parametric model is *excited* with WGN, the resulting output signal has second-order moments determined by the parameters of the model. These **stochastic processes** have special names in the literature, and are known as:

a moving average (MA) process when it is the output of an all-zero model;

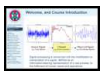
an autoregressive (AR) process when it is the output of an all-pole model;

an autoregressive moving average (ARMA) process when it is the output of an pole-zero model;

each subject to a WGN process at the input.

The parametric signal model is usually specified by normalising $d_0 = 1$ and setting the variance of the input to σ_w^2. The alternative is to specify $\sigma_w^2 = 1$ and leave d_0 arbitrary, but this isn't quite as elegant when it comes to deriving pdfs. It is also important to stress that these models assume the resulting processes are stationary, which is ensured if the corresponding systems are bounded-input, bounded-output (BIBO) stable.

5.4.2 All-pole Models



A discussion of linear signal models should begin with all-pole models because they are the easiest to analyse, and the most often used in practical applications. Assume an all-pole model of the form: *New slide*

$$H(z) = \frac{d_0}{A(z)} = \frac{d_0}{1 + \sum_{k=1}^{P} a_k\, z^{-k}} = \frac{d_0}{\prod_{k=1}^{P}(1 - p_k\, z^{-1})} \qquad \text{(M:4.2.1)}$$

where d_0 is the system gain, and P is the order of the model.

All-pole models are frequently used in signal processing applications since they are:

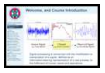
- mathematically convenient since model parameters can be estimated by solving a set of linear equations, and
- they widely parsimoniously approximate rational transfer functions, especially resonant systems.

There are various model properties of the all-pole model that are useful; these include:

1. the systems impulse response;
2. the somewhat inappropriate term called the autocorrelation of the impulse response;
3. and minimum-phase conditions.

Although the autocorrelation of the impulse response is useful to gain additional insight into aspects of the all-pole filter, it is better to consider the autocorrelation function of an AR process (i.e. the autocorrelation function of the output of an all-pole filter). However, for completeness, the details of the autocorrelation of the impulse response is included in these notes.

5.4.2.1 Frequency Response of an All-Pole Filter



New slide

The all-pole model has form:

$$H(z) = \frac{d_0}{A(z)} = \frac{d_0}{1 + \sum_{k=1}^{P} a_k\, z^{-k}} = \frac{d_0}{\prod_{k=1}^{P}(1 - p_k\, z^{-1})} \qquad \text{(M:4.2.1)}$$

and therefore its frequency response is given by:

$$H(e^{j\omega}) = \frac{d_0}{1 + \sum_{k=1}^{P} a_k\, e^{-jk\omega}} = \frac{d_0}{\prod_{k=1}^{P}(1 - p_k\, e^{-j\omega})} \qquad (5.12)$$

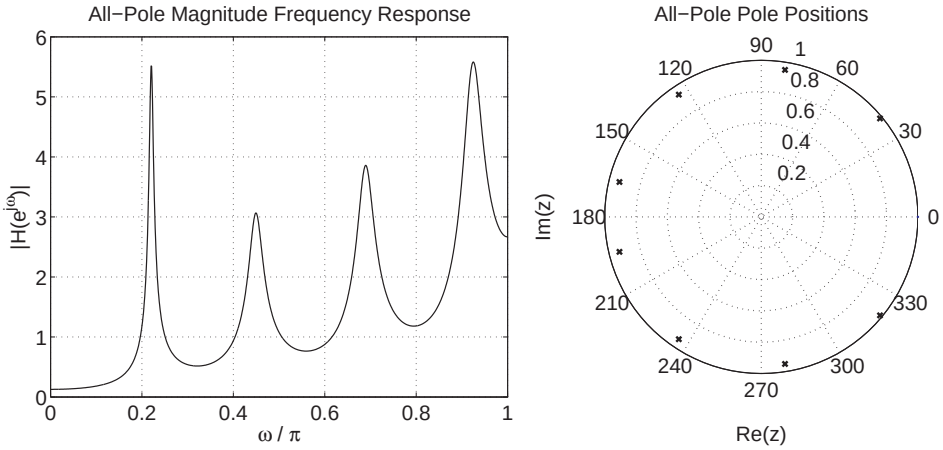
When each of the poles are written in the form $p_k = r_k e^{j\omega_k}$, then the frequency response can be written as:

$$H(e^{j\omega}) = \frac{d_0}{\prod_{k=1}^{P}(1 - r_k\, e^{-j(\omega - \omega_k)})} \qquad (5.13)$$

Hence, it can be deduced that resonances occur near the frequencies corresponding to the phase position of the poles. When the system is real, the complex-poles occur in conjugate-pairs.

Hence, the PSD of the output of an all-pole filter is given by:

$$P_{xx}(e^{j\omega}) = \sigma_w^2 \left|H(e^{j\omega})\right|^2 = \frac{G^2}{\prod_{k=1}^{P} \left|1 - r_k\, e^{-j(\omega - \omega_k)}\right|^2} \qquad (5.14)$$



(a) Magnitude frequency response. (b) Positions of poles.

Figure 5.6: The frequency response and position of the poles in an all-pole system.

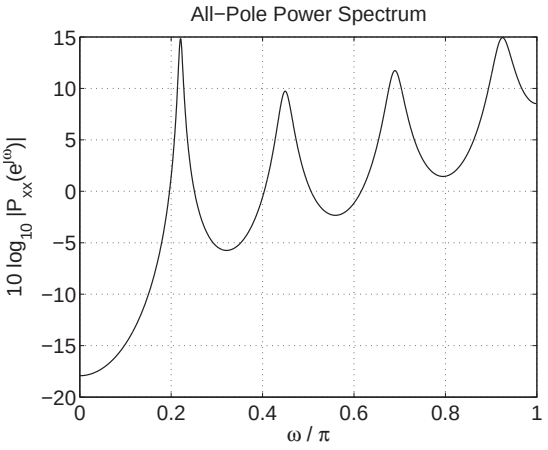


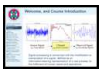
Figure 5.7: Power spectral response of an all-pole model.

where $G = \sigma_w \, d_0$ is the overall gain of the system.

Consider the all-pole model with poles at positions:

$$\{p_k\} = \{r_k \, e^{j\omega_k}\} \quad \text{where} \quad \begin{cases} \{r_k\} & = \{0.985, 0.951, 0.942, 0.933\} \\ \{\omega_k\} & = 2\pi \times \{270, 550, 844, 1131\}/2450; \end{cases} \tag{5.15}$$

The pole positions and magnitude frequency response of this system is plotted in Figure 5.6. For comparison, the PSD of the output of the system is shown in Figure 5.7.



New slide

5.4.2.2 Impulse Response of an All-Pole Filter

Recalling that the input-output difference equation for an all-pole filter is given by:

$$x(n) = -\sum_{k=1}^{P} a_k \, x(n-k) + d_0 \, w(n) \tag{M:4.1.26}$$

then the impulse response, $h(n)$, is the output of this system when the input is a delta function, $w(n) = \delta(n)$.

The impulse response of the all-pole filter satisfies the equation:

$$h(n) = -\sum_{k=1}^{P} a_k\, h(n-k) + d_0\, \delta(n) \tag{M:4.2.3}$$

The derivation in [Manolakis:2000, page 157] is somewhat verbose; nevertheless, their approach is to re-write the system function of the all-pole filter as:

$$H\,(z) + \sum_{k=1}^{P} a_k\, H\,(z)\; z^{-k} = d_0 \tag{5.16}$$

and thus by taking the inverse z-transform gives the same result as above. If $H\,(z)$ has its poles inside the unit circle, then $h[n]$ is a causal, stable sequence, and the system is **minimum-phase**.

Assuming causality, such that $h[n] = 0, \quad n < 0$ then it follows $h[-k] = 0,\ k > 0$, and therefore:

$$h[n] = \begin{cases} 0 & \text{if } n < 0 \\ d_0 & \text{if } n = 0 \\ -\sum_{k=1}^{P} a_k\, h[n-k] & \text{if } n > 0 \end{cases} \tag{M:4.2.5}$$

Thus, except for the value at $n = 0$, $h[n]$ can be obtained recursively as a linearly weighted summation of its previous values, $\{h[n-p]\,,\ p = \{1, \ldots, P\}\}$. Thus, in this sense, $h[n]$ can be *predicted*, for $n \neq 0$, with zero error from the past P past values. Thus, the coefficients $\{a_k\}$ are often referred to as **predictor coefficients**.

Finally, note that a causal $H\,(z)$ can be written as a one-sided z-transform, or infinite polynomial, $H\,(z) = \sum_{n=0}^{\infty} h[n]\; z^{-n}$. This representation implies that any finite-order, all-pole model can be represented equivalently by an infinite number of zeros, and conversely a single zero can be represented by an infinite number of poles. If the poles are inside the unit circle, then so are the corresponding zeros, and vice-versa.

5.4.2.3 Autocorrelation of the Impulse Response

The impulse response $h[n]$ of an all-pole model has infinite duration, so that its autocorrelation involves an infinite summation, which is not practical to write in *closed form* except for low-order models. However, the autocorrelation function for the all-pole model can be written as a recursive relation that relates the autocorrelation values to the model parameters. As introduced previously, the autocorrelation of the system impulse response is given by:

$$r_{hh}(l) \triangleq h(l) * h^*(-l) = \sum_{n=-\infty}^{\infty} h(n)\, h^*(n-l) \tag{5.17}$$

Multiplying both side of Equation M:4.2.3 by $h^*(n-l)$ gives and summing over all n:

$$\sum_{n=-\infty}^{\infty} \sum_{k=0}^{P} a_k\, h(n-k)\, h^*(n-l) = d_0 \sum_{n=-\infty}^{\infty} h^*(n-l)\delta(n) \tag{M:4.2.14}$$

where $a_0 = 1$. Interchanging the order of summations (as usual) in the left hand side (LHS), and setting $\hat{n} = n - k$ gives:

$$\sum_{k=0}^{P} a_k \sum_{\hat{n}=-\infty}^{\infty} h(\hat{n})\, h^*(\hat{n} - (l-k)) = d_0\, h^*(-l) \tag{5.18}$$

which can also be written as

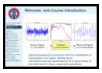
$$\sum_{k=0}^{P} a_k\, r_{hh}(l-k) = d_0\, h^*(-l) \tag{M:4.2.15}$$

Since $h(n) = 0,\ n < 0$, then $h(-l) = 0,\ l > 0$, and $h(0) = d_0$, then:

$$r_{hh}(l) = \begin{cases} d_0\, h^*(-l) - \sum_{k=1}^{P} a_k\, r_{hh}(l-k) & l < 0 \\ |d_0|^2 - \sum_{k=1}^{P} a_k\, r_{hh}(-k) & l = 0 \\ -\sum_{k=1}^{P} a_k\, r_{hh}(l-k) & l > 0 \end{cases} \tag{5.19}$$

These are recursive relationships for $r_{hh}[\ell]$ in terms of past values of the autocorrelation function.

It is also possible to write the autocorrelation in terms of the poles of the model, and to also investigate the response of the model to an impulse train (harmonic) excitation. These are not considered in this handout, but are detailed in [Manolakis:2000, Section 4.2].



5.4.2.4 All-Pole Modelling and Linear Prediction

New slide

A **linear predictor** forms an estimate, or *prediction*, $\hat{x}[n]$, of the present value of a stochastic process $x[n]$ from a linear combination of the past P samples; that is:

$$\hat{x}[n] = -\sum_{k=1}^{P} a_k\, x[n-k] \tag{M:1.4.1}$$

The coefficients $\{a_k\}$ of the linear predictor are determined by attempting to minimise some function of the **prediction error** given by:

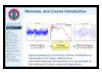
$$e(n) = x(n) - \hat{x}(n) \tag{M:1.4.2}$$

Usually the objective function is equivalent to mean-squared error (MSE), given by $E = \sum_n e^2(n)$.

Hence, the prediction error can be written as:

$$e(n) = x(n) - \hat{x}(n) = x(n) + \sum_{k=1}^{P} a_k\, x(n-k) \tag{M:4.2.50}$$

- Thus, the prediction error is equal to the excitation of the all-pole model; $e(n) = w(n)$. Clearly, **finite impulse response (FIR) linear prediction** and **all-pole modelling** are closely related.
- Many of the properties and algorithms developed for either **linear prediction** or **all-pole modelling** can be applied to the other.
- To all intents and purposes, linear prediction, **all-pole modelling**, and **AR processes** (discussed next) are equivalent terms for the same concept.



5.4.2.5 Autoregressive Processes

New slide

While **all-pole models** refer to the properties of a rational system containing only poles, **AR processes** refer to the resulting stochastic process that occurs as the result of **WGN** being applied to the input of an **all-pole filter**.

As such, the same input-output equations for all-pole models still apply although, in this case, the AR process refers to $x[n]$, whereas **all-pole modelling** would refer to the system itself, as defined by the linear difference equation and the parameters $\{a_k\}$.

Thus:

$$x[n] = -\sum_{k=1}^{P} a_k\, x[n-k] + w[n]\,, \quad w[n] \sim \mathcal{N}\left(0,\, \sigma_w^2\right) \tag{M:4.2.52}$$

The **AR process** is valid only if the corresponding **all-pole system** is stable. The autoregressive output, $x[n]$, is a stationary sequence with a mean value of zero, $\mu_x = 0$.

The autocorrelation sequence (ACS) can be calculated in a similar approach to finding the output autocorrelation and cross-correlation for linear systems.

Multiply the difference Equation M:4.2.52 through by $x^*(n-l)$ and take expectations to obtain:

$$r_{xx}(l) + \sum_{k=1}^{P} a_k\, r_{xx}(l-k) = r_{wx}(l) \tag{M:4.2.54}$$

Observing that $x(n)$ cannot depend on future values of $w(n)$ since the system is causal, then $r_{wx}(l) = \mathbb{E}\left[w(n)\, x^*(n-l)\right]$ is zero if $l > 0$, and σ_w^2 if $l = 0$.

Thus, writing Equation M:4.2.54 for $l = \{0,\, 1,\, \ldots,\, P\}$ gives:

$$r_{xx}(0) + a_1\, r_{xx}(-1) + \cdots + a_P\, r_{xx}(-P) = \sigma_w^2 \tag{5.20}$$

$$r_{xx}(1) + a_1\, r_{xx}(0) + \cdots + a_P\, r_{xx}(-P+1) = 0 \tag{5.21}$$

$$\vdots \tag{5.22}$$

$$r_{xx}(P) + a_1\, r_{xx}(P-1) + \cdots + a_P\, r_{xx}(0) = 0 \tag{5.23}$$

This can be written in matrix-vector form (noting that $r_{xx}(l) = r_{xx}^*(-l)$ and that the parameters $\{a_k\}$ are real) as:

$$\begin{bmatrix} r_{xx}(0) & r_{xx}^*(1) & \cdots & r_{xx}^*(P) \\ r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}^*(P-1) \\ \vdots & \vdots & \ddots & \vdots \\ r_{xx}(P) & r_{xx}(P-1) & \cdots & r_{xx}^*(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_P \end{bmatrix} = \begin{bmatrix} \sigma_w^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{M:4.2.56}$$

These **Yule-Walker equations** have an identical form to the **normal equations** which are a result of analysing linear prediction. The differences are minor, but the interested reader can find out more in [Therrien:1992, Chapter 8]. It is important to note that the Yule-Walker equations are linear in the parameters a_k, and there are several different efficient methods for solving them. Details, again, can be found in [Therrien:1992, Chapters 8 and 9].

5.4.2.6 Autocorrelation Function from AR parameters

In the previous section, an expression for calculating the AR coefficients given the autocorrelation values was given. But what if the AR coefficients are known, and it is desirable to calculate the autocorrelation function given these parameters. A formulation is given here. Assume that an AR process is real, such that the **Yule-Walker equations** become:

$$\begin{bmatrix} r_{xx}(0) & r_{xx}(1) & \cdots & r_{xx}(P) \\ r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(P-1) \\ \vdots & \vdots & \ddots & \vdots \\ r_{xx}(P) & r_{xx}(P-1) & \cdots & r_{xx}(0) \end{bmatrix} \hat{\mathbf{a}} = \mathbf{b} \quad \text{where} \quad \hat{\mathbf{a}} = \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_P \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} \sigma_w^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{5.24}$$

To generate the autocorrelation values from the AR parameters, it is desirable to obtain an equation of the form $\mathbf{A}\,\mathbf{r} = \mathbf{b}$, where $\begin{bmatrix} r_{xx}(0) & \cdots & r_{xx}(P) \end{bmatrix}^T$, and the matrix $\mathbf{A}$ and vector $\mathbf{b}$ are functions of the parameters $\{a_k\}$ and the input variance σ_w^2. Write the Yule-Walker equations as:

$$r_{xx}(0)\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix}\hat{\mathbf{a}} + r_{xx}(1)\begin{bmatrix} 0 & 1 & \cdots & 0 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 0 & \cdots & 1 & 0 \end{bmatrix}\hat{\mathbf{a}} + \cdots + r_{xx}(P)\begin{bmatrix} 0 & 0 & \cdots & 1 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 1 & \cdots & 0 & 0 \end{bmatrix}\hat{\mathbf{a}} = \mathbf{b} \tag{5.25}$$

By defining the $P \times P$ matrix $\mathbf{I}_{P,k}$ with ones on the kth diagonal away from the leading diagonal, and zero elsewhere, then it follows:

$$\sum_{k=0}^{P} (\mathbf{I}_{P+1,k}\,\hat{\mathbf{a}})\; r_{xx}(k) = \mathbf{b} \tag{5.26}$$

Next defining the vector $\hat{\mathbf{a}}_k = \mathbf{I}_{P+1,k}\,\hat{\mathbf{a}}$ and the matrix $\begin{bmatrix} \hat{\mathbf{a}}_0 & \cdots & \hat{\mathbf{a}}_P \end{bmatrix}$, then the matrix-vector equation

$$\mathbf{A}\,\mathbf{r} = \mathbf{b} \tag{5.27}$$

has been obtained. In low-order cases, it might be more straightforward to explicitly compute the autocorrelation functions by writing out the **Yule-Walker equations**.

All-pole models therefore have the unique property that the model parameters are completely specified by the first $P+1$ autocorrelation coefficients via a set of linear equations, as given by the equation $\mathbf{A}\,\mathbf{r} = \mathbf{b}$. An alternative way of writing this is:

$$\begin{bmatrix} \sigma_w^2 \\ a_1 \\ \vdots \\ a_P \end{bmatrix} \leftrightarrow \begin{bmatrix} r_{xx}(0) \\ \vdots \\ r_{xx}(P) \end{bmatrix} \tag{5.28}$$

Thus, the mapping of the model parameters to the autocorrelation coefficients is reversible and unique. This *correlation matching* of **all-pole models** is quite remarkable, and is not shared by **all-zero models**, and is true for **pole-zero models** only under certain conditions.

Example 5.1 (Calculating Autocorrelation Functions of All-Pole Model). Given the parameters σ_w^2, a_1, and a_2, of a second-order all-pole model, compute the autocorrelation values $r_{xx}(k)$ for $\{k = 0, 1, 2\}$.

SOLUTION. Using the results above, it follows that:

$$r_{xx}(0)\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ a_1 \\ a_2 \end{bmatrix} + r_{xx}(1)\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ a_1 \\ a_2 \end{bmatrix} + r_{xx}(2)\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sigma_w^2 \\ 0 \\ 0 \end{bmatrix} \tag{5.29}$$

or,

$$\begin{bmatrix} 1 & a_1 & a_2 \\ a_1 & 1+a_2 & 0 \\ a_2 & a_1 & 1 \end{bmatrix}\begin{bmatrix} r_{xx}(0) \\ r_{xx}(1) \\ r_{xx}(2) \end{bmatrix} = \begin{bmatrix} \sigma_w^2 \\ 0 \\ 0 \end{bmatrix} \tag{5.30}$$

Although you could try a direct version to solve this, a slightly more ad-hoc approach quickly yields a solution in this case, and is related to Gaussian elimination. Multiplying the second row by a_1 and the last row by a_2, and then subtracting them both from the first row gives:

$$\begin{bmatrix} 1-a_1^2-a_2^2 & -2a_1a_2 & 0 \\ a_1^2 & a_1(1+a_2) & 0 \\ a_2^2 & a_1\,a_2 & a_2 \end{bmatrix}\begin{bmatrix} r_{xx}(0) \\ r_{xx}(1) \\ r_{xx}(2) \end{bmatrix} = \begin{bmatrix} \sigma_w^2 \\ 0 \\ 0 \end{bmatrix} \tag{5.31}$$

It can thus be seen that the first two equations for $r_{xx}(0)$ and $r_{xx}(1)$ do not depend on $r_{xx}(2)$ and therefore, by inverting the 2 by 2 matrix, this gives:

$$\begin{bmatrix} r_{xx}(0) \\ r_{xx}(1) \end{bmatrix} = \frac{1}{a_1(1+a_2)(1-a_1^2-a_2^2)+2a_1^3\,a_2} \begin{bmatrix} a_1(1+a_2) & 2a_1\,a_2 \\ -a_1^2 & 1-a_1^2-a_2^2 \end{bmatrix} \begin{bmatrix} \sigma_w^2 \\ 0 \end{bmatrix} \tag{5.32}$$

$$= \frac{\sigma_w^2}{(1-a_1^2-a_2^2)+\frac{2a_1^2\,a_2}{1+a_2}} \begin{bmatrix} 1 \\ -\frac{a_1}{1+a_2} \end{bmatrix} \tag{5.33}$$

Moreover,

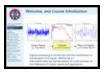
$$r_{xx}(2) = -\frac{1}{a_2} \begin{bmatrix} a_2^2 & a_1\,a_2 \end{bmatrix} \begin{bmatrix} r_{xx}(0) \\ r_{xx}(1) \end{bmatrix} = \frac{\sigma_w^2}{(1-a_1^2-a_2^2)+\frac{2a_1^2\,a_2}{1+a_2}} \left(\frac{a_1^2}{1+a_2} - a_2 \right) \tag{5.34}$$

In summary,

$$\begin{bmatrix} r_{xx}(0) \\ r_{xx}(1) \\ r_{xx}(2) \end{bmatrix} = \frac{\sigma_w^2}{(1-a_1^2-a_2^2)+\frac{2a_1^2\,a_2}{1+a_2}} \begin{bmatrix} 1 \\ -\frac{a_1}{1+a_2} \\ \frac{a_1^2}{1+a_2} - a_2 \end{bmatrix} \tag{5.35}$$

□

5.4.3 All-Zero models



New slide

Whereas **all-pole** models can capture resonant features of a particular PSD, it cannot capture *nulls* in the frequency response. These can only be modelled using a pole-zero or **all-zero** model.

The output of an all-zero model is the weighted average of delayed versions of the input signal. Thus, assume an all-zero model of the form:

$$x(n) = \sum_{k=0}^{Q} d_k\, w(n-k) \tag{M:4.3.1}$$

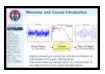
where Q is the order of the model, and the corresponding system function is given by:

$$H(z) = D(z) = \sum_{k=0}^{Q} d_k\, z^{-k} \tag{M:4.3.2}$$

Similar to the relationship between **all-pole models** and **AR processes**, **all-zero models** refer to the properties of a rational system containing only zeros, while **MA processes** refer to the resulting stochastic process that occurs as the result of **WGN** being applied to the input of an **all-zero filter**.

All-zero models are difficult to deal with since, unlike the **Yule-Walker equations** for the **all-pole model**, the solution for model parameters given the autocorrelation functions involves solving nonlinear equations, which becomes quite a complicated task.

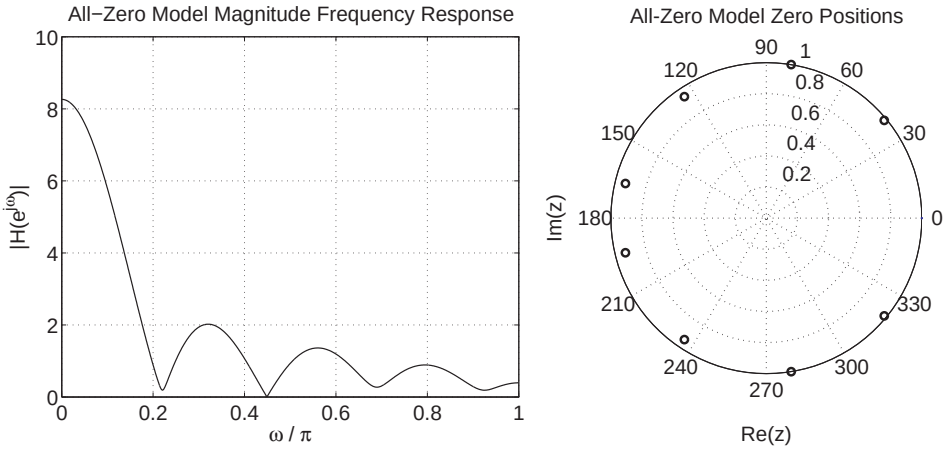
5.4.3.1 Frequency Response of an All-Zero Filter



New slide

The all-zero model has form:

$$H(z) = D(z) = \sum_{k=0}^{Q} d_k\, z^{-k} = d_0 \prod_{k=1}^{Q} \left(1 - z_k\, z^{-1}\right) \tag{5.36}$$



(a) Magnitude frequency response. (b) Positions of zeros.

Figure 5.8: The frequency response and position of the zeros in an all-zero system.

where $\{z_k\}$ are the zeros of the all-zero model. Therefore, its frequency response is given by:

$$H(e^{j\omega}) = \sum_{k=0}^{Q} d_k\, e^{-jk\omega} = d_0 \prod_{k=1}^{Q} \left(1 - z_k\, e^{-j\omega}\right) \tag{5.37}$$

When each of the zeros are written in the form $z_k = r_k e^{j\omega_k}$, then the frequency response can be written as:

$$H(e^{j\omega}) = d_0 \prod_{k=1}^{Q} \left(1 - r_k\, e^{-j(\omega-\omega_k)}\right) \tag{5.38}$$

Hence, it can be deduced that troughs or nulls occur near frequencies corresponding to the phase position of the zeros. When the system is real, the complex-zeros occur in conjugate-pairs.

Hence, the PSD of the output of an all-zero filter is given by:

$$P_{xx}(e^{j\omega}) = \sigma_w^2 \left|H(e^{j\omega})\right|^2 = G^2 \prod_{k=1}^{Q} \left|1 - r_k\, e^{-j(\omega-\omega_k)}\right|^2 \tag{5.39}$$

where $G = \sigma_w\, d_0$ is the overall gain of the system. Consider the all-zero model with zeros at positions:

$$\{z_k\} = \{r_k\, e^{j\omega_k}\} \quad \text{where} \quad \begin{cases} \{r_k\} = & \{0.985, 1, 0.942, 0.933\} \\ \{\omega_k\} = & 2\pi \times \{270, 550, 844, 1131\}/2450; \end{cases} \tag{5.40}$$

The zero positions and magnitude frequency response of this system is plotted in Figure 5.8. For comparison, the power spectral density of the output of the system is shown in Figure 5.9. Note that one of the zeros is on the unit circle, and that the frequency response at this point is zero.

5.4.3.2 Impulse Response

The impulse response of an **all-zero** model is an **FIR system** with impulse response:

$$h(n) = \begin{cases} d_n & 0 \le n \le Q \\ 0 & \text{elsewhere} \end{cases} \tag{5.41}$$

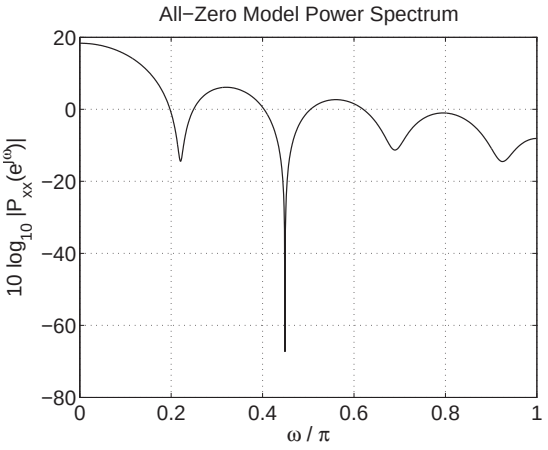


Figure 5.9: Power spectral response of an all-zero model.

5.4.3.3 Autocorrelation of the Impulse Response

Following a similar line to that shown for all-pole models, the autocorrelation of the impulse response of an all-zero system can be found.

Theorem 5.2. The autocorrelation sequence of the impulse response of an all-zero system is given by:

$$r_{hh}[\ell] = \sum_{n=-\infty}^{\infty} h[n]\ h^*[n-\ell] = \begin{cases} \sum_{k=0}^{Q-\ell} d_{k+\ell}\, d_k^* & 0 \le \ell \le Q \\ 0 & \ell > Q \end{cases} \tag{M:4.3.4}$$

and $r_{hh}^*[-\ell] = r_{hh}[\ell]$ for all ℓ.

PROOF. The autocorrelation sequence of the impulse response of an all-zero system is given by the discrete-time convolution:

$$r_{hh}[\ell] = \sum_{n=-\infty}^{\infty} h[n]\ h^*[n-\ell] \tag{5.42}$$

Considering the term $h[n]$,

$$h[n] = \begin{cases} d_n & 0 \le n \le Q \\ 0 & \text{otherwise} \end{cases} \tag{5.43}$$

or, in otherwords, $h(n) = 0$ when $n < 0$ and $n > Q$. Hence Equation 5.42 becomes:

$$r_{hh}[\ell] = \sum_{n=0}^{Q} d_n\, h^*[n-\ell] \tag{5.44}$$

Moreover, the lower-limit is constrained since

$$h^*[n-l] = \begin{cases} d_{n-\ell}^* & 0 \le n-\ell \le Q \\ 0 & \text{otherwise} \end{cases} \tag{5.45}$$

or, in otherwords, $h^*[n-l] = 0$ if $n < \ell$ and when $n > Q + \ell$. Assuming that $\ell \ge 0$, the second condition is already met by the upper-limit in Equation 5.44. Therefore, Equation 5.44 becomes:

$$r_{hh}[\ell] = \sum_{n=\ell}^{Q} d_n\, d_{n-\ell}^* \tag{5.46}$$

By substituting $k = n - \ell$, such that when $n = \{\ell,\, Q\}$, $k = \{0,\, Q - \ell\}$, then:

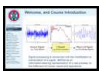
$$r_{hh}[\ell] = \sum_{k=0}^{Q-\ell} d_{k+\ell}\, d_k^*, \quad \text{for } \ell \geq 0 \tag{5.47}$$

Clearly this expression is equal to zero if $\ell > Q$. Therefore, using the result from the previous handout that $r_{hh}[\ell] = r_{hh}^*[-\ell]$, it follows:

$$r_{hh}[\ell] = \sum_{n=-\infty}^{\infty} h[n]\; h^*[n-\ell] = \begin{cases} \sum_{k=0}^{Q-\ell} d_{k+\ell}\, d_k^* & 0 \leq \ell \leq Q \\ 0 & \ell > Q \end{cases} \tag{M:4.3.4}$$

□

and $r_{hh}^*[-\ell] = r_{hh}[\ell]$ for all ℓ.



New slide

5.4.3.4 Moving-average processes

As an analogy with Section 5.4.2.5, a **MA process** refers to the stochastic process that is obtained at the output of an **all-zero filter** when a WGN sequence is applied to the input.

Thus, a MA process is an $AZ(Q)$ model with $d_0 = 1$ driven by WGN. That is,

$$x[n] = w[n] + \sum_{k=1}^{Q} d_k\, w[n-k]\,, \quad w[n] \sim \mathcal{N}\left(0,\, \sigma_w^2\right) \tag{M:4.3.9}$$

The output $x[n]$ has zero-mean, and variance of

$$\sigma_x^2 = \sigma_w^2 \left[1 + \sum_{k=1}^{Q} |d_k|^2\right] \tag{5.48}$$

The autocorrelation sequence and PSD are given by:

$$r_{xx}[\ell] = \sigma_w^2 r_{hh}[\ell] = \sigma_w^2 \sum_{k=0}^{Q-\ell} d_{k+l}\, d_k^*, \qquad \text{for } 0 \leq \ell \leq Q \tag{5.49}$$

and is zero for $\ell > Q$, with $r_{xx}[\ell] = r_{xx}^*[-\ell]$, where $d_0 = 1$, and also where $P_{xx}(e^{j\omega}) = \sigma_w^2\, |D(e^{j\omega})|^2$.

The fact that $r_{xx}[\ell] = 0$ if the samples are more than Q samples apart, means that they are therefore uncorrelated. An alternative derivation for the autocorrelation sequence for an MA process is given in the following section, Section 5.4.3.5.

5.4.3.5 Autocorrelation Function for MA Process

As stated in the previous section, using the results for the statistics of a stationary signal passed through a linear system, then the autocorrelation sequence for a MA process is given by $r_{xx}[\ell] = \sigma_w^2 r_{hh}[\ell]$, where $r_{hh}[\ell]$ is given by Equation M:4.3.4. For completeness, this section gives an alternative derivation from first principles.

Multiplying the difference equation, Equation M:4.3.1, through by $x^*[n-\ell]$ and taking expectations gives:

$$r_{xx}[\ell] = \sum_{k=0}^{Q} d_k\, r_{wx}[\ell - k] \tag{5.50}$$

Similarly, post-multiplying by $w^*[n-\ell]$ gives:

$$r_{xw}[\ell] = \sum_{k=0}^{Q} d_k \, r_{ww}[\ell - k] = \begin{cases} \sigma_w^2 \, d_\ell & 0 \le \ell \le Q \\ 0 & \text{otherwise} \end{cases} \tag{5.51}$$

since $r_{ww}[\ell] = \sigma_w^2 \, \delta(\ell)$. Recalling that $r_{wx}[\ell] = r_{xw}^*[-\ell]$, then:

$$r_{wx}[\ell] = \begin{cases} \sigma_w^2 \, d_{-\ell}^* & 0 \le -\ell \le Q \\ 0 & \text{otherwise} \end{cases} \tag{5.52}$$

with the limit $0 \le -\ell \le Q$ being equivalent to $-Q \le \ell \le 0$. Consequently,

$$r_{wx}[\ell - k] = \begin{cases} \sigma_w^2 \, d_{k-\ell}^* & 0 \le k - \ell \le Q \\ 0 & \text{otherwise} \end{cases} \tag{5.53}$$

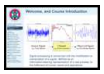
where this should be viewed as a function of k for a given ℓ. The range of non-zero values is given by $\ell \le k \le Q + \ell$. Considering $\ell > 0$, the autocorrelation sequence for an MA process is thus:

$$r_{xx}[\ell] = \sigma_w^2 \sum_{k=\ell}^{Q} d_k \, d_{k-\ell}^* = \sigma_w^2 \sum_{k=0}^{Q-\ell} d_{k+\ell} \, d_k^* \tag{5.54}$$

for $0 \le \ell \le Q$, and zero for $\ell > Q$. The last expression was been obtained by making the substitution $k \to k - \ell$. Using the relationship $r_{xx}[-\ell] = r_{xx}^*[\ell]$ gives the ACS for all values of ℓ.

Unlike AR models, is is not possible to solve for the model parameters using linear algebra techniques. It requires the solution of highly nonlinear equations, and is therefore more difficult than dealing with AR process. This, hence, is one reason why many algorithms in statistical signal processing prefer to use all-pole models over all-zero models.

5.4.4 Pole-Zero Models



New slide

Finally, the most general of LTI parametric signal models is the **pole-zero** model which, as the name suggests, is a combination of the all-pole and all-zero models, and can therefore model both resonances as well as nulls in a frequency response.

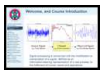
The output of a causal pole-zero model is given by the recursive input-output relationship:

$$x[n] = -\sum_{k=1}^{P} a_k \, x[n-k] + \sum_{k=0}^{Q} d_k \, w[n-k] \tag{M:4.4.1}$$

where it is assumed that the model orders $P > 0$ and $Q \ge 1$. The corresponding system function is given by:

$$H(z) = \frac{D(z)}{A(z)} = \frac{\sum_{k=0}^{Q} d_k \, z^{-k}}{1 + \sum_{k=1}^{P} a_k \, z^{-k}} \tag{5.55}$$

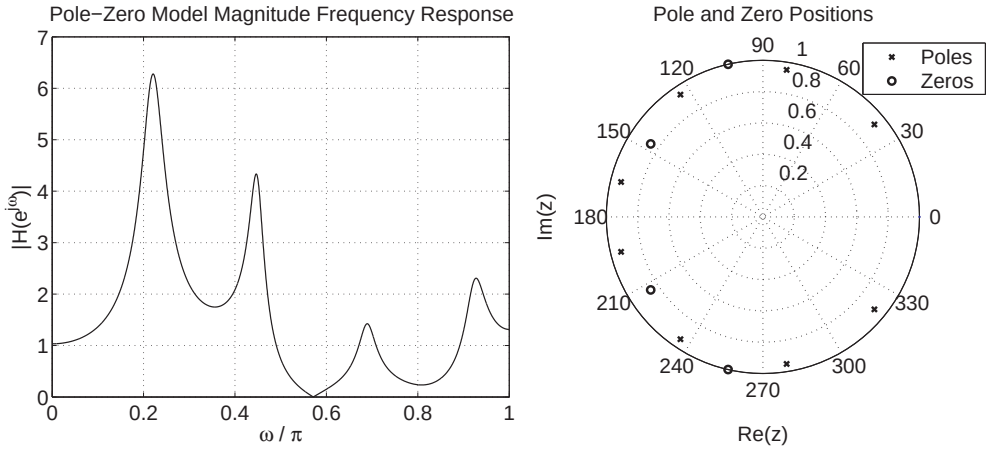
5.4.4.1 Frequency Response of an Pole-Zero Model



New slide

The pole-zero model can be written as

$$H(z) = \frac{D(z)}{A(z)} = d_0 \frac{\prod_{k=1}^{Q} (1 - z_k \, z^{-1})}{\prod_{k=1}^{P} (1 - p_k \, z^{-1})} \tag{5.56}$$



(a) Magnitude frequency response. (b) Positions of poles and zeros.

Figure 5.10: The frequency response and position of the poles and zeros in an pole-zero system.

where $\{p_k\}$ and $\{z_k\}$ are the poles and zeros of the pole-zero model. Therefore, its frequency response is given by:

$$H(e^{j\omega}) = d_0 \frac{\prod_{k=1}^{Q}(1 - z_k\, e^{-j\omega})}{\prod_{k=1}^{P}(1 - p_k\, e^{-j\omega})} \tag{5.57}$$

As before, it can be deduced that troughs or nulls occur at frequencies corresponding to the phase position of the zeros, while resonances occur at frequencies corresponding to the phase of the poles. When the system is real, the complex-poles and complex-zeros occur in conjugate-pairs.

The PSD of the output of a pole-zero filter is given by:

$$P_{xx}(e^{j\omega}) = \sigma_w^2 \left|H(e^{j\omega})\right|^2 = G^2 \frac{\prod_{k=1}^{Q}\left|1 - z_k\, e^{-j\omega}\right|^2}{\prod_{k=1}^{P}\left|1 - p_k\, e^{-j\omega}\right|^2} \tag{5.58}$$

where $G = \sigma_w\, d_0$ is the overall gain of the system.

Consider the pole-zero model with poles at positions:

$$\{p_k\} = \{r_k\, e^{j\omega_k}\} \quad \text{where} \quad \begin{cases} \{r_k\} = & \{0.925, 0.951, 0.942, 0.933\} \\ \{\omega_k\} = & 2\pi \times \{270, 550, 844, 1131\}/2450; \end{cases} \tag{5.59}$$

and zeros at:

$$\{z_k\} = \{r_k\, e^{j\omega_k}\} \quad \text{where} \quad \begin{cases} \{r_k\} = & \{1, 0.855\} \\ \{\omega_k\} = & 2\pi \times \{700, 1000\}/2450; \end{cases} \tag{5.60}$$

The pole and zero positions, and the magnitude frequency response of this system is plotted in Figure 5.10, while the PSD of the output of the system is shown in Figure 5.11. Note again that one of the zeros lies on the unit-circle, and therefore at the corresponding frequency, the frequency response is zero.

5.4.4.2 Impulse Response

The impulse response of a causal pole-zero filter can be obtained from Equation M:4.4.1 by substituting $w(n) = \delta(n)$ and $x(n) = h(n)$, such that:

$$h(n) = -\sum_{k=1}^{P} a_k\, h(n-k) + d_n, \quad n \geq 0 \tag{M:4.4.2}$$

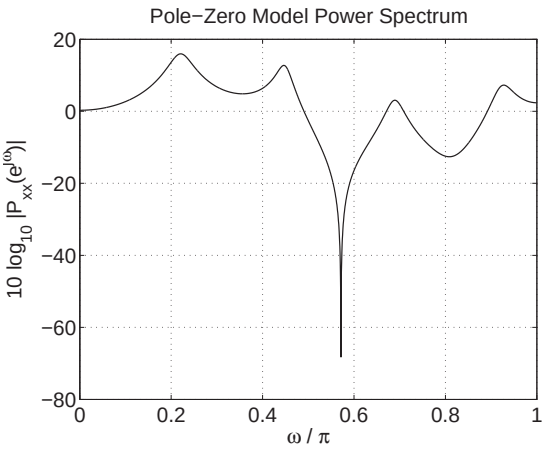


Figure 5.11: Power spectral response of an pole-zero model.

where $d_n = 0$ for $n > Q$ and $n < 0$, and $h(n) = 0$ for $n < 0$. Hence, writing this explicitly as:

$$h(n) = \begin{cases} 0 & n < 0 \\ -\sum_{k=1}^{P} a_k\, h(n-k) + d_n & 0 \le n \le Q \\ -\sum_{k=1}^{P} a_k\, h(n-k) & n > 0 \end{cases} \tag{5.61}$$

it can be seen that the impulse response obeys a *linear prediction* equation for $n > Q$. Thus, given $h(n)$ for $0 \le n \le P + Q$, the all-pole parameters $\{a_k\}$ can be calculated by using the P equations specified by $Q + 1 \le n \le P + Q$. Given the $\{a_k\}$'s, it is then possible to compute the all-zero parameters from Equation M:4.4.2 using the equations for $0 \le n \le Q$. Thus, it is clear that the first $P + Q + 1$ values of the impulse response completely specify the pole-zero model.

5.4.4.3 Autocorrelation of the Impulse Response

Multiplying both sides of Equation M:4.4.2 by $h^*(n-l)$ and summing over all n gives:

$$\sum_{n=-\infty}^{\infty} h(n)h^*(n-l) = -\sum_{k=1}^{P} a_k \sum_{n=-\infty}^{\infty} h(n-k)h^*(n-l) + \sum_{n=-\infty}^{\infty} d_n\, h^*(n-l) \tag{5.62}$$

where $d_n = 0$ if $n < 0$ or $n > Q$. Note that the order of summations in the middle term has implicitly been reordered. Using the definition for $r_{hh}(l)$ and noting that $h^*(n-l) = 0$ for $n - l < 0$ then this equation may be expressed as:

$$r_{hh}(l) = -\sum_{k=1}^{P} a_k\, r_{hh}(l-k) + \sum_{n=0}^{Q} d_n\, h^*(n-l) \tag{M:4.4.6}$$

Since the impulse response $h(n)$ is a function of the parameters $\{a_k\}$'s and $\{d_k\}$'s, then this set of equations is nonlinear in terms of these parameters. However, noting that the right hand side (RHS) of this equation is zero for $l > Q$, then:

$$\sum_{k=1}^{P} a_k\, r_{hh}(l-k) = -r_{hh}(l), \quad l > Q \tag{5.63}$$

This equation, unlike Equation M:4.4.6, is linear in the all-pole parameters $\{a_k\}$'s. Therefore, given the autocorrelation of the impulse response, the all-pole parameters can be calculated by solving

Equation 5.63 for $l \in \{Q+1, \ldots, Q+P\}$ to give:

$$\begin{bmatrix} r_{hh}(Q) & r_{hh}(Q-1) & \cdots & r_{hh}(Q+1-P) \\ r_{hh}(Q+1) & r_{hh}(Q) & \cdots & r_{hh}(Q+2-P) \\ \vdots & \vdots & \ddots & \vdots \\ r_{hh}(Q+P-1) & r_h(Q+P-2) & \cdots & r_{hh}(Q) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_P \end{bmatrix} = - \begin{bmatrix} r_{hh}(Q+1) \\ r_{hh}(Q+2) \\ \vdots \\ r_{hh}(Q+P) \end{bmatrix} \quad \text{(M:4.4.8)}$$

or, alternatively,

$$\mathbf{R}_{hh}\,\mathbf{a} = -\mathbf{r}_{hh} \quad \text{(M:4.4.9)}$$

The matrix $\mathbf{R}_{hh}$ in Equation M:4.4.9 is a non-Hermitian Toeplitz matrix; it can be solved using a variety of linear algebra techniques.

Given the all-pole parameters, it then falls to solve Equation M:4.4.6 for the all-zero parameters $\{d_k\}$'s. This is somewhat involved, but they can be found using spectral factorisation. The details are omitted from this handout, but can be found in [Therrien:1992, Section 9.1, page 509] or [Manolakis:2000, Page 178].

5.4.4.4 Autoregressive Moving-Average Processes

As with the all-pole and all-zero models, the corresponding random process associated with a pole-zero model is the **ARMA process**. This is the output of a pole-zero model, when the input of the system is driven by WGN. Hence, a causal ARMA model with model orders P and Q is defined by:

$$x(n) = -\sum_{k=1}^{P} a_k\, x(n-k) + w(n) + \sum_{k=1}^{Q} d_k\, w(n-k) \quad \text{(M:4.4.15)}$$

where $w(n) \sim \mathcal{N}(0,\, \sigma_w^2)$, the model-orders are P and Q, and the full set of model parameters are $\{\sigma_w^2,\, a_1,\, \ldots,\, a_P,\, d_1,\, \ldots,\, d_Q\}$. The output has zero-mean and variance that can be shown to equal:

$$\sigma_x^2 = -\sum_{k=1}^{P} a_k\, r_{xx}(k) + \sigma_w^2 \left[1 + \sum_{k=1}^{Q} d_k\, h(k)\right] \quad \text{(M:4.4.16)}$$

where $h(n)$ is the impulse response of the pole-zero filter. The derivation of this result is left as an exercise for the reader. The presence of $h(n)$ in the expression for the variance makes the dependence of σ_x^2 on the model parameters highly nonlinear.

Finally, in a manner similar to the derivations of the autocorrelation function for the other models, it can be shown that the autocorrelation function for the output is given by:

$$r_{xx}(l) = -\sum_{k=1}^{P} a_k\, r_{xx}(l-k) + \sigma_w^2 \left[1 + \sum_{n=l}^{Q} d_n\, h^*(n-l)\right] \quad (5.64)$$

where it has been noted that $d_0 = 1$. Similar in manner to Equation M:4.4.8 and Equation M:4.4.9 it is possible to obtain equations for finding the model parameters given the autocorrelation functions. Further details can be found in [Therrien:1992, Section 9.1, page 506]. The interested reader may wish to explore derivations of said equations.

5.5 Estimation of AR Model Parameters from Data

The Yule-Walker equations introduced earlier in this handout provide an approach for finding the model parameters for an AR process. Although a valid technique, there are two implicit assumptions that limit its use for practical problems. These assumptions are:

- That the order, P, of the model is known.
- That the correlation function, $r_{xx}[\ell]$, is known.

If these two conditions are met then, using the Yule-Walker equations, the model parameters, a_k, can be found exactly. Unfortunately, in most practical situations, *neither* of these conditions is met.

From a theoretical perspective, the first assumption that the model order is known is less of an issue than the second assumption. This is since if a larger model order than the true model order is chosen, then the excess parameters will theoretically be zero. In practice, choosing the models order is not that straightforward, and there are numerous methods for model order estimation. Model order selection criteria include names such as final prediction error (FPE), Akaike's information criterion (AIC), minimum description length (MDL), Parzen's criterion autoregressive transfer function (CAT) and B-Information criterion (BIC). There is not time in this course to discuss these techniques, although there are plenty of tutorial papers in the literature, as well as being covered by many text books.

The second assumption leads to both theoretical and practical problems since, if the correlation function is not known, it must be estimated from the data. This brings up the following questions:

1. If the correlation function is estimated, how good is the resulting estimate for the model parameters, in a statistical sense?
2. Why estimate the correlation function at all when it is the model parameters that need to be estimated?
3. What is the best procedure for this problem?

5.5.1 LS AR parameter estimation

Suppose that a particular realisation of a process that is to be modelled as an AR process is given. It is possible to estimate the correlation function as a time-average from the realisation, assuming that the process is time-ergodic, and then use these estimates in the Yule-Walker equations. The method described in this chapter effectively estimates the AR parameters in this way, although the problem is not formulated as such. Two common data-oriented methods, known as the **autocorrelation method** and the **covariance method**, are presented in this section and the next section. A description of these methods begins with the **autocorrelation method**.

Suppose linear prediction is used to model a particular realisation of a random process as accurately as possible. Thus, suppose a **linear predictor** forms an estimate, or *prediction*, $\hat{x}[n]$, of the present value of a stochastic process $x[n]$ from a linear combination of the past P samples; that is:

$$\hat{x}[n] = -\sum_{k=1}^{P} a_k \, x[n-k] \qquad \text{(M:1.4.1)}$$

Then the **prediction error** is given by:

$$e[n] = x[n] - \hat{x}[n] = x[n] + \sum_{k=1}^{P} a_k\, x[n-k] \tag{M:4.2.50}$$

Note that this is different to the WGN sequence that drives a linear system to generate an autoregressive random process; the difference is that here, the prediction error is the difference between the actual value and the predicted value of a *particular realisation* of a random process.

Writing Equation M:4.2.50 for $n \in \{n_I, \ldots, n_F\}$, in matrix-vector form:

$$\underbrace{\begin{bmatrix} e[n_I] \\ e[n_I+1] \\ \vdots \\ e[n_F] \end{bmatrix}}_{\mathbf{e}} = \underbrace{\begin{bmatrix} x[n_I] \\ x[n_I+1] \\ \vdots \\ x[n_F] \end{bmatrix}}_{\mathbf{x}} + \underbrace{\begin{bmatrix} x[n_I-1] & x[n_I-2] & \cdots & x[n_I-P] \\ x[n_I] & x[n_I-1] & \cdots & x[n_I-P+1] \\ \vdots & \vdots & \cdots & \vdots \\ x[n_F-1] & x[n_F-2] & \cdots & x[n_F-P] \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_P \end{bmatrix}}_{\mathbf{a}} \tag{5.65}$$

which can hence be written as:

$$\mathbf{e} = \mathbf{x} + \mathbf{X}\,\mathbf{a} \tag{5.66}$$

The parameters $\mathbf{a}$ can now be estimated using any of the parameter estimation techniques discussed above. Here, the least-squares estimate (LSE) is used. Thus, noting that:

$$J(\mathbf{a}) = \sum_{n=n_I}^{n_F} e^2[n] = \mathbf{e}^T\mathbf{e} \tag{5.67}$$

$$= (\mathbf{x} + \mathbf{X}\,\mathbf{a})^T (\mathbf{x} + \mathbf{X}\,\mathbf{a}) \tag{5.68}$$

$$= \mathbf{x}^T\mathbf{x} + 2\mathbf{x}^T\mathbf{X}\,\mathbf{a} + \mathbf{a}^T\mathbf{X}^T\mathbf{X}\mathbf{a} \tag{5.69}$$

where it has been noted that $\mathbf{a}^T\mathbf{X}^T\mathbf{x} = \mathbf{x}^T\mathbf{X}\,\mathbf{a}$. Hence, differentiating with respect to (w. r. t.) $\mathbf{a}$ and setting to zero gives the LSE, $\hat{\mathbf{a}}$. Noting that for real vector $\mathbf{a}$,

$$\frac{\partial}{\partial \mathbf{a}}\left(\mathbf{b}^T\,\mathbf{a}\right) = \mathbf{b} \quad \text{and} \quad \frac{\partial}{\partial \mathbf{a}}\left(\mathbf{a}^\mathbf{T}\,\mathbf{B}\mathbf{a}\right) = \left(\mathbf{B} + \mathbf{B}^T\right)\mathbf{a} \tag{5.70}$$

The reader is invited to derive these results. Hence,

$$\frac{\partial J(\mathbf{a})}{\partial \mathbf{a}} = 2\mathbf{X}^T\,\mathbf{x} + 2\mathbf{X}^T\,\mathbf{X}\,\mathbf{a} \tag{5.71}$$

where it has been noted that the matrix $\mathbf{X}^T\mathbf{X}$ is symmetric. Setting this to zero, and rearranging noting that $\mathbf{X}^T\mathbf{X}$ is of full rank, gives the LSE:

$$\mathbf{a}_{LSE} = -\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\,\mathbf{x} \tag{5.72}$$

Defining $N_p = n_F - n_I + 1$, the least-squares (LS) error is then given by:

$$J\left(\mathbf{a}_{LSE}\right) = \mathbf{x}^T\left(\mathbf{I}_{N_p} - \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\right)\mathbf{x} \tag{5.73}$$

$$= \mathbf{x}^T\left(\mathbf{x} + \mathbf{X}\,\mathbf{a}_{LSE}\right) \tag{5.74}$$

Observe the similarity of these results with those of the linear LS formulation. In fact, this derivation is identical to the LS formulation with the matrix $\mathbf{H}$ replaced by $\mathbf{X}$! There are two different methods which result from different choices of n_I and n_F. These are called the **autocorrelation method** and the **covariance method**. However, as mentioned in [Therrien:1991], these terms do not bear any relation to the statistical meanings of these terms, and so they should not be confused with the statistical definitions. The names for these methods are unfortunate, but have found a niche in signal processing, and are unlikely to be changed.

5.5.2 Autocorrelation Method

In the **autocorrelation method**, the end points are chosen as $n_I = 0$ and $n_F = N + P - 1$. Thus, the AR filter model runs over the entire length of the data, predicting some of the early points from *zero valued samples*, and predicting P additional zero values at the end. Since this method uses zeros for the data outside of the given interval, it can be thought of as applying a rectangular window to the data. For this method, the $(N + P) \times P$ data matrix $\mathbf{X}$ has the specific structure:

$$\mathbf{X} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ x[0] & 0 & \cdots & 0 \\ x[1] & x[0] & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ x[P-1] & x[P-2] & \cdots & x[0] \\ x[P] & x[P-1] & \cdots & x[1] \\ \vdots & \vdots & & \vdots \\ x[N-1] & x[N-2] & \cdots & x[N-P] \\ 0 & x[N-1] & \cdots & x[N-P+1] \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x[N-1] \end{bmatrix} \tag{T:9.112}$$

When formed into the product $\mathbf{X}^T\mathbf{X}$, this data matrix produces a Toeplitz correlation matrix; consequently, the **normal equations** may be solved very efficiently, for example using the **Levinson recursion**. Moreover, the matrix $\mathbf{X}^T\mathbf{X}$ is strictly positive definite, and thus a valid *correlation matrix*.

5.5.3 Covariance Method

An alternative method is to choose $n_I = P$ and $n_F = N - 1$. With this method, no zeros are either predicted, or used in the prediction. In other words, the limits are chosen so that the data that the AR filter operates on always remain within the measured data; no window is applied. For this method, the $(N - P) \times P$ data matrix has the specific form:

$$\mathbf{X} = \begin{bmatrix} x[P-1] & x[P-2] & \cdots & x[0] \\ x[P] & x[P-1] & \cdots & x[1] \\ \vdots & \vdots & \ddots & \vdots \\ x[N-2] & x[N-3] & \cdots & x[N-P-1] \end{bmatrix} \tag{T:9.113}$$

A variation of this method called the **prewindowed covariance method** chooses $n_I = 0$ and $n_F = N-1$, and results in a data matrix that consists of the first N rows of Equation T:9.112. Moreover, the **postwindowed covariance method** chooses $n_I = P$ and $n_F = N + P - 1$. In the **autocorrelation method**, the data is said to be both *prewindowed* and *postwindowed*.

With the covariance method, or the prewindowed covariance method, the resulting correlation matrix is positive semidefinite, but it is *not* Toeplitz. Thus, the Yule-Walker equations are more difficult to solve. Moreover, the resulting AR model *may not be stable*, since the poles corresponding to the estimated parameters may not lie within the unit circle. Nevertheless, unstable cases rarely seem to occur in practice, and the covariance method is often preferred because it makes use of only the measured data. This avoids any bias in the estimation of the AR filter coefficients. In addition, with some mild conditions, the method can be shown to be equivalent to the maximum-likelihood estimate (MLE).

Example 5.2 ([Therrien:1991, Example 9.6, Page 539]). It is desired to estimate the parameters of a second-order AR model for the sequence $\{x[n]\}_0^4 = \{1, -2, 3, -4, 5\}$ by using both the autocorrelation and covariance methods.

SOLUTION. Applying both methods as requested:

Autocorrelation Method The data matrix can be obtained from Equation T:9.112, and is given by:

$$\mathbf{X} = \begin{bmatrix} 0 & 0 \\ x[0] & 0 \\ x[1] & x[0] \\ x[2] & x[1] \\ x[3] & x[2] \\ x[4] & x[3] \\ 0 & x[4] \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ -2 & 1 \\ 3 & -2 \\ -4 & 3 \\ 5 & -4 \\ 0 & 5 \end{bmatrix} \tag{5.75}$$

Hence, it can be shown that:

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 0 & 1 & -2 & 3 & -4 & 5 & 0 \\ 0 & 0 & 1 & -2 & 3 & -4 & 5 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ -2 & 1 \\ 3 & -2 \\ -4 & 3 \\ 5 & -4 \\ 0 & 5 \end{bmatrix} \tag{5.76}$$

$$= \begin{bmatrix} 55 & -40 \\ -40 & 55 \end{bmatrix} \tag{5.77}$$

Note that the matrix is Toeplitz. The least squares Yule-Walker equations can then be found by solving:

$$\mathbf{a}_{LSE} = -\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{x} \tag{5.78}$$

$$= -\begin{bmatrix} 55 & -40 \\ -40 & 55 \end{bmatrix}^{-1} \begin{bmatrix} 0 & 1 & -2 & 3 & -4 & 5 & 0 \\ 0 & 0 & 1 & -2 & 3 & -4 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 3 \\ -4 \\ 5 \\ 0 \\ 0 \end{bmatrix} \tag{5.79}$$

Solving these equations gives:

$$\mathbf{a}_{LSE} = \begin{bmatrix} \frac{232}{285} \\ \frac{34}{285} \end{bmatrix} \approx \begin{bmatrix} 0.8140 \\ 0.1193 \end{bmatrix} \tag{5.80}$$

The LS error is then given by:

$$J\left(\mathbf{a}_{LSE}\right) = \mathbf{x}^T\left(\mathbf{x} + \mathbf{X}\mathbf{a}_{LSE}\right) = 25.54 \tag{5.81}$$

Hence, the **prediction error variance** is estimated as:

$$\sigma_e^2 = \frac{J\left(\mathbf{a}_{LSE}\right)}{N} = \frac{25.54}{7} = 3.64 \tag{5.82}$$

Covariance Method Next, apply the covariance method to the same problem. Since the AR filter stays entirely within the data, the error is evaluated from $n = 2$ to $n = 4$. The data matrix is therefore:

$$\mathbf{X} = \begin{bmatrix} x[1] & x[0] \\ x[2] & x[1] \\ x[3] & x[2] \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 3 & -2 \\ -4 & 3 \end{bmatrix} \tag{5.83}$$

Notice that, in this data matrix, not all the the data has been used, since $x[4]$ does not appear. Hence, the correlation matrix is given by:

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} -2 & 3 & -4 \\ 1 & -2 & 3 \end{bmatrix} \begin{bmatrix} -2 & 1 \\ 3 & -2 \\ -4 & 3 \end{bmatrix} = \begin{bmatrix} 29 & -20 \\ -20 & 14 \end{bmatrix} \tag{5.84}$$

This matrix is not Toeplitz. The LSE estimate is therefore:

$$\mathbf{a}_{LSE} = -\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{x} \tag{5.85}$$

$$= -\begin{bmatrix} 29 & -20 \\ -20 & 14 \end{bmatrix}^{-1} \begin{bmatrix} -2 & 3 & -4 \\ 1 & -2 & 3 \end{bmatrix} \begin{bmatrix} 3 \\ -4 \\ 5 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \tag{5.86}$$

Moreover, the LS error is then given by:

$$J\left(\mathbf{a}_{LSE}\right) = \mathbf{x}^T\left(\mathbf{x} + \mathbf{X}\mathbf{a}_{LSE}\right) = 0 \tag{5.87}$$

Hence, the **prediction error variance** is estimated as:

$$\sigma_e^2 = \frac{J\left(\mathbf{a}_{LSE}\right)}{N} = \frac{0}{3} = 0 \tag{5.88}$$

Evidently, this filter predicts the data perfectly. Indeed, if the prediction error, $e[n]$, is computed over the chosen range $n = 2$ to $n = 4$, it is found to be zero at every point. The price to be paid for this perfect prediction, however, is an unstable AR model. The transfer function for this AR model can be written as:

$$H(z) = \frac{1}{1 + 2z^{-1} + z^{-2}} = \frac{1}{(1 + z^{-1})^2} \tag{5.89}$$

□

which has a double pole at $z = -1$. Therefore, a bounded-input into this filter can potentially produce an unbounded-output. Further, any errors in computation of the model coefficients can easily put a pole outside of the unit circle.