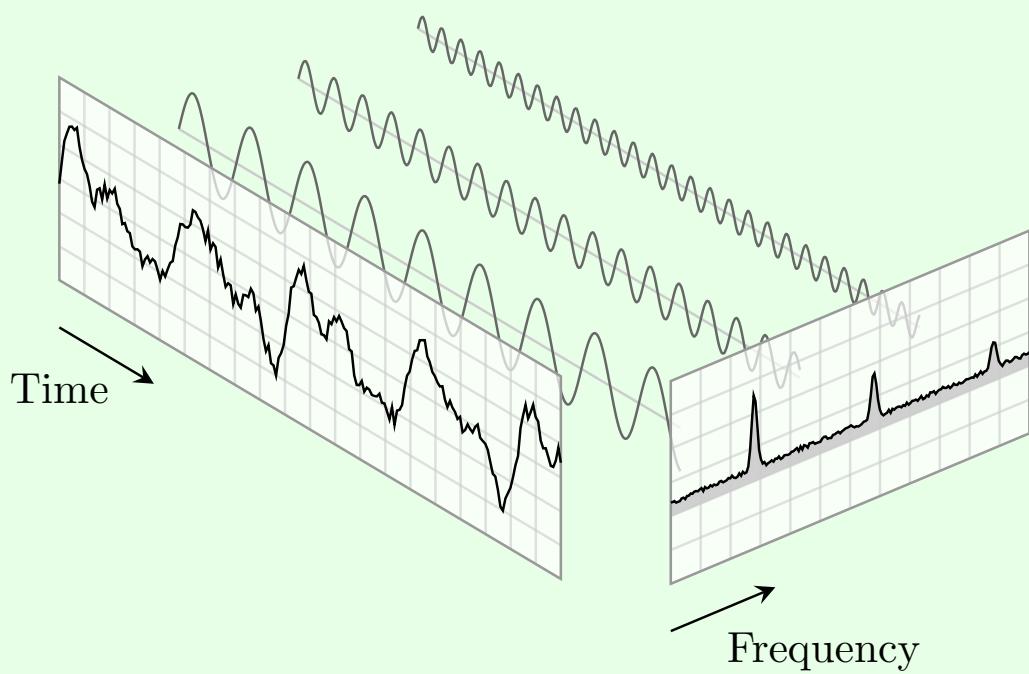

**LECTURE NOTES
OF
DIGITAL SIGNAL PROCESSING**

COLLECTION OF THE LECTURES NOTES OF PROFESSOR FEDERICA BATTISTI.

EDITED BY
ARDINO ROCCO
ACADEMIC YEAR 2020-2021



Abstract

In this document I have tried to reorder the notes of the Digital Signal Processing course held by Professor Federica Battisti at DEI of the University of Padua during the first semester of the 2020-21 academic year of the master's degree in Physics of Data.

The notes are fully integrated with the material provided by the professor in the Moodle platform. In addition, I will integrate them, as best as possible, with the books recommended by the professor.

There may be formatting errors, wrong marks, missing exponents and even missing parts, since I'm still working on them. If you find errors or if you have any suggestions, let me know (you can send an e-mail at rocco.ardino@studenti.unipd.it, labeled with **DSP::TYPO/SUGGESTION**) and I will correct/integrate them, so that this document can be a good study support. However, these notes are not to be intended as a substitute of the lectures held by the professor or of lecture notes made by other people.

Padova, Sunday 3rd January, 2021
Rocco Ardino

Contents

1	Introduction	1
1.1	What is a signal?	1
1.2	Is sampling of a continuous signal always possible?	3
1.3	Discrete time and amplitude	3
1.4	Advantages and disadvantages of digital signals	5
2	Discrete-Time signals	7
2.1	Time-Domain representation and basic definitions	7
2.1.1	Categorization of discrete-time signals	8
2.1.2	Norm of a discrete-time signal	9
2.2	Operations on sequences	10
2.2.1	Basic operations	10
2.2.2	Ensamble averaging	11
2.2.3	Samping rate alteration	12
2.3	Classification of sequences	13
2.4	Basic signals	16
2.4.1	Unit sample sequence	17
2.4.2	Unit step sequence	17
2.4.3	Real sinusoidal sequence	17
2.4.4	Exponential sequence	17
3	Discrete-Time systems	21
3.1	Examples of discrete-time systems	21
3.1.1	Accumulator system	21
3.1.2	M -point moving average system	22
3.1.3	Exponentially weighted running average system	23
3.1.4	Linear interpolation system	23
3.1.5	Median system	24
3.2	Classification of discrete-time systems	25
3.2.1	Linear systems	25
3.2.2	Shift-Invariant systems	27
3.2.3	Causal systems	28
3.2.4	Stable systems	29
3.2.5	Passive and lossless systems	29
3.3	Impulse and step response	29
3.4	Convolution sum	31
3.5	Simple interconnection schemes	35
3.5.1	Cascade connection	35
3.5.2	Parallel connection	36
3.6	Stablility and causality conditions	38
3.6.1	Stability condition of an LTI discrete-time system	38
3.6.2	Causality condition of an LTI discrete-time system	39

3.7	Classification of LTI discrete-time systems	40
3.8	Correlation and autocorrelation	41
3.8.1	Properties of autocorrelation and cross-correlation	43
3.8.2	Normalized forms of correlation	44
3.8.3	Particular cases of correlation computation	44
4	Fourier Analysis	47
4.1	Continuous-time Fourier transform	47
4.2	Discrete-time Fourier transform	49
4.3	Linear convolution using DTFT	58
4.4	The frequency response	60
4.5	The concept of filtering	64
4.6	Discrete Fourier Transform	66
4.6.1	Matrix relations	69
4.6.2	DTFT from DFT by interpolation and sampling of DTFT	70
4.6.3	DFT properties	72
4.6.4	Circular shift and convolution of a sequence	72
4.7	DFT of real sequences	77
4.8	Linear convolution using the DFT	79
5	The z-transform	85
5.1	Definition of z -transform	85
5.2	Rational z -transforms	88
5.2.1	Finite-length sequence z -transform	89
5.2.2	Right-sided sequence z -transform	89
5.2.3	Left-sided sequence z -transform	90
5.2.4	Two-sided sequence z -transform	90
5.3	Inverse z -transform	91
5.4	z -transform properties	93
5.5	Stability condition	95
6	Filter design	99
6.1	Preliminary concepts and definitions	99
6.1.1	Phase delay	99
6.1.2	Group delay	99
6.1.3	Types of transfer function	100
6.1.4	Bounded real transfer function	102
6.1.5	Allpass transfer function	103
6.1.6	Phase characteristics	106
6.1.7	Frequency response from transfer function	108
6.2	Simple digital filters	110
6.2.1	Lowpass FIR digital filters	110
6.2.2	Highpass FIR digital filters	112
6.2.3	Lowpass IIR digital filters	113
6.2.4	Highpass IIR digital filters	114
6.2.5	Bandpass IIR digital filters	115
6.2.6	Bandstop IIR digital filters	117
6.2.7	Higher-Order IIR digital filters	117
6.3	Linear-phase FIR transfer functions	118
6.3.1	Symmetric impulse response with odd length	120
6.3.2	Symmetric impulse response with even length	122
6.3.3	Antisymmetric impulse response with odd length	123
6.3.4	Antisymmetric impulse response with even length	124

6.3.5	General form of frequency response	124
6.3.6	Zero locations	126
6.4	Comb filters	128
6.5	Digital filter structures	130
6.5.1	Block diagram representation	131
6.5.2	The delay-free loop problem	133
6.5.3	Canonic and non-canonic structures	134
6.5.4	Equivalent structures	135
6.6	Basic FIR digital filter structures	136
6.6.1	Direct form FIR filter structures	136
6.6.2	Cascade form FIR filter structures	136
6.6.3	Polyphase FIR structures	137
6.6.4	Linear-phase FIR structures	139
6.6.5	Tapped delay line	140
6.7	Basic IIR digital filter structures	140
6.7.1	Direct form IIR filter structures	141
6.7.2	Cascade form IIR filter structures	143
6.7.3	Parallel form IIR filter structures	145

Bibliography**147**

Chapter 1

Introduction

In this course we will enter into the world of Digital Signal Processing, trying to understand the theoretical fundations of this field with several examples. In particular, we will follow this outline of arguments:

- What is Digital Signal Processing?
- Discrete-time signals.
- Signals and Hilbert spaces.
- Fourier analysis.
- Discrete-time filters.
- The z-transform.
- Filter design, in particular A/D and D/A conversions and the design of a digital communication system.

Concerning the study material, along with the teaching professor's notes and slides, several textbooks are suggested to integrate the study material:

- [1] *Signal Processing for Communication*, P. Prandoni and M. Vetterli.
- [2] *Digital Signal Processing*, A. V. Oppenheim and R. W. Schafer.
- [3] *Digital Signal Processing: A Computer-Based Approach*, S. K. Mitra.

1.1 What is a signal?

Let us start with a definition of what we are going to deal with for most of our time. A signal is the description of the evolution of a physical phenomenon. For example, we can have:

- **continuous-time signals**, such as voltage, current, temperature and speed, whose spectrum is sketched in Figure 1.1;
- **discrete-time signals**, such as minimum/maximum temperature, lap intervals in races, sampled continuous signals, with an example spectrum showed in Figure 1.2.

Signals may have to be transformed in order to:

Lecture 1.
Tuesday 29th
September, 2020.

Signal definition
and types of signals

Operations on
signals

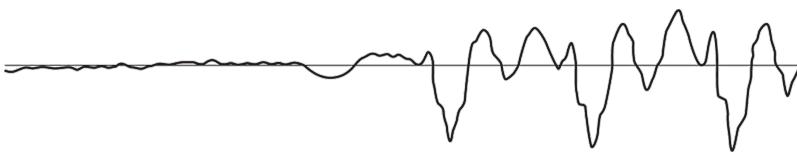


Figure 1.1: Example of continuous-time signal spectrum.

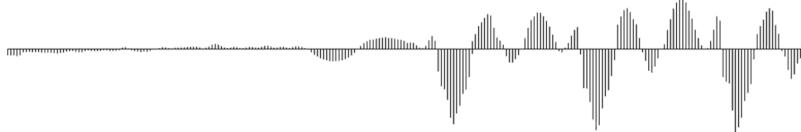


Figure 1.2: Example of discrete-time signal spectrum, obtained from sampling of the signal in Figure 1.1.

- be analyzed: this is necessary to **understand** the informations carried by the signal itself. For example, we can amplify or filter out embedded information, detect patterns, prepare the signal to survive a transmission channel and much more;
- be synthesized: this is necessary to **create** a signal to contain a specific information. Some common examples in this case are the cellphones, tv, radio and synthetic music.

Therefore, to do these operations we need specific methods to measure, characterize, model and simulate transmission channels, but also mathematical tools that split common channels and transformations into easily manipulated building blocks.

From analog to digital signals

Now, we enter more into the detail of digital signals. When we are dealing with analog signals, we can record them but we can hardly find a function that can describe them. In order to have an idea of the difficulty of this task, let us look at the signal in Figure 1.3: how can we describe this function? This is where the digital part comes into practice.

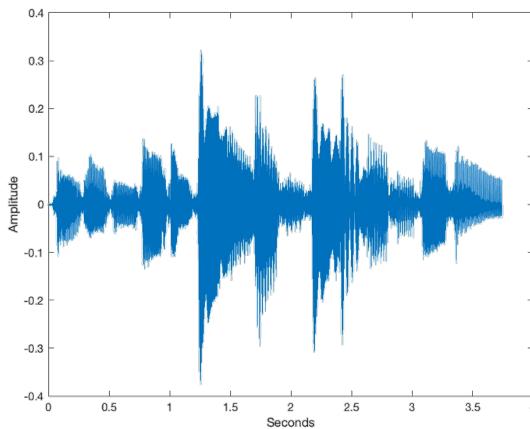


Figure 1.3: Example of analog signal.

Mathematical definition of digital signals

A continuous-time (analog) signal is defined along a continuum of time and is thus represented by a continuous independent variable. Discrete-time signals are defined at discrete times, and thus, the independent variable has discrete values. Therefore, the ladders are represented as sequences of numbers x , in which the n^{th} number in the sequence is denoted $x[n]$, and it is formally written as:

$$x = \{x[n]\}, \quad -\infty < n < \infty \quad (1.1)$$

where n is an integer. An example is showed in Figure 1.4.

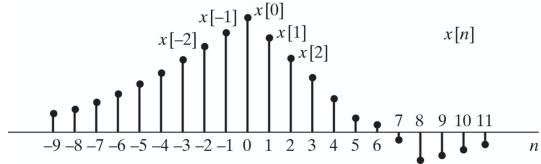


Figure 1.4: Example of graphic representation of a discrete-time signal.

Digital signal
sampled from
periodic signals

Such sequences can arise from periodic sampling of an analog signal $x_a(t)$. In this case the numeric value of the n^{th} number in the sequence is equal to the value of the analog signal, namely $x_a(t)$, at time nT :

$$x[n] = x_a(nT), \quad -\infty < n < \infty \quad (1.2)$$

where T is the sampling period.

1.2 Is sampling of a continuous signal always possible?

At this point we ask ourselves a question: when we represent an analog signal as a digital one, are we loosing information? The question has been answered by C. Shannon and H. Nyquist: under specific conditions, the analog and the digital representations of a signal are equivalent. Here we come to the sampling theorem. An analog signal can be represented as a linear combination of cardinal sine (denoted by sinc and plotted in Figure 1.5) functions shifted ad scaled by the values of the discrete time sequence. Mathematically, this translates into:

$$x(t) = \sum_{n=-\infty}^{\infty} x[n] \operatorname{sinc}\left(\frac{t - nT_S}{T_S}\right) \quad (1.3)$$

Sampling theorem

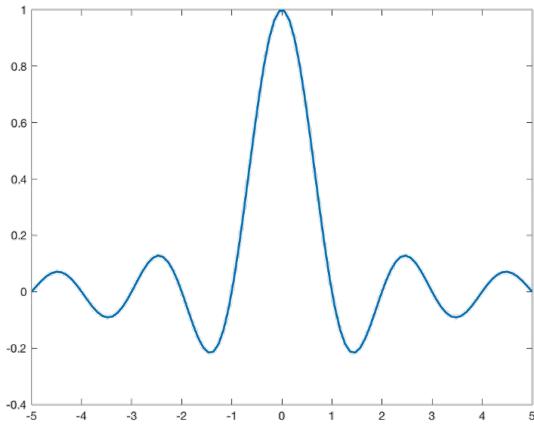


Figure 1.5: Cardinal sine sinc function.

Utility of Fourier
transform

It is important to determine the value of T_S so that the sampling theorem holds. The Fourier transform helps us understanding how fast the signal moves and consequently guides us to the selection of T_S .

1.3 Discrete time and amplitude

Till now we discussed the discretization of time. What if we discretize also the amplitude of the signal? In this case each sample can take predetermined values

in a countable set. Moreover, as consequence, we can always map the level of the sample onto an integer. An example of a sampled sin function is showed in Figure 1.6.

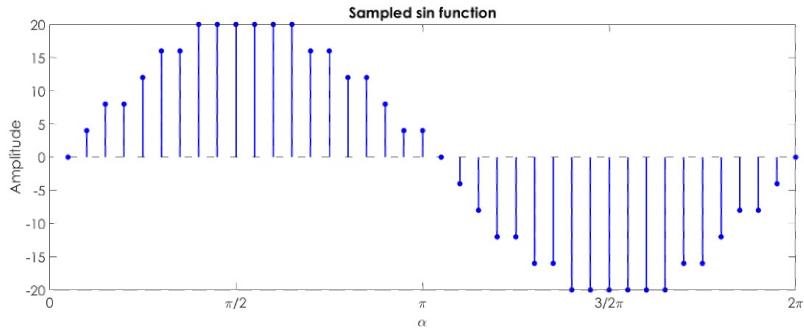


Figure 1.6: Sampled sin function.

It is of paramount importance to discretize both time and amplitude. By this way, we can ease tasks like storage, processing and transmission. In particular, let us describe the ladder.

Signal transmission

Transmission is one of the fields that most benefits from discrete signal representation. We give an idea of this fact through an example. Let us consider the channel scheme in Figure 1.7 and a sinusoidal analog signal, which is attenuated and acquires a noise component, as represented in Figure 1.9

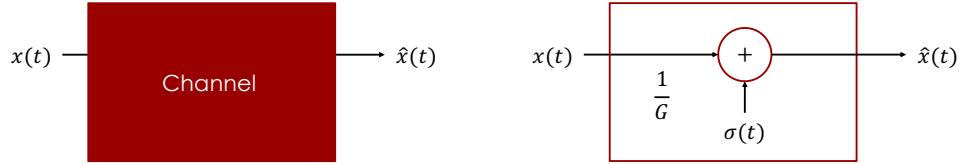


Figure 1.7: Scheme of an example of transmission channel.

We can counteract this problem by trying to undo the errors introduced by the signal (only the gain):

$$\hat{x}_1(t) = G \left[\frac{x(t)}{G} + \sigma(t) \right] = x(t) + G\sigma(t) \quad (1.4)$$

By this way, we also amplify the noise. Sometimes this is unavoidable, but what we can do is to amplify the signal as near to the place where it is acquired as possible, so that the noise is amplified before it increases in the transmission to the place where it is analyzed. In particular, noise amplification becomes more relevant if we put in cascade several transmission blocks, as represented in Figure 1.8. In this case, the output of the transmission chain reads:

$$\hat{x}_N(t) = x(t) + NG\sigma(t) \quad (1.5)$$

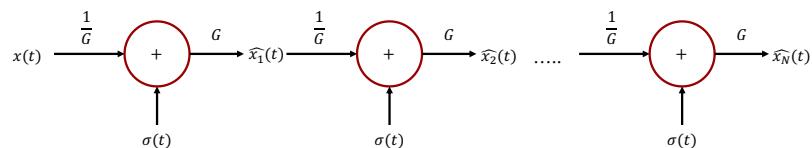


Figure 1.8: Scheme of an example of a cascade of transmission blocks.

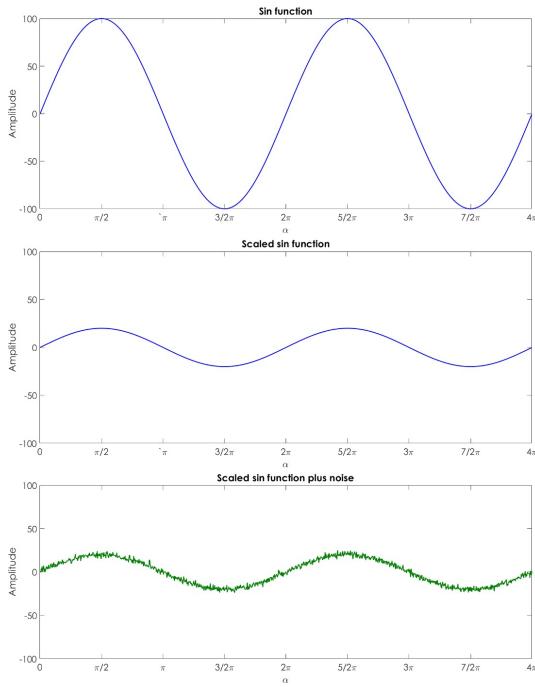


Figure 1.9: Sinusoidal analog signal (top), sinusoidal attenuated signal (center), sinusoidal attenuated noisy signal (bottom).

This discussion was mainly related to analog signals. Now, if we take into account the transmission of a digital signal, we have to perform other operations, such as introducing a thresholding operator to discretize the signal. With respect to the previous example, the new transmission block is showed in Figure 1.10 and the output reads:

$$\hat{x}_1(t) = \text{sign} [x(t) + G\sigma(t)] \quad (1.6)$$

which is in practice a square wave.

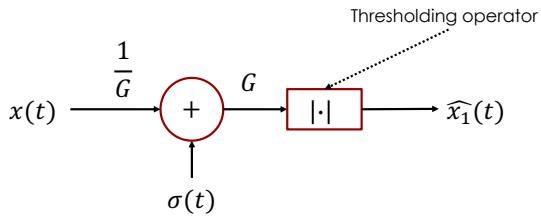


Figure 1.10: Scheme of an example of transmission channel for digital signals.

1.4 Advantages and disadvantages of digital signals

To conclude this introduction, we have still to present the main advantage of digital signals, but also their drawbacks:

Pro and cons

- ✓ noise is easy to control after initial quantization;
- ✓ highly linear (within limited dynamic range);
- ✓ complex algorithms fit into a single chip;
- ✓ flexibility, parameters can easily be varied in software;

- ✓ digital processing is insensitive to component tolerances, aging, environmental conditions, electromagnetic interference;
- ✗ discrete-time processing artifacts (aliasing);
- ✗ a significantly large amount of power can be required (battery, cooling);
- ✗ digital clock and switching cause interference.

Chapter 2

Discrete-Time signals

In this Chapter we define more formally the concept of the discrete-time signal and some other properties useful for their description. Historically, discrete-time signals have often been introduced as the discretized version of continuous-time signals, i.e. as the sampled values of analog quantities. For this reason, many of the derivations proceeded within the framework of an underlying continuous-time reality.

Lecture 2.
Thursday 1st
October, 2020.

2.1 Time-Domain representation and basic definitions

Signals are represented as **sequences of numbers**, also called **samples**. Typically, the value of a signal sample is denoted as $x[n]$, with n being an integer in the range. So, $x[n]$ is defined only for $n \in \mathbb{Z}$ and undefined otherwise, namely for non-integer values of n .

Sample sequence

In the introduction we have already presented the common notation for discrete-time signals, namely $\{x[n]\}$, with $-\infty < n < \infty$. Sometimes the numerical values of the samples are explicitly written inside the curly braces, for example:

$$\{x[n]\} = \{\dots, -0.2, \textcolor{red}{2.2}, 1.1, 0.2, -3.7, 2.9, \dots\} \quad (2.1)$$

where the element in red is the sample at time index $n = 0$.

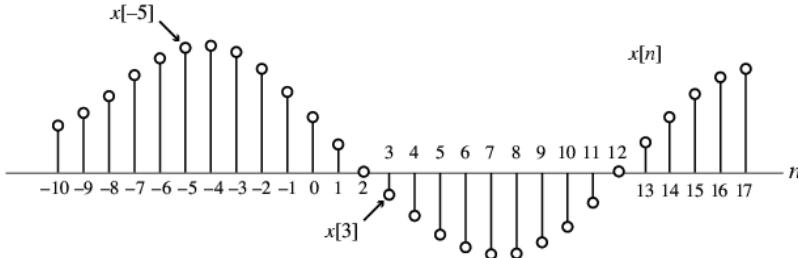


Figure 2.1: Example of sample sequence.

In some applications, a discrete-time sequence $\{x[n]\}$ may be generated by **periodically sampling** a continuous-time signal $x_a(t)$ at uniform intervals of time. An example is showed in Figure 2.2. Here, the n^{th} sample is given by:

Periodic sampling

$$x[n] = x_a(t)_{t=nT} = x_a(nT), \quad n = \dots, -2, -1, 0, 1, 2, \dots \quad (2.2)$$

The spacing T between two consecutive samples is called the **sampling interval** or **sampling period**. The reciprocal of sampling interval T , denoted as F_T , is called the **sampling frequency**:

Sampling period
and frequency

$$F_T = \frac{1}{T} \quad (2.3)$$

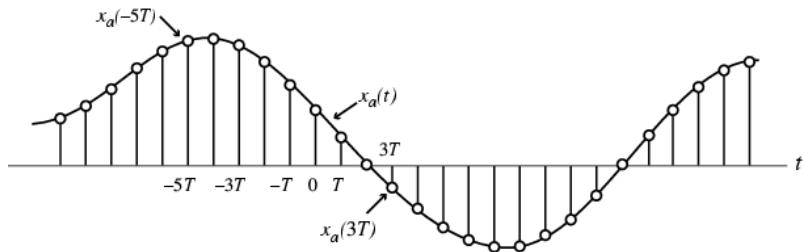


Figure 2.2: Example of periodically sampled sequence.

Real and complex sequences

The physical unit of the sampling frequency is “cycles per second”, or **Hertz** (Hz), if T is in seconds. Whether or not the sequence $\{x[n]\}$ has been obtained by sampling, the quantity $x[n]$ is called the n^{th} sample of the sequence. Moreover, $\{x[n]\}$ is called **real sequence** if the n^{th} sample is real for all values of n . Otherwise, $\{x[n]\}$ is called **complex sequence**.

In particular, a complex sequence can be written as:

$$\{x[n]\} = \{x_{\text{re}}[n]\} + j\{x_{\text{im}}[n]\} \quad (2.4)$$

where $\{x_{\text{re}}[n]\}$ and $\{x_{\text{im}}[n]\}$ are the real and imaginary parts of $x[n]$. The complex conjugate sequence of $\{x[n]\}$ reads:

$$\{x^*[n]\} = \{x_{\text{re}}[n]\} - j\{x_{\text{im}}[n]\} \quad (2.5)$$

Note that the braces are often ignored to denote a sequence if there is no ambiguity.

Example 1: Real and complex sequences

- $\{x[n]\} = \{\cos(0.25n)\}$ is a real sequence;
- $\{y[n]\} = \{e^{j0.3n}\}$ is a complex sequence.

We can write:

$$\{y[n]\} = \underbrace{\{\cos(0.3n)\}}_{\{y_{\text{re}}[n]\}} + j \underbrace{\{\sin(0.3n)\}}_{\{y_{\text{im}}[n]\}} \quad (2.6)$$

Then, we consider the complex conjugate of $\{y[n]\}$:

$$\{w[n]\} = \{y^*[n]\} = \{\cos(0.3n)\} - j\{\sin(0.3n)\} = \{e^{-j0.3n}\} \quad (2.7)$$

2.1.1 Categorization of discrete-time signals

Continuous and discrete-valued signals

We can have two types of discrete-time signals:

- sampled-data signals in which the samples are **continuous-valued**;
- digital signals in which the samples are **discrete-valued**.

Signals in a practical digital signal processing system are digital signals obtained by quantizing the sample values either by rounding or truncation. An example of visualization is showed in Figure 2.3.

Finite and infinite-length discrete-time signals

Again, a discrete-time signal may be a **finite-length** or an **infinite-length** sequence. In the first case, the sequence is defined only for a finite time interval, namely for $N_1 \leq n \leq N_2$, where $-\infty < N_1$, $N_2 < \infty$ and $N_1 \leq N_2$. The length or duration of the ladder is $N = N_2 - N_1 + 1$.

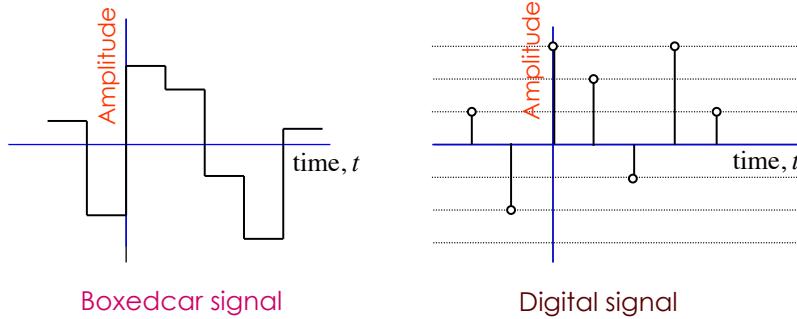


Figure 2.3: Different way to visualize a signal.

A length- N sequence is often referred to as an N -point sequence. In general, the length of a finite-length sequence can be increased by using the zero-padding, namely by appending it with zeros. For example:

$$x_p[n] = \begin{cases} n^2 & -3 \leq n \leq 4 \\ 0 & 5 \leq n \leq 8 \end{cases} \quad (2.8)$$

is a finite-length sequence of length 12 obtained by zero-padding $x[n] = n^2$, $-3 \leq n \leq 4$, with 4 zero-valued samples.

There is another categorization: a discrete-time signal can be **right-sided** or **left-sided**. A right-sided sequence $x[n]$ has zero-valued samples for $n < N_1$. Moreover, if $N_1 \geq 0$, a right-sided sequence is called a **causal sequence**. On the other side, a left-sided sequence $x[n]$ has zero-valued samples for $n > N_2$ and if $N_2 \leq 0$, it is an **anti-causal sequence**. An example of both the types is showed in Figure 2.4.

Right and
left-sided
discrete-time
signals and
(anti)-causal
sequences

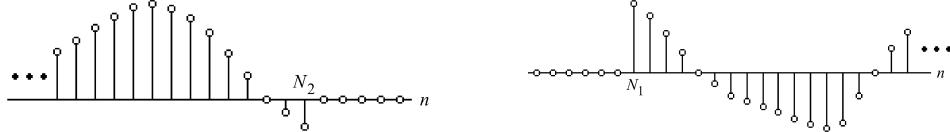


Figure 2.4: Examples of a left-sided sequence (left) and a right-sided sequence (right).

2.1.2 Norm of a discrete-time signal

When dealing with the analysis of discrete signal, the concept of norm is frequent. In particular, the L_p -norm of a signal reads:

$$\|x\|_p = \left(\sum_{n=-\infty}^{\infty} |x[n]|^p \right)^{\frac{1}{p}} \quad (2.9)$$

Norm of a
discrete-time signal

where p is a positive integer. The ladder definition provides an estimate of the size of the signal. The value of p is typically 1, 2 or ∞ :

- $p = 1$: $\|x\|_1$ is the absolute value of $\{x[n]\}$;
- $p = 2$: $\|x\|_2$ is the Root-Mean-Squared (RMS) value of $\{x[n]\}$;
- $p = \infty$: $\|x\|_\infty$ is the peak absolute value of $\{x[n]\}$, namely $\|x\|_\infty = |x|_{\max}$.

We given an example to understand the application of the norm.

Example 2: Norm of a signal

Let $\{y[n]\}$, $0 \leq n \leq N - 1$, be an approximation of $\{x[n]\}$, $0 \leq n \leq N - 1$. An estimate of the relative error is given by the ratio of the L_2 -norm of the difference signal and the L_2 -norm of $\{x[n]\}$:

Relative error

$$E_{\text{rel}} = \left(\frac{\left(\sum_{n=0}^{N-1} |y[n] - x[n]|^2 \right)^{\frac{1}{p}}}{\left(\sum_{n=0}^{N-1} |x[n]|^2 \right)^{\frac{1}{p}}} \right) \quad (2.10)$$

2.2 Operations on sequences

In order to apply a certain operation on a discrete-time signal we can employ a discrete-time system. If the latter is single-input and single-output, it operates on an input sequence, according to some prescribed rules, and develops another sequence, called the output sequence, with more desirable properties. This process is schematized in Figure 2.5.

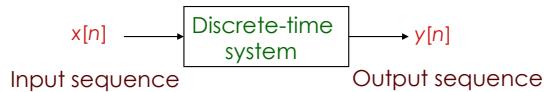


Figure 2.5: Scheme of operation on an input sequence, returning an output sequence.

2.2.1 Basic operations

In most cases, the operation defining a particular discrete-time system is composed of some basic operations, which we are going to list and study:

*Product operation
and windowing*

- **product** (modulation) operation: $y[n] = x[n] \cdot w[n]$

It is employed to form a finite-length sequence from an infinite-length sequence by multiplying the latter with a finite-length sequence called **window sequence**;

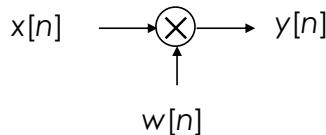


Figure 2.6: Scheme of product operation.

Addition operation

- **addition** operation: $y[n] = x[n] + w[n]$

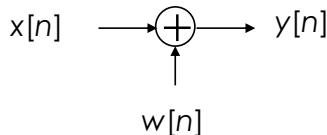
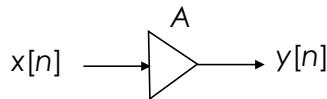


Figure 2.7: Scheme of addition operation.

*Multiplication
operation*

- **multiplication** operation: $y[n] = A \cdot x[n]$

**Figure 2.8:** Scheme of multiplication operation.

- **time-shifting** operation: $y[n] = x[n - N]$, $N \in \mathbb{Z}$

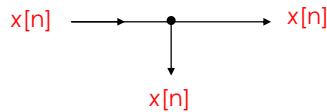
If $N > 0$, it is a delaying opeartion. In particular, we have the unit delay with $y[n] = x[n - 1]$. If $N > 0$, it is an advance operation, with the particular case of unit advance $y[n] = x[n + 1]$;

Time-shifting operation

**Figure 2.9:** Schemes of delaying time-shift (left) and delaying time advance (right) operations.

- **time-reversal** operation: $y[n] = x[-n]$
- **branching** operation: it provides multiple copies of a sequence.

Time-reversal operation
Branching operation

**Figure 2.10:** Scheme of branching operation.

When applying these operations, some caution has to be kept. In fact, operations on two or more sequences can be carried out if all sequences involved are of the same length and defined for the same range of the time index n . However, if the sequences are not of the same length, in some situations, this problem can be circumvented by **appending zero-valued samples** to the sequence(s) of smaller lengths to make all sequences have the same range of the time index.

2.2.2 Ensemble averaging

This composite operation is a very simple application of the addition operation, useful to improve the quality of measured data corrupted by an additive random noise. In some cases, actual uncorrupted data vector \vec{s} remains essentially the same from one mesurement to next, while the additive noise vector is random and not reproducible. Let us denote with \vec{d}_i the noise vector corrupting the i^{th} measurement of the uncorrupted data vector \vec{s} :

$$\vec{x}_i = \vec{s} + \vec{d}_i \quad (2.11)$$

The average data vector, called the **ensemble average**, obtained after k measurements is given by:

$$\vec{x}_{\text{avg}} = \frac{1}{k} \sum_{i=1}^k \vec{x}_i = \frac{1}{k} \sum_{i=1}^k (\vec{s} + \vec{d}_i) = \vec{s} + \frac{1}{k} \sum_{i=1}^k \vec{d}_i \quad (2.12)$$

Ennable averaging to improve data quality

For large values of k , \vec{x}_{avg} is usually a reasonable replica of the desired data vector \vec{s} . An example of results obtained by the application of this operation is showed in the plots in Figure 2.11.

Another more general example of ensable average is showed by the scheme in Figure 2.12, where the figuring block operation reads:

$$y[n] = \alpha_1 x[n] + \alpha_2 x[n - 1] + \alpha_3 x[n - 2] + \alpha_4 x[n - 3] \quad (2.13)$$

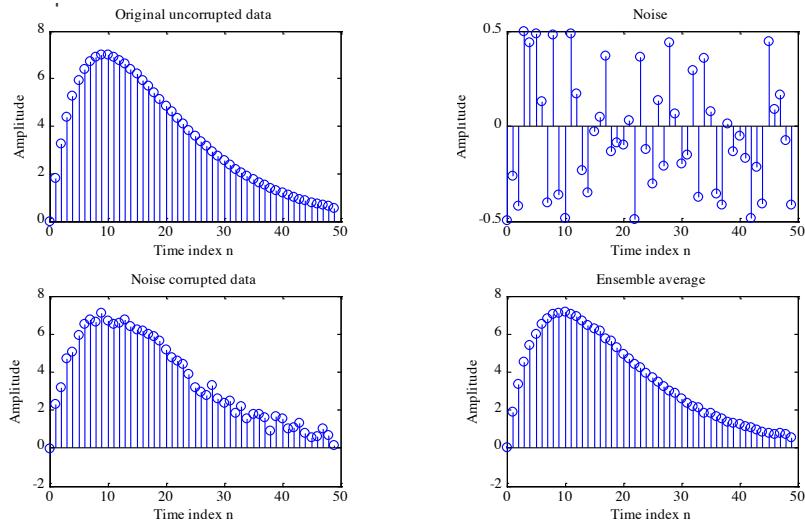


Figure 2.11: Example of results obtained by the application of the ensemble average.

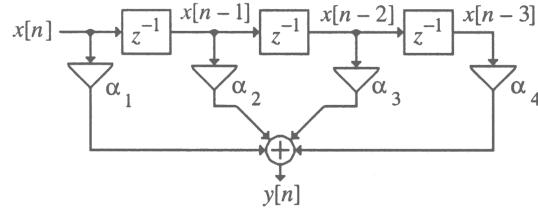


Figure 2.12: Ensemble average block, performing the operation in Eq. 2.13.

2.2.3 Sampling rate alteration

This is another more complex operation employed to generate a new sequence $y[n]$ with a sampling rate F'_T higher or lower than that of the sampling rate F_T of a given sequence $x[n]$. The sampling rate alteration ratio reads:

$$R = \frac{F'_T}{F_T} \quad (2.14)$$

Sampling rate alteration

Interpolation and decimation

Up-sampler

Down-sampler

In particular:

- if $R > 1$, the process is called **interpolation**;
- if $R < 1$, the process is called **decimation**.

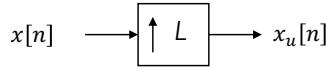
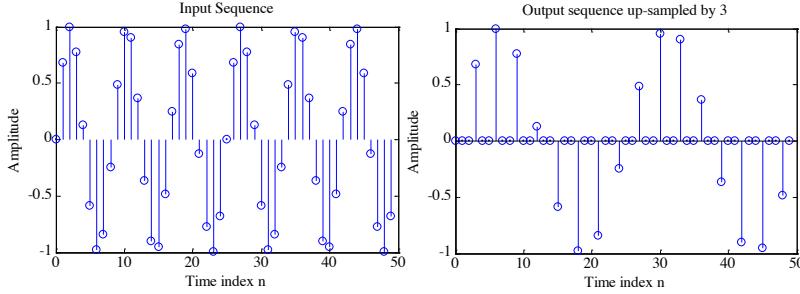
In the case of **up-sampling** by an integer factor $L > 1$, $L - 1$ equidistant zero-valued samples are inserted by the up-sampler between each two consecutive samples of the input sequence $x[n]$:

$$x_u[n] = \begin{cases} x\left[\frac{n}{L}\right] & n = 0, \pm L, \pm 2L, \dots \\ 0 & \text{otherwise} \end{cases} \quad (2.15)$$

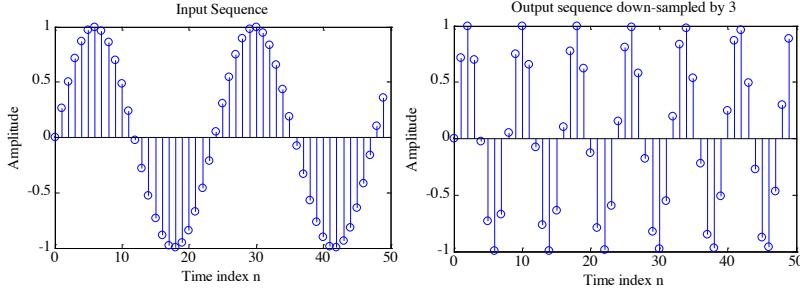
The scheme of this operation is showed in Figure 2.13. An example of the results obtained by the application to a certain input sequence are showed in the plots in Figure 2.14.

On the other hand, in the case of **down-sampling** by an integer factor $M > 1$, every M^{th} samples of the input sequence are kept and $M - 1$ in-between samples are removed:

$$y[n] = x[nM] \quad (2.16)$$

**Figure 2.13:** Scheme of up-sampling operation.**Figure 2.14:** Example of results obtained by the application of the up-sampling.

The scheme of this operation is showed in Figure 2.15. An example of the results obtained by the application to a certain input sequence are showed in the plots in Figure 2.16.

**Figure 2.15:** Scheme of down-sampling operation.**Figure 2.16:** Example of results obtained by the application of the down-sampling.

2.3 Classification of sequences

Several classifications of discrete-time sequences are possible, based on certain features of the sequences themselves.

A first classification is based on the symmetry of the sequence. In fact, we can have **conjugate-symmetric** sequences, namely satisfying

$$x[n] = x^*[-n] \quad (2.17)$$

If $x[n]$ is real, then it is an **even sequence**.

Another possibility is a **conjugate-antisymmetric** sequence, namely satisfying:

$$x[n] = -x^*[-n] \quad (2.18)$$

Again, if $x[n]$ is real, then it is an **odd sequence**.

It follows from the definition that for a conjugate-symmetric sequence $\{x[n]\}$, $x[0]$ must be a real number. Likewise, for a conjugate antisymmetric sequence $\{y[n]\}$,

*Classification
based on symmetry*

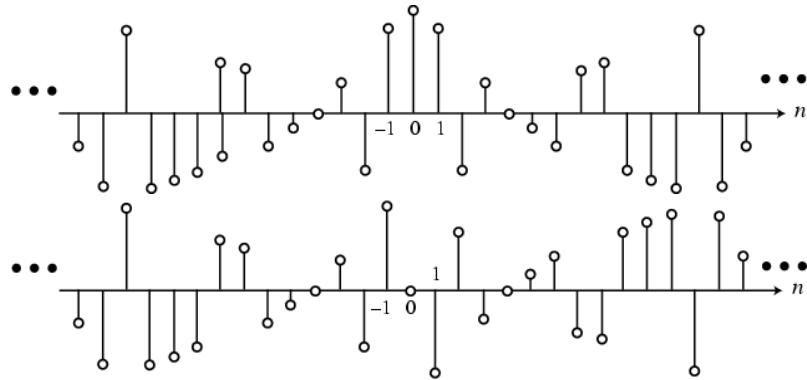


Figure 2.17: Example of even (top) and odd (bottom) sequences.

$y[0]$ must be an imaginary number. Another consequence is that for an odd sequence $\{w[n]\}$, $w[0] = 0$.

Any complex sequence can be expressed as a sum of its conjugate-symmetric part and its conjugate-antisymmetric part:

$$x[n] = x_{\text{cs}}[n] + x_{\text{ca}}[n] \quad (2.19)$$

where:

$$x_{\text{cs}}[n] = \frac{1}{2}(x[n] + x^*[-n]) \quad (2.20)$$

$$x_{\text{ca}}[n] = \frac{1}{2}(x[n] - x^*[-n]) \quad (2.21)$$

Example 3: Classification based on symmetry

We consider the length-7 sequence defined for $-3 \leq n \leq 3$ and its conjugate and time reversed versions:

$$\{g[n]\} = \{0, 1 + j4, -2 + j3, 4 - j2, -5 - j6, -j2, 3\} \quad (2.22)$$

$$\{g^*[n]\} = \{0, 1 - j4, -2 - j3, 4 + j2, -5 + j6, j2, 3\} \quad (2.23)$$

$$\{g^*[-n]\} = \{3, j2, -5 + j6, 4 + j2, -2 - j3, 1 - j4, 0\} \quad (2.24)$$

Therefore:

$$\{g_{\text{cs}}[n]\} = \{1.5, 0.5 + j3, -3.5 + j4.5, 4, -3.5 - j4.5, 0.5 - j3, 1.5\} \quad (2.25)$$

$$\{g_{\text{ca}}[n]\} = \{-1.5, 0.5 + j, 1.5 - j1.5, -j2, -1.5 - j1.5, -0.5 - j, 1.5\} \quad (2.26)$$

It can be easily verified that $g_{\text{cs}}[n] = g_{\text{cs}}^*[-n]$ and $g_{\text{ca}}[n] = -g_{\text{ca}}^*[-n]$.

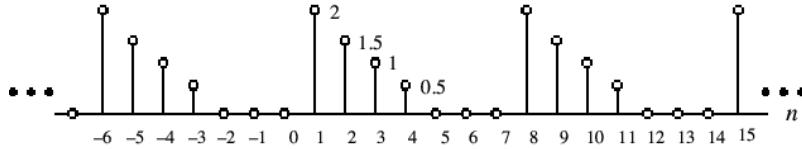
Now, specializing the previous discussion to real sequences, any of them can be expressed as a sum of its even part and its odd part:

$$x[n] = x_{\text{ev}}[n] + x_{\text{od}}[n] \quad (2.27)$$

where:

$$x_{\text{ev}}[n] = \frac{1}{2}(x[n] + x[-n]) \quad (2.28)$$

$$x_{\text{od}}[n] = \frac{1}{2}(x[n] - x[-n]) \quad (2.29)$$

**Figure 2.18:** Example of periodic sequence.

A sequence $\tilde{x}[n]$ satisfying $\tilde{x}[n] = \tilde{x}[n + kN]$ is called a **periodic sequence** with a period N where N is a positive integer and k is any integer. The smallest value of N satisfying $\tilde{x}[n] = \tilde{x}[n + kN]$ is called the **fundamental period**. A sequence not satisfying the periodicity condition is called an **aperiodic sequence**.

Before introducing the classification based on energy and power of the signal, we have to introduce the previous concepts.

The **total energy** of a sequence $x[n]$ is defined by:

Definition of energy

$$E_x = \sum_{n=-\infty}^{\infty} |x[n]|^2 \quad (2.30)$$

An infinite length sequence with finite sample values may or may not have finite energy. On the other hand, a finite length sequence with finite sample values has finite energy.

The **average power** of an aperiodic sequence is defined by:

Definition of average power

$$P_x = \lim_{k \rightarrow \infty} \frac{1}{2k+1} \sum_{n=-k}^k |x[n]|^2 \quad (2.31)$$

We can also define the energy of a sequence $x[n]$ over a finite interval $-k \leq n \leq k$ as:

$$E_{x,k} = \sum_{n=-k}^k |x[n]|^2 \quad (2.32)$$

Then, the averaged power reads:

$$P_x = \lim_{k \rightarrow \infty} \frac{1}{2k+1} E_{x,k} \quad (2.33)$$

The average power of a periodic sequence $\tilde{x}[n]$ with a period N is given by:

$$P_x = \frac{1}{N} \sum_{n=0}^{N-1} |\tilde{x}[n]|^2 \quad (2.34)$$

The average power of an infinite-length sequence may be finite or infinite.

We come now to the classification based on the energy and the power:

Classification based on energy and power

- an infinite energy signal with finite average power is called a **power signal**. As example, one can think about a periodic sequence, which has a finite average power but infinite energy;
- a finite energy signal with zero average power is called an **energy signal**. As example, one can think about a finite-length sequence which has finite energy but zero average power.

Example 4: Classification based on energy and power

We consider the causal sequence defined by:

$$x[n] = \begin{cases} 3(-1)^n & n \geq 0 \\ 0 & n < 0 \end{cases} \quad (2.35)$$

Note that $x[n]$ has infinite energy and its average power is given by:

$$P_x = \lim_{k \rightarrow \infty} \frac{1}{2k+1} \left(9 \sum_{n=0}^k 1 \right) = \lim_{k \rightarrow \infty} \frac{9(k+1)}{2k+1} = 4.5 \quad (2.36)$$

Bounded sequences

We move now to other types of classifications. A sequence $x[n]$ is said to be **bounded** if:

$$|x[n]| \leq B_x < \infty \quad (2.37)$$

Absolutely summable sequences

For example, the sequence $x[n] = \cos(0.3\pi n)$ is bounded since the cosine is bounded between -1 and 1 .

A sequence $x[n]$ is said to be **absolutely summable** if:

$$\sum_{n=-\infty}^{\infty} |x[n]| < \infty \quad (2.38)$$

Example 5: Absolute summable sequences

The sequence:

$$y[n] = \begin{cases} 0.3^n & n \geq 0 \\ 0 & n < 0 \end{cases} \quad (2.39)$$

is absolutely summable since:

$$\sum_{n=0}^{\infty} |0.3^n| = \frac{1}{1-0.3} \approx 1.42857 < \infty \quad (2.40)$$

Square summable sequences

A sequence $x[n]$ is said to be **square summable** if:

$$\sum_{n=-\infty}^{\infty} |x[n]|^2 < \infty \quad (2.41)$$

For example, the sequence:

$$h[n] = \frac{\sin(0.4n)}{\pi n} \quad (2.42)$$

is square summable but no absolutely summable.

2.4 Basic signals

In this Section, we list some of the most common discrete-time sequences that we find in this field.

2.4.1 Unit sample sequence

The **unit sample** sequence is defined as:

$$\delta[n] = \begin{cases} 0 & n \neq 0 \\ 1 & n = 0 \end{cases} \quad (2.43)$$

and it is visualized in Figure 2.19.

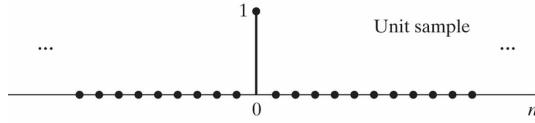


Figure 2.19: Unit sample sequence.

The importance of this type of sequence stem from the fact that an arbitrary sequence can be represented as a sum of scaled, delayed impulses:

$$x[n] = \sum_{k=-\infty}^{\infty} x[k]\delta[n-k] \quad (2.44)$$

2.4.2 Unit step sequence

The **unit step** sequence is defined as:

Unit step sequence

$$u[n] = \begin{cases} 0 & n \geq 0 \\ 1 & n < 0 \end{cases} \quad (2.45)$$

and it is visualized in Figure 2.20.

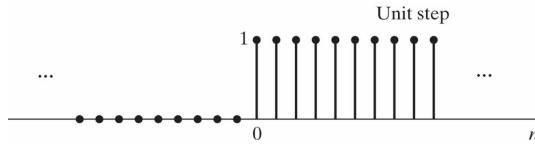


Figure 2.20: Unit step sequence.

The unit step can be expressed in terms of unit samples as:

$$u[n] = \sum_{k=0}^{\infty} \delta[n-k] \quad (2.46)$$

2.4.3 Real sinusoidal sequence

The **real sinusoidal** sequence is defined as:

Real sinusoidal sequence

$$x[n] = A \cos(\omega_0 n + \varphi) \quad (2.47)$$

where A is the amplitude, ω_0 is the angular frequency, and φ is the phase of $x[n]$. It is visualized in Figure 2.21.

2.4.4 Exponential sequence

The **exponential** sequence is defined as:

Exponential sequence

$$x[n] = A\alpha^n, \quad -\infty < n < \infty \quad (2.48)$$

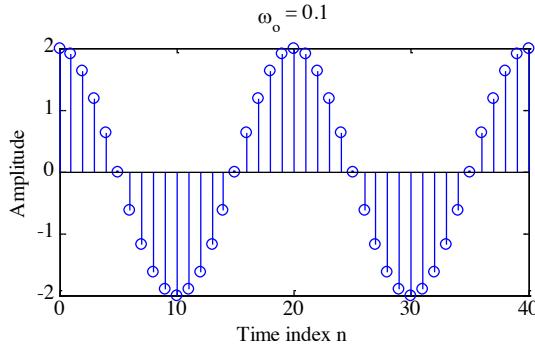


Figure 2.21: Real sinusoidal sequence.

where A and α are real or complex numbers. If we rewrite:

$$\alpha = e^{\sigma_0 + j\omega_0} \quad (2.49)$$

$$A = |A|e^{j\varphi} \quad (2.50)$$

then we can express:

$$x[n] = |A|e^{j\varphi}e^{(\sigma_0 + j\omega_0)n} = x_{\text{re}}[n] + jx_{\text{im}}[n] \quad (2.51)$$

where:

$$x_{\text{re}}[n] = |A|e^{\sigma_0 n} \cos(\omega_0 n + \varphi) \quad (2.52)$$

$$x_{\text{im}}[n] = |A|e^{\sigma_0 n} \sin(\omega_0 n + \varphi) \quad (2.53)$$

$x_{\text{re}}[n]$ and $x_{\text{im}}[n]$ of a complex exponential sequence are real sinusoidal sequences with constant ($\sigma_0 = 0$), growing ($\sigma_0 > 0$) and decaying ($\sigma_0 < 0$) amplitudes for $n > 0$. This is visualized in Figure 2.22.

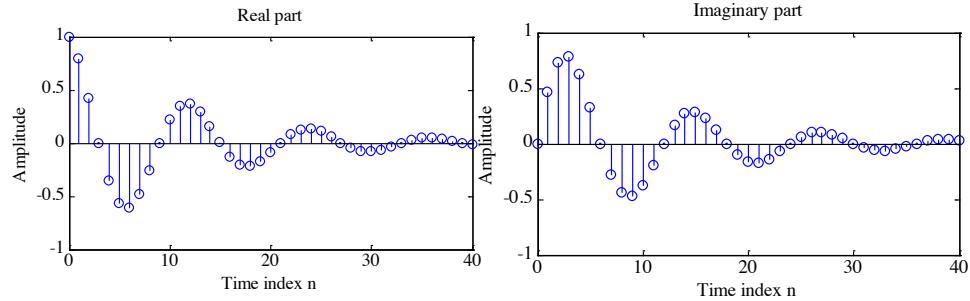


Figure 2.22: Real and imaginary part of an exponential sequence.

If we consider a real exponential sequence:

$$x[n] = A\alpha^n, \quad -\infty < n < \infty \quad (2.54)$$

where A and α are real numbers, we get results similar to the ones in Figure 2.23. Note that the sinusoidal sequence $A \cos(\omega_0 n + \varphi)$ and the complex exponential sequence $B e^{j\omega_0 n}$ are periodic sequences of period N if $\omega_0 \pi = 2\pi r$, where N and r are positive integers. The smallest value of N satisfying the latter equality is the fundamental period of the sequence, as we have already seen previously. To verify this fact, we consider:

$$x_1[n] = \cos(\omega_0 n + \varphi) \quad (2.55)$$

$$x_2[n] = \cos(\omega_0(n + N) + \varphi) \quad (2.56)$$

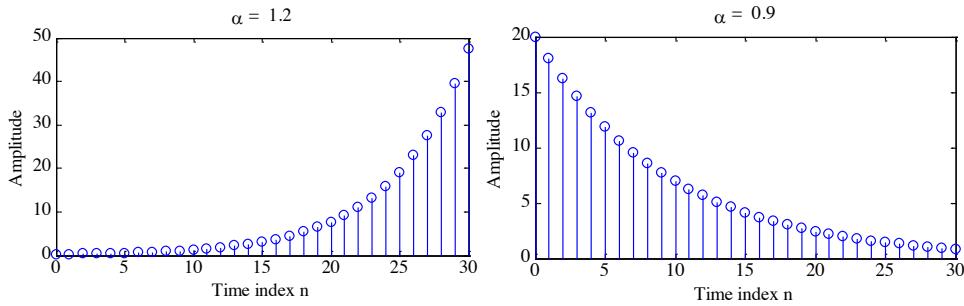


Figure 2.23: Amplitude of the real exponential sequence in Eq. 2.54, for $\alpha = 1.2$ (left) and $\alpha = 0.9$ (right).

Now:

$$\begin{aligned} x_2[n] &= \cos(\omega_0(n + N) + \varphi) \\ &= \cos(\omega_0n + \varphi) \cos(\omega_0N) - \sin(\omega_0n + \varphi) \sin(\omega_0N) \end{aligned} \quad (2.57)$$

which will be equal to $\cos(\omega_0n + \varphi) = x_1[n]$ only if $\sin(\omega_0N) = 0$ and $\cos(\omega_0N) = 1$. These two conditions are met if and only if $\omega_0N = 2\pi r$ or $\frac{2\pi}{\omega_0} = \frac{N}{r}$. If $\frac{2\pi}{\omega_0}$ is a non-integer rational number, then the period will be a multiple of $\frac{2\pi}{\omega_0}$, otherwise the sequence is aperiodic.

From the previous discussion, we can extract two properties:

- let us consider $x[n] = e^{j\omega_1 n}$ and $y[n] = e^{j\omega_2 n}$, with $0 \leq \omega_1 \leq \pi$ and $2\pi k \leq \omega_2 < 2\pi(k+1)$, where k is any positive integer. If $\omega_2 = \omega_1 + 2\pi k$, then $x[n] = y[n]$. Thus, $x[n]$ and $y[n]$ are indistinguishable;
- the frequency of oscillation of $A \cos(\omega_0)$ increases as ω_0 increases from 0 to π , and then decreases as ω_0 increases from π to 2π . Thus frequencies in the neighborhood of $\omega = 0$ are called **low frequencies**, whereas, frequencies in the neighborhood of π are called **high frequencies**.

Chapter 3

Discrete-Time systems

In this Chapter we will introduce the concept of discrete-time system in the framework of Hilbert spaces and we will study how linear algebra can be extremely useful when dealing with this kind of systems.

A discrete-time system processes a given input sequence $x[n]$ to generate an output sequence $y[n]$ with more desirable properties. In most applicationns, the discrete-time system is a single-input and single-output system, just like in the scheme in Figure 3.1.

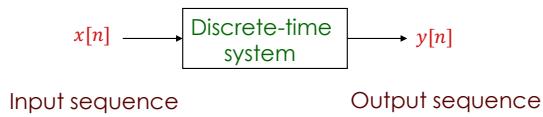


Figure 3.1: Scheme of a single-input single-output discrete-time system.

Lecture 3.
Tuesday 6th
October, 2020.
Discrete-time
system

3.1 Examples of discrete-time systems

Some already seen examples of discrete-time systems are the modulator and the adder, which are 2-input 1-output systems, and the multiplier, the unit delay, the unit advance, which are 1-input 1-output. In this Section, we study other types of systems.

3.1.1 Accumulator system

The output of the **accumulator** $y[n]$ at time instant n is the sum of the input sample $x[n]$ at time instant n and the previous output $y[n - 1]$ at time instant $n - 1$, which is the sum of all previous input sample values from $-\infty$ to $n - 1$. Mathematically, this reads:

$$\begin{aligned} y[n] &= \sum_{\ell=-\infty}^n x[\ell] = \sum_{\ell=-\infty}^{n-1} x[\ell] \\ &= y[n - 1] + x[n] \end{aligned} \tag{3.1}$$

Accumulator
system

So, the system cumulatively adds, namely it accumulates all input sample values. The input-output relation can also be rewritten in the form:

$$\begin{aligned} y[n] &= \sum_{\ell=-\infty}^{-1} x[\ell] + \sum_{\ell=0}^n x[\ell] \\ &= y[-1] + \sum_{\ell=0}^n x[\ell], \quad n \geq 0 \end{aligned} \tag{3.2}$$

This form is used for a causal input sequence, in which case $y[-1]$ is called the initial condition.

3.1.2 *M*-point moving average system

M-point moving average system

The ***M*-point moving average** is used to smooth random variations in data. In fact, the operation it applies reads:

$$y[n] = \frac{1}{M} \sum_{k=0}^{M-1} x[n - k] \quad (3.3)$$

In most applications, the data $x[n]$ is a bounded sequence, so also the *M*-point average $y[n]$ is a bounded sequence. Moreover, if there is no bias in the measurements, an improved estimate of the noisy data is obtained by simply increasing M .

From a complexity point of view, the direct implementation of this system requires $M - 1$ additions, 1 division and the storage of the $M - 1$ past input data samples. However, a more efficient implementation can be developed if we write:

Recursive *M*-point moving average system

$$\begin{aligned} y[n] &= \frac{1}{M} \left(\sum_{\ell=0}^{M-1} x[n - \ell] + x[n - M] - x[n - M] \right) \\ &= \frac{1}{M} \left(\sum_{\ell=1}^M x[n - \ell] + x[n] - x[n - M] \right) \\ &= \frac{1}{M} \left(\sum_{\ell=0}^{M-1} x[n - 1 - \ell] + x[n] - x[n - M] \right) \\ &= y[n - 1] + \frac{1}{M} (x[n] - x[n - M]) \end{aligned} \quad (3.4)$$

In this case, the computation of the *M*-point moving average system is recursive and it requires only 2 additions and 1 division.

Let us consider an application using this system.

Example 6: *M*-point moving average system

We consider a signal $s[n]$ corrupted by a noise $d[n]$, namely:

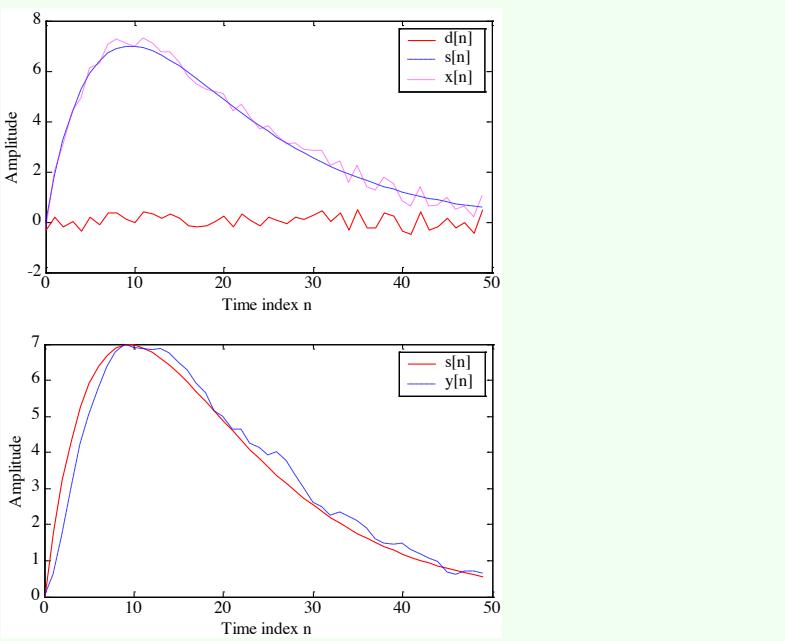
$$x[n] = s[n] + d[n] \quad (3.5)$$

In particular:

$$s[n] = 2[n(0.9)^n] \quad (3.6)$$

$$d[n] = \text{random signal} \quad (3.7)$$

The results obtained by applying the moving average are showed below:



3.1.3 Exponentially weighted running average system

In this case, the computation of the **exponentially weighted running average** requires only 2 additions, 1 multiplication and the storage of the previous running average and it does not require the storage of past input data samples. Mathematically, the operation applied reads:

$$y[n] = \alpha y[n - 1] + x[n], \quad 0 < \alpha < 1 \quad (3.8)$$

In particular, for $0 < \alpha < 1$, the exponentially weighted average filter places more emphasis on current data samples and less emphasis on past data samples. In fact:

$$\begin{aligned} y[n] &= \alpha(\alpha y[n - 2] + x[n - 1]) + x[n] \\ &= \alpha^2 y[n - 2] + \alpha x[n - 1] + x[n] \\ &= \alpha^2(\alpha y[n - 3] + x[n - 2]) + \alpha x[n - 1] + x[n] \\ &= \alpha^3 y[n - 3] + \alpha^2 x[n - 2] + \alpha x[n - 1] + x[n] \end{aligned} \quad (3.9)$$

3.1.4 Linear interpolation system

The **linear interpolation** is employed to estimate the sample values between pairs of adjacent samples of a discrete-time system. A clear example of how this stuff works is showed in Figure 3.2, where the so called factor-of-4 interpolation is employed.

Now, we explain how it works mathematically. If we consider the factor-of-2 interpolator, the operation it performs reads:

$$y[n] = x_u[n] + \frac{1}{2}(x_u[n - 1] + x_u[n + 1]) \quad (3.10)$$

Again, the operation performed by a factor-of-3 interpolator reads:

$$y[n] = x_u[n] + \frac{1}{3}(x_u[n - 1] + x_u[n + 2]) + \frac{2}{3}(x_u[n - 2] + x_u[n + 1]) \quad (3.11)$$

and so on and so forth. The application of a factor-of-2 interpolator to an image can be observed in Figure 3.3. This is a simple example of compression and decompression of informations.

Exponentially weighted running average system

Linear interpolation system

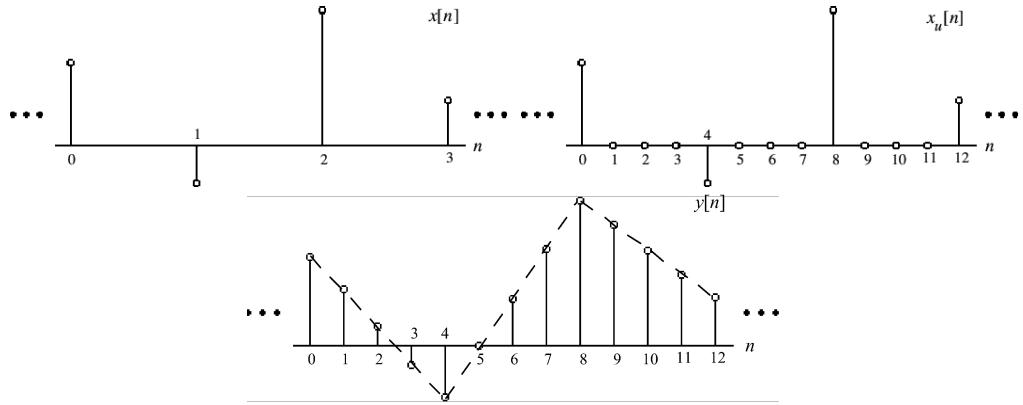


Figure 3.2: Result of the factor-of-4 interpolation.

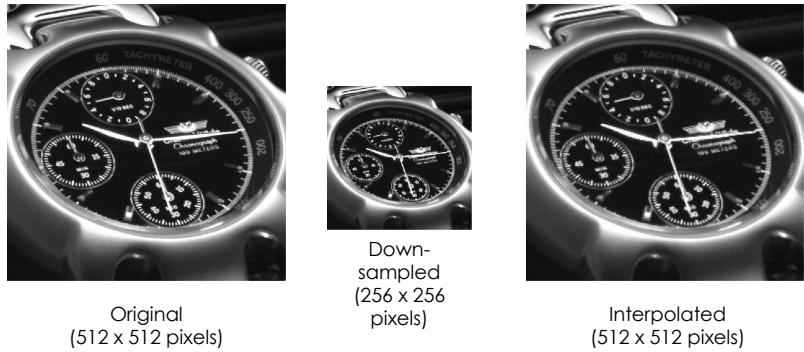


Figure 3.3: Application of a factor-of-2 interpolator to a watch image, which is down-sampled before applying the operation.

3.1.5 Median system

Median system

The **median** of a set of $(2k + 1)$ numbers is the number such that k numbers from the set have values greater than this number and the other k numbers have smaller values. The median can be determined by rank-ordering the numbers in the set by their values and choosing the number in the middle.

Example 7: Median system

If we consider the set of numbers:

$$\{2, -3, 10, 5, -1\} \quad (3.12)$$

the rank-order set is given by:

$$\{-3, -1, 2, 5, 10\} \quad (3.13)$$

and hence, the median is 2

Now, a median filter can be implemented by sliding a window of odd length over the input sequence $\{x[n]\}$ one sample at a time. The output $y[n]$ at the instant n is the median value of the samples inside the window centered at n . A possible application is found in removing additive random noise, which shows up as sudden large errors in the corrupted signal. Some plots of the result of this application are showed in Figure 3.4.

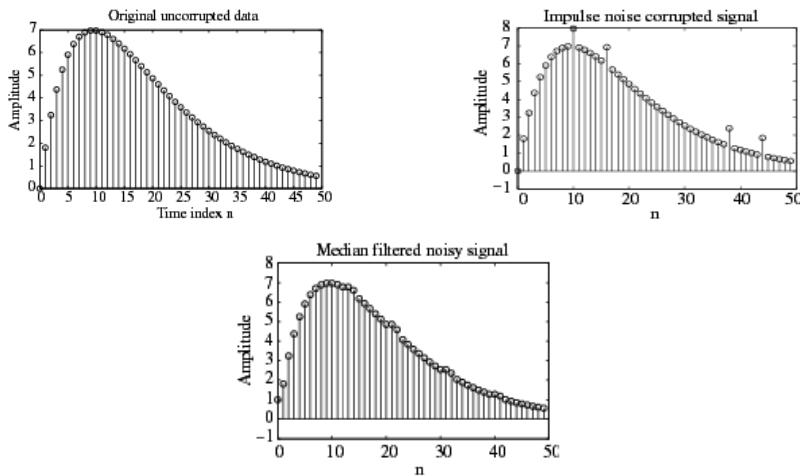


Figure 3.4: Example of results of the application of median filtering to a corrupted signal.

3.2 Classification of discrete-time systems

After having sketched some common examples, in this Section we present the classification of discrete-system based on their features and we enter into the detail by studying each of them. We can have:

- linear systems;
- shift-invariant systems;
- causal systems;
- stable systems;
- passive and lossless systems.

3.2.1 Linear systems

Definition 1: Linear systems

Definition of linear system

If $y_1[n]$ is the output due to an input $x_1[n]$ and $y_2[n]$ is the output due to an input $x_2[n]$, then for an input:

$$x[n] = \alpha x_1[n] + \beta x_2[n] \quad (3.14)$$

the output is given by:

$$y[n] = \alpha y_1[n] + \beta y_2[n] \quad (3.15)$$

This property must hold for any arbitrary constants α and β and for all possible inputs $x_1[n]$ and $x_2[n]$ in order to have a **linear system**.

Let us return to the accumulator example. Here we show that this is a linear system. In fact, if we consider the accumulator outputs for the input sequences $x_1[n]$ and

*Linear systems:
the accumulator*

$x_2[n]$, which read respectively:

$$y_1[n] = \sum_{\ell=-\infty}^n x_1[\ell] \quad (3.16)$$

$$y_2[n] = \sum_{\ell=-\infty}^n x_2[\ell] \quad (3.17)$$

then the output for an input $x[n] = \alpha x_1[n] + \beta x_2[n]$ reads:

$$\begin{aligned} y[n] &= \sum_{\ell=-\infty}^n (\alpha x_1[\ell] + \beta x_2[\ell]) \\ &= \alpha \sum_{\ell=-\infty}^n x_1[\ell] + \beta \sum_{\ell=-\infty}^n x_2[\ell] \\ &= \alpha y_1[n] + \beta y_2[n] \end{aligned} \quad (3.18)$$

If we consider tha accumulator with causal input applied at $n = 0$, the outputs $y_1[n]$ and $y_2[n]$ for inputs $x_1[n]$ and $x_2[n]$ are given by:

$$y_1[n] = y_1[-1] + \sum_{\ell=0}^n x_1[\ell] \quad (3.19)$$

$$y_2[n] = y_2[-1] + \sum_{\ell=0}^n x_2[\ell] \quad (3.20)$$

Moreover, the output $y[n]$ for an input $\alpha x_1[n] + \beta x_2[n]$ is given by:

$$y[n] = y[-1] + \sum_{\ell=0}^n (\alpha x_1[\ell] + \beta x_2[\ell]) \quad (3.21)$$

and, again, for $\alpha y_1[n] + \beta y_2[n]$:

$$\begin{aligned} y[n] &= \alpha y_1[n] + \beta y_2[n] \\ &= \alpha \left(y_1[-1] + \sum_{\ell=0}^n x_1[\ell] \right) + \beta \left(y_2[-1] + \sum_{\ell=0}^n x_2[\ell] \right) \\ &= (\alpha y_1[-1] + \beta y_2[-1]) + \left(\alpha \sum_{\ell=0}^n x_1[\ell] + \beta \sum_{\ell=0}^n x_2[\ell] \right) \end{aligned} \quad (3.22)$$

Thus, if $y[-1] = \alpha y_1[-1] + \beta y_2[-1]$:

$$y[n] = \alpha y_1[n] + \beta y_2[n] \quad (3.23)$$

For the causal accumulator to be linear, the condition $y[-1] = \alpha y_1[-1] + \beta y_2[-1]$ must gold for all initial conditions $y[-1]$, $y_1[-1]$, $y_2[-1]$, and for all constants α and β . This condition cannot be satisfied unless the accumulator is initially at rest with zero initial condition. Instead, for non-zero initial condition, the system is non-linear.

*Non-linear systems:
the median filter*

Let us move to the median filter example in the contest of linear systems discussion. The median filter described earlier is a non-linear dscrete-time system. To show this, we consider an example with a window length of 3. The output of the filter for an input:

$$\{x_1[n]\} = \{3, 4, 5\}, \quad 0 \leq n \leq 2 \quad (3.24)$$

is:

$$\{y_1[n]\} = \{3, 4, 4\}, \quad 0 \leq n \leq 2 \quad (3.25)$$

Again, the output for an input:

$$\{x_2[n]\} = \{2, -1, -1\}, \quad 0 \leq n \leq 2 \quad (3.26)$$

is:

$$\{y_2[n]\} = \{0, -1, -1\}, \quad 0 \leq n \leq 2 \quad (3.27)$$

However, the output for the sum of the previous inputs:

$$\{x[n]\} = \{x_1[n] + x_2[n]\} \quad (3.28)$$

is:

$$\{y[n]\} = \{3, 4, 3\} \neq \{y_1[n] + y_2[n]\} \quad (3.29)$$

Hence, the median filter is a non-linear discrete-time system.

3.2.2 Shift-Invariant systems

Definition of shift-invariant system

Definition 2: Shift-invariant systems

A **shift-invariant system** is a discrete-time system such that, if $y_1[n]$ is the response to an input $x_1[n]$, then the response to an input:

$$x[n] = x_1[n - n_0] \quad (3.30)$$

is:

$$y[n] = y_1[n - n_0] \quad (3.31)$$

where n_0 is any positive or negative integer. This relation must hold for any arbitrary input and its corresponding output.

In the case of sequences and systems with indices n related to discrete instants of time, the shift-invariance property is called **time-invariance property**. The ladder ensures that for a specified input, the output is independent of the time the input is being applied.

Example 8: Shift-invariant system

We consider the up-sampler with an input-output relation given by:

$$x_u[n] \begin{cases} x\left[\frac{n}{L}\right] & n = 0, \pm L, \pm 2L, \dots \\ 0 & \text{otherwise} \end{cases} \quad (3.32)$$

For an input $x[n] = x_1[n - n_0]$, the output $x_{1,u}[n]$ is given by:

$$\begin{aligned} x_{1,u}[n] &= \begin{cases} x\left[\frac{n}{L}\right] & n = 0, \pm L, \pm 2L, \dots \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} x\left[\frac{n-Ln_0}{L}\right] & n = 0, \pm L, \pm 2L, \dots \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (3.33)$$

However, from the definition of the up-sampler:

$$x_u[n - n_0] = \begin{cases} x\left[\frac{n-n_0}{L}\right] & n = n_0, n_0 \pm L, n_0 \pm 2L, \dots \\ 0 & \text{otherwise} \end{cases} \neq x_{1,u}[n] \quad (3.34)$$

Hence, the up-sampler is a **time-varying system**.

*Linear
time-invariant
(LTI) systems*

Returning to the main discussion, if a system satisfies both the linearity and the time-invariance properties, it is known as **linear time-invariant (LTI) system**. LTI systems are mathematically easy to analyze and characterize, and consequently, easy to design. Indeed, highly useful signal processing algorithms have been developed utilizing this class of systems over the last several decades.

3.2.3 Causal systems

Definition of causal system

Definition 3: Causal systems

A **causal system** is a discrete-time system such that the n_0^{th} output sample $y[n_0]$ depends only on the input samples $x[n]$ for $n \leq n_0$ and does not depend on input samples for $n > n_0$.

Now, let $y_1[n]$ and $y_2[n]$ be the responses of a causal discrete-time system to the inputs $x_1[n]$ and $x_2[n]$, respectively. Then:

$$x_1[n] = x_2[n], \quad n < N \quad (3.35)$$

implies also that:

$$y_1[n] = y_2[n], \quad n < N \quad (3.36)$$

Therefore, for a causal system, changes in output samples do not precede changes in the input samples.

Example 9: Causal systems

We give here some examples of causal systems:

$$y[n] = \alpha_1 x[n] + \alpha_2 x[n - 1] + \alpha_3 x[n - 2] + \alpha_4 x[n - 3] \quad (3.37)$$

$$y[n] = b_0 x[n] + b_1 x[n - 1] + b_2 x[n - 2] + \alpha_1 y[n - 1] + \alpha_2 y[n - 2] \quad (3.38)$$

$$y[n] = y[n - 1] + x[n] \quad (3.39)$$

We give here some examples of non-causal systems as well:

$$y[n] = x_u[n] + \frac{1}{2}(x_u[n - 1] + x_u[n + 1]) \quad (3.40)$$

$$y[n] = x_u[n] + \frac{1}{3}(x_u[n - 1] + x_u[n + 2]) + \frac{2}{3}(x_u[n - 2] + x_u[n + 1]) \quad (3.41)$$

Note that a non-causal system can be implemented as a causal system by delaying the output by an appropriate number of samples. For example, a causal implementation of the factor-of-2 interpolator is given by:

$$y[n] = x_u[n - 1] + \frac{1}{2}(x_u[n - 2] + x_u[n]) \quad (3.42)$$

3.2.4 Stable systems

Definition 4: Stable systems

There are various definitions of stability. We consider here the **bounded-input, bounded-output (BIBO) stability**. If $y[n]$ is the response to an input $x[n]$ and if:

$$|x[n]| \leq B_x \quad \forall n \quad (3.43)$$

then:

$$|y[n]| \leq B_y \quad \forall n \quad (3.44)$$

With our conventions, a **stable system** is a discrete-time system satisfying this property.

Example 10: Stable systems

The M -point moving average filter is BIBO stable. In fact, for a bounded input $|x[n]| \leq B_x$, we have:

$$|y[n]| = \left| \frac{1}{M} \sum_{k=0}^{M-1} x[n-k] \right| \leq \frac{1}{M} \sum_{k=0}^{M-1} |x[n-k]| \leq \frac{1}{M} (MB_x) \leq B_x \quad (3.45)$$

3.2.5 Passive and lossless systems

Definition 5: Passive and lossless systems

A discrete-time system is defined to be **passive** if, for every finite-energy input $x[n]$, the output $y[n]$ has, at most, the same energy, namely:

$$\sum_{n=-\infty}^{\infty} |y[n]|^2 \leq \sum_{n=-\infty}^{\infty} |x[n]|^2 < \infty \quad (3.46)$$

For a **lossless** system, the inequality in Eq. 3.46 is satisfied with an equal sign for every output.

Example 11: Passive and lossless systems

We consider the discrete-time system defined by $y[n] = \alpha x[n - N]$ with N a positive integer. Its output energy is given by:

$$\sum_{n=-\infty}^{\infty} |y[n]|^2 = |\alpha|^2 \sum_{n=-\infty}^{\infty} |x[n]|^2 \quad (3.47)$$

Hence, it is a passive system if $|\alpha| < 1$ and a lossless system if $|\alpha| = 1$.

Definition of stable system

Definition of passive and lossless system

Lecture 5.
Tuesday 13th
October, 2020.

3.3 Impulse and step response

Linearity and time-invariance taken together have an incredibly powerful consequence on the behavior of a system. Indeed, an LTI system turns out to be completely characterized by its response to the input $x[n] = \delta[n]$. More in general, the response

Impulse and step responses

of a discrete-time system to a unit sample sequence $\{\delta[n]\}$ is called the unit sample response or, simply, the **impulse response**, and it is denoted by $\{h[n]\}$. On the other hand, the response of a discrete-time system to a unit step sequence $\{\mu[n]\}$ is called the unit step response, or simply, the **step response**, and is denoted by $\{s[n]\}$.

Example 12: Impulse response

The impulse response of the system:

$$y[n] = \alpha_1 x[n] + \alpha_2 x[n - 1] + \alpha_3 x[n - 2] + \alpha_4 x[n - 3] \quad (3.48)$$

is obtained by setting $x[n] = \delta[n]$, resulting in:

$$h[n] = \alpha_1 \delta[n] + \alpha_2 \delta[n - 1] + \alpha_3 \delta[n - 2] + \alpha_4 \delta[n - 3] \quad (3.49)$$

The impulse response is thus a finite-length sequence of length 4 given by:

$$\{h[n]\} = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\} \quad (3.50)$$

Example 13: Impulse response of accumulator

The impulse response of the discrete-time accumulator:

*Impulse response
of accumulator*

$$y[n] = \sum_{\ell=-\infty}^{\infty} x[\ell] \quad (3.51)$$

is obtained by setting $x[n] = \delta[n]$, resulting in:

$$h[n] = \sum_{\ell=-\infty}^{\infty} \delta[\ell] = \mu[n] \quad (3.52)$$

Example 14: Impulse response of interpolator

The impulse response of the factor-of-2 interpolator:

*Impulse response
of interpolator*

$$y[n] = x_u[n] + \frac{1}{2}(x_u[n - 1] + x_u[n + 1]) \quad (3.53)$$

is obtained by setting $x_u[n] = \delta[n]$ and is given by:

$$h[n] = \delta[n] + \frac{1}{2}(\delta[n - 1] + \delta[n + 1]) \quad (3.54)$$

The impulse response is thus a finite-length sequence of length 3:

$$\{h[n]\} = \{0.5, 1, 0.5\} \quad (3.55)$$

*Input-output
realationship*

As we have already pointed out, a consequence of the linear, time-invariance property is that an LTI discrete system is completely characterized by its impulse response. This means that, knowing the impulse response, one can compute the output of the system for any arbitrary input. Let us explain this fact with an example. We consider an LTI discrete-time system and we denote with $h[n]$ its impulse response. We compute its output $y[n]$ for the input:

$$x[n] = 0.5\delta[n + 2] + 1.5\delta[n - 1] - \delta[n - 2] + 0.75\delta[n - 5] \quad (3.56)$$

As the system is linear, we can compute its outputs for each member of the input separately and add the individual outputs to determine $y[n]$. Since the system is time-invariant:

Input \longrightarrow Output

$$\delta[n+2] \longrightarrow h[n+2] \quad (3.57)$$

$$\delta[n-1] \longrightarrow h[n-1] \quad (3.58)$$

$$\delta[n-2] \longrightarrow h[n-2] \quad (3.59)$$

$$\delta[n-5] \longrightarrow h[n-5] \quad (3.60)$$

Likewise, as the system is linear:

Input \longrightarrow Output

$$0.5\delta[n+2] \longrightarrow 0.5h[n+2] \quad (3.61)$$

$$1.5\delta[n-1] \longrightarrow 1.5h[n-1] \quad (3.62)$$

$$-\delta[n-2] \longrightarrow -h[n-2] \quad (3.63)$$

$$0.75\delta[n-5] \longrightarrow 0.75h[n-5] \quad (3.64)$$

Hence, because of the linearity property we get:

$$y[n] = 0.5h[n+2] + 1.5h[n-1] - h[n-2] + 0.75h[n-5] \quad (3.65)$$

Now, any arbitrary input sequence $x[n]$ can be expressed as a linear combination of delayed and advanced unit sample sequences in the form:

$$x[n] = \sum_{k=-\infty}^{\infty} x[k]\delta[n-k] \quad (3.66)$$

The response of the LTI system to an input $x[k]\delta[n-k]$ will be $x[k]h[n-k]$. Hence, the response $y[n]$ to an input like Eq. 3.66 will be:

$$y[n] = \sum_{k=-\infty}^{\infty} x[k]h[n-k] \quad (3.67)$$

which can be alternately written as:

$$y[n] = \sum_{k=-\infty}^{\infty} x[n-k]h[k] \quad (3.68)$$

3.4 Convolution sum

The summation in Eqs. 3.67 and 3.68 requires a separate discussion since it is a very common operation in the field of Digital Signal Processing. It is called **convolution sum** of the sequences $x[n]$ and $h[n]$ and represented compactly as:

Convolution sum

$$y[n] = x[n] \circledast h[n] \quad (3.69)$$

and it has the following properties:

Properties of convolution sum

- **commutative:**

$$x[n] \circledast h[n] = h[n] \circledast x[n] \quad (3.70)$$

- associative:

$$(x[n] \otimes h[n]) \otimes y[n] = x[n] \otimes (h[n] \otimes y[n]) \quad (3.71)$$

- distributive:

$$x[n] \otimes (h[n] + y[n]) = x[n] \otimes h[n] + x[n] \otimes y[n] \quad (3.72)$$

Operations of convolution sum

It is also possible to give an interpretation of the convolution sum, which is schematized in Figure 3.5 with the following operations:

- time-reverse $h[k]$ to form $h[-k]$;
- shift $h[-k]$ to the right by n sampling periods if $n > 0$ or shift to the left by n sampling periods if $n < 0$ to form $h[n - k]$;
- form the product $v[k] = x[k]h[n - k]$;
- sum all samples of $v[k]$ to develop the n^{th} sample of $y[n]$ of the convolution sum.

The computation of an output sample using the convolution sum is simply a sum of products, therefore, it involves fairly simple operations such as additions, multiplications and delays.

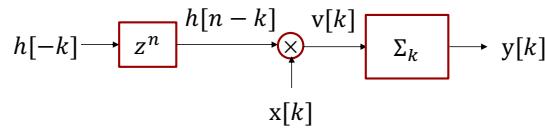


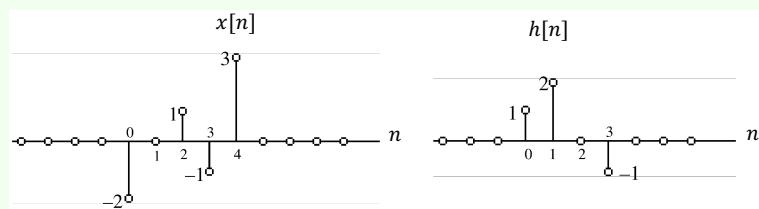
Figure 3.5: Scheme of convolution sum operations.

In practice, if either the input or the impulse response is of finite length, the convolution sum can be used to compute the output sample as it involves a finite sum of products. Moreover, if both the input sequence and the impulse response sequence are of finite length, the output sequence is also of finite length. If both of them are of infinite length, the convolution sum cannot be used to compute the output. In particular, for systems characterized by an infinite impulse response sequence, an alternate time-domain description involving a finite sum of products will be considered.

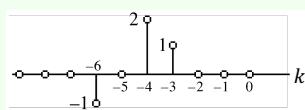
Example 15: Characterization of LTI discrete-time system

We develop the sequence $y[n]$ generated by the convolution of the sequences $x[n]$ and $h[n]$ showed below.

Example of characterization of LTI system



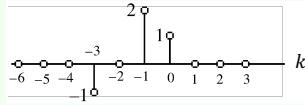
As can be seen from the shifted time-reversed version $\{h[n - k]\}$ for $n < 0$, showed below for $n = -3$, for any value of the sample index k , the k^{th} sample of either $\{x[k]\}$ or $\{h[n - k]\}$ is zero.



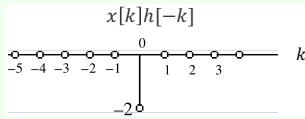
As a result, for $n < 0$, the product of the k^{th} samples of $\{x[k]\}$ and $\{h[n - k]\}$ is always zero and hence:

$$y[n] = 0, \quad n < 0 \quad (3.73)$$

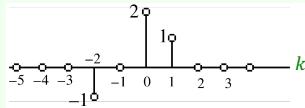
We consider now the computation of $y[0]$. The sequence $\{h[-k]\}$ is showed below.



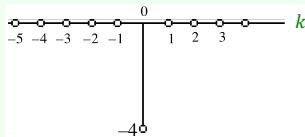
The product sequence $\{x[k]h[-k]\}$ is plotted below which has a single non-zero sample $x[0]h[0]$ for $k = 0$.



Thus $y[0] = x[0]h[0] = -2$. For the computation of $y[1]$, we shift $\{h[-k]\}$ to the right by one sample period to form $\{h[1 - k]\}$ as showed below.



The product sequence $\{x[k]h[1 - k]\}$ is showed below as well.



Hence $y[1] = x[0]h[1] + x[1]h[0] = -4 + 0 = -4$. To calculate $y[2]$, we form $\{h[2 - k]\}$ and the product sequence $\{x[k]h[2 - k]\}$. Hence, $y[2] = x[0]h[2] + x[1]h[1] + x[2]h[0] = 0 + 0 + 1 = 1$. Continuing the process, we get:

$$y[3] = x[0]h[3] + x[1]h[2] + x[2]h[1] + x[3]h[0] = 2 + 0 + 0 + 1 = 3 \quad (3.74)$$

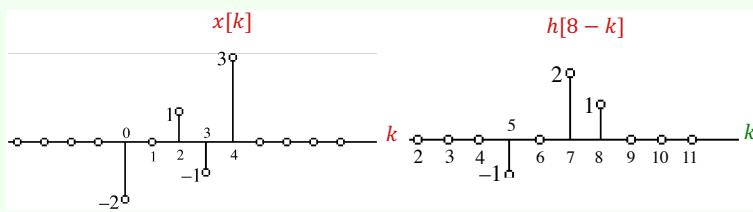
$$y[4] = x[1]h[3] + x[2]h[2] + x[3]h[1] + x[4]h[0] = 0 + 0 - 2 + 3 = 1 \quad (3.75)$$

$$y[5] = x[2]h[3] + x[3]h[2] + x[4]h[1] = -1 + 0 + 6 = 5 \quad (3.76)$$

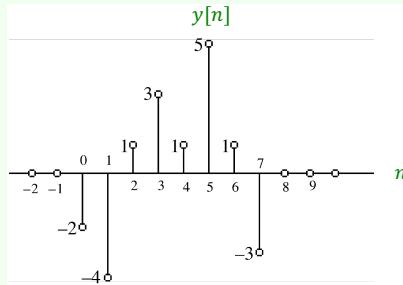
$$y[6] = x[3]h[3] + x[4]h[2] = 1 + 0 = 1 \quad (3.77)$$

$$y[7] = x[4]h[3] = -3 \quad (3.78)$$

From the plot of $\{h[n - k]\}$ for $n > 7$ and the plot of $\{x[k]\}$, it can be seen that there is no overlap between these two sequences. As a result, $y[n] = 0$ for $n > 7$.



Finally, the sequence $\{y[n]\}$ generated by the convolution sum is showed below.



We have to note that the sum of indices of each sample product inside the convolution sum is equal to the index of the sample being generated by the convolution operation. For example, the computation of $y[3]$ in the previous example involves the products $x[0]h[3]$, $x[1]h[2]$, $x[2]h[1]$ and $x[3]h[0]$, and so, the sum of indices in each of these products is equal to 3.

Again, in the same example, we can observe that the convolution of a sequence $\{x[n]\}$ of length 5 with a sequence $\{h[n]\}$ of length 4 resulted in a sequence $\{y[n]\}$ of length 8. In general, if the lengths of the two sequences being convolved are M and N , then the sequence generated by the convolution is of length $M + N - 1$.

Taking again into account the previous example, we can notice a certain pattern in the calculation of $y[n]$. From this intuition, we introduce the **tabular method**, that can be used to convolve two finite-length sequences in a clean and intuitive way. We present it by considering the particular case of the convolution of $\{g[n]\}$, $0 \leq n \leq 3$, with $\{h[n]\}$, $0 \leq n \leq 2$, generating the sequence $y[n] = g[n] \circledast h[n]$. The samples of $\{g[n]\}$ and $\{h[n]\}$ are then multiplied using the conventional multiplication method without any carry operation. The result of the entire procedure is showed in Table 3.1.

$n:$	0	1	2	3	4	5
$g[n]:$	$g[0]$	$g[1]$	$g[2]$	$g[3]$		
$h[n]:$	$h[0]$	$h[1]$	$h[2]$			
	$g[0]h[0]$	$g[1]h[0]$	$g[2]h[0]$	$g[3]h[0]$		
		$g[0]h[1]$	$g[1]h[1]$	$g[2]h[1]$	$g[3]h[1]$	
			$g[0]h[2]$	$g[0]h[2]$	$g[0]h[2]$	$g[0]h[2]$
$y[n]:$	$y[0]$	$y[1]$	$y[2]$	$y[3]$	$y[4]$	$y[5]$

Table 3.1: Results obtained by applying the tabular method for convolution sum computation.

The samples $y[n]$ generated by the convolution sum are obtained by adding the entries in the column above each sample:

$$y[0] = g[0]h[0] \quad (3.79)$$

$$y[1] = g[1]h[0] + g[0]h[1] \quad (3.80)$$

$$y[2] = g[2]h[0] + g[1]h[1] + g[0]h[2] \quad (3.81)$$

$$y[3] = g[3]h[0] + g[2]h[1] + g[1]h[2] \quad (3.82)$$

$$y[4] = g[3]h[1] + g[2]h[2] \quad (3.83)$$

$$y[5] = g[3]h[2] \quad (3.84)$$

The method can also be applied to convolve two finite-length two-sided sequences. In

this case, a decimal point is first placed to the right of the sample with the time index $n = 0$ for each sequence. Next, the convolution is computed ignoring the location of the decimal point. Finally, the decimal point is inserted according to the rules of conventional multiplication. The sample immediately to the left of the decimal point is then located at the time index $n = 0$.

3.5 Simple interconnection schemes

Now, we move to some examples of interconnection schemes of discrete-time systems. There are two main possibilities:

- **cascade connection;**
- **parallel connection.**

3.5.1 Cascade connection

The impulse response $h[n]$ of the cascade of two LTI discrete-time systems with impulse responses $h_1[n]$ and $h_2[n]$ is given by the scheme in Figure 3.6.

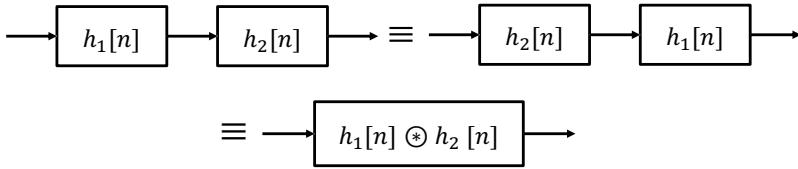


Figure 3.6: Scheme of the cascade connection of two LTI discrete-time systems.

Note that the ordering of the systems in the cascade has no effect on the overall impulse response because of the commutative property of convolution. Moreover, the cascade connection of two stable systems is stable and the cascade of two passive (lossless) systems is passive (lossless).

An application of the cascade connection can be found in the development of an inverse system. If the cascade connection satisfies the relation $h_1[n] \otimes h_2[n]$, then the LTI system $h_1[n]$ is said to be the inverse of $h_2[n]$ and viceversa. Consequently, an application of the inverse system concept is in the recovery of a signal $x[n]$ from its distorted version $\hat{x}[n]$ appearing at the output of a transmission channel. If the impulse response of the channel is known, then $x[n]$ can be recovered by designing an inverse system of the channel as represented in Figure 3.7.

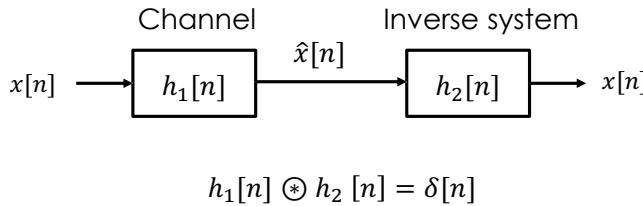


Figure 3.7: Scheme of the cascade connection for signal recovering.

Example 16: Cascade connection and inverse system

We consider the discrete-time accumulator with an impulse response $\mu[n]$. Its inverse system satisfies the condition:

$$\mu[n] \otimes h_2[n] = \delta[n] \quad (3.85)$$

Examples of simple interconnection schemes

Cascade connection scheme

Applications of the cascade connection

Backward difference system

It follows from Eq. 3.85 that:

$$h_2[n] = 0, \quad n < 0 \quad (3.86)$$

$$h_2[0] = 1 \quad (3.87)$$

and that:

$$\sum_{\ell=0}^n h_2[\ell] = 0, \quad n \geq 1 \quad (3.88)$$

Thus, the impulse response of the inverse system of the discrete-time accumulator is given by:

$$h_2[n] = \delta[n] - \delta[n - 1] \quad (3.89)$$

which is called a **backward difference system**.

3.5.2 Parallel connection

Parallel connection scheme

The impulse response $h[n]$ of the parallel of two LTI discrete-time systems with impulse responses $h_1[n]$ and $h_2[n]$ is given by $h[n] = h_1[n] + h_2[n]$ and it is schematized in Figure 3.8.

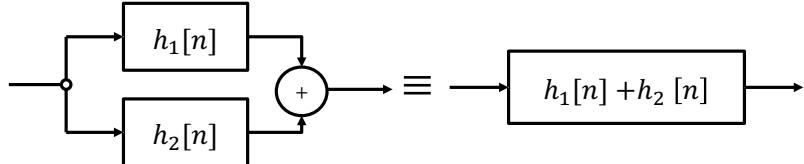


Figure 3.8: Scheme of the parallel connection of two LTI discrete-time systems.

Now, we give a simple example of interconnection scheme.

Example 17: Interconnection schemes

We consider the discrete-time system where:

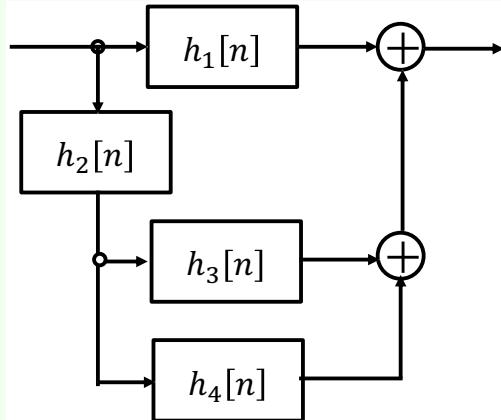
$$h_1[n] = \delta[n] + 0.5\delta[n - 1] \quad (3.90)$$

$$h_2[n] = 0.5\delta[n] - 0.25\delta[n - 1] \quad (3.91)$$

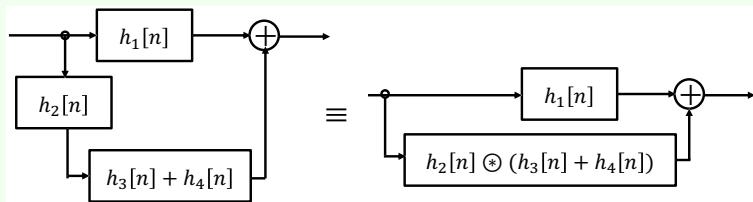
$$h_3[n] = 2.0\delta[n] \quad (3.92)$$

$$h_4[n] = -2(0.5)^n \mu[n] \quad (3.93)$$

which can be schematized as in the diagram below.



The block diagram can be further simplified.



The overall impulse response $h[n]$ is given by:

$$\begin{aligned} h[n] &= h_1[n] + h_2[n] \otimes (h_3[n] + h_4[n]) \\ &= h_1[n] + h_2[n] \otimes h_3[n] + h_2[n] \otimes h_4[n] \end{aligned} \quad (3.94)$$

Now, we can rewrite:

$$h_2[n] \otimes h_3[n] = \left(\frac{1}{2} \delta[n] - \frac{1}{4} \delta[n-1] \right) \otimes (2\delta[n]) = \delta[n] - \frac{1}{2} \delta[n-1] \quad (3.95)$$

an:

$$\begin{aligned} h_2[n] \otimes h_4[n] &= \left(\frac{1}{2} \delta[n] - \frac{1}{4} \delta[n-1] \right) \otimes \left(-2 \left(\frac{1}{2} \right)^n \mu[n] \right) \\ &= - \left(\frac{1}{2} \right)^n \mu[n] + \frac{1}{2} \left(\frac{1}{2} \right)^{n-1} \mu[n-1] \\ &= - \left(\frac{1}{2} \right)^n \mu[n] + \left(\frac{1}{2} \right)^n \mu[n-1] \\ &= - \left(\frac{1}{2} \right)^n \delta[n] \\ &= - \delta[n] \end{aligned} \quad (3.96)$$

Therefore, we come to the final result:

$$\begin{aligned} h[n] &= \delta[n] + \frac{1}{2} \delta[n-1] + \delta[n] - \frac{1}{2} \delta[n-1] \\ &= \delta[n] \end{aligned} \quad (3.97)$$

3.6 Stability and causality conditions

Now, we deepen the discussion on the stability and the causality of discrete time system. We have already defined these concepts, but now we enter more into the detail.

Stability condition of an LTI system

BIBO stability of LTI system

3.6.1 Stability condition of an LTI discrete-time system

In the previous discussion we have chosen among the numerous definitions of stability the convention of **BIBO stability condition**. We recall that a discrete-time system is BIBO stable if and only if the output sequence $\{y[n]\}$ remains bounded for all bounded input sequences $\{x[n]\}$. We specialize this definition to an LTI system.

Corollary 1: BIBO stability of LTI system

An LTI discrete-time system is BIBO stable if and only if its impulse response sequence $\{h[n]\}$ is absolutely summable, namely:

$$S = \sum_{n=-\infty}^{\infty} |h[n]| < \infty \quad (3.98)$$

Proof. We assume that $h[n]$ is a real sequence. Since the input sequence $x[n]$ is bounded, we have:

$$|x[n]| \leq B_x < \infty \quad (3.99)$$

Therefore:

$$\begin{aligned} |y[n]| &= \left| \sum_{k=-\infty}^{\infty} h[k]x[n-k] \right| \\ &\leq \sum_{k=-\infty}^{\infty} |h[k]| |x[n-k]| \\ &\leq B_x \sum_{k=-\infty}^{\infty} |h[k]| = B_x S \end{aligned} \quad (3.100)$$

Thus, $S < \infty$ implies $|y[n]| \leq B_y < \infty$, indicating that $y[n]$ is also bounded. To prove the converse, we assume that $y[n]$ is bounded, i.e. $|y[n]| \leq B_y$. Then, we consider the input given by:

$$x[n] = \text{sign}(h[-n]) \quad (3.101)$$

where $\text{sign}(c) = +1$ if $c \geq 0$ and $\text{sign}(c) = -1$ if $c < 0$. Note also that, since $|x[n]| = 1$, $\{x[n]\}$ is bounded. For this input, $y[n]$ at $n = 0$ is:

$$y[0] = \sum_{k=-\infty}^{\infty} \text{sign}(h[k])h[k] = S \leq B_y < \infty \quad (3.102)$$

Therefore, $|y[n]| \leq B_y$ implies $S < \infty$. ■

We further study the BIBO stability for an LTI system through an example.

Example 18: BIBO stability of LTI systems

We consider a causal LTI discrete-time system with an impulse response:

$$h[n] = (\alpha)^n \mu[n] \quad (3.103)$$

For this system, we have:

$$S = \sum_{n=-\infty}^{\infty} |\alpha^n| \mu[n] = \sum_{n=0}^{\infty} |\alpha^n| = \frac{1}{1 - |\alpha|} \quad (3.104)$$

Therefore, $S < \infty$ if $|\alpha| < 1$, for which the system is BIBO stable. If $|\alpha| \geq 1$, the system is not BIBO stable.

3.6.2 Causality condition of an LTI discrete-time system

As before, the causality condition has already been discussed before but here we focus on the LTI case. Let $x_1[n]$ and $x_2[n]$ be two input sequences with:

$$x_1[n] = x_2[n], \quad n \leq n_0 \quad (3.105)$$

The corresponding output samples at $n = n_0$ of an LTI system with an impulse response $\{h[n]\}$ are then given by:

$$y_1[n_0] = \sum_{k=-\infty}^{\infty} h[k]x_1[n_0 - k] = \sum_{k=0}^{\infty} h[k]x_1[n_0 - k] + \sum_{k=-\infty}^{-1} h[k]x_1[n_0 - k] \quad (3.106)$$

$$y_2[n_0] = \sum_{k=-\infty}^{\infty} h[k]x_2[n_0 - k] = \sum_{k=0}^{\infty} h[k]x_2[n_0 - k] + \sum_{k=-\infty}^{-1} h[k]x_2[n_0 - k] \quad (3.107)$$

If the LTI system is also causal, then:

$$y_1[n] = y_2[n] \quad (3.108)$$

As $x_1[n] = x_2[n]$ for $n \leq n_0$:

$$\sum_{k=0}^{\infty} h[k]x_1[n_0 - k] = \sum_{k=0}^{\infty} h[k]x_2[n_0 - k] \quad (3.109)$$

This implies:

$$\sum_{k=\infty}^{-1} h[k]x_1[n_0 - k] = \sum_{k=\infty}^{-1} h[k]x_2[n_0 - k] \quad (3.110)$$

As $x_1[n] \neq x_2[n]$ for $n > n_0$, the only way the condition in Eq. 3.110 will hold is if both sums are equal to zero, which is satisfied if:

$$h[k] = 0, \quad k < 0 \quad (3.111)$$

So, an LTI discrete-time system is causal if and only if its impulse response $\{h[n]\}$ is a causal sequence.

Causality condition of an LTI system

3.7 Classification of LTI discrete-time systems

Classification of LTI systems

Finite-dimensional LTI systems

In this Section we list some of the possible ways through which an LTI discrete-time system can be classified.

An important subclass of LTI discrete-time systems is characterized by a **linear constant coefficient difference equation** of the form:

$$\sum_{k=0}^N d_k y[n-k] = \sum_{k=0}^M p_k x[n-k] \quad (3.112)$$

where $x[n]$ and $y[n]$ are, respectively, the input and the output of the system, while $\{d_k\}$ and $\{p_k\}$ are constants characterizing the system. The order of the system is given by $\max(N, M)$, which is the order of the difference equation.

It is possible to implement an LTI system characterized by a constant coefficient difference equation as here the computation involves two finite sums of products. If we assume the system to be causal, then the output $y[n]$ can be recursively computed using:

$$y[n] = - \sum_{k=1}^N \frac{d_k}{d_0} y[n-k] + \sum_{k=0}^M \frac{p_k}{d_0} x[n-k] \quad (3.113)$$

provided $d_0 \neq 0$. $y[n]$ can be computed for all $n \geq 0$, knowing $x[n]$ and the initial conditions $y[n_0-1], y[n_0-2], \dots, y[n_0-N]$.

Classification on impulse response length: FIR and IIR systems

A common classification is based on the **impulse response length**. If $h[n]$ is of finite length, namely $h[n] = 0$ for $n < N_1$ and $n > N_2$, with $N_1 < N_2$, then it is known as a **finite impulse response (FIR)** discrete-time system. The convolution sum description here reads:

$$y[n] = \sum_{k=N_1}^{N_2} h[k] x[n-k] \quad (3.114)$$

The output $y[n]$ of an FIR LTI discrete-time system can be computed directly from the convolution sum as it is a finite sum of products. Some examples are the moving-average system and the linear interpolators.

If the impulse response is of infinite length, then it is known as an **infinite impulse response (IIR)** discrete-time system. The class of IIR systems we are interested in this course is characterized by linear constant coefficient difference equations. For example, the discrete-time accumulator defined by $y[n] = y[n-1] + x[n]$ is seen to be an IIR system. Let us consider another important example.

Example 19: IIR systems: integrator

The numerical integration formulas that are used to numerically solve integrals of the form:

$$y(t) = \int_0^t x(\tau) d\tau \quad (3.115)$$

can be shown to be characterized by linear constant coefficient difference equations, and hence, are examples of IIR systems.

If we divide the interval of integration into n equal parts of length T , then the

previous integral can be rewritten as:

$$y(nT) = y((n-1)T) + \int_{(n-1)T}^{nT} x(\tau) d\tau \quad (3.116)$$

where we have set $t = nT$ and used the notation:

$$y(nT) = \int_0^{nT} x(\tau) d\tau \quad (3.117)$$

Using the trapezoidal method we can write:

$$\int_{(n-1)T}^{nT} x(\tau) d\tau = \frac{T}{2} \{x((n-1)T) + x(nT)\} \quad (3.118)$$

Hence, a numerical representation of the definite integral is given by:

$$y(nT) = y((n-1)T) + \frac{T}{2} \{x((n-1)T) + x(nT)\} \quad (3.119)$$

Now, let $y[n] = y(nT)$ and $x[n] = x(nT)$. Then, Eq. 3.119 reduces to:

$$y[n] = y[n-1] + \frac{T}{2} \{x[n] + x[n-1]\} \quad (3.120)$$

which is recognized as the difference equation representation of a first-order IIR discrete-time system.

Another important classification is based on the **output calculation process**. We can have:

- **non-recursive systems**, where the output can be calculated sequentially, knowing only the present and past input samples;
- **recursive systems**, where the output computation involves past output samples in addition to the present and past input samples.

Classification on the output calculation process: recursive and non-recursive systems

Lastly, there is a classification based on the **nature of the coefficients**. We can have:

- **real systems**, in which the impulse response samples are real valued;
- **complex systems**, in which the impulse response samples are complex valued.

Classification on the coefficients: real and complex systems

3.8 Correlation and autocorrelation

There are applications where it is necessary to compare one reference signal with one or more signals to determine the similarity between the pair and to determine additional information based on the similarity. For example, in digital communications, a set of data symbols are represented by a set of unique discrete-time sequences. If one of these sequences has been transmitted, the receiver has to determine which particular sequence has been received by comparing the received signal with every member of possible sequences from the set. Similarly, in radar and sonar applications, the received signal reflected from the target is a delayed version of the transmitted signal and by measuring the delay, one can determine the location of the target. The

Correlation of signals

detection problem gets more complicated in practice, as often the received signal is corrupted by additive random noise.

A measure of similarity between a pair of energy signals, $x[n]$ and $y[n]$, is given by the **cross-correlation sequence** defined by:

$$r_{xy}[\ell] = \sum_{n=-\infty}^{\infty} x[n]y[n-\ell], \quad \ell = 0, \pm 1, \pm 2, \dots \quad (3.121)$$

The parameter ℓ is called **lag** and it indicates the time-shift between the pair of signals. So, $y[n]$ is said to be shifted by ℓ samples to the right with respect to the reference sequence $x[n]$ for positive values of ℓ , and shifted by ℓ samples to the left for negative values of ℓ . The ordering of the subscripts xy in the definition of $r_{xy}[\ell]$ specifies that $x[n]$ is the reference sequence which remains fixed in time, while $y[n]$ is being shifted with respect to $x[n]$. If $y[n]$ is made the reference signal and we shift $x[n]$ with respect to $y[n]$, then the corresponding cross-correlation sequence is given by:

$$\begin{aligned} r_{yx}[\ell] &= \sum_{n=-\infty}^{\infty} y[n]x[n-\ell] \\ &= \sum_{m=-\infty}^{\infty} y[m+\ell]x[m] \\ &= r_{xy}[-\ell] \end{aligned} \quad (3.122)$$

Autocorrelation sequence

Thus, $r_{yx}[\ell]$ is obtained by time-reversing $r_{xy}[\ell]$.

The **autocorrelation sequence** of $x[n]$ is given by:

$$r_{xx}[\ell] = \sum_{n=-\infty}^{\infty} x[n]x[n-\ell] \quad (3.123)$$

obtained by setting $y[n] = x[n]$ in the definition of the cross-correlation sequence $r_{xy}[\ell]$. Note that:

$$r_{xx}[0] = \sum_{n=-\infty}^{\infty} (x[n])^2 = E_x \quad (3.124)$$

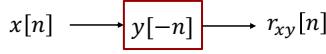
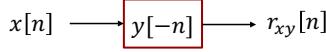
namely the energy of the signal $x[n]$.

From the relation $r_{yx}[\ell] = r_{xy}[\ell]$, it follows that $r_{xx}[\ell] = r_{xx}[-\ell]$, implying that $r_{xx}[\ell]$ is an even function for real $x[n]$. Moreover, an examination of Eq. 3.121 reveals that the expression for the cross-correlation looks quite similar to that of the linear convolution. This similarity is much clearer if we rewrite the expression for the cross-correlation as:

$$r_{xy}[\ell] = \sum_{n=-\infty}^{\infty} x[n]y[-(\ell-n)] = x[\ell] \circledast y[-\ell] \quad (3.125)$$

So, the cross-correlation of $y[n]$ with the reference signal $x[n]$ can be computed by processing $x[n]$ with an LTI discrete-system of impulse response $y[-n]$, as showed in the scheme in Figure 3.9.

Likewise, the autocorrelation of $x[n]$ can be computed by processing $x[n]$ with an LTI discrete-time system of impulse response $x[-n]$, as showed in the scheme in Figure 3.10.

**Figure 3.9:** Scheme of cross-correlation computation.**Figure 3.10:** Scheme of autocorrelation computation.

3.8.1 Properties of autocorrelation and cross-correlation

Now, we consider two finite-energy sequences $x[n]$ and $y[n]$. The energy of the combined sequence $ax[n] + y[n - \ell]$ is also finite and non-negative, i.e.:

$$\begin{aligned} \sum_{n=-\infty}^{\infty} (ax[n] + y[n - \ell])^2 &= a^2 \sum_{n=-\infty}^{\infty} (x[n])^2 \\ &\quad + 2a \sum_{n=-\infty}^{\infty} x[n]y[n - \ell] \\ &\quad + \sum_{n=-\infty}^{\infty} (y[n - \ell])^2 \\ &\geq 0 \end{aligned} \tag{3.126}$$

Properties of autocorrelation and cross-correlation sequences

Thus:

$$a^2r_{xx}[0] + 2ar_{xy}[\ell] + r_{yy}[0] \geq 0 \tag{3.127}$$

where $r_{xx}[0] = E_x > 0$ and $r_{yy}[0] = E_y > 0$. We can rewrite Eq. 3.126 as:

$$\begin{bmatrix} a & 1 \\ r_{xy}[\ell] & r_{yy}[0] \end{bmatrix} \begin{bmatrix} r_{xx}[0] & r_{xy}[\ell] \\ r_{xy}[\ell] & r_{yy}[0] \end{bmatrix} \begin{bmatrix} a \\ 1 \end{bmatrix} \geq 0 \tag{3.128}$$

for any finite value of a . Or, in other words, the matrix in Eq. 3.128 is positive semidefinite, which means:

$$r_{xx}[0]r_{yy}[0] - r_{xy}^2[\ell] \geq 0 \tag{3.129}$$

or, equivalently:

$$|r_{xy}[\ell]| \leq \sqrt{r_{xx}[0]r_{yy}[0]} = \sqrt{E_x E_y} \tag{3.130}$$

The inequality in Eq. 3.130 provides an upper bound for the cross-correlation samples. If we set $y[n] = x[n]$, then the inequality reduces to:

$$|r_{xy}[\ell]| \leq r_{xx}[0] = E_x \tag{3.131}$$

Thus, at zero lag ($\ell = 0$), the sample value of the autocorrelation sequence has its maximum value.

Now, we consider the case:

$$y[n] = \pm bx[n - N] \tag{3.132}$$

where N is an integer and $b > 0$ is an arbitrary number. In this case:

$$E_y = b^2 E_x \tag{3.133}$$

Therefore:

$$\sqrt{E_x E_y} = \sqrt{b^2 E_x^2} = b E_x \tag{3.134}$$

Using the result of Eq. 3.130, we get:

$$-br_{xx}[0] \leq r_{xy}[\ell] \leq br_{xx}[0] \tag{3.135}$$

*Normalized forms
of correlation*

3.8.2 Normalized forms of correlation

Normalized forms of autocorrelation and cross-correlation are given by:

$$\rho_{xx}[\ell] = \frac{r_{xx}[\ell]}{r_{xx}[0]} \quad (3.136)$$

$$\rho_{xy}[\ell] = \frac{r_{xy}[\ell]}{\sqrt{r_{xx}[0]r_{yy}[0]}} \quad (3.137)$$

They are often used for convenience in comparing and displaying. Note also that:

$$|\rho_{xx}[\ell]| \leq 1 \quad (3.138)$$

$$|\rho_{xy}[\ell]| \leq 1 \quad (3.139)$$

independently of the range of values of $x[n]$ and $y[n]$.

3.8.3 Particular cases of correlation computation

*Correlation
computation for
power signals*

The cross correlation sequence for a pair of power signals, $x[n]$ and $y[n]$, is defined as:

$$r_{xy}[\ell] = \lim_{k \rightarrow \infty} \frac{1}{2k+1} \sum_{n=-k}^k x[n]y[n-\ell] \quad (3.140)$$

The autocorrelation sequence of a power signal $x[n]$ is given by:

$$r_{xx}[\ell] = \lim_{k \rightarrow \infty} \frac{1}{2k+1} \sum_{n=-k}^k x[n]x[n-\ell] \quad (3.141)$$

*Correlation
computation for
periodic signals*

The cross-correlation sequence for a pair of periodic signals of period N , $\tilde{x}[n]$ and $\tilde{y}[n]$, is defined as:

$$r_{\tilde{x}\tilde{y}}[\ell] = \frac{1}{N} \sum_{n=0}^{N-1} \tilde{x}[n]\tilde{y}[n-\ell] \quad (3.142)$$

The autocorrelation sequence of a periodic signal $\tilde{x}[n]$ of period N is given by:

$$r_{\tilde{x}\tilde{x}}[\ell] = \frac{1}{N} \sum_{n=0}^{N-1} \tilde{x}[n]\tilde{x}[n-\ell] \quad (3.143)$$

Note that both $r_{\tilde{x}\tilde{y}}[\ell]$ and $r_{\tilde{x}\tilde{x}}[\ell]$ are also periodic signals with a period N . The periodicity property of the autocorrelation sequence can be exploited to determine the period of a periodic signal that may have been corrupted by an additive random disturbance.

Now, let $\tilde{x}[n]$ be a periodic signal corrupted by the random noise $d[n]$ resulting in the signal:

$$w[n] = \tilde{x}[n] + d[n] \quad (3.144)$$

which is observed for $0 \leq n \leq M-1$, where $M \gg N$. The autocorrelation of $w[n]$ is

given by:

$$\begin{aligned}
r_{ww}[\ell] &= \frac{1}{M} \sum_{n=0}^{M-1} w[n]w[n - \ell] \\
&= \frac{1}{M} \sum_{n=0}^{M-1} (\tilde{x}[n] + d[n])(\tilde{x}[n - \ell] + d[n - \ell]) \\
&= \frac{1}{M} \sum_{n=0}^{M-1} \tilde{x}[n]\tilde{x}[n - \ell] + \frac{1}{M} \sum_{n=0}^{M-1} d[n]d[n - \ell] \\
&\quad + \frac{1}{M} \sum_{n=0}^{M-1} \tilde{x}[n]d[n - \ell] + \frac{1}{M} \sum_{n=0}^{M-1} d[n]\tilde{x}[n - \ell] \\
&= r_{\tilde{x}\tilde{x}}[\ell] + r_{dd}[\ell] + r_{\tilde{x}d}[\ell] + r_{d\tilde{x}}[\ell]
\end{aligned} \tag{3.145}$$

In Eq. 3.145, $r_{\tilde{x}\tilde{x}}[\ell]$ is a periodic sequence with a period N and hence it will have peaks at $\ell = 0, N, 2N, \dots$, with the same amplitudes as ℓ approaches M . As $\tilde{x}[n]$ and $d[n]$ are not correlated, samples of cross-correlation sequences $r_{\tilde{x}d}[\ell]$ and $r_{d\tilde{x}}[\ell]$ are likely to be very small with respect to the amplitudes of $r_{\tilde{x}\tilde{x}}[\ell]$. The autocorrelation $r_{dd}[\ell]$ of $d[n]$ will show a peak at $\ell = 0$ with other samples having rapidly decreasing amplitudes with increasing values of $|\ell|$. Hence, peaks of $r_{ww}[\ell]$ for $\ell > 0$ are essentially due to the peaks of $r_{\tilde{x}\tilde{x}}[\ell]$ and they can be used to determine whether $\tilde{x}[n]$ is a periodic sequence and also its period N , if the peaks occur at periodic intervals.

Chapter 4

Fourier Analysis

In this Chapter (and in the following one), we introduce various ways to analyze sequences and discrete-time systems. They range from the analytical to the computational and are all variations of the Fourier transform. But, why do Fourier methods play such a prominent role in Digital Signal Processing? The answer is simple and it is because they are based on eigensequences of LTI systems (convolution operators). So, in this Chapter we will introduce the discrete-time Fourier transform (DTFT), namely the Fourier transform for infinite-length discrete-time signals. In the following Chapter, we will then extend the discussion to the so-called z -transform.

Lecture 8.
Thursday 22nd
October, 2020.

4.1 Continuous-time Fourier transform

Let us start with the basic definitions concerning this very important tool.

Definition 6: Fourier transform of a continuous-time signal

The **CTFT** of a continuous-time signal $x_a(t)$ is given by:

$$X_a(j\Omega) = \int_{-\infty}^{\infty} x_a(t)e^{-j\Omega t} dt \quad (4.1)$$

often referred to as the Fourier spectrum or simply the spectrum of the continuous-time signal.

Definition of Fourier transform of continuous-time signals

Definition 7: Inverse Fourier transform of a continuous-time signal

The **inverse CTFT** of a Fourier transform $X_a(j\Omega)$ is given by:

$$x_a(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X_a(j\Omega)e^{+j\Omega t} d\Omega \quad (4.2)$$

often referred to as the Fourier integral.

Definition of inverse Fourier transform of continuous-time signals

A CTFT pair will be denoted as:

$$x_a(t) \longleftrightarrow X_a(j\Omega) \quad (4.3)$$

Note that Ω is real and denotes the continuous-time angular frequency variable in radians. In general, the CTFT is a complex function of Ω in the range $-\infty < \Omega < \infty$. It can be expressed in the polar form as:

$$X_a(j\Omega) = |X_a(j\Omega)|e^{j\theta_a(\Omega)} \quad (4.4)$$

where $\theta_a(\Omega) = \arg \{X_a(j\Omega)\}$. The quantity $|X_a(j\Omega)|$ is called the **magnitude spectra**.

Magnitude and phase spectra

Dirichlet conditions

trum and the quantity $\theta_a(\Omega)$ is called the **phase spectrum**. Both spectra are real function of Ω and in general the CTFT $X_a(j\Omega)$ exists if $x_a(t)$ satisfies the **Dirichlet conditions**:

- the signal $x_a(t)$ has a finite number of discontinuities and a finite number of maxima and minima in any finite interval;
- the signal is absolutely integrable, i.e.:

$$\int_{-\infty}^{\infty} |x_a(t)| dt < \infty \quad (4.5)$$

If the Dirichlet conditions are satisfied, then:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} X_a(j\Omega) e^{+j\Omega t} d\Omega \quad (4.6)$$

converges to $x_a(t)$ except at values of t where $x_a(t)$ has discontinuities. Moreover, it can be shown that if $x_a(t)$ is absolutely integrable, then proving the existence of the CTFT reduces to proving:

$$|X_a(t\Omega)| < \infty \quad (4.7)$$

Energy density spectrum of continuous-time signal

Now, we focus on the energy density spectrum of a continuous-time signal. The **total energy** E_x of a finite energy continuous-time complex signal $x_a(t)$ is given by:

$$\begin{aligned} E_x &= \int_{-\infty}^{\infty} |x_a(t)|^2 dt \\ &= \int_{-\infty}^{\infty} x_a(t)x_a^*(t) dt \\ &= \int_{-\infty}^{\infty} x_a(t) \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} X_a^*(j\Omega) e^{-j\Omega t} d\Omega \right] dt \end{aligned} \quad (4.8)$$

Interchanging the order of the integration we get:

$$\begin{aligned} E_x &= \frac{1}{2\pi} \int_{-\infty}^{\infty} X_a^*(j\Omega) \left[\int_{-\infty}^{\infty} x_a(t) e^{-j\Omega t} dt \right] d\Omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} X_a^*(j\Omega) X_a(j\Omega) d\Omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} |X_a(j\Omega)|^2 d\Omega \end{aligned} \quad (4.9)$$

Parseval's relation for finite-energy continuous-time signals

Energy density spectrum

Hence:

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |X_a(j\Omega)|^2 d\Omega \quad (4.10)$$

The above relation is more commonly known as the **Parseval's relation** for finite-energy continuous-time signals. The quantity $|X_a(j\Omega)|^2$ is called the **energy density spectrum** of $x_a(t)$ and it is usually denoted as:

$$S_{xx}(\Omega) = |X_a(j\Omega)|^2 \quad (4.11)$$

Energy over a specified range of frequencies

The **energy over a specified range of frequencies** $\Omega_a \leq \Omega \leq \Omega_b$ can be computed using:

$$E_{x,r} = \frac{1}{2\pi} \int_{\Omega_a}^{\Omega_b} S_{xx}(\Omega) d\Omega \quad (4.12)$$

In the case of a full-band, finite-energy, continuous-time signal, the spectrum occupies the whole frequency range $-\infty \leq \Omega \leq \infty$. On the other hand, a **band-limited continuous-time signal** has a spectrum that is limited to a portion of the frequency range $-\infty \leq \Omega \leq \infty$. In particular, an **ideal band-limited signal** has a spectrum that is zero outside a finite frequency range $\Omega_a \leq |\Omega| \leq \Omega_b$ and it can be computed using:

$$X_a(j\Omega) = \begin{cases} 0 & 0 \leq |\Omega| < \Omega_a \\ 0 & \Omega_b < |\Omega| < \infty \end{cases} \quad (4.13)$$

However, an ideal band-limited signal cannot be generated in practice.

Band-limited signals are classified according to the frequency range where most of the signal's is concentrated:

- a **lowpass** continuous-time signal has a spectrum occupying the frequency range $0 < |\Omega| \leq \Omega_p < \infty$, where Ω_p is called the **bandwidth** of the signal;
- a **highpass** continuous-time signal has a spectrum occupying the frequency range $0 < \Omega_p \leq |\Omega| < \infty$, where the bandwidth of the signal is from Ω_p to ∞ ;
- a **bandpass** continuous-time signal has a spectrum occupying the frequency range $0 < \Omega_L \leq |\Omega| \leq \Omega_H < \infty$, where the bandwidth is $\Omega_H - \Omega_L$.

4.2 Discrete-time Fourier transform

Let us introduce the definition of the tool of Fourier transform for discrete-time signals.

(Ideal)
Band-limited
continuous-time
signals

Classification of
band-limited
signals

Definition of
discrete-time
Fourier transform

Definition 8: Discrete-time Fourier transform

The discrete-time Fourier transform (DTFT) $X(e^{j\omega})$ of a sequence $x[n]$ is given by:

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x[n] e^{-j\omega n} \quad (4.14)$$

where in general $X(e^{j\omega})$ is a complex function of the real variable ω and can be rewritten as:

$$X(e^{j\omega}) = X_{\text{re}}(e^{j\omega}) + jX_{\text{im}}(e^{j\omega}) \quad (4.15)$$

$X_{\text{re}}(e^{j\omega})$ and $X_{\text{im}}(e^{j\omega})$ are respectively, the **real** and **imaginary parts** of $X(e^{j\omega})$, and are real functions of ω . $X(e^{j\omega})$ can alternately be expressed as:

$$X(e^{j\omega}) = |X(e^{j\omega})| e^{j\theta(\omega)} \quad (4.16)$$

where $\theta(\omega) = \arg \{X(e^{j\omega})\}$. $|X(e^{j\omega})|$ and $\arg \{X(e^{j\omega})\}$ are called respectively **magnitude function** and **phase function**. Both quantities are again real functions of ω . In many applications, the DTFT is called the **Fourier spectrum**. Likewise, $|X(e^{j\omega})|$ and $\theta(\omega)$ are called respectively the **magnitude** and **phase spectra**.

Magnitude and
phase
functions/spectra

For a real sequence $x[n]$, $|X(e^{j\omega})|$ and $X_{\text{re}}(e^{j\omega})$ are even functions of ω , whereas, $\theta(\omega)$ and $X_{\text{im}}(e^{j\omega})$ are odd functions of ω . Note also that $X(e^{j\omega}) = |X(e^{j\omega})| e^{j\theta(\omega+2\pi k)} =$

Considerations on
parity

Principal value of phase function

$|X(e^{j\omega})|e^{j\theta(\omega)}$ for any integer k . This means that the phase function $\theta(\omega)$ cannot be uniquely specified for any DTFT. Unless otherwise stated, we shall assume that the phase function $\theta(\omega)$ is restricted to the range of values $-\pi \leq \theta(\omega) < \pi$, called the **principal value**.

Example 20: DTFT of the unit sample sequence

The DTFT of the unit sample sequence $\delta[n]$ is given by:

DTFT of the unit sample sequence

$$\Delta(e^{j\omega}) = \sum_{n=-\infty}^{\infty} \delta[n]e^{-j\omega n} = \delta[0] = 1 \quad (4.17)$$

Example 21: DTFT of a causal sequence

Consider the causal sequence:

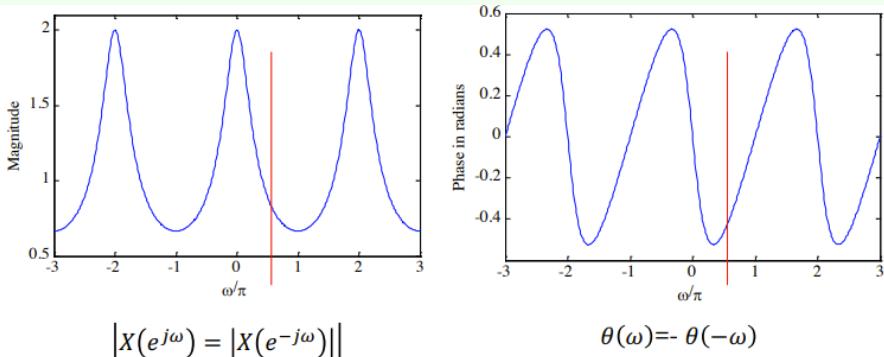
DTFT of a causal sequence

$$x[n] = \alpha^n \mu[n], \quad |\alpha| < 1 \quad (4.18)$$

Its DTFT is given by:

$$\begin{aligned} X(e^{j\omega}) &= \sum_{n=-\infty}^{\infty} \alpha^n \mu[n]e^{-j\omega n} \\ &= \sum_{n=0}^{\infty} \alpha^n e^{-j\omega n} \\ &= \sum_{n=0}^{\infty} (\alpha e^{-j\omega})^n \\ &= \frac{1}{1 - \alpha e^{-j\omega}} \end{aligned} \quad (4.19)$$

as $|\alpha e^{-j\omega}| = |\alpha| < 1$. If we take for example $\alpha = 0.5$, we get the plot below for the magnitude and phase of the DTFT.



We continue the discussion on DTFT with the following derivation about the properties of continuity and periodicity. We also present a way to compute the Fourier coefficients.

Corollary 2: Parity and periodicity of DTFT

The DTFT $X(e^{j\omega})$ of a sequence $x[n]$ is a continuous function of ω . It is also a

Continuity and periodicity of DTFT

periodic function of ω with a period 2π :

$$X(e^{j(\omega+2\pi k)}) = \sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n}e^{-j2\pi kn} = \sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n} = X(e^{j\omega}) \quad (4.20)$$

Therefore:

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n} \quad (4.21)$$

represents the Fourier series representation of the periodic function. As a result, the Fourier coefficients $x[n]$ can be computed from $X(e^{j\omega})$ using the Fourier integral:

$$x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega})e^{j\omega n} d\omega \quad (4.22)$$

Proof. Consider:

$$x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\sum_{\ell=-\infty}^{\infty} x[\ell]e^{-j\omega\ell} \right) e^{j\omega n} d\omega \quad (4.23)$$

The order of integration and summation can be interchanged if the summation inside the brackets converges uniformly, i.e. if $X(e^{j\omega})$ exists. Then:

$$\begin{aligned} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\sum_{\ell=-\infty}^{\infty} x[\ell]e^{-j\omega\ell} \right) e^{j\omega n} d\omega &= \sum_{\ell=-\infty}^{\infty} x[\ell] \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j\omega(n-\ell)} d\omega \right) \\ &= \sum_{\ell=-\infty}^{\infty} x[\ell] \frac{\sin(\pi(n-\ell))}{\pi(n-\ell)} \\ &= \sum_{\ell=-\infty}^{\infty} x[\ell] \delta[n-\ell] \\ &= x[n] \end{aligned} \quad (4.24)$$

For the convergence condition, an infinite series of the form:

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n} \quad (4.25)$$

may or may not converge. Therefore, let us consider:

$$X_k(e^{j\omega}) = \sum_{n=-k}^k x[n]e^{-j\omega n} \quad (4.26)$$

Then, for uniform convergence of $X_k(e^{j\omega})$ we want:

$$\lim_{k \rightarrow \infty} X_k(e^{j\omega}) = X(e^{j\omega}) \quad (4.27)$$

Now, if $x[n]$ is an absolutely summable sequence, i.e., if $\sum_{n=-\infty}^{\infty} |x[n]| < \infty$, then:

$$|X(e^{j\omega})| = \left| \sum_{n=-k}^k x[n]e^{-j\omega n} \right| \leq \sum_{n=-k}^k |x[n]| < \infty \quad (4.28)$$

for all values of ω . Thus, the absolute summability of $x[n]$ is a sufficient condition for the existence of the DTFT $X(e^{j\omega})$. ■

Let us study some examples about the absolute summability condition.

Example 22: Absolute summability condition

The sequence $x[n] = \alpha^n \mu[n]$ for $|\alpha| < 1$ is absolutely summable as:

$$\sum_{n=-k}^k |\alpha^n| \mu[n] = \sum_{n=0}^{\infty} |\alpha^n| = \frac{1}{1 - |\alpha|} < \infty \quad (4.29)$$

and its DTFT $X(e^{j\omega})$ therefore converges to $\frac{1}{1 - \alpha e^{j\omega}}$ uniformly.

Note that since:

$$\sum_{n=-\infty}^{\infty} |x[n]|^2 \leq \left(\sum_{n=-\infty}^{\infty} |x[n]| \right)^2 \quad (4.30)$$

an absolutely summable sequence has always a finite energy. However, a finite-energy sequence is not necessarily absolutely summable.

Example 23: Absolute summability condition

The sequence:

$$x[n] = \begin{cases} \frac{1}{n} & n \geq 1 \\ 0 & n \leq 0 \end{cases} \quad (4.31)$$

has finite energy equal to:

$$E_x = \sum_{n=1}^{\infty} \left(\frac{1}{n} \right)^2 = \frac{\pi^2}{6} \quad (4.32)$$

but $x[n]$ is not absolutely summable.

Not absolutely summable sequences and mean square convergence

In order to represent a finite energy sequence $x[n]$ that is not absolutely summable by a DTFT $X(e^{j\omega})$, it is necessary to consider a **mean-square convergence** of $X(e^{j\omega})$:

$$\lim_{k \rightarrow \infty} \int_{-\pi}^{\pi} |X(e^{j\omega}) - X_k(e^{j\omega})|^2 d\omega = 0 \quad (4.33)$$

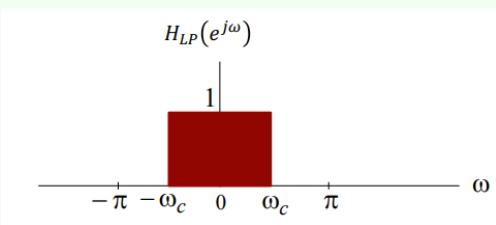
The total energy of the error $X(e^{j\omega}) - X_k(e^{j\omega})$ must approach zero at each value of ω as k goes to ∞ . In such a case, the absolute value of the error $|X(e^{j\omega}) - X_k(e^{j\omega})|$ may not go to zero as k goes to ∞ and the DTFT is no longer bounded. Let us consider an example.

Example 24: Mean-square convergence

Consider the DTFT:

$$H_{LP}(e^{j\omega}) = \begin{cases} 1 & 0 \leq |\omega| \leq \omega_c \\ 0 & \omega < |\omega| \leq \pi \end{cases} \quad (4.34)$$

showed in the plot below.



The inverse DTFT of $H_{LP}(e^{j\omega})$ is given by:

$$h_{LP}[n] = \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} e^{j\omega n} d\omega = \frac{1}{2\pi} \left(\frac{e^{j\omega_c n}}{jn} - \frac{e^{-j\omega_c n}}{jn} \right) = \frac{\sin(\omega_c n)}{\pi n} \quad (4.35)$$

for $-\infty < n < \infty$. The energy of $h_{LP}[n]$ is given by $\frac{\omega_c}{\pi}$. So, $h_{LP}[n]$ is a finite-energy sequence, but it is not absolutely summable. In fact, the result:

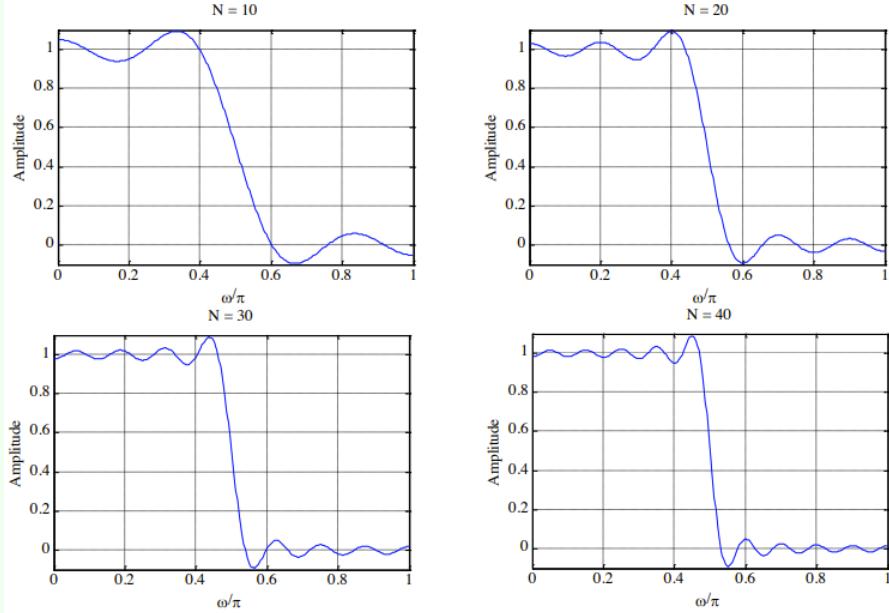
$$\sum_{n=-k}^{n=k} h_{LP}[n] e^{-j\omega n} = \sum_{n=-k}^{n=k} \frac{\sin(\omega_c n)}{\pi n} e^{-j\omega n} \quad (4.36)$$

does not uniformly converge to $H_{LP}(e^{j\omega})$ for all values of ω , but converges to $H_{LP}(e^{j\omega})$ in the mean-square sense.

The mean-square convergence property of the sequence $h_{LP}[n]$ can be further illustrated by examining the plot of the function:

$$H_{LP,k}(e^{j\omega}) = \sum_{n=-k}^{k} \frac{\sin(\omega_c n)}{\pi n} e^{-j\omega n} \quad (4.37)$$

for various values of k , as showed below.



As can be seen from these plots, independently of the value of k there are ripples in the plot of $H_{LP,k}(e^{j\omega})$ around both sides of the point $\omega = \omega_c$. The number of ripples increases as k increases with the height of the largest ripple remaining the same for all values of k . As k goes to ∞ , the condition:

$$\lim_{k \rightarrow \infty} \int_{-\pi}^{\pi} |H_{LP}(e^{j\omega}) - H_{LP,k}(e^{j\omega})|^2 d\omega = 0 \quad (4.38)$$

Gibbs phenomenon

DTFT for not absolutely nor square summable sequences

Dirac delta function

holds indicating the convergence of $H_{LP,k}(e^{j\omega})$ to $H_{LP}(e^{j\omega})$. The oscillatory behaviour of $H_{LP,k}(e^{j\omega})$ approximating $H_{LP}(e^{j\omega})$ in the mean-square sense at a point of discontinuity is known as the **Gibbs phenomenon**.

The DTFT can also be defined for a certain class of sequences which are neither absolutely summable nor square summable. Examples of such sequences are the unit step sequence $\mu[n]$, the sinusoidal sequence $\cos(\omega_0 n + \varphi)$ and the exponential sequence $A\alpha^n$. For this type of sequences, a DTFT representation is possible using the **Dirac delta function** $\delta(\omega)$.

A Dirac delta function $\delta(\omega)$ is a function of ω with infinite height, zero width, and unit area. It is the limiting form of a unit area pulse function p_Δ as Δ goes to zero satisfying:

$$\lim_{\Delta \rightarrow 0} \int_{-\infty}^{\infty} p_\Delta(\omega) d\omega = \int_{-\infty}^{\infty} \delta(\omega) d\omega \quad (4.39)$$

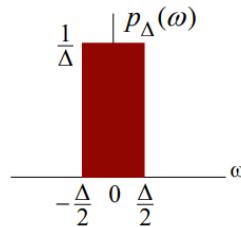


Figure 4.1: Plot and area of $p_\Delta(\omega)$ function.

Example 25: Dirac δ application

Consider the complex exponential sequence:

$$x[n] = e^{j\omega_0 n} \quad (4.40)$$

Its DTFT is given by:

$$X(e^{j\omega}) = \sum_{k=-\infty}^{\infty} 2\pi\delta(\omega - \omega_0 + 2k\pi) \quad (4.41)$$

where $\delta(\omega)$ is an impulse function of ω and $-\pi \leq \omega_0 \leq \pi$. The function in Eq. 4.41 is periodic in ω with a period 2π and it is called **periodic impulse train**. In order to verify that $X(e^{j\omega})$ given above is indeed the DTFT of $x[n] = e^{j\omega_0 n}$ we compute the inverse DTFT of $X(e^{j\omega})$:

$$\begin{aligned} x[n] &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{k=-\infty}^{\infty} 2\pi\delta(\omega - \omega_0 + 2\pi k) e^{j\omega n} d\omega \\ &= \int_{-\pi}^{\pi} \delta(\omega - \omega_0) e^{j\omega n} d\omega \\ &= e^{j\omega_0 n} \end{aligned} \quad (4.42)$$

where we have used the sampling property of the impulse function $\delta(\omega)$.

Periodic impulse train

Last but not least, we list in Table 4.1 a set of commonly used DTFT pairs.

Sequence	DTFT
$\delta[n]$	1
1	$\sum_{k=-\infty}^{\infty} 2\pi\delta(\omega + 2\pi k)$
$e^{j\omega_0 n}$	$\sum_{k=-\infty}^{\infty} 2\pi\delta(\omega - \omega_0 + 2\pi k)$
$\mu[n]$	$\frac{1}{1 - e^{-j\omega}} + \sum_{k=-\infty}^{\infty} \pi\delta(\omega + 2\pi k)$
$\alpha^n \mu[n] \quad (\alpha < 1)$	$\frac{1}{1 - \alpha e^{-j\omega}}$

Table 4.1: Commonly used DTFT pairs.

There are a number of important properties of the DTFT that are useful in signal processing applications. These are listed here in Figures 4.2, 4.3, 4.4 without proof, since it is quite straightforward to derive them. We illustrate the applications of some of the DTFT properties.

DTFT properties

Sequence	Discrete-Time Fourier Transform
$x[n]$	$X(e^{j\omega})$
$x[-n]$	$X(e^{-j\omega})$
$x^*[-n]$	$X^*(e^{j\omega})$
$\text{Re}\{x[n]\}$	$X_{\text{cs}}(e^{j\omega}) = \frac{1}{2}\{X(e^{j\omega}) + X^*(e^{-j\omega})\}$
$j\text{Im}\{x[n]\}$	$X_{\text{ca}}(e^{j\omega}) = \frac{1}{2}\{X(e^{j\omega}) - X^*(e^{-j\omega})\}$
$x_{\text{cs}}[n]$	$X_{\text{re}}(e^{j\omega})$
$x_{\text{ca}}[n]$	$jX_{\text{im}}(e^{j\omega})$

Figure 4.2: DTFT symmetry relations for a complex sequence $x[n]$.

Lecture 9.
Tuesday 27th
October, 2020.

Example 26: DTFT properties

We determine the DTFT $Y(e^{j\omega})$ of:

$$y[n] = (n+1)\alpha^n \mu[n], \quad |\alpha| < 1 \quad (4.43)$$

Let $x[n] = \alpha^n \mu[n]$, with $\alpha < 1$. We can therefore write:

$$y[n] = nx[n] + x[n] \quad (4.44)$$

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

$$X(e^{j\omega}) = \frac{1}{1 - \alpha e^{-j\omega}} \quad (4.45)$$

Using the differentiation property of the DTFT, we observe that the DTFT of

Sequence	Discrete-Time Fourier Transform
$x[n]$	$X(e^{j\omega}) = X_{\text{re}}(e^{j\omega}) + jX_{\text{im}}(e^{j\omega})$
$x_{\text{ev}}[n]$	$X_{\text{re}}(e^{j\omega})$
$x_{\text{od}}[n]$	$jX_{\text{im}}(e^{j\omega})$
Symmetry relations	
$X(e^{j\omega}) = X^*(e^{-j\omega})$	
$X_{\text{re}}(e^{j\omega}) = X_{\text{re}}(e^{-j\omega})$	
$X_{\text{im}}(e^{j\omega}) = -X_{\text{im}}(e^{-j\omega})$	
$ X(e^{j\omega}) = X(e^{-j\omega}) $	
$\arg\{X(e^{j\omega})\} = -\arg\{X(e^{-j\omega})\}$	

Figure 4.3: DTFT symmetry realtions for a real sequence $x[n]$.

Type of Property	Sequence	Discrete-Time Fourier Transform
	$g[n]$	$G(e^{j\omega})$
	$h[n]$	$H(e^{j\omega})$
Linearity	$\alpha g[n] + \beta h[n]$	$\alpha G(e^{j\omega}) + \beta H(e^{j\omega})$
Time-shifting	$g[n - n_0]$	$e^{-j\omega n_0} G(e^{j\omega})$
Frequency-shifting	$e^{j\omega_0 n} g[n]$	$G\left(e^{j(\omega - \omega_0)}\right)$
Differentiation in frequency	$ng[n]$	$j \frac{dG(e^{j\omega})}{d\omega}$
Convolution	$g[n] \circledast h[n]$	$G(e^{j\omega})H(e^{j\omega})$
Modulation	$g[n]h[n]$	$\frac{1}{2\pi} \int_{-\pi}^{\pi} G(e^{j\theta})H(e^{j(\omega - \theta)}) d\theta$
Parseval's relation	$\sum_{n=-\infty}^{\infty} g[n]h^*[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} G(e^{j\omega})H^*(e^{j\omega}) d\omega$	

Figure 4.4: DTFT general properties.

$nx[n]$ is given by:

$$j \frac{dX(e^{j\omega})}{d\omega} = j \frac{d}{d\omega} \left(\frac{1}{1 - \alpha e^{-j\omega}} \right) = \frac{\alpha e^{-j\omega}}{(1 - \alpha e^{-j\omega})^2} \quad (4.46)$$

Next, using the linearity property of the DTFT, we arrive at:

$$Y(e^{j\omega}) = \frac{\alpha e^{-j\omega}}{(1 - \alpha e^{-j\omega})^2} + \frac{1}{1 - \alpha e^{-j\omega}} = \frac{1}{(1 - \alpha e^{-j\omega})^2} \quad (4.47)$$

Example 27: DTFT properties

We determine the DTFT $V(e^{j\omega})$ of the sequence $v[n]$, defined by:

$$d_0 v[n] + d_1 v[n - 1] = p_0 \delta[n] + p_1 \delta[n - 1] \quad (4.48)$$

The DTFT of $\delta[n]$ is 1. Using the time-shifting property of the DTFT, we ob-

serve that the DTFT of $\delta[n - 1]$ is $e^{-j\omega}$ and the DTFT of $v[n - 1]$ is $e^{-j\omega}V(e^{j\omega})$. Using the linearity property, we then obtain the frequency-domain representation of $d_0v[n] + d_1v[n - 1]$ as:

$$d_0V(e^{j\omega}) + d_1e^{-j\omega}V(e^{j\omega}) = p_0 + p_1e^{-j\omega} \quad (4.49)$$

Solving the above equation we get:

$$V(e^{j\omega}) = \frac{p_0 + p_1e^{-j\omega}}{d_0 + d_1e^{-j\omega}} \quad (4.50)$$

As in the previous Section for the CTFT, we move to the discussion on the energy density spectrum for the DTFT. The **total energy** of a finite-energy sequence $g[n]$ is given by:

$$E_g = \sum_{n=-\infty}^{\infty} |g[n]|^2 \quad (4.51)$$

From Parseval's relation we observe that:

$$E_g = \sum_{n=-\infty}^{\infty} |g[n]|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |G(e^{j\omega})|^2 d\omega \quad (4.52)$$

The quantity:

$$S_{gg}(\omega) = |G(e^{j\omega})|^2 \quad (4.53)$$

is called the **energy density spectrum**. The area under this curve in the range $-\pi \leq \omega \leq \pi$ divided by 2π is the energy of the sequence.

As before, we also discuss about the case of band-limited signals, but in the discrete-time version. Since the spectrum of a discrete-time signal is a periodic function of ω with a period 2π , a full-band signal has a spectrum occupying the frequency range $-\pi \leq \omega \leq \pi$. A **band-limited discrete-time signal** has a spectrum that is limited to a portion of the frequency range $-\pi \leq \omega \leq \pi$.

An **ideal band-limited discrete-time signal** has a spectrum that is zero outside a frequency range $0 < \omega_a \leq |\omega| \leq \omega_b < \pi$, that is:

$$X(e^{j\omega}) = \begin{cases} 0 & 0 \leq |\omega| < \omega_a \\ 0 & \omega_b < |\omega| < \pi \end{cases} \quad (4.54)$$

However, an ideal band-limited discrete-time signal cannot be generated in practice. A classification of a band-limited discrete-time signal is based on the frequency range where most of the signal energy is concentrated and it is analog to the classification given in the previous Section (lowpass, highpass and bandpass).

Energy density spectrum

*(Ideal)
Band-limited
discrete-time
signals*

Example 28: Band-limited discrete-time signals

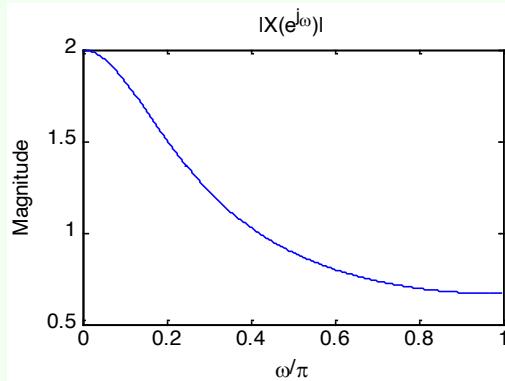
Consider the sequence:

$$x[n] = (0.5)^n \mu[n] \quad (4.55)$$

The DTFT is:

$$X(e^{j\omega}) = \frac{1}{1 - 0.5e^{-j\omega}} \quad (4.56)$$

and the magnitude spectrum is showed below.



It can be showed that 80% of the energy of this lowpass signal is contained in the frequency range $0 \leq |\omega| \leq 0.5081\pi$. Hence, we can define the 80% bandwidth to be 0.5081π radians.

Returning to the energy density spectrum, we consider some other examples introducing also the concept of band-limited signals.

Example 29: Energy density spectrum

We compute the energy of the sequence:

$$h_{LP}[n] = \frac{\sin(\omega_c n)}{\pi n}, \quad -\infty < n < \infty \quad (4.57)$$

Here:

$$\sum_{n=-\infty}^{\infty} |h_{LP}[n]|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |H_{LP}(e^{j\omega})|^2 d\omega \quad (4.58)$$

where:

$$H_{LP}(e^{j\omega}) = \begin{cases} 1 & 0 \leq |\omega| \leq \omega_c \\ 0 & \omega_c < |\omega| \leq \pi \end{cases} \quad (4.59)$$

Therefore:

$$\sum_{n=-\infty}^{\infty} |h_{LP}[n]|^2 = \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} d\omega = \frac{\omega_c}{\pi} < \infty \quad (4.60)$$

Hence, $h_{LP}[n]$ is a finite-energy lowpass sequence.

4.3 Linear convolution using DTFT

Convolution theorem

An important property of the DTFT is given by the **convolution theorem**.

Theorem 1: Convolution theorem

If $y[n] = x[n] \circledast h[n]$, then the DTFT $Y(e^{j\omega})$ of $y[n]$ is given by:

$$Y(e^{j\omega}) = X(e^{j\omega})H(e^{j\omega}) \quad (4.61)$$

Linear convolution using DTFT

An implication of this result is that the linear convolution $y[n]$ of the sequences $x[n]$ and $h[n]$ can be performed as follows:

- compute the DTFTs $X(e^{j\omega})$ and $H(e^{j\omega})$ of the sequences $x[n]$ and $h[n]$, respectively;
- form the DTFT $Y(e^{j\omega}) = X(e^{j\omega})H(e^{j\omega})$;
- compute the inverse DTFT $y[n]$ of $Y(e^{j\omega})$.

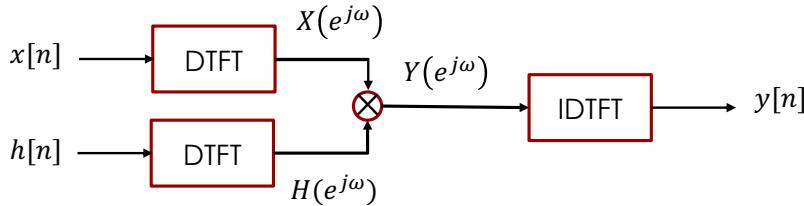


Figure 4.5: Scheme of the computation of linear convolution $y[n]$ of the sequences $x[n]$ and $h[n]$.

Note that in numerical computation, when the computed phase function is outside the range $[-\pi, \pi]$, the phase is computed modulo 2π , to bring the computed value to this range. Thus, the phase functions of some sequences exhibit discontinuities of 2π radians in the plot. For example, there is a discontinuity of 2π at $\omega = 0.72$ in the phase response below:

$$X(e^{j\omega}) = \frac{0.008 - 0.033e^{-j\omega} + 0.05e^{-2j\omega} - 0.033e^{-3j\omega} + 0.008e^{-4j\omega}}{1 + 2.37e^{-j\omega} + 2.7e^{-2j\omega} + 1.6e^{-3j\omega} + 0.41e^{-4j\omega}} \quad (4.62)$$

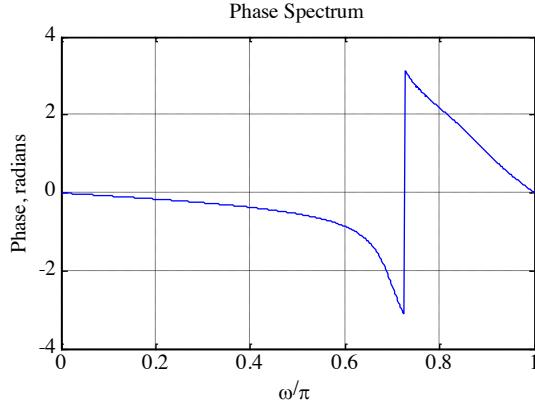


Figure 4.6: Discontinuity in the phase response of Eq. 4.62.

In such cases, often an alternate type of phase function that is continuous function of ω is derived from the original phase function by removing the discontinuities of 2π . Process of discontinuity removal is called **unwrapping the phase** and the **unwrapped phase function** will be denoted as $\theta_c(\omega)$.

Unwrapped phase function

For example, the unwrapped phase function of the DTFT in Eq. 4.62 is showed in Figure 4.7.

The conditions under which the phase function will be a continuous function of ω is next derived. Now consider:

$$\ln X(e^{j\omega}) = \ln |X(e^{j\omega})| + j\theta(\omega) \quad (4.63)$$

where:

$$\theta(\omega) = \arg \{X(e^{j\omega})\} \quad (4.64)$$

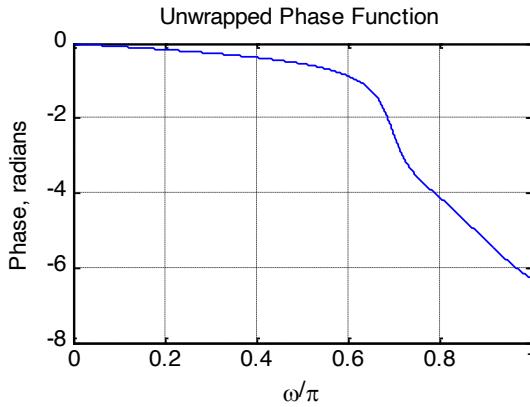


Figure 4.7: Unwrapped phase function of Eq. 4.62.

From $\ln X(e^{j\omega})$ we can also compute $\frac{d \ln X(e^{j\omega})}{d\omega}$:

$$\frac{d \ln X(e^{j\omega})}{d\omega} = \frac{d \ln |X(e^{j\omega})|}{d\omega} + j \frac{d\theta(\omega)}{d\omega} \quad (4.65)$$

Thus, $\frac{d\theta(\omega)}{d\omega}$ is given by the imaginary part of:

$$\frac{1}{X(e^{j\omega})} \left[\frac{dX_{\text{re}}(e^{j\omega})}{d\omega} + j \frac{dX_{\text{im}}(e^{j\omega})}{d\omega} \right] \quad (4.66)$$

Hence:

$$\frac{d\theta(\omega)}{d\omega} = \frac{1}{|X(e^{j\omega})|^2} \left[X_{\text{re}}(e^{j\omega}) \frac{dX_{\text{im}}(e^{j\omega})}{d\omega} - X_{\text{im}}(e^{j\omega}) \frac{dX_{\text{re}}(e^{j\omega})}{d\omega} \right] \quad (4.67)$$

Definition of phase function from its derivate

The phase function can thus be defined unequivocally by its derivative:

$$\theta(\omega) = \int_0^\omega \left[\frac{d\theta(\eta)}{d\eta} \right] d\eta \quad (4.68)$$

with the constraint $\theta(0) = 0$.

The phase function defined by Eq. 4.68 is called the **unwrapped phase function** of $X(e^{j\omega})$ and it is a continuous function of ω . Therefore, $\ln X(e^{j\omega})$ exists. Moreover, the phase function will be an odd function of ω if:

$$\frac{1}{\pi} \int_0^{2\pi} \left[\frac{d\theta(\eta)}{d\eta} \right] d\eta = 0 \quad (4.69)$$

If the above constraint is not satisfied, then the computed phase function will exhibit absolute jumps greater than π .

4.4 The frequency response

Most discrete-time signals encountered in practice can be represented as a linear combination of a very large, maybe infinite, number of sinusoidal discrete-time signals of different angular frequencies. Thus, knowing the response of the LTI system to a single sinusoidal signal, we can determine its response to more complicated signals by making use of the superposition property.

Output of LTI systems for eigenfunction signals

An important property of an LTI system is that for certain types of input signals, called **eigenfunctions**, the output signal is the input signal multiplied by a complex constant. We consider here one such eigenfunction as the input.

Consider the LTI discrete-time system with an impulse response $\{h[n]\}$. Its input-output relationship in the time-domain is given by the convolution sum:

$$y[n] = \sum_{k=-\infty}^{\infty} h[k]x[n-k] \quad (4.70)$$

If the input is of the form:

$$x[n] = e^{j\omega n}, \quad -\infty < n < \infty \quad (4.71)$$

then it follows that the output is given by:

$$y[n] = \sum_{k=-\infty}^{\infty} h[k]x[n-k] = \left(\sum_{k=-\infty}^{\infty} h[k]e^{-j\omega k} \right) e^{j\omega n} \quad (4.72)$$

Now, let:

$$H(e^{j\omega}) = \sum_{k=-\infty}^{\infty} h[k]e^{-j\omega k} \quad (4.73)$$

Then we can write:

$$y[n] = H(e^{j\omega})e^{j\omega n} \quad (4.74)$$

Thus for a complex exponential input signal $e^{j\omega n}$, the output of an LTI discrete-time system is also a complex exponential signal of the same frequency multiplied by a complex constant $H(e^{j\omega})$. Thus, $e^{j\omega n}$ is an eigenfunction of the system.

The quantity $H(e^{j\omega})$ is called the **frequency response** of the LTI discrete-time system. $H(e^{j\omega})$ provides a frequency-domain description of the system and is precisely the **DTFT of the impulse response** $\{h[n]\}$ of the system. $H(e^{j\omega})$, in general, is a complex function of ω with a period 2π . It can be expressed in terms of its real and imaginary parts:

$$H(e^{j\omega}) = H_{\text{re}}(e^{j\omega}) + jH_{\text{im}}(e^{j\omega}) \quad (4.75)$$

or, in terms of its magnitude and phase:

$$H(e^{j\omega}) = |H(e^{j\omega})|e^{j\theta(\omega)} \quad (4.76)$$

where:

$$\theta(\omega) = \arg \{H(e^{j\omega})\} \quad (4.77)$$

The function $|H(e^{j\omega})|$ is called the **magnitude response** and the function $\theta(\omega)$ is called the **phase response** of the LTI discrete-time system. Design specifications for the LTI discrete-time system, in many applications, are given in terms of the magnitude response or the phase response or both.

In some cases, the magnitude function is specified in decibels as:

$$g(\omega) = 20 \log_{10} |H(e^{j\omega})| \text{ dB} \quad (4.78)$$

where $G(\omega)$ is called the **gain function**. The negative of the gain function $A(\omega) = -G(\omega)$ is called the **attenuation or loss function**.

Note that magnitude and phase functions are real functions of ω , whereas the frequency response is a complex function of ω . If the impulse response $h[n]$ is real then it follows that the magnitude function is an even function of ω :

$$|H(e^{j\omega})| = |H(e^{-j\omega})| \quad (4.79)$$

and the phase function is an odd function of ω :

$$\theta(\omega) = -\theta(-\omega) \quad (4.80)$$

Likewise, for a real impulse response $h[n]$, $H_{\text{re}}(e^{j\omega})$ is even and $H_{\text{im}}(e^{j\omega})$ is odd.

Frequency response

Magnitude and phase responses

Decibel notation

Gain and attenuation (or loss) functions

Example 30: M-point moving average filter

Consider the M -point moving average filter with an impulse response given by:

Frequency response
of M -point moving
average filter

$$h[n] = \begin{cases} \frac{1}{M} & 0 \leq n \leq M-1 \\ 0 & \text{otherwise} \end{cases} \quad (4.81)$$

Its frequency response is then given by:

$$H(e^{j\omega}) = \frac{1}{M} \sum_{n=0}^{M-1} e^{-j\omega n} \quad (4.82)$$

Performing all the calculations:

$$\begin{aligned} H(e^{j\omega}) &= \frac{1}{M} \left(\sum_{n=0}^{\infty} e^{-j\omega n} - \sum_{n=M}^{\infty} e^{-j\omega n} \right) \\ &= \frac{1}{M} \left(\sum_{n=0}^{\infty} e^{-j\omega n} \right) (1 - e^{-jM\omega}) \\ &= \frac{1}{M} \frac{1 - e^{-jM\omega}}{1 - e^{-j\omega}} \\ &= \frac{1}{M} \frac{\sin(\frac{M\omega}{2})}{\sin(\frac{\omega}{2})} e^{-j\frac{(M-1)\omega}{2}} \end{aligned} \quad (4.83)$$

Thus, the magnitude response of the M -point moving average filter is given by:

$$|H(e^{j\omega})| = \left| \frac{1}{M} \frac{\sin(\frac{M\omega}{2})}{\sin(\frac{\omega}{2})} \right| \quad (4.84)$$

and the phase response is given by

$$\theta(\omega) = -\frac{(M-1)\omega}{2} + \pi \sum_{k=1}^{\lfloor \frac{M}{2} \rfloor} \mu \left[\omega - \frac{2\pi k}{M} \right] \quad (4.85)$$

Note that the frequency response also determines the **steady-state response** of an LTI discrete-time system to a sinusoidal input.

Example 31: Steady-state response

Steady-state
response of an LTI
system to a
sinusoidal input

We determine the steady-state output $y[n]$ of a real coefficient LTI discrete-time system with a frequency response $H(e^{j\omega})$ for an input:

$$x[n] = A \cos(\omega_0 n + \varphi), \quad -\infty < n < \infty \quad (4.86)$$

We can express the input $x[n]$ as:

$$x[n] = g[n] + g^*[n] \quad (4.87)$$

where:

$$g[n] = \frac{1}{2} A e^{j\varphi} e^{j\omega_0 n} \quad (4.88)$$

Now the output of the system for an input $e^{j\omega_0 n}$ is simply $H(e^{j\omega_0})e^{j\omega_0 n}$. Because of linearity, the response $v[n]$ to an input $g[n]$ is given by:

$$v[n] = \frac{1}{2}Ae^{j\varphi}H(e^{j\omega_0})e^{j\omega_0 n} \quad (4.89)$$

Likewise, the output $v^*[n]$ to the input $g^*[n]$ is:

$$v^*[n] = \frac{1}{2}Ae^{-j\varphi}H(e^{-j\omega_0})e^{-j\omega_0 n} \quad (4.90)$$

Combining the last two equations we get:

$$\begin{aligned} y[n] &= v[n] + v^*[n] \\ &= \frac{1}{2}Ae^{j\varphi}H(e^{j\omega_0})e^{j\omega_0 n} + \frac{1}{2}Ae^{-j\varphi}H(e^{-j\omega_0})e^{-j\omega_0 n} \\ &= \frac{1}{2}A|H(e^{j\omega_0})|\left\{e^{j\theta(\omega_0)}e^{j\varphi}e^{j\omega_0 n} + e^{-j\theta(\omega_0)}e^{-j\varphi}e^{-j\omega_0 n}\right\} \\ &= A|H(e^{j\omega_0})|\cos(\omega_0 n + \theta(\omega_0) + \varphi) \end{aligned} \quad (4.91)$$

Thus, the output $y[n]$ has the same sinusoidal waveform as the input with two differences:

- the amplitude is multiplied by $|H(e^{j\omega_0})|$, the value of the magnitude function at $\omega = \omega_0$;
- the output has a phase lag relative to the input by an amount $\theta(\omega_0)$, the value of the phase function at $\omega = \omega_0$.

The expression for the steady-state response developed earlier assumes that the system is initially relaxed before the application of the input $x[n]$. In practice, excitation $x[n]$ to a discrete-time system is usually a right-sided sequence applied at some sample index $n = n_0$. Now, we develop the expression for the output for such an input.

Without any loss of generality, assume $x[n] = 0$ for $n < 0$. From the input-output relation in Eq. 4.70, we observe that for an input:

$$x[n] = e^{j\omega n}\mu[n] \quad (4.92)$$

the output is given by:

$$y[n] = \left(\sum_{k=-\infty}^{\infty} h[k]e^{j\omega(n-k)} \right) \mu[n] = \left(\sum_{k=-\infty}^{\infty} h[k]e^{-j\omega k} \right) e^{j\omega n} \mu[n] \quad (4.93)$$

The output for $n < 0$ is $y[n] = 0$, while for $n \geq 0$ it is given by:

$$\begin{aligned} y[n] &= \left(\sum_{k=0}^{\infty} h[k]e^{-j\omega k} \right) e^{j\omega n} - \left(\sum_{k=n+1}^{\infty} h[k]e^{-j\omega k} \right) e^{j\omega n} \\ &= H(e^{j\omega})e^{j\omega n} - \left(\sum_{k=n+1}^{\infty} h[k]e^{-j\omega k} \right) e^{j\omega n} \end{aligned} \quad (4.94)$$

Steady-state and transient responses

The first term on the RHS is the same as that obtained when the input is applied at $n = 0$ to an initially relaxed system and it is the **steady-state response**:

$$y_{\text{sr}}[n] = H(e^{j\omega})e^{j\omega n} \quad (4.95)$$

The second term on the RHS is called the **transient response**:

$$y_{\text{tr}}[n] = - \left(\sum_{k=n+1}^{\infty} h[k] e^{-j\omega k} \right) e^{j\omega n} \quad (4.96)$$

To determine the effect of the above term on the total output response, we observe:

$$|y_{\text{tr}}[n]| = \left| \sum_{k=n+1}^{\infty} h[k] e^{-j\omega(k-n)} \right| \leq \sum_{k=n+1}^{\infty} |h[k]| \leq \sum_{k=0}^{\infty} |h[k]| \quad (4.97)$$

For a causal, stable LTI IIR discrete-time system, $h[n]$ is absolutely summable. As a result, the transient response $y_{\text{tr}}[n]$ is a bounded sequence. Moreover, as $n \rightarrow \infty$:

$$\sum_{k=n+1}^{\infty} |h[k]| \rightarrow 0 \quad (4.98)$$

and hence, the transient response decays to zero as n gets very large.

For a causal FIR LTI discrete-time system with an impulse response $h[n]$ of length $N + 1$, $h[n] = 0$ for $n > N$. Hence, $y_{\text{tr}}[n] = 0$ for $n > N - 1$. Here the output reaches the steady-state value $y_{\text{sr}}[n] = H(e^{j\omega})e^{j\omega n}$ at $n = N$.

4.5 The concept of filtering

Digital filters and importance of Fourier transform

One application of an LTI discrete-time system is to pass certain frequency components in an input sequence without any distortion (if possible) and to block other frequency components. Such systems are called **digital filters** and one of the main subjects of discussion in this course.

The key to the filtering process is the Fourier transform

$$x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} H(e^{j\omega}) e^{j\omega n} d\omega \quad (4.99)$$

Frequency-selective filters

It expresses an arbitrary input as a linear weighted sum of an infinite number of exponential sequences, or equivalently, as a linear weighted sum of sinusoidal sequences. Thus, by appropriately choosing the values of the magnitude function $|H(e^{j\omega})|$ of the LTI digital filter at frequencies corresponding to the frequencies of the sinusoidal components of the input, some of these components can be selectively and heavily attenuated or filtered with respect to the others.

To understand the mechanism behind the design of **frequency-selective filters**, consider a real-coefficient LTI discrete-time system characterized by a magnitude function:

$$|H(e^{j\omega})| \approx \begin{cases} 1 & |\omega| \leq \omega_c \\ 0 & \omega_c < |\omega| \leq \pi \end{cases} \quad (4.100)$$

We apply to the system an input:

$$x[n] = A \cos(\omega_1 n) + B \cos(\omega_2 n), \quad 0 < \omega_1 < \omega_c < \omega_2 < \pi \quad (4.101)$$

Because of linearity, the output of this system is of the form:

$$y[n] = A |H(e^{j\omega_1})| \cos(\omega_1 n + \theta(\omega_1)) + B |H(e^{j\omega_2})| \cos(\omega_2 n + \theta(\omega_2)) \quad (4.102)$$

As $|H(e^{j\omega_1})| \approx 1$ and $|H(e^{j\omega_2})| \approx 0$, the output reduces to:

$$y[n] \approx A |H(e^{j\omega_1})| \cos(\omega_1 n + \theta(\omega_1)) \quad (4.103)$$

Thus, the system acts like a lowpass filter.

Now we consider an example of design of a very simple digital filter.

Example 32: Design of a simple digital filter

The input consists of a sum of two sinusoidal sequences of angular frequencies 0.1 rad/sample and 0.4 rad/sample. We need to design a highpass filter that will pass the high-frequency component of the input but block the low-frequency component.

For simplicity, assume the filter to be an FIR filter of length 3 with an impulse response:

$$h[0] = h[2] = \alpha \quad (4.104)$$

$$h[1] = \beta \quad (4.105)$$

The convolution sum description of this filter is then given by:

$$\begin{aligned} y[n] &= h[0]x[n]h[1]x[n-1] + h[2]x[n-2] \\ &= \alpha x[n] + \beta x[n-1] + \alpha x[n-2] \end{aligned} \quad (4.106)$$

$y[n]$ and $x[n]$ are, respectively, the output and the input sequences.

The design objective is to choose suitable values of α and β so that the output is a sinusoidal sequence with a frequency of 0.4 rad/sample.

Now, the frequency response of the FIR filter is given by:

$$\begin{aligned} H(e^{j\omega}) &= h[0] + h[1]e^{-j\omega} + h[2]e^{-j2\omega} \\ &= \alpha(1 + e^{-j2\omega}) + \beta e^{-j\omega} \\ &= 2\alpha\left(\frac{e^{j\omega} + e^{-j\omega}}{2}\right)e^{-j\omega} + \beta e^{-j\omega} \\ &= (2\alpha \cos \omega + \beta)e^{-j\omega} \end{aligned} \quad (4.107)$$

The magnitude and phase functions are:

$$|H(e^{j\omega})| = 2\alpha \cos \omega + \beta \quad (4.108)$$

$$\theta(\omega) = -\omega \quad (4.109)$$

In order to block the low-frequency component, the magnitude function at $\omega = 0.1$ should be equal to zero. Likewise, to pass the high-frequency component, the magnitude function at $\omega = 0.4$ should be equal to one. Thus, the two conditions that must be satisfied are:

$$|H(e^{j0.1})| = 2\alpha \cos(0.1) + \beta = 0 \quad (4.110)$$

$$|H(e^{j0.4})| = 2\alpha \cos(0.4) + \beta = 1 \quad (4.111)$$

Solving the above two equations we get:

$$\alpha = -6.76195 \quad (4.112)$$

$$\beta = 13.456335 \quad (4.113)$$

$$(4.114)$$

Thus the output-input relation of the FIR filter is given by:

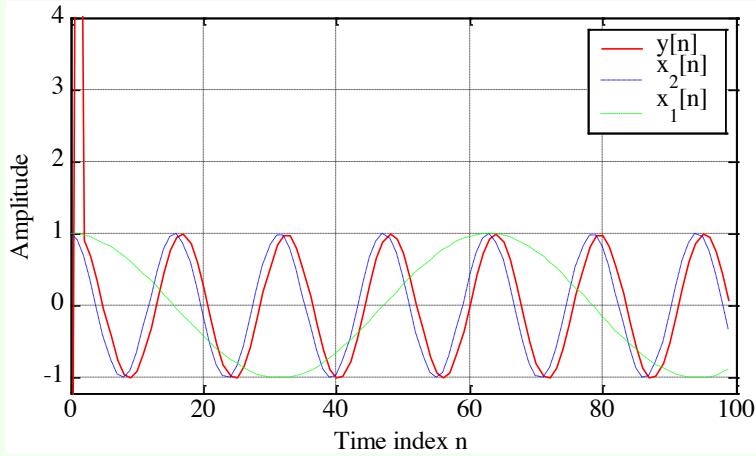
$$y[n] = -6.76195(x[n] + x[n-2]) + 13.456335x[n-1] \quad (4.115)$$

where the input is:

$$x[n] = \{\cos(0.1n) + \cos(0.4n)\}\mu[n] \quad (4.116)$$

A plot of the signals of interests is showed below.

Example of design of a simple digital filter



The first seven samples of the output are showed below as well.

n	$\cos(0.1n)$	$\cos(0.4n)$	$x[n]$	$y[n]$
0	1.0	1.0	2.0	-13.52390
1	0.9950041	0.9210609	1.9160652	13.956333
2	0.9800665	0.6967067	1.6767733	0.9210616
3	0.9553364	0.3623577	1.3176942	0.6967064
4	0.9210609	-0.0291995	0.8918614	0.3623572
5	0.8775825	-0.4161468	0.4614357	-0.0292002
6	0.8253356	-0.7373937	0.0879419	-0.4161467

From this table, it can be seen that, neglecting the least significant digit:

$$y[n] = \cos(0.4(n-1)), \quad n \geq 2 \quad (4.117)$$

Computation of the present value of the output requires the knowledge of the present and two previous input samples. Hence, the first two output samples, $y[0]$ and $y[1]$, are the result of assumed zero input sample values at $n = -1$ and $n = -2$. Therefore, first two output samples constitute the transient part of the output. Since the impulse response is of length 3, the steady-state is reached at $n = N = 2$. Note also that the output is delayed version of the high-frequency component $\cos(0.4n)$ of the input, and the delay is one sample period.

Lecture 12.
Thursday 5th
November, 2020.

4.6 Discrete Fourier Transform

We have discussed the DTFT for a discrete-time function given by:

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n} \quad (4.118)$$

and the inverse DTFT:

$$x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega})e^{j\omega n} d\omega \quad (4.119)$$

Drawbacks of DTFT

The pair and their properties and applications have some limitations. The input signal is usually aperiodic and may be finite in length.

Moreover, we often do not have an infinite amount of data which is required by DTFT. For example in a computer we cannot calculate uncountable infinite (continuum) of

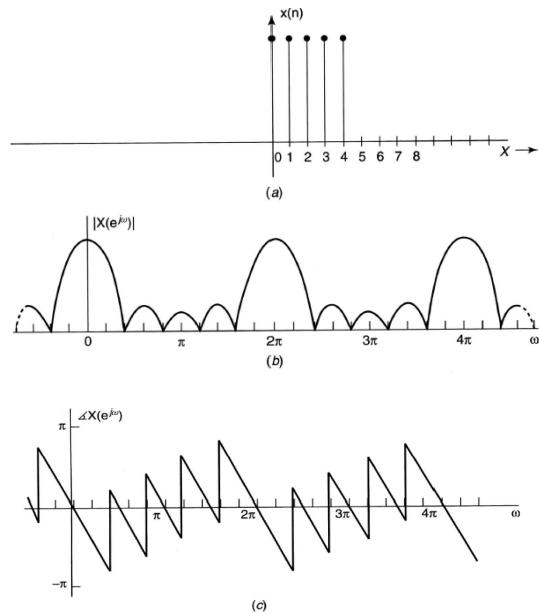


Figure 4.8: In order from top to bottom, a finite-length signal, its magnitude spectrum, its phase spectrum.

frequencies as required by DTFT. Thus, we use DTF to look at finite segment of data. We only observe the data through a window:

Discrete Fourier transform as solution

$$x_0[n] = x[n]w_R[n] \quad (4.120)$$

$$w_R[n] = \begin{cases} 1 & n = 0, 1, \dots, N-1 \\ 0 & \text{otherwise} \end{cases} \quad (4.121)$$

In this case, the $x_0[n]$ is just a sampled data between $n = 0, n = N-1$ (so, N points). The solution to our problems is given by the Discrete Fourier Transform (DFT).

Definition of discrete Fourier transform

Definition 9: Discrete Fourier Transform (DFT)

The simplest relation between a length- N sequence $x[n]$, defined for $0 \leq n \leq N-1$, and its DTFT $X(e^{j\omega})$ is obtained by uniformly sampling on the ω -axis between $0 \leq \omega \leq 2\pi$ at $\omega_k = \frac{2\pi k}{N}$, for $0 \leq k \leq N-1$. From the definition of the DTFT we thus have:

$$X[k] = [X(e^{j\omega})]_{\omega=\frac{2\pi k}{N}} = \sum_{n=0}^{N-1} x[n]e^{-j2\pi k \frac{n}{N}} \quad (4.122)$$

Note that $X[k]$ is also a length- N sequence in the frequency domain and it is called the **Discrete Fourier Transform (DFT)** of the sequence $x[n]$. Using the notation $W_N = e^{-j\frac{2\pi}{N}}$, the DFT is usually expressed as:

$$X[k] = \sum_{n=0}^{N-1} x[n]W_N^{kn}, \quad 0 \leq k \leq N-1 \quad (4.123)$$

Definition of inverse discrete Fourier transform

Definition 10: Inverse Discrete Fourier Transform (IDFT)

The **Inverse Discrete Fourier Transform (IDFT)** is given by:

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] W_N^{-kn}, \quad 0 \leq n \leq N-1 \quad (4.124)$$

Validation of the DFT definition

To verify the above expression, we multiply both sides of the above equation by $W_N^{\ell n}$ and sum the result from $n = 0$ to $n = N - 1$, resulting in:

$$\begin{aligned} \sum_{n=0}^{N-1} x[n] W_N^{\ell n} &= \sum_{n=0}^{N-1} \left(\frac{1}{N} \sum_{k=0}^{N-1} X[k] W_N^{-kn} \right) W_N^{\ell n} \\ &= \frac{1}{N} \sum_{n=0}^{N-1} \sum_{k=0}^{N-1} X[k] W_N^{-(k-\ell)n} \\ &= \frac{1}{N} \sum_{k=0}^{N-1} \sum_{n=0}^{N-1} X[k] W_N^{-(k-\ell)n} \end{aligned} \quad (4.125)$$

Making use of the identity:

$$\sum_{n=0}^{N-1} W_N^{-(k-\ell)n} = \begin{cases} N & k - \ell = rN, \quad r \in \mathbb{Z} \\ 0 & \text{otherwise} \end{cases} \quad (4.126)$$

we observe that the RHS of the last equation is equal to $X[\ell]$. Hence:

$$\sum_{n=0}^{N-1} x[n] W_N^{\ell n} = X[\ell] \quad (4.127)$$

Examples of discrete Fourier transform calculation

Example 33: Discrete Fourier Transform

Consider the length- N sequence:

$$x[n] = \begin{cases} 1 & n = 0 \\ 0 & 1 \leq n \leq N-1 \end{cases} \quad (4.128)$$

Its N -point DFT is given by:

$$X[k] = \sum_{n=0}^{N-1} x[n] W_N^{kn} = x[0] W_N^0 = 1 \quad (4.129)$$

with $0 \leq k \leq N - 1$.

Example 34: Discrete Fourier Transform

Consider the length- N sequence:

$$y[n] = \begin{cases} 1 & n = m \\ 0 & 0 \leq n \leq m-1, \quad m+1 \leq n \leq N-1 \end{cases} \quad (4.130)$$

Its N -point DFT is given by:

$$Y[k] = \sum_{n=0}^{N-1} y[n]W_N^{kn} = y[m]W_N^{km} = W_N^{km} \quad (4.131)$$

with $0 \leq k \leq N - 1$.

Example 35: Discrete Fourier Transform

Consider the length- N sequence defined for $0 \leq n \leq N - 1$:

$$g[n] = \cos\left(\frac{2\pi rn}{N}\right), \quad 0 \leq r \leq N - 1 \quad (4.132)$$

Using trigonometric identities, we can rewrite:

$$g[n] = \frac{1}{2}\left(e^{j2\pi r \frac{n}{N}} + e^{-j2\pi r \frac{n}{N}}\right) = \frac{1}{2}(W_N^{-rn} + W_N^{rn}) \quad (4.133)$$

The N -point DFT of $g[n]$ is thus given by:

$$G[k] = \sum_{n=0}^{N-1} g[n]W_N^{kn} = \frac{1}{2}\left(\sum_{n=0}^{N-1} W_N^{-(r-k)n} + \sum_{n=0}^{N-1} W_N^{(r+k)n}\right) \quad (4.134)$$

with $0 \leq k \leq N - 1$. Making use of the identity:

$$\sum_{n=0}^{N-1} W_N^{-(k-\ell)n} = \begin{cases} N & k - \ell = rN, r \in \mathbb{Z} \\ 0 & \text{otherwise} \end{cases} \quad (4.135)$$

we get:

$$G[k] = \begin{cases} \frac{N}{2} & k = r \\ \frac{N}{2} & k = N - r \\ 0 & \text{otherwise} \end{cases} \quad (4.136)$$

with $0 \leq k \leq N - 1$.

4.6.1 Matrix relations

The DFT analysis and synthesis formulas can now be easily expressed in a familiar matrix notation. More precisely, the DFT samples previously defined in Eq. 4.123 can be expressed in matrix form as:

$$\mathbf{X} = \mathbf{D}_N \mathbf{x} \quad (4.137)$$

where:

$$\mathbf{X} = \begin{bmatrix} X[0] \\ X[1] \\ \vdots \\ X[N-1] \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix} \quad (4.138)$$

and \mathbf{D}_N is the $N \times N$ **DFT matrix** given by:

DFT matrix form

DFT matrix

$$\mathbf{D}_N = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & W_N^1 & W_N^2 & \cdots & W_N^{(N-1)} \\ 1 & W_N^2 & W_N^4 & \cdots & W_N^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & W_N^{(N-1)} & W_N^{2(N-1)} & \cdots & W_N^{(N-1)^2} \end{bmatrix} \quad (4.139)$$

Likewise, the IDFT relation given by:

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] W_N^{-kn}, \quad 0 \leq n \leq N-1 \quad (4.140)$$

can be expressed in matrix form as:

$$\mathbf{x} = \mathbf{D}_N^{-1} \mathbf{X} \quad (4.141)$$

Inverse DFT matrix

where \mathbf{D}_N^{-1} is the $N \times N$ **inverse DFT matrix**, given by:

$$\mathbf{D}_N^{-1} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & W_N^{-1} & W_N^{-2} & \cdots & W_N^{-(N-1)} \\ 1 & W_N^{-2} & W_N^{-4} & \cdots & W_N^{-2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & W_N^{-(N-1)} & W_N^{-2(N-1)} & \cdots & W_N^{-(N-1)^2} \end{bmatrix} = \frac{1}{N} \mathbf{D}_N^* \quad (4.142)$$

4.6.2 DTFT from DFT by interpolation and sampling of DTFT

DTFT from DFT interpolation

The N -point DFT $X[k]$ of a length- N sequence $x[n]$ is simply the frequency samples of its DTFT $X(e^{j\omega})$ evaluated at N uniformly spaced frequency points:

$$\omega = \omega_k = \frac{2\pi k}{N}, \quad 0 \leq k \leq N-1 \quad (4.143)$$

Given the N -point DFT $X[k]$ of a length- N sequence $x[n]$, its DTFT $X(e^{j\omega})$ can be uniquely determined from $X[k]$. Thus:

$$\begin{aligned} X(e^{j\omega}) &= \sum_{n=0}^{N-1} x[n] e^{-j\omega n} \\ &= \sum_{n=0}^{N-1} \left[\frac{1}{N} \sum_{k=0}^{N-1} X[k] W_N^{-kn} \right] e^{-j\omega n} \\ &= \frac{1}{N} \sum_{k=0}^{N-1} X[k] \underbrace{\sum_{n=0}^{N-1} e^{-j(\omega - \frac{2\pi k}{N})n}}_S \end{aligned} \quad (4.144)$$

To develop a compact expression for the sum S , let $r = e^{-j(\omega - \frac{2\pi k}{N})}$. Then, we can rewrite the sum as:

$$S = \sum_{n=0}^{N-1} r^n \quad (4.145)$$

From Eq. 4.145:

$$\begin{aligned} rS &= \sum_{n=1}^N r^n = 1 + \sum_{n=1}^{N-1} r^n r^N - 1 \\ &= \sum_{n=0}^{N-1} r^n + r^N - 1 = S + r^N - 1 \end{aligned} \quad (4.146)$$

or, equivalently:

$$S - rS = (1 - r)S = 1 - r^N \quad (4.147)$$

Hence:

$$S = \frac{1 - r^N}{1 - r} = \frac{1 - e^{-j(\omega n - 2\pi k)}}{1 - e^{-j(\omega - \frac{2\pi k}{N})}} = \frac{\sin\left(\frac{\omega N 2\pi k}{2}\right)}{\sin\left(\frac{\omega N - 2\pi k}{2N}\right)} e^{-j\left(\frac{\omega - 2\pi k}{N}\right)\left(\frac{N-1}{2}\right)} \quad (4.148)$$

Therefore:

$$X(e^{j\omega}) = \frac{1}{N} \sum_{k=0}^{N-1} X[k] \frac{\sin\left(\frac{\omega N 2\pi k}{2}\right)}{\sin\left(\frac{\omega N - 2\pi k}{2N}\right)} e^{-j\left(\frac{\omega - 2\pi k}{N}\right)\left(\frac{N-1}{2}\right)} \quad (4.149)$$

Now, we study the inverse operation. We consider a sequence $x[n]$ with a DTFT $X(e^{j\omega})$. We sample $X(e^{j\omega})$ at N equally spaced points $\omega_k = \frac{2\pi k}{N}$, $0 \leq k \leq N-1$, developing the N frequency samples $\{X(e^{j\omega_k})\}$. These N frequency samples can be considered as an N -point DFT $Y[k]$ whose N -point inverse DFT is a length- N sequence $y[n]$. Now:

$$X(e^{j\omega}) = \sum_{\ell=-\infty}^{\infty} x[\ell] e^{-j\omega\ell} \quad (4.150)$$

Thus:

$$Y[k] = X(e^{j\omega_k}) = X(e^{j\frac{2\pi k}{N}}) = \sum_{\ell=-\infty}^{\infty} x[\ell] e^{-j2\pi k \frac{\ell}{N}} = \sum_{\ell=-\infty}^{\infty} x[\ell] W_N^{k\ell} \quad (4.151)$$

The inverse DFT of $Y[k]$ yields:

$$\begin{aligned} y[n] &= \frac{1}{N} \sum_{k=0}^{N-1} Y[k] W_N^{-kn} \\ &= \frac{1}{N} \sum_{k=0}^{N-1} \sum_{\ell=-\infty}^{\infty} x[\ell] W_N^{k\ell} W_N^{-kn} \\ &= \sum_{\ell=-\infty}^{\infty} x[\ell] \left[\sum_{k=0}^{N-1} W_N^{-k(n-\ell)} \right] \\ &= \sum_{m=-\infty}^{\infty} x[n + mN] \end{aligned} \quad (4.152)$$

with $0 \leq n \leq N-1$, where in the last passage the identity in Eq. 4.135 is employed. Thus, $y[n]$ is obtained from $x[n]$ by adding an infinite number of shifted replicas of $x[n]$, with each replica shifted by an integer multiple of N sampling instants, and observing the sum only for the interval $0 \leq n \leq N-1$.

To apply the last result to finite-length sequences, we assume that the samples outside the specified range are zeros. Thus, if $x[n]$ is a length- M sequence with $M \leq N$, then $y[n] = x[n]$ for $0 \leq n \leq N-1$. If $M > N$, there is a time-domain aliasing of samples of $x[n]$ in generating $y[n]$, and $x[n]$ cannot be recovered from $y[n]$.

Sampling the DTFT

Aliasing

Example 36: Aliasing

Let $x[n] = \{0, 1, 2, 3, 4, 5\}$. By sampling its DTFT $X(e^{j\omega})$ at $\omega_k = \frac{2\pi k}{4}$, with $0 \leq k \leq 3$, and then applying a 4-point IDFT to these samples, we arrive at the sequence $y[n]$ given by:

$$y[n] = x[n] + x[n+4] + x[n-4], \quad 0 \leq n \leq 3 \quad (4.153)$$

We get $y[n] = \{4, 6, 2, 3\}$. However, $x[n]$ cannot be recovered from $y[n]$.

DFT properties

4.6.3 DFT properties

Like the DTFT, the DFT also satisfies a number of properties that are useful in signal processing applications. Some of these properties are essentially identical to those of the DTFT, while some others are somewhat different. A summary of the DFT properties are given in Figures 4.9, 4.10 and 4.11.

Length- N Sequence	N -point DFT
$x[n]$	$X[k]$
$x^*[n]$	$X^*[(-k)_N]$
$x^*[(-n)_N]$	$X^*[k]$
$\text{Re}\{x[n]\}$	$X_{\text{pcs}}[k] = \frac{1}{2}\{X[(k)_N] + X^*[(-k)_N]\}$
$j \text{Im}\{x[n]\}$	$X_{\text{pca}}[k] = \frac{1}{2}\{X[(k)_N] - X^*[(-k)_N]\}$
$x_{\text{pcs}}[n]$	$\text{Re}\{X[k]\}$
$x_{\text{pca}}[n]$	$j \text{Im}\{X[k]\}$

Note: $x_{\text{pcs}}[n]$ and $x_{\text{pca}}[n]$ are the periodic conjugate-symmetric and periodic conjugate-antisymmetric parts of $x[n]$, respectively. Likewise, $X_{\text{pcs}}[k]$ and $X_{\text{pca}}[k]$ are the periodic conjugate-symmetric and periodic conjugate-antisymmetric parts of $X[k]$, respectively.

Figure 4.9: Symmetry relations of DFT for a complex sequence $x[n]$.

Length- N Sequence	N -point DFT
$x[n]$	$X[k] = \text{Re}\{X[k]\} + j \text{Im}\{X[k]\}$
$x_{\text{pe}}[n]$	$\text{Re}\{X[k]\}$
$x_{\text{po}}[n]$	$j \text{Im}\{X[k]\}$
Symmetry relations	$X[k] = X^*[(-k)_N]$ $\text{Re } X[k] = \text{Re } X[(-k)_N]$ $\text{Im } X[k] = -\text{Im } X[(-k)_N]$ $ X[k] = X[(-k)_N] $ $\arg X[k] = -\arg X[(-k)_N]$

Note: $x_{\text{pe}}[n]$ and $x_{\text{po}}[n]$ are the periodic even and periodic odd parts of $x[n]$, respectively.

Figure 4.10: Symmetry relations of DFT for a real sequence $x[n]$.

4.6.4 Circular shift and convolution of a sequence

After having listed some of the most common and useful properties of DFT, we focus on the operations of circular shift and circular convolution. Let us start from the

Type of Property	Length- N Sequence	N -point DFT
	$\frac{g[n]}{h[n]}$	$\frac{G[k]}{H[k]}$
Linearity	$\alpha g[n] + \beta h[n]$	$\alpha G[k] + \beta H[k]$
Circular time-shifting	$g[\langle n - n_0 \rangle_N]$	$W_N^{kn_0} G[k]$
Circular frequency-shifting	$W_N^{-k_0 n} g[n]$	$G[\langle k - k_0 \rangle_N]$
Duality	$G[n]$	$N g[\langle -k \rangle_N]$
N -point circular convolution	$\sum_{m=0}^{N-1} g[m]h[\langle n - m \rangle_N]$	$G[k]H[k]$
Modulation	$g[n]h[n]$	$\frac{1}{N} \sum_{m=0}^{N-1} G[m]H[\langle k - m \rangle_N]$

Parseval's relation	$\sum_{n=0}^{N-1} x[n] ^2 = \frac{1}{N} \sum_{k=0}^{N-1} X[k] ^2$
---------------------	---

Figure 4.11: General properties of DFT.

former.

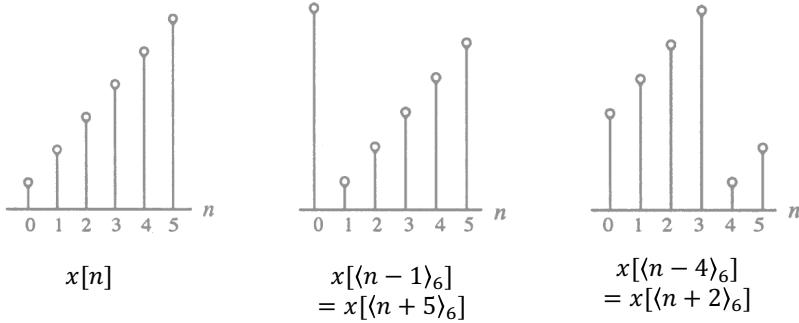
This property is analogous to the time-shifting property of the DTFT but with a difference. We inspect it by considering length- N sequences defined for $0 \leq n \leq N-1$. The sample values of such sequences are equal to zero for values of $n < 0$ and $n \geq N$. If $x[n]$ is such a sequence, then for any arbitrary integer n_0 the shifted sequence $x_1[n] = x[n - n_0]$ is no longer defined for the range $0 \leq n \leq N-1$. We thus need to define another type of shift that will always keep the shifted sequence in the range $0 \leq n \leq N-1$. The desired shift, called the **circular shift**, is defined using a **modulo operation**:

$$x_c[n] = x[\langle n - n_0 \rangle_N] \quad (4.154)$$

For $n_0 > 0$ (**right circular shift**), the above equation implies:

$$x_c[n] = \begin{cases} x[n - n_0] & n_0 \leq n \leq N-1 \\ x[N - n_0 + n] & 0 \leq n \leq n_0 \end{cases} \quad (4.155)$$

An illustration of the concept of circular shift is showed in Figure 4.12. As it is possible to observe, a right circular shift by n_0 is equivalent to a left circular shift by $N - n_0$ sample periods. A circular shift by an integer number n_0 greater than N is equivalent to a circular shift by $\langle n_0 \rangle_N$.

**Figure 4.12:** Illustration of circular shift.

Now, we pass to the circular convolution. This operation is analogous to the linear convolution, but with a difference. Consider two length- N sequences, $g[n]$ and $h[n]$, respectively. Their linear convolution results in a length- $(2N-1)$ sequence $y_L[n]$

Circular shift of a sequence

Circular shift and modulo operation

Circular convolution of a sequence

given by:

$$y_L[n] = \sum_{m=0}^{N-1} g[m]h[n-m], \quad 0 \leq n \leq 2N-2 \quad (4.156)$$

In computing $y_L[n]$, we have assumed that both length- N sequences have been zero-padded to extend their lengths to $2N-1$. The longer form of $y_L[n]$ results from the time-reversal of the sequence $h[n]$ and its linear shift to the right. The first nonzero value of $y_L[n]$ is $y_L[0] = g[0]h[0]$ and the last nonzero value is $y_L[2N-2] = g[N-1]h[N-1]$.

To develop a convolution-like operation resulting in a length- N sequence $y_C[n]$, we need to define a **circular time-reversal**, and then apply a **circular time-shift**. Resulting operation, called a **circular convolution**, is defined by:

$$y_C[n] = \sum_{m=0}^{N-1} g[m]h[\langle n-m \rangle_N], \quad 0 \leq n \leq N-1 \quad (4.157)$$

Since the operation defined involves two length- N sequences, it is often referred to as an **N -point circular convolution**, denoted as:

$$y[n] = g[n] \circledast h[n] \quad (4.158)$$

We remark that the circular convolution is commutative, namely:

$$g[n] \circledast h[n] = h[n] \circledast g[n] \quad (4.159)$$

Now, we study an example to understand how this operation works.

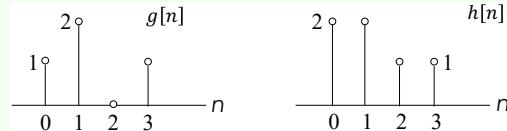
Example 37: Circular convolution

Here, we compute the 4-point circular convolution of the two length-4 sequences:

$$g[n] = \{1, 2, 0, 1\} \quad (4.160)$$

$$h[n] = \{2, 2, 1, 1\} \quad (4.161)$$

as sketched in the figure below.



The result is a length-4 sequence $y_C[n]$ given by:

$$y_C[n] = g[n] \circledast h[n] = \sum_{m=0}^3 g[m]h[\langle n-m \rangle_4], \quad 0 \leq n \leq 3 \quad (4.162)$$

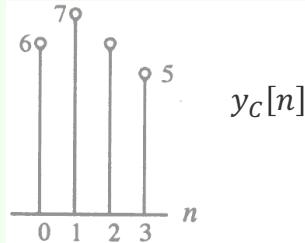
We can observe that:

$$\begin{aligned} y_C[0] &= \sum_{m=0}^3 g[m]h[\langle -m \rangle_4] \\ &= g[0]h[0] + g[1]h[3] + g[2]h[2] + g[3]h[1] \\ &= (1 \cdot 2) + (2 \cdot 1) + (0 \cdot 1) + (1 \cdot 2) = 6 \end{aligned} \quad (4.163)$$

$$\begin{aligned}
y_C[1] &= \sum_{m=0}^3 g[m]h[\langle 1 - m \rangle_4] \\
&= g[0]h[1] + g[1]h[0] + g[2]h[3] + g[3]h[2] \\
&= (1 \cdot 2) + (2 \cdot 2) + (0 \cdot 1) + (1 \cdot 1) = 7
\end{aligned} \tag{4.164}$$

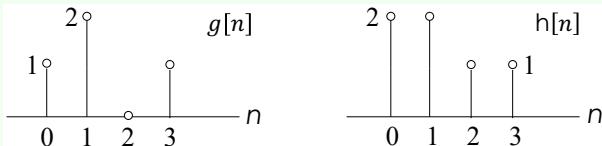
$$\begin{aligned}
y_C[2] &= \sum_{m=0}^3 g[m]h[\langle 2 - m \rangle_4] \\
&= g[0]h[2] + g[1]h[1] + g[2]h[0] + g[3]h[3] \\
&= (1 \cdot 1) + (2 \cdot 2) + (0 \cdot 2) + (1 \cdot 1) = 6
\end{aligned} \tag{4.165}$$

$$\begin{aligned}
y_C[3] &= \sum_{m=0}^3 g[m]h[\langle 3 - m \rangle_4] \\
&= g[0]h[3] + g[1]h[2] + g[2]h[1] + g[3]h[0] \\
&= (1 \cdot 1) + (2 \cdot 1) + (0 \cdot 2) + (1 \cdot 2) = 5
\end{aligned} \tag{4.166}$$



Example 38: Circular convolution

Here, we compute the 4-point circular convolution of the two length-4 sequences in the figure below:



The 4-point DFT $G[k]$ of $g[n]$ is given by:

$$\begin{aligned}
G[k] &= g[0] + g[1]e^{-j\frac{2\pi k}{4}} + g[2]e^{-j\frac{4\pi k}{4}} + g[3]e^{-j\frac{6\pi k}{4}} \\
&= 1 + 2e^{-j\frac{\pi k}{2}} + e^{-j\frac{3\pi k}{2}}, \quad 0 \leq k \leq 3
\end{aligned} \tag{4.167}$$

Therefore:

$$G[0] = 1 + 2 + 1 = 4 \tag{4.168}$$

$$G[1] = 1 - j2 + j = 1 - j \tag{4.169}$$

$$G[2] = 1 - 2 - 1 = -2 \tag{4.170}$$

$$G[3] = 1 + j2 - j = 1 + j \tag{4.171}$$

Likewise:

$$\begin{aligned}
H[k] &= h[0] + h[1]e^{-j\frac{2\pi k}{4}} + h[2]e^{-j\frac{4\pi k}{4}} + h[3]e^{-j\frac{6\pi k}{4}} \\
&= 2 + 2e^{-j\frac{\pi k}{2}} + e^{-j\pi k} + e^{-j\frac{3\pi k}{2}}, \quad 0 \leq k \leq 3
\end{aligned} \tag{4.172}$$

Hence, we get:

$$H[0] = 2 + 2 + 1 + 1 = 6 \quad (4.173)$$

$$H[1] = 2 - j2 - 1 + j = 1 - j \quad (4.174)$$

$$H[2] = 2 - 2 + 1 - 1 = 0 \quad (4.175)$$

$$H[3] = 2 + j2 - 1 - j = 1 + j \quad (4.176)$$

The two 4-point DFTs can also be computed using the matrix relations:

$$\begin{bmatrix} G[0] \\ G[1] \\ G[2] \\ G[3] \end{bmatrix} = \mathbf{D}_4 \begin{bmatrix} g[0] \\ g[1] \\ g[2] \\ g[3] \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -j & -1 & j \\ 1 & -1 & 1 & -1 \\ 1 & j & -1 & -j \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 1-j \\ -2 \\ 1+j \end{bmatrix} \quad (4.177)$$

$$\begin{bmatrix} H[0] \\ H[1] \\ H[2] \\ H[3] \end{bmatrix} = \mathbf{D}_4 \begin{bmatrix} h[0] \\ h[1] \\ h[2] \\ h[3] \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -j & -1 & j \\ 1 & -1 & 1 & -1 \\ 1 & j & -1 & -j \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 1-j \\ 0 \\ 1+j \end{bmatrix} \quad (4.178)$$

where \mathbf{D}_4 is the 4-point DFT matrix. If $Y_C[k]$ denotes the 4-point DFT of $y_C[n]$, we observe $Y_C[k] = G[k]H[k]$ for $0 \leq k \leq 3$, so:

$$\begin{bmatrix} Y_C[0] \\ Y_C[1] \\ Y_C[2] \\ Y_C[3] \end{bmatrix} = \begin{bmatrix} G[0]H[0] \\ G[1]H[1] \\ G[2]H[2] \\ G[3]H[3] \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 24 \\ -j2 \\ 0 \\ j2 \end{bmatrix} \quad (4.179)$$

and the 4-point inverse DFT of $Y_C[k]$ yields:

$$\begin{bmatrix} y_C[0] \\ y_C[1] \\ y_C[2] \\ y_C[3] \end{bmatrix} = \frac{1}{4} \mathbf{D}_4^* \begin{bmatrix} Y_C[0] \\ Y_C[1] \\ Y_C[2] \\ Y_C[3] \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & j & -1 & -j \\ 1 & -1 & 1 & -1 \\ 1 & -j & -1 & j \end{bmatrix} \begin{bmatrix} 24 \\ -j \\ 0 \\ j2 \end{bmatrix} = \begin{bmatrix} 6 \\ 7 \\ 6 \\ 5 \end{bmatrix} \quad (4.180)$$

Example 39: Circular convolution

Now, let us extend the two length-4 sequences to length-7 by appending each with three zero-valued samples, namely:

$$g_e[n] = \begin{cases} g[n] & 0 \leq n \leq 3 \\ 0 & 4 \leq n \leq 6 \end{cases} \quad (4.181)$$

$$h_e[n] = \begin{cases} h[n] & 0 \leq n \leq 3 \\ 0 & 4 \leq n \leq 6 \end{cases} \quad (4.182)$$

We next determine the 7-point circular convolution of $g_e[n]$ and $h_e[n]$:

$$y[n] = \sum_{m=0}^6 g_e[n]h_e[\langle n-m \rangle_7], \quad 0 \leq n \leq 6 \quad (4.183)$$

Therefore:

$$\begin{aligned} y[0] &= g_e[0]h_e[0] + g_e[1]h_e[6] + g_e[3]h_e[4] + g_e[4]h_e[3] + g_e[5]h_e[2] + g_e[6]h_e[1] \\ &= g[0]h[0] = 1 \cdot 2 = 2 \end{aligned} \quad (4.184)$$

Continuing the process, we arrive at:

$$y[1] = g_e[0]h_e[1] + g_e[1]h_e[0] = 6 \quad (4.185)$$

$$y[2] = g_e[0]h_e[2] + g_e[1]h_e[1] + g_e[2]h_e[0] = 5 \quad (4.186)$$

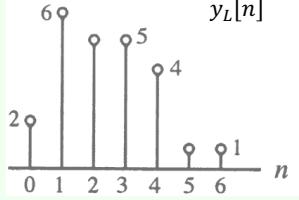
$$y[3] = g_e[0]h_e[3] + g_e[1]h_e[2] + g_e[2]h_e[1] + g_e[3]h_e[0] = 5 \quad (4.187)$$

$$y[4] = g_e[1]h_e[3] + g_e[2]h_e[2] + g_e[3]h_e[1] = 4 \quad (4.188)$$

$$y[5] = g_e[2]h_e[3] + g_e[3]h_e[2] = 1 \quad (4.189)$$

$$y[6] = g_e[3]h_e[3] = 1 \quad (4.190)$$

So, $y[n]$ is precisely the sequence $y_L[n]$ obtained by a linear convolution of $g[n]$ and $h[n]$.



In general, the N -point circular convolution can be written in matrix form as:

$$\begin{bmatrix} y_C[0] \\ y_C[1] \\ y_C[2] \\ \vdots \\ y_C[N-1] \end{bmatrix} = \begin{bmatrix} h[0] & h[n-1] & \cdots & h[1] \\ h[1] & [0] & \cdots & h[2] \\ h[2] & h[1] & \cdots & h[3] \\ \vdots & \vdots & \ddots & \vdots \\ 1 & h[n-2] & \cdots & h[0] \end{bmatrix} \begin{bmatrix} g[0] \\ g[1] \\ g[2] \\ \vdots \\ g[N-1] \end{bmatrix} \quad (4.191)$$

Note that the elements of each diagonal of the $N \times N$ matrix are equal. We call such a matrix a **circulant matrix**.

4.7 DFT of real sequences

In most practical applications, sequences of interest are real. In such cases, the symmetry properties of the DFT can be exploited to make the DFT computations more efficient.

Let $g[n]$ and $h[n]$ be two length- N real sequences with $G[k]$ and $H[k]$ denoting their respective N -point DFTs. These two N -point DFTs can be computed efficiently using a single N -point DFT. Now, define a complex length- N sequence:

$$x[n] = g[n] + jh[n] \quad (4.192)$$

Hence, $g[n] = \text{Re}\{x[n]\}$ and $h[n] = \text{Im}\{x[n]\}$. Let $X[k]$ denote the N -point DFT of $x[n]$. Then, we arrive at:

$$G[k] = \frac{1}{2} \{X[k] + X^*[\langle -k \rangle_N]\} \quad (4.193)$$

$$H[k] = \frac{1}{2j} \{X[k] - X^*[\langle -k \rangle_N]\} \quad (4.194)$$

Note that for $0 \leq k \leq N-1$:

$$X^*[\langle -k \rangle_N] = X^*[\langle N-k \rangle_N] \quad (4.195)$$

Matrix form of circular convolution and circulant matrix

Lecture 13.
Tuesday 10th
November, 2020.

DFT of real sequences

Example 40: DFT of real sequences

We compute the 4-point DFTs of the two real sequences $g[n]$ and $h[n]$:

$$g[n] = \{1, 2, 0, 1\} \quad (4.196)$$

$$h[n] = \{2, 2, 1, 1\} \quad (4.197)$$

Then $x[n] = g[n] + jh[n]$ is given by:

$$x[n] = \{1 + j2, 2 + j2, 1 + j\} \quad (4.198)$$

Its DFT $X[k]$ is:

$$\begin{bmatrix} X[0] \\ X[1] \\ X[2] \\ X[3] \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -j & -1 & j \\ 1 & -1 & 1 & -1 \\ 1 & j & -1 & -1 \end{bmatrix} \begin{bmatrix} 1 + j2 \\ 2 + j2 \\ j \\ 1 + j \end{bmatrix} = \begin{bmatrix} 4 + j6 \\ 2 \\ -2 \\ j2 \end{bmatrix} \quad (4.199)$$

From the above:

$$X^*[k] = \{4 - j6, 2, -2, -j2\} \quad (4.200)$$

Hence:

$$X^*[\langle 4 - k \rangle_4] = \{4 - j6, -j2, -2, 2\} \quad (4.201)$$

Therefore:

$$G[k] = \{4, 1 - j, -2, 1 + j\} \quad (4.202)$$

$$H[k] = \{6, 1 - j, 0, 1 + j\} \quad (4.203)$$

verifying the results derived in Lecture 12.

Now, let $v[n]$ be a length- $2N$ real sequence with a $2N$ -point DFT $V[k]$. Then, we define two length- N real sequences $g[n]$ and $h[n]$ as follows. Let $G[k]$ and $H[k]$ denote their respective N -point DFTs:

$$\begin{cases} g[n] = v[2n] \\ h[n] = v[2n + 1] \end{cases} \quad 0 \leq n \leq N \quad (4.204)$$

We define a length- N complex sequence $x[n] = g[n] + jh[n]$ with an N -point DFT $X[k]$. Then, as showed earlier:

$$G[k] = \frac{1}{2} \{X[k] + X^*[\langle -k \rangle_N]\} \quad (4.205)$$

$$H[k] = \frac{1}{2j} \{X[k] - X^*[\langle -k \rangle_N]\} \quad (4.206)$$

Now, for $0 \leq k \leq 2N - 1$:

$$\begin{aligned}
V[k] &= \sum_{n=0}^{2N-1} v[n]W_{2N}^{nk} \\
&= \sum_{n=0}^{N-1} v[2n]W_{2N}^{2nk} + \sum_{n=0}^{N-1} v[2n+1]W_{2N}^{(2n+1)k} \\
&= \sum_{n=0}^{N-1} g[n]W_N^{nk} + \sum_{n=0}^{N-1} W_N^{nk}W_{2N}^k \\
&= \sum_{n=0}^{N-1} g[n]W_N^{nk} + W_{2N}^k \sum_{n=0}^{N-1} h[n]W_N^{nk} \\
&= G[\langle k \rangle_N] + W_{2N}^k H[\langle k \rangle_N]
\end{aligned} \tag{4.207}$$

Example 41: DFT of real sequences

Let us determine the 8-point DFT $V[k]$ of the length-8 real sequence:

$$v[n] = \{1, 2, 2, 2, 0, 1, 1, 1\} \tag{4.208}$$

We form two length-4 real sequences as follows:

$$g[n] = v[2n] = \{1, 2, 0, 1\} \tag{4.209}$$

$$h[n] = v[2n+1] = \{2, 2, 1, 1\} \tag{4.210}$$

Now:

$$V[k] = G[\langle k \rangle_4] + W_8^k H[\langle k \rangle_4] \quad 0 \leq k \leq 7 \tag{4.211}$$

Substituting the values of the 4-point DFTs $G[k]$ and $H[k]$ computed earlier, we get:

$$V[0] = G[0] + H[0] = 4 + 6 = 10 \tag{4.212}$$

$$V[1] = G[1] + W_1^0 H[1] = (1 - j) + e^{-j\frac{\pi}{4}}(1 - j) = 1 - j2.4142 \tag{4.213}$$

$$V[2] = G[2] + W_2^0 H[2] = -2 + e^{-j\frac{3\pi}{4}} \cdot 0 = -2 \tag{4.214}$$

$$V[3] = G[3] + W_3^0 H[3] = (1 + j) + e^{-j\frac{3\pi}{4}}(1 + j) = 1 - j0.4142 \tag{4.215}$$

$$V[4] = G[0] + W_4^0 H[0] = 4 + e^{-j\pi} \cdot 6 = -2 \tag{4.216}$$

$$V[5] = G[1] + W_5^0 H[1] = (1 - j) + e^{-j\frac{5\pi}{4}}(1 - j) = 1 + j0.4142 \tag{4.217}$$

$$V[6] = G[2] + W_6^0 H[2] = -2 + e^{-j\frac{3\pi}{2}} \cdot 0 = -2 \tag{4.218}$$

$$V[7] = G[3] + W_7^0 H[3] = (1 + j) + e^{-j\frac{7\pi}{4}}(1 + j) = 1 + j2.4142 \tag{4.219}$$

4.8 Linear convolution using the DFT

Linear convolution is a key operation in many signal processing applications. Since a DFT can be efficiently implemented using FFT algorithms, it is of particular interest to develop methods for the implementation of linear convolution using the DFT.

For the purpose, let $g[n]$ and $h[n]$ be two finite-length sequences of length N and M , respectively. Moreover, we denote with $L = N + M + 1$. Then, we define two

Linear convolution using the DFT

length- L sequences:

$$g_e[n] = \begin{cases} g[n] & 0 \leq n \leq N-1 \\ 0 & N \leq n \leq L-1 \end{cases} \quad (4.220)$$

$$h_e[n] = \begin{cases} h[n] & 0 \leq n \leq M-1 \\ 0 & M \leq n \leq L-1 \end{cases} \quad (4.221)$$

Then:

$$y_L[n] = g[n] \circledast h[n] = y_C[n] = g_e[n] \odot h_e[n] \quad (4.222)$$

The corresponding implementation scheme is illustrated in Figure 4.13.

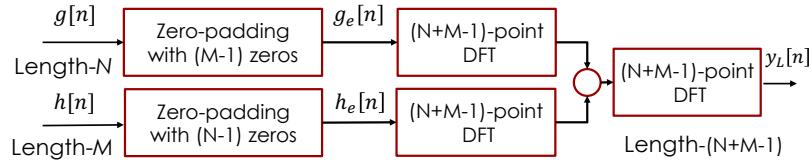


Figure 4.13: Scheme of linear convolution of two finite-length sequences.

We next consider the **DFT-based implementation** of:

$$y[n] = \sum_{\ell=0}^{M-1} h[\ell]x[n-\ell] = h[n] \circledast x[n] \quad (4.223)$$

where $h[n]$ is a finite-length sequence of length M and $x[n]$ is an infinite length (or a finite length sequence of length much greater than M). We first segment $x[n]$, assumed to be a causal sequence here without any loss of generality, into a set of contiguous finite-length subsequences $x_m[n]$ of length N each:

$$x[n] = \sum_{m=0}^{\infty} x_m[n - mN] \quad (4.224)$$

where:

$$x_m[n] = \begin{cases} x[n + mN] & 0 \leq n \leq N-1 \\ 0 & \text{otherwise} \end{cases} \quad (4.225)$$

Thus, we can write:

$$y[n] = h[n] \circledast x[n] = \sum_{m=0}^{\infty} y_m[n - mN] \quad (4.226)$$

where:

$$y_m = h[n] \circledast x_m[n] \quad (4.227)$$

Since $h[n]$ is of length M and $x_m[n]$ is of length N , the linear convolution $h[n] \circledast x_m[n]$ is of length $N + M - 1$.

As a result, the desired linear convolution $y[n] = h[n] \circledast x[n]$ has been broken up into a **sum of infinite number of short-length linear convolutions** of length $N + M - 1$ each:

$$y_m[n] = h[n] \circledast x_m[n] \quad (4.228)$$

Each of these short convolutions can be implemented using the DFT-based method discussed earlier, where now the DFTs (and the inverse DFT) are computed on the basis of $N + M - 1$ points.

There is one more subtlety to take care of before we can implement: the calculation in Eq. 4.226 using the DFT-based approach. Let us explain it proceeding in order:

- the first convolution in Eq. 4.226, namely $y_0 = h[n] \circledast x_0[n]$, is of length $N + M - 1$ and it is defined for $0 \leq n \leq N + M - 2$;
 - the second short convolution, namely $y_1[n] = h[n] \circledast x_1[n]$, is also of length $N + M - 1$ but it is defined for $N \leq n \leq 2N + M - 2$;
 - likewise, the third short convolution, namely $y_2[n] = h[n] \circledast x_2[n]$, is also of length $N + M - 1$ but it is defined for $2N \leq b \leq 3N + M - 2$, and so on and so forth;
 - in these particular cases, we can observe that there is an overlap of $M - 1$ samples between $h[n] \circledast x_1[n]$ and $h[n] \circledast x_2[n]$.

In general, there will be an overlap of $M - 1$ samples between the samples of the short convolutions $h[n] \circledast x_{r-1}[n]$ and $h[n] \circledast x_r[n]$ for $(r - 1)N \leq n \leq rN + M - 1$. This process is illustrated in Figures 4.14 and 4.15 for $M = 5$ and $N = 7$. In these cases, $y[n]$ obtained by a linear convolution of $x[n]$ and $h[n]$ is given by:

$$y[n] = y_0[n] \quad 0 \leq n \leq 6 \quad (4.229)$$

$$y[n] = y_0[n] + y_1[n - 7] \quad 7 \leq n \leq 10 \quad (4.230)$$

$$y[n] = y_1[n - 7] \quad 11 \leq n \leq 13 \quad (4.231)$$

$$y[n] = y_1[n-7] + y_2[n-14] \quad 14 \leq n \leq 17 \quad (4.232)$$

$$y[n] = y_2[n - 14] \quad 18 \leq n \leq 20 \quad (4.233)$$

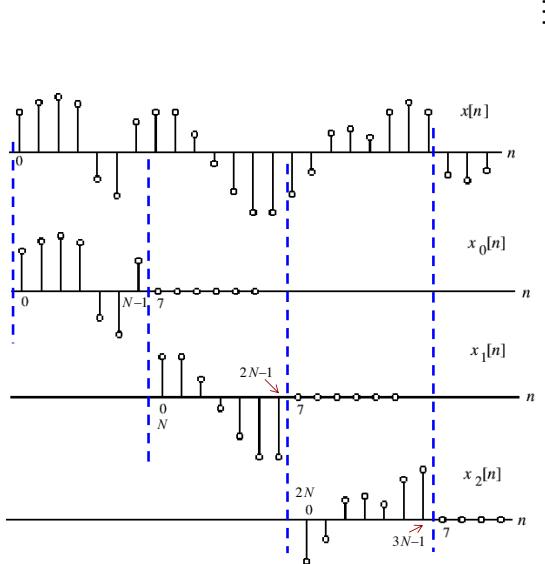


Figure 4.14: Overlap-add method for $M = 5$ and $N = 7$.

The above procedure is called the **overlap-add method** since the results of the short linear convolutions overlap and the overlapped portions are added to get the correct final result.

In implementing the overlap-add method using the DFT, we need to compute two $(N + M - 1)$ -point DFTs and one $(N + M - 1)$ -point IDFT since the overall linear convolution was expressed as a sum of short-length linear convolutions of length $N + M - 1$ each. It is possible to implement the overall linear convolution by performing instead circular convolution of length shorter than $N + M - 1$. To this end, it is necessary to segment $x[n]$ into overlapping blocks $x_m[n]$, keep the terms of the circular convolution of $h[n]$ with $x_m[n]$ that corresponds to the terms obtained by a linear convolution of $h[n]$ and $x_m[n]$, and throw away the other parts of the circular convolution.

Overlap-add method

Implementation using circular convolution

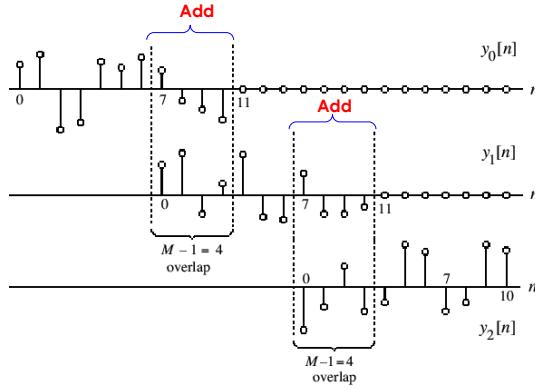


Figure 4.15: Overlap-add method for $M = 5$ and $N = 7$.

To understand the correspondence between the linear and circular convolutions, consider a length-4 sequence $x[n]$ and a length-3 sequence $h[n]$. Let $y_L[n]$ denote the result of a linear convolution of $x[n]$ with $h[n]$. The six samples of $y_L[n]$ are given by:

$$y_L[0] = h[0]x[0] \quad (4.234)$$

$$y_L[1] = h[0]x[1] + h[1]x[0] \quad (4.235)$$

$$y_L[2] = h[0]x[2] + h[1]x[1] + h[2]x[0] \quad (4.236)$$

$$y_L[3] = h[0]x[3] + h[1]x[2] + h[2]x[1] \quad (4.237)$$

$$y_L[4] = h[1]x[3] + h[2]x[2] \quad (4.238)$$

$$y_L[5] = h[2]x[3] \quad (4.239)$$

If we append $h[n]$ with a single zero-valued sample and convert it into a length-4 sequence $h_e[n]$, the 4-point circular convolution $y_C[n]$ of $h_e[n]$ and $x[n]$ is given by:

$$y_C[0] = h[0]x[0] + h[1]x[3] + h[2]x[2] \quad (4.240)$$

$$y_C[1] = h[0]x[1] + h[1]x[0] + h[2]x[3] \quad (4.241)$$

$$y_C[2] = h[0]x[2] + h[1]x[1] + h[2]x[0] \quad (4.242)$$

$$y_C[3] = h[0]x[3] + h[1]x[2] + h[2]x[1] \quad (4.243)$$

If we compare the expressions for the samples of $y_L[n]$ with the samples of $y_C[n]$, we observe that the first 2 terms of $y_C[n]$ do not correspond to the first 2 terms of $y_L[n]$, whereas the last 2 terms of $y_C[n]$ are precisely the same as the third and the forth terms of $y_L[n]$.

In general, if we consider the N -point circular convolution of a length- M sequence $h[n]$ with a length- N sequence $x[n]$ with $N > M$, the first $M - 1$ samples of the circular convolution are incorrect and are rejected. The remaining $N - M + 1$ samples correspond to the correct samples of the linear convolution of $h[n]$ with $x[n]$.

Now we consider an infinitely long or very long sequence $x[n]$. We break it up as a collection of smaller length (length-4) overlapping sequences $x_m[n]$ as $x_m[n] = x[n + 2m]$, with $0 \leq n \leq 3$, $0 \leq m \leq \infty$. Next, we form:

$$w_m[n] = h[n] \odot x_m[n] \quad (4.244)$$

or, equivalently

$$w_m[0] = h[0]x_m[0] + h[1]x_m[3] + h[2]x_m[2] \quad (4.245)$$

$$w_m[1] = h[0]x_m[1] + h[1]x_m[0] + h[2]x_m[3] \quad (4.246)$$

$$w_m[2] = h[0]x_m[2] + h[1]x_m[1] + h[2]x_m[0] \quad (4.247)$$

$$w_m[3] = h[0]x_m[3] + h[1]x_m[2] + h[2]x_m[1] \quad (4.248)$$

Computing the above for $m = 0, 1, 2, 3, \dots$, and substituting the values of $x_m[n]$, we arrive at:

$$\begin{aligned}
 w_0[0] &= h[0]x[0] + h[1]x[3] + h[2]x[2] && \leftarrow \text{Reject} \\
 w_0[1] &= h[0]x[1] + h[1]x[0] + h[2]x[3] && \leftarrow \text{Reject} \\
 w_0[2] &= h[0]x[2] + h[1]x[1] + h[2]x[0] = y[2] && \leftarrow \text{Save} \\
 w_0[3] &= h[0]x[3] + h[1]x[2] + h[2]x[1] = y[3] && \leftarrow \text{Save} \\
 w_1[0] &= h[0]x[2] + h[1]x[5] + h[2]x[4] && \leftarrow \text{Reject} \\
 w_1[1] &= h[0]x[3] + h[1]x[2] + h[2]x[5] && \leftarrow \text{Reject} \\
 w_1[2] &= h[0]x[4] + h[1]x[3] + h[2]x[2] = y[4] && \leftarrow \text{Save} \\
 w_1[3] &= h[0]x[5] + h[1]x[4] + h[2]x[3] = y[5] && \leftarrow \text{Save} \\
 w_2[0] &= h[0]x[4] + h[1]x[5] + h[2]x[6] && \leftarrow \text{Reject} \\
 w_2[1] &= h[0]x[5] + h[1]x[4] + h[2]x[7] && \leftarrow \text{Reject} \\
 w_2[2] &= h[0]x[6] + h[1]x[5] + h[2]x[4] = y[6] && \leftarrow \text{Save} \\
 w_2[3] &= h[0]x[7] + h[1]x[6] + h[2]x[5] = y[7] && \leftarrow \text{Save}
 \end{aligned}$$

It should be noted that to determine $y[0]$ and $y[1]$ we need to form $x_{-1}[n]$, setting $x_{-1}[0] = x_{-1}[1] = 0$, $x_{-1}[2] = x[0]$, $x_{-1}[3] = x[1]$, and compute:

$$w_{-1}[n] = h[n] \circledast x_{-1}[n] \quad 0 \leq n \leq 3 \quad (4.249)$$

then reject $w_{-1}[0]$ and $w_{-1}[1]$, and save $w_{-1}[2] = y[0]$ and $w_{-1}[3] = y[1]$.

In general, let $h[n]$ be a length- N sequence. Let $x_m[n]$ denote the m^{th} section of an infinitely long sequence $x[n]$ of length N and defined by:

$$x_m[n] = x[n + m(N - m + 1)] \quad 0 \leq n \leq N - 1 \quad (4.250)$$

with $M < N$. Let $w_m[n] = h[n] \circledast x_m[n]$. Then, we reject the first $M - 1$ samples of $w_m[n]$ and “about” the remaining $M - M + 1$ samples of $w_m[n]$ to form $y_L[n]$, namely the linear convolution of $h[n]$ and $x[n]$. If $y_m[n]$ denotes the saved portion of $w_m[n]$, i.e.:

$$y_m[n] = \begin{cases} 0 & 0 \leq n \leq M - 2 \\ w_m[n] & M - 1 \leq n \leq N - 2 \end{cases} \quad (4.251)$$

then:

$$y_L[n + m(N - M + 1)] = y_m[n] \quad M - 1 \leq n \leq N - 1 \quad (4.252)$$

The approach is called **overlap-save method** since the input is segmented into overlapping sections and parts of the results of the circular convolutions are saved and abutted to determine the linear convolution result. The process is illustrated in Figure 4.16.

*General case using
circular
convolution*

*Overlap-save
method*

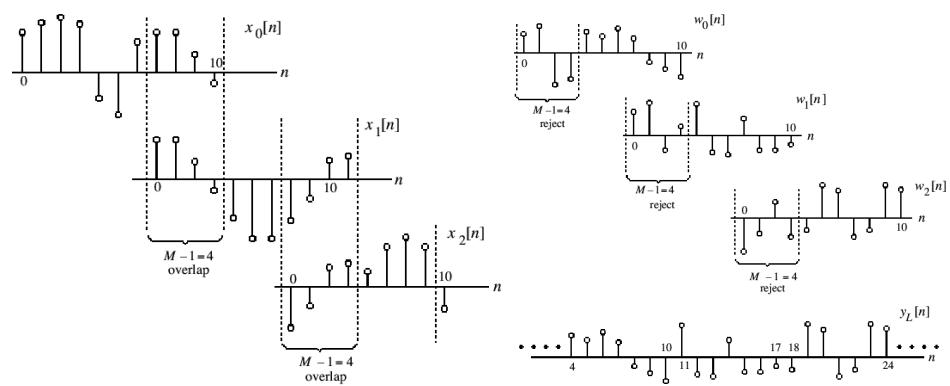


Figure 4.16: Illustration of the overlap-save method.

Chapter 5

The z -transform

We have seen that the DTFT provides a frequency-domain representation of discrete-time signals and LTI discrete-time systems. However, because of the convergence condition in many cases, the DTFT of a sequence may not exist. As a result, it is not possible to make use of such frequency-domain characterization in these cases. A possible solution and alternative is a generalization of the DTFT, which leads to the z -transform. The ladder may exist for many sequences for which the DTFT does not exist. Moreover, the use of z -transform techniques permits simple but powerful algebraic manipulations. Consequently, the z -transform has become an important tool in the analysis and design of digital filters

5.1 Definition of z -transform

Let us begin the discussion on this topic with the definition of the main tool.

Definition 11: z -transform

For a given sequence $g[n]$, its **z -transform** $G(z)$ is defined as:

$$G(z) = \sum_{n=-\infty}^{\infty} g[n]z^{-n} \quad (5.1)$$

where $z = \text{Re}[z] + j \text{Im}[z]$ is a complex variable.

If we let $z = re^{j\omega}$, then the z -transform reduces to:

$$G(re^{j\omega}) = \sum_{n=-\infty}^{\infty} g[n]r^{-n}e^{-j\omega n} \quad (5.2)$$

Eq. 5.2 can be interpreted as the DTFT of the modified sequence $\{g[n]r^{-n}\}$. For $r = 1$ (i.e., $|z| = 1$), the z -transform reduces to its DTFT, provided the ladder exists. Like the DTFT, there are conditions on the convergence of the infinite series like:

$$\sum_{n=-\infty}^{\infty} g[n]z^{-n} \quad (5.3)$$

For a given sequence, the set R of values of z for which its z -transform converges is called the **region of convergence (ROC)**.

From our earlier discussion on the uniform convergence of the DTFT, it follows that the series:

$$G(re^{j\omega}) = \sum_{n=-\infty}^{\infty} g[n]r^{-n}e^{-j\omega n} \quad (5.4)$$

Lecture 14.
Thursday 12th
November, 2020.

*Problems of DTFT
and z -transform as
alternative*

*Definition of the
 z -transform*

*Connections to the
DTFT*

*Region of
convergence (ROC)
of the z -transform*

converges if $\{g[n]r^{-n}\}$ is absolutely summable, i.e. if:

$$\sum_{n=-\infty}^{\infty} |g[n]r^{-n}| < \infty \quad (5.5)$$

In general, the ROC R of a z -transform of a sequence $g[n]$ is an annular region of the z -plane, namely:

$$R_{g^-} < |z| < R_{g^+} \quad (5.6)$$

where $0 \leq R_{g^-} < R_{g^+} < \infty$.

*Examples of
 z -transform
calculation*

Example 42: z -transform calculation

We determine the z -transform $X(z)$ of the causal sequence $x[n] = \alpha^n \mu[n]$ and its ROC. Now:

$$X(z) = \alpha^n \mu[n] z^{-n} = \sum_{n=0}^{\infty} \alpha^n z^{-n} \quad (5.7)$$

The above power series converges to:

$$X(z) = \frac{1}{1 - \alpha z^{-1}} \quad |\alpha z^{-1}| < 1 \quad (5.8)$$

Therefore, the ROC is the annular region $|z| > |\alpha|$.

Example 43: z -transform calculation

The z -transform $\mu(z)$ of the unit step sequence $\mu[n]$ can be obtained from:

$$X(z) = \frac{1}{1 - \alpha z^{-1}} \quad |\alpha z^{-1}| < 1 \quad (5.9)$$

By setting $\alpha = 1$:

$$\mu(z) = \frac{1}{1 - z^{-1}} \quad |z^{-1}| < 1 \quad (5.10)$$

Therefore, the ROC is the annular region $1 < |z| < \infty$. Note that the unit step sequence $\mu[n]$ is not absolutely summable, and hence its DTFT does not converge uniformly.

Example 44: z -transform calculation

Consider the anti-causal sequence:

$$y[n] = -\alpha^n \mu[-n - 1] \quad (5.11)$$

Its z -transform is given by:

$$\begin{aligned} Y(z) &= -\sum_{n=-\infty}^{-1} \alpha^n z^{-n} = -\sum_{m=1}^{\infty} \alpha^{-m} z^m \\ &= -\alpha^{-1} z \sum_{m=0}^{\infty} \alpha^{-m} z^m = -\frac{-\alpha^{-1} z}{1 - \alpha z^{-1}} \\ &= \frac{1}{1 - \alpha z^{-1}} \end{aligned} \quad (5.12)$$

for $|\alpha^{-1} z| < 1$. Therefore, the ROC is the annular region $|z| < |\alpha|$.

Note that the z -transforms of the two sequences $\alpha^n \mu[n]$ and $-\alpha^n \mu[-n-1]$ are identical even though the two parent sequences are different. The only way a unique sequence can be associated with a z -transform is by specifying its ROC.

Another important point is that the DTFT $G(e^{j\omega})$ of a sequence $g[n]$ converges uniformly if and only if the ROC of the z -transform $G(z)$ of $g[n]$ includes the unit circle. However, the existence of the DTFT does not always imply the existence of the z -transform.

*Connection
between uniform
convergence of the
DTFT and the
ROC*

Example 45: z -transform calculation

The finite energy sequence:

$$h_{LP}[n] = \frac{\sin(\omega_c n)}{\pi n} \quad -\infty < n < \infty \quad (5.13)$$

has a DTFT given by:

$$H_{LP}(e^{j\omega}) = \begin{cases} 1 & 0 \leq |\omega| \leq \omega_c \\ 0 & \omega_c < |\omega| \leq \pi \end{cases} \quad (5.14)$$

which converges in the mean-square sense. However, $h_{LP}[n]$ does not have a z -transform as it is not absolutely summable for any value of r .

Some commonly used z -transform pairs are listed in Figure 5.1.

*Commonly used
 z -transform pairs*

Sequence	z -Transform	ROC
$\delta[n]$	1	All values of z
$\mu[n]$	$\frac{1}{1 - z^{-1}}$	$ z > 1$
$\alpha^n \mu[n]$	$\frac{1}{1 - \alpha z^{-1}}$	$ z > \alpha $
$(r^n \cos \omega_o n) \mu[n]$	$\frac{1 - (r \cos \omega_o) z^{-1}}{1 - (2r \cos \omega_o) z^{-1} + r^2 z^{-2}}$	$ z > r$
$(r^n \sin \omega_o n) \mu[n]$	$\frac{(r \sin \omega_o) z^{-1}}{1 - (2r \cos \omega_o) z^{-1} + r^2 z^{-2}}$	$ z > r$

Figure 5.1: Common z -transform pairs.

Rational
 z -transforms

5.2 Rational z -transforms

In the case of the LTI discrete-time systems we are concerned with in this course, all the pertinent z -transforms are **rational functions** of z^{-1} , that is, they are fractions of two polynomials in z^{-1} :

$$G(z) = \frac{P(z)}{D(z)} = \frac{p_0 + p_1 z^{-1} + \cdots + p_{M-1} z^{-(M-1)} + p_M z^{-M}}{d_0 + d_1 z^{-1} + \cdots + d_{N-1} z^{-(N-1)} + d_N z^{-N}} \quad (5.15)$$

The degree of the numerator polynomial $P(z)$ is M and the degree of the denominator polynomial $D(z)$ is N . An alternate representation of a rational z -transform is as a ratio of two polynomials in z :

$$G(z) = z^{(N-M)} \frac{p_0 z^M + \cdots + p_{M-1} z + p_M}{d_0 z^N + \cdots + d_{N-1} z + d_N} \quad (5.16)$$

Rational
 z -transform in
factored form

Again, a rational z -transform can be alternately written in **factored form** as:

$$G(z) = \frac{\frac{p_0}{N} \prod_{\ell=1}^M (1 - \xi_\ell z^{-1})}{\prod_{\ell=1}^N (1 - \lambda_\ell z^{-1})} = z^{(N-M)} \frac{\frac{p_0}{N} \prod_{\ell=1}^M (z - \xi_\ell)}{\prod_{\ell=1}^N (z - \lambda_\ell)} \quad (5.17)$$

Zeros and poles of
rational
 z -transform

In particular, we have the following quantities of interest:

- $z = \xi_\ell$, roots of the numerator polynomial. These values of z are known as the **zeros** of $G(z)$;
- $z = \lambda_\ell$, roots of the denominator polynomial. These values of z are known as the **poles** of $G(z)$.

Example 46: ROC of a rational z -transform

The z -transform $H(z)$ of the sequence $h[n] = (-0.6)^n \mu[n]$ is given by:

$$H(z) = \frac{1}{1 + 0.6z^{-1}} \quad |z| > 0.6 \quad (5.18)$$

Here the ROC is just outside the circle going through the point $z = -0.6$.

Example 47: Zeros and poles

The z -transform:

$$\mu(z) = \frac{1}{1 - z^{-1}} \quad |z| > 1 \quad (5.19)$$

has a zero at $z = 0$ and a pole at $z = 1$.

Log-magnitude
plot of z -transform
for physical
interpretation

A physical interpretation of the concepts of poles and zeros can be given by plotting the **log-magnitude** $20 \log_{10} |G(z)|$ as showed in Figure 5.2 for:

$$G(z) = \frac{1 - 2.4z^{-1} + 2.88z^{-2}}{1 - 0.8z^{-1} + 0.64z^{-2}} \quad (5.20)$$

Observe that the **magnitude plot** exhibits very large peaks around the points $z = 0.4 \pm j0.6928$, which are the poles of $G(z)$. It also exhibits very narrow and deep wells around the location of the zeros at $z = 1.2 \pm j1.2$.

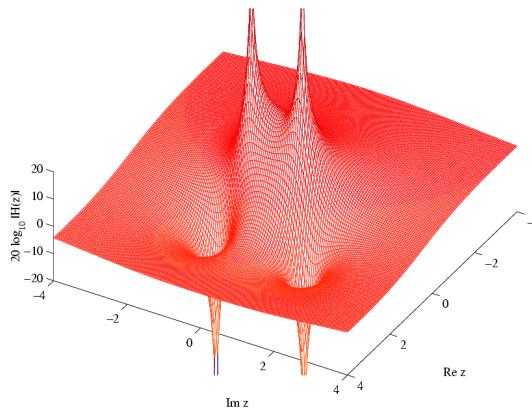


Figure 5.2: Log-magnitude plot for $G(z)$ in Eq. 5.20.

Now, we remark that the ROC of a z -transform is an important concept. Without its knowledge, there is no unique relationship between a sequence and its z -transform. Hence, **the z -transform must always be specified with its ROC**. Moreover, there is a relationship between the ROC of the z -transform of the impulse response of a causal LTI discrete-time system and its BIBO stability.

Another important distinction is that a sequence can be one of the following types:

- **finite-length;**
- **right-sided;**
- **left-sided;**
- **two-sided.**

In general, the ROC depends on the type of the sequence of interest and we show it in the following Subsections.

5.2.1 Finite-length sequence z -transform

We consider a **finite-length sequence** $g[n]$ defined for $-M \leq n \leq N$, where M and N are non-negative integers and $|g[n]| < \infty$. Its z -transform is given by:

$$G(z) = \sum_{n=-M}^N g[n]z^{-n} = \frac{\sum_{n=0}^{N+M} g[n-M]z^{N+M-n}}{z^N} \quad (5.21)$$

Note that $G(z)$ has M zeros and N poles. As can be seen from the expression for $G(z)$, the z -transform of a finite-length bounded sequence converges everywhere in the z -plane except possibly at $z = 0$ and/or at $z = \infty$.

5.2.2 Right-sided sequence z -transform

A **right-sided sequence** with nonzero sample values for $n \geq 0$ is sometimes called a **causal sequence**. So, we consider a causal sequence $u_1[n]$. Its z -transform is given by:

$$U_1(z) = \sum_{n=0}^{\infty} u_1[n]z^{-n} \quad (5.22)$$

It can be showed that $U_1(z)$ converges exterior to a circle with $|z| = R_1$, including the point $z = \infty$.

Importance of the ROC of z -transform

ROC dependency on the type of sequence

z -transform of finite-length sequence

z -transform of right-sided sequence

On the other hand, a right-sided sequence $u_2[n]$ with nonzero sample values only for $n \geq -M$ with M non-negative has a z -transform $U_2(z)$ with M poles at $z = \infty$. The ROC of $U_2(z)$ is exterior to a circle $|z| = R_2$, excluding the point $z = \infty$.

5.2.3 Left-sided sequence z -transform

z-transform of left-sided sequence

A **left-sided sequence** with nonzero sample values for $n \leq 0$ is sometimes called **anticausal sequence**. So, we consider an anticausal sequence $v_1[n]$. Its z -transform is given by:

$$V_1(z) = \sum_{n=-\infty}^0 v_1[n]z^{-n} \quad (5.23)$$

It can be showed that $V_1(z)$ converges interior to a circle $|z| = R_3$, including the point $z = 0$.

On the other hand, a left-sided sequence with nonzero sample values only for $n \leq N$ with N non-negative has a z -transform $V_2(z)$ with N poles at $z = 0$. The ROC of $V_2(z)$ is interior to a circle $|z| = R_4$, excluding the point $z = 0$.

5.2.4 Two-sided sequence z -transform

z-transform of two-sided sequence

The z -transform of a **two-sided sequence** $w[n]$ can be expressed as:

$$W(z) = \sum_{n=-\infty}^{\infty} w[n]z^{-n} = \sum_{n=0}^{\infty} w[n]z^{-n} + \sum_{n=-\infty}^{-1} w[n]z^{-n} \quad (5.24)$$

The first term on the RHS can be interpreted as the z -transform of a right-sided sequence and it thus converges exterior to the circle $|z| = R_5$. The second term of the RHS can be interpreted as the z -transform of a left-sided sequence and it thus converges interior to the circle $|z| = R_6$. If $R_5 < R_6$, there is an overlapping ROC given by $R_5 < |z| < R_6$. If $R_5 > R_6$, there is no overlap and the z -transform does not exist.

In particular, let us consider as example the two-sided sequence:

$$u[n] = \alpha^n \quad (5.25)$$

where α can be either real or complex. Its z -transform is given by:

$$U(z) = \sum_{n=-\infty}^{\infty} \alpha^n z^{-n} = \sum_{n=0}^{\infty} \alpha^n z^{-n} + \sum_{n=-\infty}^{-1} \alpha^n z^{-n} \quad (5.26)$$

The first term on the RHS converges for $|z| > |\alpha|$, whereas the second term converges for $|z| < |\alpha|$. There is no overlap between these two regions, hence the z -transform of $u[n] = \alpha^n$ does not exist.

The ROC of a rational z -transform cannot contain any pole (since it is infinite at a pole) and is bounded by the poles. To show the latter statement, we assume that the z -transform $X(z)$ has simple poles at $z = \alpha$ and $z = \beta$. We also assume that the corresponding sequence $x[n]$ is a right-sided sequence. Then, $x[n]$ has the form:

$$x[n] = (r_1 \alpha^n + r_2 \beta^n) \mu[n - N_0] \quad |\alpha| < |\beta| \quad (5.27)$$

where N_0 is a positive or negative integer. Now, the z -transform of the right-sided sequence $\gamma^n \mu[n - N_0]$ exists if:

$$\sum_{n=N_0}^{\infty} |\gamma^n z^{-n}| < \infty \quad (5.28)$$

for some z . The condition in Eq. 5.28 holds for $|z| > |\gamma|$, but not for $|z| \leq |\gamma|$. Therefore, the z -transform of Eq. 5.27 has a ROC defined by $|\beta| < |z| \leq \infty$. Likewise, the z -transform of a left-sided sequence:

$$x[n] = (r_1\alpha^n + r_2\beta^n)\mu[-n - N_0] \quad |\alpha| < |\beta| \quad (5.29)$$

has a ROC defined by $0 \leq |z| < |\alpha|$.

5.3 Inverse z -transform

Firstly, we recall that, for $z = re^{j\omega}$, the z -transform $G(z)$ given by:

$$G(z) = \sum_{n=-\infty}^{\infty} g[n]z^{-n} = \sum_{n=-\infty}^{\infty} g[n]r^{-n}e^{-j\omega n} \quad (5.30)$$

Inverse z -transform using DTFT analogy

is the DTFT of the modified sequence $g[n]r^{-n}$. Accordingly, the inverse DTFT is thus given by:

$$g[n]r^{-n} = \frac{1}{2\pi} \int_{-\pi}^{\pi} G(re^{j\omega})e^{j\omega n} d\omega \quad (5.31)$$

By making a change of variable $z = re^{j\omega}$, the previous equation can be converted into a **contour integral** given by:

$$g[n] = \frac{1}{2\pi j} \oint_{C'} G(z)z^{n-1} dz \quad (5.32)$$

Inverse z -transform through contour integral

where C' is a counterclockwise contour of integration defined by $|z| = r$. But the integral remains unchanged when it is replaced with any contour C encircling the point $z = 0$ in the ROC of $G(z)$. The contour integral can be evaluated using the **Cauchy's residue theorem** resulting in:

$$g[n] = \sum_C \text{Res}[G(z)z^{n-1}] \quad (5.33)$$

Cauchy's residue theorem

Eq. 5.33 needs to be evaluated at all the values of n , but this is not pursued here.

Now, a rational z -transform $G(z)$ with a causal inverse transform $g[n]$ has an ROC that is exterior to a circle. Here, it is more convenient to express $G(z)$ in a **partial-fraction expansion form** and then determine $g[n]$ by summing the inverse transforms of the individual simpler terms in the expansion. Therefore, a rational $G(z)$ can be expressed as:

$$G(z) = \frac{P(z)}{D(z)} = \frac{\sum_{i=0}^M p_i z^{-i}}{\sum_{i=0}^N d_i z^{-i}} \quad (5.34)$$

z-transform in partial fraction expansion form

If $M \geq N$, then $G(z)$ can be re-expressed as:

$$G(z) = \sum_{\ell=0}^{M-N} \eta_\ell z^{-\ell} + \frac{P_1(z)}{D(z)} \quad (5.35)$$

where the degree of $P_1(z)$ is less than N . The rational function $\frac{P_1(z)}{D(z)}$ is called a **proper fraction**. To develop the proper fraction part $\frac{P_1(z)}{D(z)}$ from $G(z)$, a **long division** of $P(z)$ by $D(z)$ should be carried out in a reverse order until the remainder polynomial $P_1(z)$ is of lower degree than that of the denominator $D(z)$.

Proper fractions and long division technique

Example 48: Inverse transform by partial-fraction expansion

Consider:

$$G(z) = \frac{2 + 0.8z^{-1} + 0.5z^{-2} + 0.3z^{-3}}{1 + 0.8z^{-1} + 0.2z^{-2}} \quad (5.36)$$

By long division in reverse order we arrive at:

$$G(z) = -3.5 + 1.5z^{-1} + \underbrace{\frac{5.5 + 2.1z^{-1}}{1 + 0.8z^{-1} + 0.2z^{-2}}}_{\text{Proper fraction}} \quad (5.37)$$

In most practical cases, the rational z -transform of interest $G(z)$ is a proper fraction with simple poles. Let the poles of $G(z)$ be at $z = \lambda_k$, with $1 \leq k \leq N$. A partial-fraction expansion of $G(z)$ is then of the form:

$$G(z) = \sum_{\ell=1}^N \left(\frac{\rho_\ell}{1 - \lambda_\ell z^{-1}} \right) \quad (5.38)$$

*Residues in
partial-fraction
expansion*

The constants ρ_ℓ in the partial-fraction expansion are called the **residues** and are given by:

$$\rho_\ell = [(1 - \lambda_\ell z^{-1})G(z)]_{z=\lambda_\ell} \quad (5.39)$$

Each term of the sum in partial-fraction expansion has a ROC given by $|z| > |\lambda_\ell|$ and thus has an inverse transform of the form $\rho_\ell(\lambda_\ell)^n \mu[n]$. Therefore, the inverse transform $g[n]$ of $G(z)$ is given by:

$$g[n] = \sum_{\ell=1}^N \rho_\ell (\lambda_\ell)^n \mu[n] \quad (5.40)$$

Note that the approach in Eq. 5.40 with a slight modification can also be used to determine the inverse of a rational z -transform of a non-causal sequence.

Example 49: Inverse transfrom of a causal sequence

Let the z -transform $H(z)$ of a causal sequence $h[n]$ be given by:

$$H(z) = \frac{z(z+2)}{(z-0.2)(z+0.6)} = \frac{1+2z^{-1}}{(1-0.2z^{-1})(1+0.6z^{-1})} \quad (5.41)$$

A partial-fraction expansion of $H(z)$ is then of the form:

$$H(z) = \frac{\rho_1}{1-0.2z^{-1}} + \frac{\rho_2}{1+0.6z^{-1}} \quad (5.42)$$

Now:

$$\rho_1 = [(1-0.2z^{-1})H(z)]_{z=0.2} = \left[\frac{1+2z^{-1}}{1+0.6z^{-1}} \right]_{z=0.2} = 2.75 \quad (5.43)$$

$$\rho_2 = [(1+0.6z^{-1})H(z)]_{z=-0.6} = \left[\frac{1+2z^{-1}}{1-0.2z^{-1}} \right]_{z=-0.6} = -1.75 \quad (5.44)$$

Hence:

$$H(z) = \frac{2.75}{1-0.2z^{-1}} - \frac{1.75}{1+0.6z^{-1}} \quad (5.45)$$

The inverse transform of Eq. 5.45 is therefore given by:

$$h[n] = 2.75(0.2)^n \mu[n] - 1.75(-0.6)^n \mu[n] \quad (5.46)$$

In case $G(z)$ has multiple poles, the partial-fraction expansion is of slightly different form. Let the pole at $z = v$ be of multiplicity L and the remaining $N - L$ poles be simple and at $z = \lambda_\ell$, for $1 \leq \ell \leq N - L$. Then, the partial-fraction expansion of $G(z)$ is of the form:

$$G(z) = \sum_{\ell=0}^{M-N} \eta_\ell z^{-\ell} + \sum_{\ell=1}^{N-L} \frac{\rho_\ell}{1 - \lambda_\ell z^{-1}} + \sum_{i=1}^L \frac{\gamma_i}{(1 - vz^{-1})^i} \quad (5.47)$$

where the constants γ_i are computed using:

$$\gamma_i = \frac{1}{(L-i)!(-v)^{L-i}} \frac{d^{L-i}}{dz^{L-i}} [(1-vz^{-1})G(z)]_{z=v} \quad 1 \leq i \leq L \quad (5.48)$$

The residues ρ_ℓ are calculated as before.

Partial-fraction expansion with poles of higher multiplicity

5.4 z -transform properties

A list of properties of the z -transform is showed in Figure 5.3.

Properties of z -transform

Property	Sequence	z -Transform	ROC
	$g[n]$	$G(z)$	\mathcal{R}_g
	$h[n]$	$H(z)$	\mathcal{R}_h
Conjugation	$g^*[n]$	$G^*(z^*)$	\mathcal{R}_g
Time-reversal	$g[-n]$	$G(1/z)$	$1/\mathcal{R}_g$
Linearity	$\alpha g[n] + \beta h[n]$	$\alpha G(z) + \beta H(z)$	Includes $\mathcal{R}_g \cap \mathcal{R}_h$
Time-shifting	$g[n - n_0]$	$z^{-n_0} G(z)$	\mathcal{R}_g , except possibly the point $z = 0$ or ∞
Multiplication by an exponential sequence	$\alpha^n g[n]$	$G(z/\alpha)$	$ \alpha \mathcal{R}_g$
Differentiation of $G(z)$	$ng[n]$	$-z \frac{dG(z)}{dz}$	\mathcal{R}_g , except possibly the point $z = 0$ or ∞
Convolution	$g[n] \circledast h[n]$	$G(z)H(z)$	Includes $\mathcal{R}_g \cap \mathcal{R}_h$
Modulation	$g[n]h[n]$	$\frac{1}{2\pi j} \oint_C G(v)H(z/v)v^{-1} dv$	Includes $\mathcal{R}_g \mathcal{R}_h$
Parseval's relation		$\sum_{n=-\infty}^{\infty} g[n]h^*[n] = \frac{1}{2\pi j} \oint_C G(v)H^*(1/v^*)v^{-1} dv$	

Note: If \mathcal{R}_g denotes the region $R_{g-} < |z| < R_{g+}$ and \mathcal{R}_h denotes the region $R_{h-} < |z| < R_{h+}$, then $1/\mathcal{R}_g$ denotes the region $1/R_{g+} < |z| < 1/R_{g-}$ and $\mathcal{R}_g \mathcal{R}_h$ denotes the region $R_{g-} R_{h-} < |z| < R_{g+} R_{h+}$.

Figure 5.3: Properties of z -transform.

Examples for z -transform properties

Now, we present some examples with cases where the properties can be usefully applied.

Example 50: z -transform properties

Consider the two-sided sequence:

$$v[n] = \alpha^n \mu[n] - \beta^n \mu[-n-1] \quad (5.49)$$

Let $x[n] = \alpha^n \mu[n]$ and $y = -\beta^n \mu[-n-1]$ with $X(z)$ and $Y(z)$ denoting, respectively, their z -transforms. Now:

$$X(z) = \frac{1}{1 - \alpha z^{-1}} \quad |z| > |\alpha| \quad (5.50)$$

$$Y(z) = \frac{1}{1 - \beta z^{-1}} \quad |z| < |\beta| \quad (5.51)$$

Using the linearity property we arrive at:

$$V(z) = X(z) + Y(z) = \frac{1}{1 - \alpha z^{-1}} + \frac{1}{1 - \beta z^{-1}} \quad (5.52)$$

The ROC of $V(z)$ is given by the overlap regions of $|z| > |\alpha|$ and $|z| < |\beta|$. We have that:

- if $|\alpha| < |\beta|$, then there is an overlap and the ROC is an annular region $|\alpha| < |z| < |\beta|$;
- if $|\alpha| > |\beta|$, then there is no overlap and $V(z)$ does not exist.

Example 51: z -transform properties

We determine the z -transform and its ROC of the causal sequence:

$$x[n] = r^n (\cos(\omega_0 n)) \mu[n] \quad (5.53)$$

We can express $x[n] = v[n] + v^*[n]$, where:

$$v[n] = \frac{1}{2} r^n e^{j\omega_0 n} \mu[n] = \frac{1}{2} \alpha^n \mu[n] \quad (5.54)$$

The z -transform of $v[n]$ is given by:

$$V(z) = \frac{1}{2} \frac{1}{1 - \alpha z^{-1}} = \frac{1}{2} \frac{1}{1 - re^{j\omega_0 z^{-1}}} \quad |z| > |\alpha| = r \quad (5.55)$$

Using the conjugation property, we obtain the z -transform of $v^*[n]$ as:

$$V^*(z^*) = \frac{1}{2} \frac{1}{1 - \alpha^* z^{-1}} = \frac{1}{2} \frac{1}{1 - re^{-j\omega_0 z^{-1}}} \quad |z| > |\alpha| \quad (5.56)$$

Finally, using the linearity property we get:

$$X(z) = V(z) + V^*(z^*) = \frac{1}{2} \left(\frac{1}{1 - re^{j\omega_0 z^{-1}}} + \frac{1}{1 - re^{-j\omega_0 z^{-1}}} \right) \quad (5.57)$$

or:

$$X(z) = \frac{1 - (r \cos \omega_0) z^{-1}}{1 - (2r \cos \omega_0) z^{-1} + r^2 z^{-2}} \quad |z| > r \quad (5.58)$$

Example 52: z -transform properties

We determine the z -transform $Y(z)$ and the ROC of the sequence:

$$y[n] = (n + 1) \alpha^n \mu[n] \quad (5.59)$$

We can write $y[n] = nx[n] + x[n]$ where:

$$x[n] = \alpha^n \mu[n] \quad (5.60)$$

Now, the z -transform $X(z)$ of $x[n] = \alpha^n \mu[n]$ is given by:

$$X(z) = \frac{1}{1 - \alpha z^{-1}} \quad |z| > |\alpha| \quad (5.61)$$

Using the differentiation property, we arrive at the z -transform of $nx[n]$ as:

$$-z \frac{dX(z)}{dz} = \frac{\alpha z^{-1}}{1 - \alpha z^{-1}} \quad |z| > |\alpha| \quad (5.62)$$

Using the linearity property we finally obtain:

$$Y(z) = \frac{1}{1 - \alpha z^{-1}} + \frac{\alpha z^{-1}}{(1 - \alpha z^{-1})^2} = \frac{1}{(1 - \alpha z^{-1})^2} \quad |z| > |\alpha| \quad (5.63)$$

Lecture 17.
Tuesday 24th
November, 2020.

5.5 Stability condition

We move to the discussion about stability, which represents a very important topic. We have already seen that a causal LTI digital filter is BIBO stable if and only if its impulse response $h[n]$ is absolutely summable, i.e.:

$$S = \sum_{n=-\infty}^{\infty} |h[n]| < \infty \quad (5.64)$$

We now develop a stability condition in terms of the pole locations of the transfer function $H(z)$.

The ROC of the z -transform $H(z)$ of the impulse response sequence $h[n]$ is defined by values of $|z| = r$ for which $h[n]r^{-n}$ is absolutely summable. Thus, if the ROC includes the unit circle $|z| = 1$, then the digital filter is stable, and viceversa. In addition, for a stable and causal digital filter for which $h[n]$ is a right-sided sequence, the ROC will include the unit circle and entire z -plane including the point $z = \infty$. Note that a FIR digital filter with bounded impulse response is always stable. On the other hand, a IIR filter may be unstable if not designed properly. In addition, an originally stable IIR filter characterized by infinite precision coefficients may become unstable when coefficients get quantized due to implementation.

*Stability condition
in terms of the
pole locations*

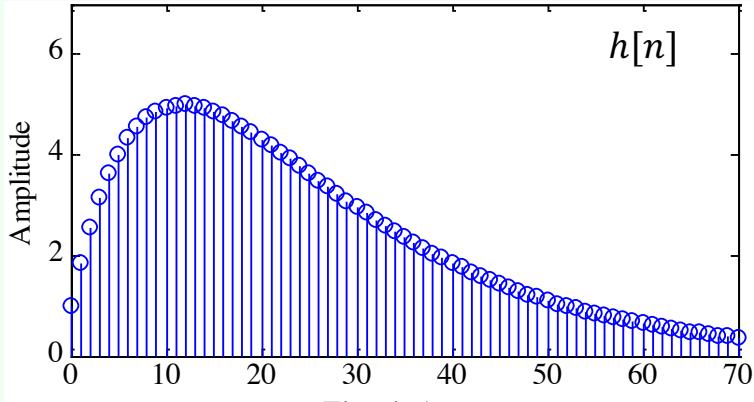
*Problem of
coefficients
quantization*

Example 53: Stability condition

We consider the causal IIR transfer function:

$$H(z) = \frac{1}{1 - 1.845z^{-1} + 0.850586z^{-2}} \quad (5.65)$$

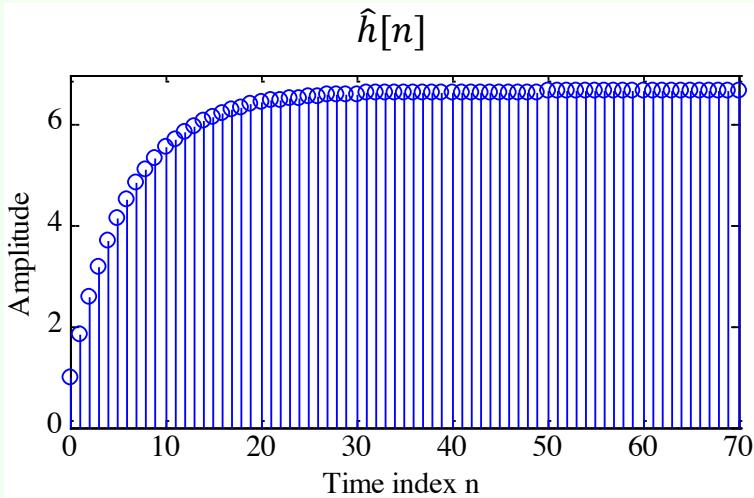
The plot of the impulse response coefficients is showed below. As can be seen from the plot, the impulse response coefficient $h[n]$ decays rapidly to zero value as n increases.



The absolute summability condition of $h[n]$ is satisfied. Hence, $H(z)$ is a stable transfer function. Now, consider the case when the transfer function coefficients are rounded to values with 2 digits after the decimal point:

$$\hat{H}(z) = \frac{1}{1 - 1.85z^{-1} + 0.85z^{-2}} \quad (5.66)$$

A plot of the impulse response of $\hat{h}[n]$ is showed below.



In this case, the impulse coefficient $\hat{h}[n]$ increases rapidly to a constant value as n increases. Hence, the absolute summability condition of $\hat{h}[n]$ is violated. Thus, $\hat{H}[z]$ is an unstable transfer function.

Stability testing of an IIR transfer function

The stability testing of an IIR transfer function is therefore an important problem. In most cases it is difficult to compute the infinite sum in Eq. 5.64. For a causal IIR transfer function, the sum S can be computed approximately as:

$$S_k = \sum_{n=0}^{k-1} |h[n]| \quad (5.67)$$

The partial sum is computed for increasing values of k until the difference between a series of consecutive values of S_k is smaller than some arbitrarily chosen small number, which is typically 10^{-6} . For a transfer function of very high order this approach may not be satisfactory. An alternate, easy-to-test, stability condition is developed next. Let us consider the causal IIR digital filter with a rational transfer function $H(z)$

given by:

$$H(z) = \frac{\sum_{k=0}^M p_k z^{-k}}{\sum_{k=0}^N d_k z^{-k}} \quad (5.68)$$

Its impulse response $\{h[n]\}$ is a right-sided sequence. The ROC of $H(z)$ is exterior to a circle going through the pole furthest from $z = 0$. But stability requires that $\{h[n]\}$ is absolutely summable. This in turn implies that the DTFT $H(e^{j\omega})$ of $\{h[n]\}$ exists. Now, if the ROC of the z -transform $H(z)$ includes the unit circle, then:

$$H(e^{j\omega}) = [H(z)]_{z=e^{j\omega}} \quad (5.69)$$

In conclusion, all the poles of a causal stable transfer function $H(z)$ must be strictly inside the unit circle, as showed in Figure 5.4.

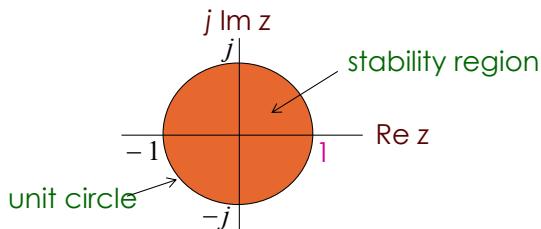


Figure 5.4: Stability condition (shaded area).

Example 54: Stability condition

The factored form of the previous example transfer function is:

$$H(z) = \frac{1}{1 - 1.845z^{-1} + 0.850586z^{-2}} = \frac{1}{(1 - 0.902z^{-1})(1 - 0.943z^{-1})} \quad (5.70)$$

which has a real pole at $z = 0.902$ and a real pole at $z = 0.943$. Since both poles are inside the unit circle, $H(z)$ is BIBO stable.

On the other hand, the factored form of $\hat{H}(z)$ is:

$$\hat{H}(z) = \frac{1}{1 - 1.85z^{-1} + 0.85z^{-2}} = \frac{1}{(1 - z^{-1})(1 - 0.85z^{-1})} \quad (5.71)$$

which has a real pole on the unit circle at $z = 1$ and the other pole inside the unit circle. Since not both the poles are inside the unit circle, $\hat{H}(z)$ is unstable.

Chapter 6

Filter design

In the previous Chapters we have seen an overview on both the nature of discrete-time signals and on the tools that can be used to analyze their properties. Now, we can focus on the study of the fundamental building blocks of any Digital Signal Processing system and on the design of more complex composite systems. We will go on by studying several examples to understand how the theory can be applied in practice.

6.1 Preliminary concepts and definitions

Before starting with the discussion of filter design, we have to introduce and recap some other useful concepts for the following discussion.

6.1.1 Phase delay

If the input $x[n]$ to an LTI system $H(e^{j\omega})$ is a sinusoidal signal of frequency ω_0 , i.e.:

$$x[n] = A \cos(\omega_0 n + \varphi), \quad -\infty < n < \infty \quad (6.1)$$

then the output $y[n]$ is also a sinusoidal signal of the same frequency ω_0 , but lagging in phase by $\theta(\omega_0)$ radians:

$$y[n] = A |H(e^{j\omega_0})| \cos(\omega_0 n + \theta(\omega_0) + \varphi), \quad -\infty < n < \infty \quad (6.2)$$

We can rewrite the output expression as:

$$y[n] = A |H(e^{j\omega_0})| \cos(\omega_0(n - \tau_p(\omega_0) + \varphi)) \quad (6.3)$$

where:

Phase delay

$$\tau_p(\omega_0) = -\frac{\theta(\omega_0)}{\omega_0} \quad (6.4)$$

is called the **phase delay**. The minus sign in front indicates a phase lag. Thus, the output $y[n]$ is a **time-delayed version** of the input $x[n]$. In general, $y[n]$ will not be a delayed replica of $x[n]$ unless the phase delay $\tau_p(\omega_0)$ is an integer.

6.1.2 Group delay

When the input is composed of many sinusoidal components with different frequencies that are not harmonically related, each component will go through different phase delays. In this case, the signal delay is determined using the **group delay**, defined by:

Group delay

$$\tau_g(\omega) = -\frac{d\theta(\omega)}{d\omega} \quad (6.5)$$

In defining the group delay, it is assumed that the phase function is unwrapped so that its derivatives exist.

Example 55: Phase and group delay

The phase function of the FIR filter:

$$y[n] = \alpha x[n] + \beta x[n-1] + \gamma x[n-2] \quad (6.6)$$

is:

$$\theta(\omega) = -\omega \quad (6.7)$$

Hence, its group delay is given by $\tau_g(\omega) = 1$.

6.1.3 Types of transfer function

Classification on the length of the impulse response

Other classifications

Passband and stopband

Cutoff frequency

The time-domain classification of an LTI digital transfer function sequence is based on the length of its impulse response. We can have:

- **Finite Impulse Response (FIR)** transfer functions;
- **Infinite Impulse Response (IIR)** transfer functions.

In the case of digital transfer functions with frequency-selective frequency responses, there are two other types of classifications:

- a classification based on the **shape of the magnitude function** $|H(e^{j\omega})|$;
- a classification based on the **form of the phase function** $\theta(\omega)$.

One common classification is based on an ideal magnitude response. A digital filter designed to pass signal components of certain frequencies without distortion should have a frequency response equal to one at these frequencies, and should have a frequency response equal to zero at all other frequencies.

We focus now to the case of ideal filters. In general, the range of frequencies where the frequency response takes the value of one is called the **passband**. The range of frequencies where the frequency response takes the value of zero is called the **stopband**. In this context, the frequency responses of the four popular types of ideal digital filters with real impulse response coefficients are showed in Figure 6.1.

In particular, the passband and stopband of those filters are listed in Table 6.1. The frequencies ω_c , ω_{c1} and ω_{c2} are called the **cutoff frequencies**. An ideal filter has a magnitude response equal to one in the passband and zero in the stopband, and has a zero phase everywhere.

Type	Passband	Stopband
Lowpass	$0 \leq \omega \leq \omega_c$	$\omega_c < \omega \leq \pi$
Highpass	$\omega_c \leq \omega \leq \pi$	$0 \leq \omega < \omega_c$
Bandpass	$\omega_{c1} \leq \omega \leq \omega_{c2}$	$0 \leq \omega < \omega_{c1}$ and $\omega_{c2} < \omega \leq \pi$
Stopband	$0 \leq \omega \leq \omega_{c1}$ and $\omega_{c2} \leq \omega \leq \pi$	$\omega_{c1} < \omega < \omega_{c2}$

Table 6.1: Passband and stopband of the four popular types of ideal digital filters.

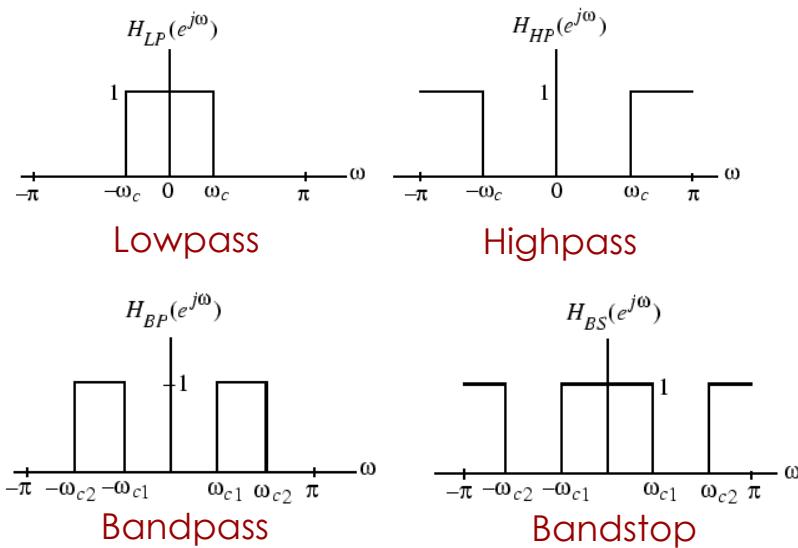


Figure 6.1: Frequency responses of the four popular types of ideal digital filters with real impulse response coefficients.

Let us pickle an example. Earlier in the course we derived the inverse DTFT of the frequency response $H_{LP}(e^{j\omega})$ of the ideal lowpass filter:

$$h_{LP}[n] = \frac{\sin(\omega_c n)}{\pi n}, \quad -\infty < n < \infty \quad (6.8)$$

We have also showed that the impulse response in Eq. 6.8 is not absolutely summable, and hence, the corresponding transfer function is not BIBO stable. Also, $h_{LP}[n]$ is not causal and is of doubly infinite length. The remaining three ideal filters are also characterized by doubly infinite, non-causal impulse responses and are not absolutely summable. Thus, the ideal filters with the ideal “brick wall” frequency responses cannot be realized with finite dimensional LTI filter.

To develop stable and realizable transfer functions, the ideal frequency response specifications are relaxed by including a **transition band** between the passband and the stopband. This permits the magnitude response to decay slowly from its maximum value in the passband to the zero value in the stopband. Moreover, the magnitude response is allowed to vary by a small amount both in the passband and the stopband. Typical magnitude response specifications of a lowpass filter are showed in Figure 6.2.

Transition band
between passband
and stopband

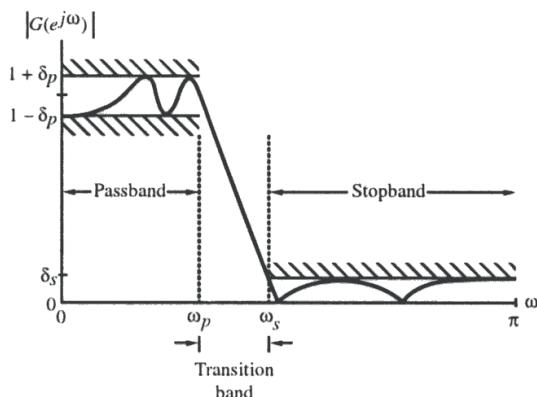


Figure 6.2: Typical magnitude response specifications of a lowpass filter.

Bounded real transfer function

6.1.4 Bounded real transfer function

A causal stable real-coefficient transfer function $H(z)$ is defined as a **bounded real (BR) transfer function** if:

$$|H(e^{j\omega})| \leq 1 \quad \forall \omega \quad (6.9)$$

Now, let $x[n]$ and $y[n]$ denote, respectively, the input and output of a digital filter characterized by a BR transfer function $H(z)$ with $X(e^{j\omega})$ and $Y(e^{j\omega})$ denoting their DTFTs. Then, the condition in Eq. 6.9 implies that:

$$|Y(e^{j\omega})|^2 \leq |X(e^{j\omega})|^2 \quad (6.10)$$

Integrating Eq. 6.10 from $-\pi$ to π and applying Parseval's relation, we get:

$$\sum_{n=-\infty}^{\infty} y[n]^2 \leq \sum_{n=-\infty}^{\infty} x[n]^2 \quad (6.11)$$

Thus, for all finite-energy inputs, the output energy is less than or equal to the input energy implying that a digital filter characterized by a BR transfer function can be viewed as a **passive structure**. If $|H(e^{j\omega})| = 1$, then the output energy is equal to the input energy, and such a digital filter is therefore a **lossless system**.

A causal stable real-coefficient transfer function $H(z)$ with $|H(e^{j\omega})| = 1$ is thus called a **lossless bounded real (LBR) transfer function**. The BR and LBR transfer functions are the keys to the realization of digital filters with low coefficient sensitivity.

Lossless bounded real transfer function

Example 56: Bounded real transfer function

Consider the causal stable IIR transfer function:

$$H(z) = \frac{k}{1 - \alpha z^{-1}}, \quad 0 < |\alpha| < 1 \quad (6.12)$$

where k is a real constant. Its square-magnitude function is given by:

$$|H(e^{j\omega})|^2 = [H(z)H(z^{-1})]_{z=e^{j\omega}} = \frac{k^2}{(1 + \alpha^2) - 2\alpha \cos \omega} \quad (6.13)$$

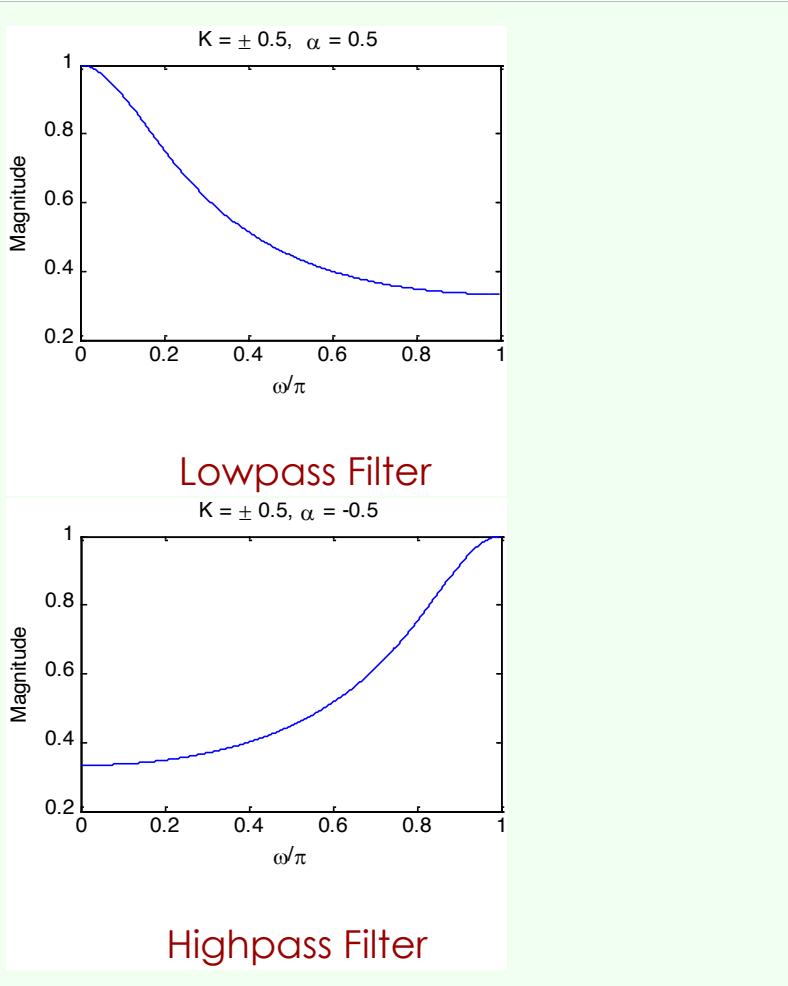
The maximum value of $|H(e^{j\omega})|^2$ is obtained when $2\alpha \cos \omega$ in the denominator is a maximum and the minimum value is obtained when $2\alpha \cos \omega$ is a minimum. For $\alpha > 0$, the maximum value of $2\alpha \cos \omega$ is equal to 2α at $\omega = 0$, and the minimum value is -2α at $\omega = \pi$.

Thus, for $\alpha > 0$, the maximum value of $|H(e^{j\omega})|^2$ is equal to $\frac{k^2}{(1-\alpha)^2}$ at $\omega = 0$ and the minimum value is equal to $\frac{k^2}{(1+\alpha)^2}$ at $\omega = \pi$.

On the other hand, for $\alpha < 0$, the maximum value of $2\alpha \cos \omega$ is equal to -2α at $\omega = \pi$ and the minimum value is equal to 2α at $\omega = 0$. Here, the maximum value of $|H(e^{j\omega})|^2$ is equal to $\frac{k^2}{(1-\alpha)^2}$ at $\omega = \pi$, and the minimum value is equal to $\frac{k^2}{(1+\alpha)^2}$ at $\omega = 0$. Hence, the maximum value can be made equal to 1 by choosing $k = \pm(1 - \alpha)$, in which case the minimum value becomes $\frac{(1-\alpha)^2}{(1+\alpha)^2}$. Hence:

$$H(z) = \frac{k}{1 - \alpha z^{-1}}, \quad 0 < |\alpha| < 1 \quad (6.14)$$

is a BR function for $k = \pm(1 - \alpha)$. Plots of the magnitude function for $\alpha = \pm 0.5$ with values of k chosen to make $H(z)$ a BR function are showed below.



6.1.5 Allpass transfer function

Definition 12: Allpass transfer function

Allpass transfer function definition

An IIR transfer function $A(z)$ with unity magnitude response for all frequencies, i.e.:

$$|A(e^{j\omega})|^2 = 1 \quad \forall \omega \quad (6.15)$$

is called an **allpass transfer function**.

An M^{th} order causal real-coefficient allpass transfer function is of the form:

$$A_M(z) = \pm \frac{d_M + d_{M-1}z^{-1} + \dots + d_1z^{-M+1} + z^{-M}}{1 + d_1z^{-1} + \dots + d_{M-1}z^{-M+1} + d_Mz^{-M}} \quad (6.16)$$

If we denote the denominator polynomials of $A_M(z)$ as $D_M(z)$:

$$D_M(z) = 1 + d_1z^{-1} + \dots + d_{M-1}z^{-M+1} + d_Mz^{-M} \quad (6.17)$$

then it follows that $A_M(z)$ can be written as:

$$A_M(z) = \pm \frac{z^{-M} D_M(z^{-1})}{D_M(z)} \quad (6.18)$$

Note that if $z = re^{j\omega}$ is a pole of a real coefficient allpass transfer function, then it has a zero at $z = \frac{1}{r}e^{-j\varphi}$.

Mirror-image polynomial and symmetry

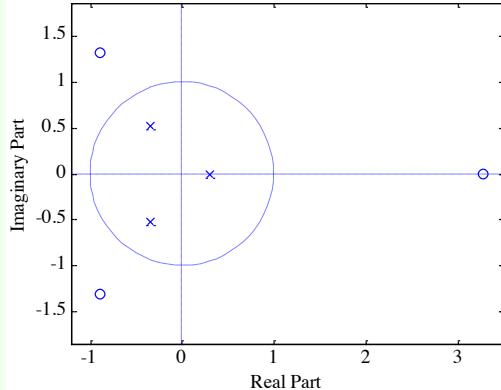
The numerator of a real-coefficient allpass transfer function is said to be the **mirror-image polynomial** of the denominator, and vice versa. We shall use the notation $\tilde{D}_M(z)$ to denote the mirror-image polynomial of a degree- M polynomial $D_M(z)$, i.e.:

$$\tilde{D}_M(z) = z^{-M} D_M(z^{-1}) \quad (6.19)$$

The expression in Eq. 6.18 implies that the poles and zeros of a real-coefficient allpass function exhibit **mirror-image symmetry** in the z -plane.

Example 57: Allpass transfer function

$$A_3(z) \frac{-0.2 + 0.81z^{-1} + 0.4z^{-2} + z^{-3}}{1 + 0.4z^{-1} + 0.18z^{-2} - 0.2z^{-3}} \quad (6.20)$$



To show that $|A_M(e^{j\omega})|^2 = 1$, we observe that:

$$A_M(z^{-1}) = \pm \frac{z^M D_M(z)}{D_M(z^{-1})} \quad (6.21)$$

Therefore:

$$A_M(z) A_M(z^{-1}) = \frac{z^{-M} D_M(z^{-1})}{D_M(z)} \frac{z^M D_M(z)}{D_M(z^{-1})} \quad (6.22)$$

Hence:

$$|A_M(e^{j\omega})|^2 = [A_M(z) A_M(z^{-1})]_{z=e^{j\omega}} = 1 \quad (6.23)$$

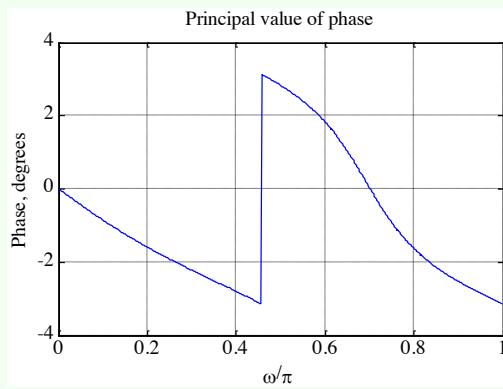
Now, the poles of a causal stable transfer function must lie inside the unit circle in the z -plane. Hence, all the zeros of a causal stable allpass transfer function must lie outside the unit circle in a mirror-image symmetry with its poles situated inside the unit circle.

Example 58: Allpass transfer function

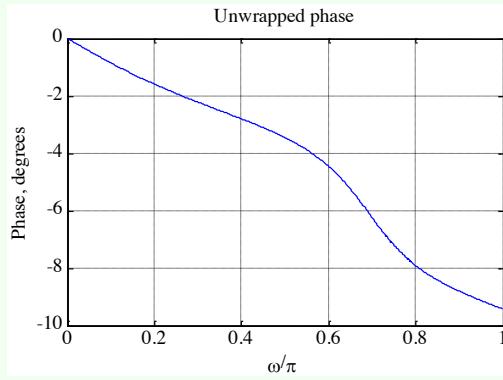
Consider the 3rd order allpass function:

$$A_3(z) = \frac{-0.2 + 0.18z^{-1} + 0.4z^{-2} + z^{-3}}{1 + 0.4z^{-1} + 0.18z^{-2} - 0.2z^{-3}} \quad (6.24)$$

The principal value of phase is showed in the plot below. Note the discontinuity by the amount of 2π in the phase $\theta(\omega)$.



If we unwrap the phase by removing the discontinuity, we arrive at the unwrapped phase function $\theta_c(\omega)$ indicated in the plot below. Note that the unwrapped phase function is a continuous function of ω .



We list now the properties of allpass transfer function:

- a causal stable real-coefficient allpass transfer function is a lossless bounded real (LBR) function or, equivalently, a causal stable allpass filter is a lossless structure;
- the magnitude function of a stable allpass function $A(z)$ satisfies:

$$|A(z)| \begin{cases} < 1 & |z| > 1 \\ = 1 & |z| = 1 \\ > 1 & |z| < 1 \end{cases} \quad (6.25)$$

- let $\tau(\omega)$ denote the group delay function of an allpass filter $A(z)$:

$$\tau(\omega) = -\frac{d}{d\omega}[\theta_c(\omega)] \quad (6.26)$$

The unwrapped phase function $\theta_c(\omega)$ of a stable allpass function is a monotonically decreasing function of ω so that $\tau(\omega)$ is everywhere positive in the range $0 < \omega < \pi$. The group delay of an M^{th} order stable real-coefficient allpass transfer function satisfies:

$$\int_0^\pi \tau(\omega) d\omega = M\pi \quad (6.27)$$

A simple but often used application of an allpass filter is as a **delay equalizer**. Let $G(z)$ be the transfer function of a digital filter designed to meet a prescribed

*Properties of
allpass transfer
function*

*Magnitude
function*

Group delay

*Delay equalizer
application*

magnitude response. The non-linear phase response of $G(z)$ can be corrected by cascading it with an allpass filter $A(z)$ so that the overall cascade has a constant group delay in the band of interest. Since $|A_M(e^{j\omega})| = 1$, we have:

$$|G(e^{j\omega})A(e^{j\omega})| = |G(e^{j\omega})| \quad (6.28)$$

The overall group delay is given by the sum of the group delays of $G(z)$ and $A(z)$.

6.1.6 Phase characteristics

Classification on phase characteristics

Zero-phase filtering

A second classification of a transfer function is with respect to its **phase characteristics**. In many applications, it is necessary that the digital filter designed does not distort the phase of the input signal components for frequencies in the passband. One way to avoid any phase distortion is to make the frequency response of the filter real and non-negative, i.e., to design the filter with a **zero-phase characteristic**. However, it is not possible to design a causal digital filter with a zero phase. For non-real-time processing of real-valued input signals of finite length, **zero-phase filtering** can be very simply implemented by relaxing the causality requirement. One zero-phase filtering scheme is sketched in Figure 6.3. It is easy to verify the ladder in the frequency domain.

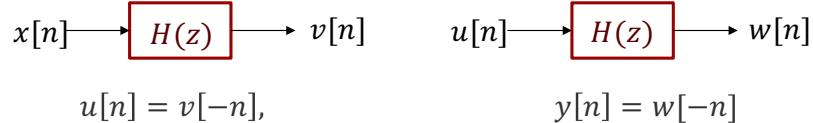


Figure 6.3: A possible zero-phase filtering scheme.

Let $X(e^{j\omega})$, $V(e^{j\omega})$, $U(e^{j\omega})$, $W(e^{j\omega})$ and $Y(e^{j\omega})$ denote the DTFTs of $x[n]$, $v[n]$, $u[n]$, $w[n]$ and $y[n]$, respectively. Making use of the symmetry relations, we arrive at the relations between various DTFTs:

$$V(e^{j\omega}) = H(e^{j\omega})X(e^{j\omega}) \quad (6.29)$$

$$W(e^{j\omega}) = H(e^{j\omega})U(e^{j\omega}) \quad (6.30)$$

$$U(e^{j\omega}) = V^*(e^{j\omega}) \quad (6.31)$$

$$Y(e^{j\omega}) = W^*(e^{j\omega}) \quad (6.32)$$

Combining the above equations we get:

$$\begin{aligned} Y(e^{j\omega}) &= W^*(e^{j\omega}) \\ &= H^*(e^{j\omega})U^*(e^{j\omega}) \\ &= H^*(e^{j\omega})V(e^{j\omega}) \\ &= H^*(e^{j\omega})H(e^{j\omega})X(e^{j\omega}) \\ &= |H(e^{j\omega})|^2 X(e^{j\omega}) \end{aligned} \quad (6.33)$$

The most general type of a filter with a linear phase has a frequency response given by:

$$H(e^{j\omega}) = e^{-j\omega D} \quad (6.34)$$

which has a linear phase from $\omega = 0$ to $\omega = 2\pi$. Note also that:

$$|H(e^{j\omega})| = 1 \quad (6.35)$$

and:

$$\tau(\omega) = -\frac{d}{d\omega}[\theta_c(\omega)] = D \quad (6.36)$$

The output $y[n]$ of this filter to an input $x[n] = Ae^{j\omega n}$ is then given by:

$$y[n] = Ae^{-j\omega D}e^{j\omega n} = Ae^{j\omega(n-D)} \quad (6.37)$$

If $x_a(t)$ and $y_a(t)$ represent the continuous-time signals whose sampled versions, sampled at $t = nT$, are $x[n]$ and $y[n]$ given above, then the delay between $x_a(t)$ and $y_a(t)$ is precisely the group delay of amount D . If D is an integer, then $y[n]$ is identical to $x[n]$, but delayed by D samples. If D is not an integer, $y[n]$, being delayed by a fractional part, is not identical to $x[n]$. In the latter case, the waveform of the underlying continuous-time output is identical to the waveform of the underlying continuous-time input and delayed D units of time.

If it is desired to pass input signal components in a certain frequency range undistorted in both magnitude and phase, then the transfer function should exhibit a unity magnitude response and a linear-phase response in the band of interest. In Figure 6.4 the frequency response of a lowpass filter with a linear-phase characteristic in the passband is showed. Since the signal components in the stopband are blocked, the phase response in the stopband can be of any shape.

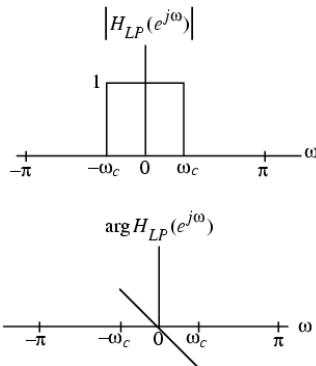


Figure 6.4: Frequency response of a lowpass filter with a linear-phase characteristic in the passband.

Example 59: Linear-phase transfer function

We determine now the impulse response of an ideal lowpass filter with a linear phase response:

$$H_{LP}(e^{j\omega}) = \begin{cases} e^{-j\omega n_0} & 0 < |\omega| < \omega_c \\ 0 & \omega_c \leq |\omega| \leq \pi \end{cases} \quad (6.38)$$

Applying the frequency-shifting property of the DTFT to the impulse response of an ideal zero-phase lowpass filter we arrive at:

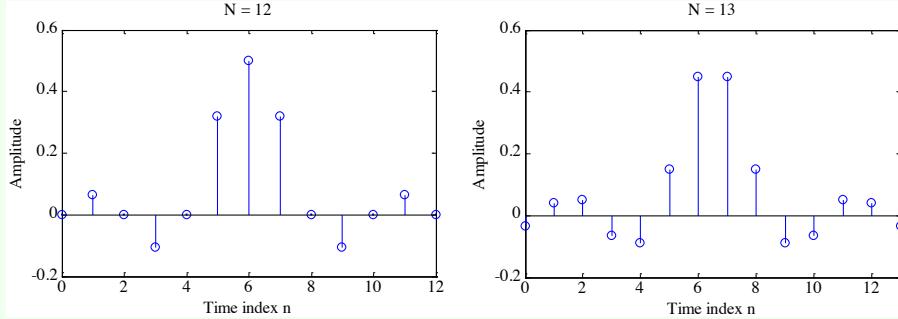
$$h_{LP}[n] = \frac{\sin(\omega(n - n_0))}{\pi(n - n_0)} \quad -\infty < n < \infty \quad (6.39)$$

As before, the above filter is noncausal and of doubly infinite length, and hence, unrealizable. By truncating the impulse response to a finite number of terms, a realizable FIR approximation to the ideal lowpass filter can be developed. The truncated approximation may or may not exhibit linear phase, depending on the value of n_0 chosen.

If we choose $n_0 = \frac{N}{2}$ with N a positive integer, the truncated and shifted approximation:

$$\hat{h}_{LP}[n] = \frac{\sin\left(\omega_c\left(n - \frac{N}{2}\right)\right)}{\pi\left(n - \frac{N}{2}\right)} \quad 0 \leq n \leq N \quad (6.40)$$

will be a length $N + 1$ causal linear-phase FIR filter. In the plot below the filter coefficients are showed. They are obtained using the function sinc for two different values of N .



Because of the symmetry of the impulse response coefficients as indicated in the two plots, the frequency response of the truncated approximation can be expressed as:

Zero-phase
response or
amplitude response

$$\hat{H}_{LP}(e^{j\omega}) = \sum_{n=0}^N \hat{h}_{LP}[n] e^{-j\omega n} = e^{-j\omega \frac{N}{2}} \tilde{H}_{LP}(\omega) \quad (6.41)$$

where $\tilde{H}_{LP}(\omega)$, called the **zero-phase response** or **amplitude response**, is a real function of ω .

Lecture 19.
Thursday 3rd
December, 2020.

Frequency response
from transfer
function

6.1.7 Frequency response from transfer function

If the ROC of the transfer function $H(z)$ includes the unit circle, then the **frequency response** $H(e^{j\omega})$ of the LTI digital filter can be obtained simply as follows:

$$H(e^{j\omega}) = [H(z)]_{z=e^{j\omega}} \quad (6.42)$$

For a real coefficient transfer function $H(z)$ it can be showed that:

$$|H(e^{j\omega})|^2 = H(e^{j\omega})H^*(e^{j\omega}) = H(e^{j\omega})H(e^{-j\omega}) = [H(z)H(z^{-1})]_{z=e^{j\omega}} \quad (6.43)$$

For a stable rational transfer function in the form:

$$H(z) = \frac{p_0}{d_0} z^{(N-M)} \frac{\prod_{k=1}^M (z - \xi_k)}{\prod_{k=1}^N (z - \lambda_k)} \quad (6.44)$$

Factored form of
frequency response

the **factored form of the frequency response** is given by:

$$H(e^{j\omega}) = \frac{p_0}{d_0} e^{j\omega(N-M)} \frac{\prod_{k=1}^M (e^{j\omega} - \xi_k)}{\prod_{k=1}^N (e^{j\omega} - \lambda_k)} \quad (6.45)$$

It is convenient to visualize the contributions of the zero ($z - \xi_k$) and the pole factor ($z - \lambda_k$) from the factored form of the frequency response. The **magnitude function** is given by:

$$|H(e^{j\omega})| = \left| \frac{p_0}{d_0} \frac{\prod_{k=1}^M |e^{j\omega} - \xi_k|}{\prod_{k=1}^N |e^{j\omega} - \lambda_k|} \right| \quad (6.46)$$

The **phase response for a rational transfer function** is of the form:

$$\arg \{ H(e^{j\omega}) \} = \arg \left(\frac{p_0}{d_0} \right) + \omega(N - M) + \sum_{k=1}^M \arg (e^{j\omega} - \xi_k) - \sum_{k=1}^N \arg (e^{j\omega} - \lambda_k) \quad (6.47)$$

The **magnitude-squared function of a real-coefficient transfer function** can be computed using:

$$|H(e^{j\omega})|^2 = \left| \frac{p_0}{d_0} \right|^2 \frac{\prod_{k=1}^M (e^{j\omega} - \xi_k)(e^{j\omega} - \xi_k^*)}{\prod_{k=1}^N (e^{j\omega} - \lambda_k)(e^{j\omega} - \lambda_k^*)} \quad (6.48)$$

From the factored form of the frequency response in Eq. 6.45 it is convenient to develop a geometric interpretation of the frequency response computation from the pole-zero plot as ω varies from 0 to 2π on the unit circle. The geometric interpretation can be used to obtain a sketch of the response as a function of the frequency. A typical factor in the factored form of the frequency response is given by $(e^{j\omega} - \rho e^{j\varphi})$, where $\rho e^{j\varphi}$ is a zero, if it is zero factor, or it is a pole, if it is a pole factor. As showed in Figure 6.5, in the z -plane the factor $(e^{j\omega} - \rho e^{j\varphi})$ represents a vector starting at the point $z = \rho e^{j\varphi}$ and ending on the unit circle at $z = e^{j\omega}$.

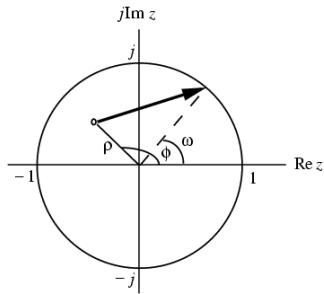


Figure 6.5: Factor $(e^{j\omega} - \rho e^{j\varphi})$ in the z -plane.

As ω is varied from 0 to 2π , the tip of the vector moves counterclockwise from the point $z = 1$, tracing the unit circle, and back to the point $z = 1$.

Magnitude function

Phase response for a rational transfer function

Magnitude-squared function of a real-coefficient transfer function

Geometric interpretation of frequency response

As indicated by the modulus of $H(e^{j\omega})$, in Eq. 6.46, the magnitude response $|H(e^{j\omega})|$ at a specific value of ω is given by the product of the magnitudes of all zero vectors divided by the product of the magnitudes of all pole vectors. Likewise, from Eq. 6.47, we observe that the phase response at a specific value of ω is obtained by adding the phase of the term $\frac{p_0}{d_0}$ and the linear-phase term $\omega(N - M)$ to the sum of the angles of the zero vectors minus the angles of the pole vectors. Thus, an approximate plot of the magnitude and phase responses of the transfer function of an LTI digital filter can be developed by examining the pole and zero locations.

Now, a zero (pole) vector has the smallest magnitude when $\omega = \varphi$. To highly attenuate signal components in a specified frequency range, we need to place zeros very close to or on the unit circle in this range. Likewise, to highly emphasize signal components in a specified frequency range, we need to place poles very close to or on the unit circle in this range.

6.2 Simple digital filters

Simple digital filters

Later in the course we shall review various methods of designing frequency-selective filters satisfying prescribed specifications. We now describe several low-order FIR and IIR digital filters with reasonable selective frequency responses that often are satisfactory in a number of applications. FIR digital filters considered here have integer-valued impulse response coefficients. These filters are employed in a number of practical applications, primarily because of their simplicity, which makes them amenable to inexpensive hardware implementations.

6.2.1 Lowpass FIR digital filters

Example of lowpass FIR digital filter

The simplest lowpass FIR digital filter is the **2-point moving-average filter** given by:

$$H_0(z) = \frac{1}{2}(1 + z^{-1}) = \frac{z + 1}{2z} \quad (6.49)$$

The transfer function in Eq. 6.49 has a zero at $z = -1$ and a pole at $z = 0$. Note that here the pole vector has a unity magnitude for all values of ω . On the other hand, as ω increases from 0 to π , the magnitude of the zero vector decreases from a value of 2, namely the diameter of the unit circle, to 0. Hence, the magnitude response $|H_0(e^{j\omega})|$ is a monotonically decreasing function of ω from $\omega = 0$ to $\omega = \pi$. The maximum value of the magnitude function is 1 at $\omega = 0$, and the minimum value is 0 at $\omega = \pi$, i.e.:

$$|H_0(e^{j0})| = 1 \quad (6.50)$$

$$|H_0(e^{j\pi})| = 0 \quad (6.51)$$

The frequency response of the above filter is given by:

$$H_0(e^{j\omega}) = e^{-j\frac{\omega}{2}} \cos\left(\frac{\omega}{2}\right) \quad (6.52)$$

The magnitude response $|H_0(e^{j\omega})| = \cos\left(\frac{\omega}{2}\right)$ can be seen to be a monotonically decreasing function of ω , as showed in Figure 6.6.

The frequency $\omega = \omega_c$ at which:

$$|H_0(e^{j\omega_c})| = \frac{1}{\sqrt{2}} H_0(e^{j0}) \quad (6.53)$$

is of practical interest since here the gain $G(\omega_c)$ in dB is given by:

$$G(\omega_c) = 20 \log_{10} |H(e^{j\omega_c})| = 20 \log_{10} |H(e^{j0})| - 20 \log_{10} \sqrt{2} \approx -3 \text{ dB} \quad (6.54)$$

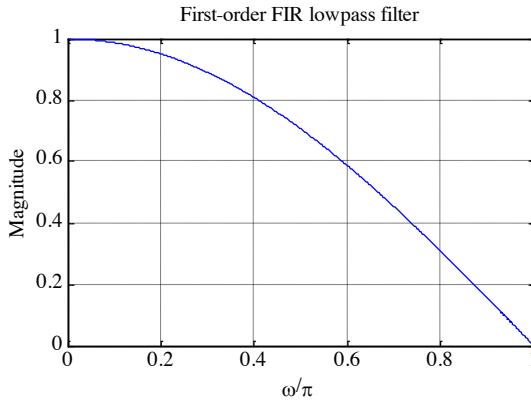


Figure 6.6: Magnitude response of the FIR lowpass filter.

since the dc gain $G(0) = 20 \log_{10} |H(e^{j0})| = 0$.

Thus, the gain $G(\omega)$ at $\omega = \omega_c$ is approximately 3 dB less than the gain at $\omega = 0$. As a result, ω_c is called the **3-dB cutoff frequency**. To determine the value of ω_c , we set:

$$|H_0(e^{j\omega_c})|^2 = \cos^2\left(\frac{\omega_c}{2}\right) = \frac{1}{2} \quad (6.55)$$

which yields $\omega_c = \frac{\pi}{2}$.

The 3-dB cutoff frequency ω_c can be considered as the passband edge frequency. As a result, for the filter $H_0(z)$ the passband width is approximately $\frac{\pi}{2}$. The stopband is from $\frac{\pi}{2}$ to π . Note that $H_0(z)$ has a zero at $z = -1$ or $\omega = \pi$, which is in the stopband of the filter.

3-dB cutoff frequency

A **cascade** of the simple FIR filter:

$$H_0(z) = \frac{1}{2}(1 + z^{-1}) \quad (6.56)$$

Lowpass FIR filter cascade

results in an improved lowpass frequency response as showed in Figure 6.7 for a cascade of 3 sections.

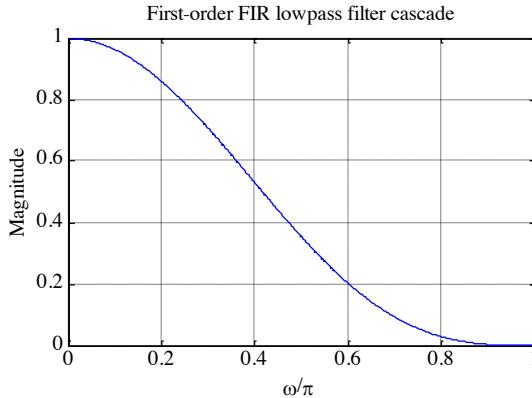


Figure 6.7: Magnitude response of the FIR lowpass filter cascade.

The 3-dB cutoff frequency of a cascade of M sections is given by:

$$\omega_c = 2 \arccos\left(2^{-\frac{1}{2M}}\right) \quad (6.57)$$

For $M = 3$, the above yields $\omega_c = 0.302\pi$. Thus, the cascade of first-order sections yields a sharper magnitude response but at the expense of a decrease in the width of the passband.

A better approximation to the ideal lowpass filter is given by a higher-order moving-average filter. Signals with rapid fluctuations in sample values are generally associated with high-frequency components. These are essentially removed by a moving-average filter resulting in a smoother output waveform.

6.2.2 Highpass FIR digital filters

Example of highpass FIR digital filter

The simplest highpass FIR filter is obtained from the simplest lowpass FIR filter by replacing z with $-z$. This results in:

$$H_1(z) = \frac{1}{2}(1 - z^{-1}) \quad (6.58)$$

The corresponding frequency response is given by:

$$H_1(e^{j\omega}) = j e^{-j\frac{\omega}{2}} \sin\left(\frac{\omega}{2}\right) \quad (6.59)$$

whose magnitude response is showed in Figure 6.8.

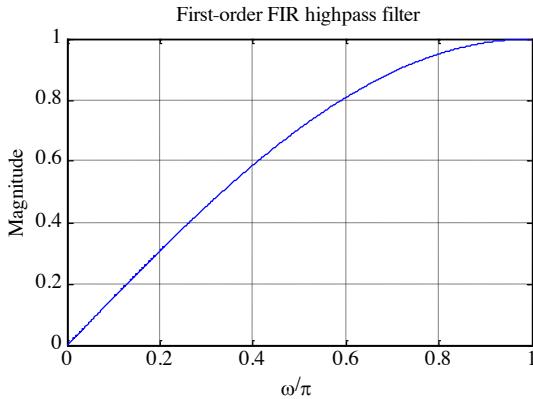


Figure 6.8: Magnitude response of the FIR highpass filter.

The monotonically increasing behavior of the magnitude function can again be demonstrated by examining the pole-zero pattern of the transfer function $H_1(z)$. The highpass transfer function $H_1(z)$ has a zero at $z = 1$ or $\omega = 0$, which is in the stopband of the filter.

Improved highpass magnitude response can be obtained by cascading several sections of the first-order highpass filter. Alternately, a **higher-order highpass filter** of the form:

$$H_1(z) = \frac{1}{M} \sum_{n=0}^{M-1} (-1)^n z^{-n} \quad (6.60)$$

Higher-order highpass filter

Application in MTI radars

is obtained by replacing z with $-z$ in the transfer function of a moving average filter. An application of the FIR highpass filters is in **moving-target-indicator (MTI) radars**. In these radars, interfering signals, called **clutters**, are generated from fixed objects in the path of the radar beam. The clutter, generated mainly from ground echoes and weather returns, has frequency components near zero frequency (dc). The clutter can be removed by filtering the radar return signal through a two-pulse canceler, which is the first-order FIR highpass filter. For a more effective removal it may be necessary to use a **three-pulse canceler** obtained by cascading two two-pulse cancelers:

$$H_1(z) = \frac{1}{2}(1 - z^{-1}) \quad (6.61)$$

6.2.3 Lowpass IIR digital filters

We have already showed that the first-order causal IIR transfer function:

$$H(z) = \frac{k}{1 - \alpha z^{-1}} \quad 0 < \alpha < 1 \quad (6.62)$$

has a lowpass magnitude response for $\alpha > 0$. An improved lowpass magnitude response is obtained by adding a factor $(1 + z^{-1})$ to the numerator of the transfer function:

$$H(z) = \frac{k(1 + z^{-1})}{1 - \alpha z^{-1}} \quad 0 < \alpha < 1 \quad (6.63)$$

This forces the magnitude response to have a zero at $\omega = \pi$ in the stopband of the filter. On the other hand, the first-order causal IIR transfer function:

$$H(z) = \frac{k}{1 - \alpha z^{-1}} \quad -1 < \alpha < 0 \quad (6.64)$$

has a highpass magnitude response for $\alpha < 0$. However, the modified transfer function obtained with the addition of a factor $(1 + z^{-1})$ to the numerator:

$$H(z) = \frac{k(1 + z^{-1})}{1 - \alpha z^{-1}} \quad -1 < \alpha < 0 \quad (6.65)$$

exhibits a lowpass magnitude response. The **modified first-order lowpass transfer function** for both positive and negative values of α is then given by:

$$H_{LP}(z) = \frac{k(1 + z^{-1})}{1 - \alpha z^{-1}} \quad 0 < |\alpha| < 1 \quad (6.66)$$

As ω increases from 0 to π , the magnitude of the zero vector decreases from a value of 2 to 0. The maximum values of the magnitude function is $\frac{2k}{(1-\alpha)}$ at $\omega = 0$ and the minimum value is 0 at $\omega = \pi$, i.e.:

$$|H_{LP}(e^{j\omega})| = \frac{2k}{(1 - \alpha)} \quad (6.67)$$

$$|H_{LP}(e^{j\pi})| = 0 \quad (6.68)$$

Therefore, $|H_{LP}(e^{j\omega})|$ is a monotonically decreasing function of ω from $\omega = 0$ to $\omega = \pi$.

For most applications, it is usual to have a dc gain of 0 dB, that is to have the maximum magnitude $|H(e^{j0})| = 1$. To this end, we choose $k = \frac{1-\alpha}{2}$, resulting in the **first-order IIR lowpass transfer function**:

$$H_{LP}(z) = \frac{1 - \alpha}{2} \frac{(1 + z^{-1})}{1 - \alpha z^{-1}} \quad 0 < |\alpha| < 1 \quad (6.69)$$

The transfer function in Eq. 6.69 has a zero at i.e., at $\omega = \pi$, which is in the stopband. It has also a real pole at $z = \alpha$.

As ω increases from 0 to π , the magnitude of the zero vector decreases from a value of 2 to 0, whereas, for a positive value of α , the magnitude of the pole vector increases from a value of $1 - \alpha$ to $1 + \alpha$. The maximum value of the magnitude function is 1 at $\omega = 0$, and the minimum value is 0 at $\omega = \pi$, i.e. $|H_{LP}(e^{j0})| = 1$, $|H_{LP}(e^{j\pi})| = 0$. Therefore, $H_{LP}(e^{j\omega})$ is a monotonically decreasing function of ω from $\omega = 0$ to $\omega = \pi$ as indicated in Figure 6.9.

Example of lowpass IIR digital filter

Modified first-order lowpass transfer function

First-order IIR lowpass transfer function

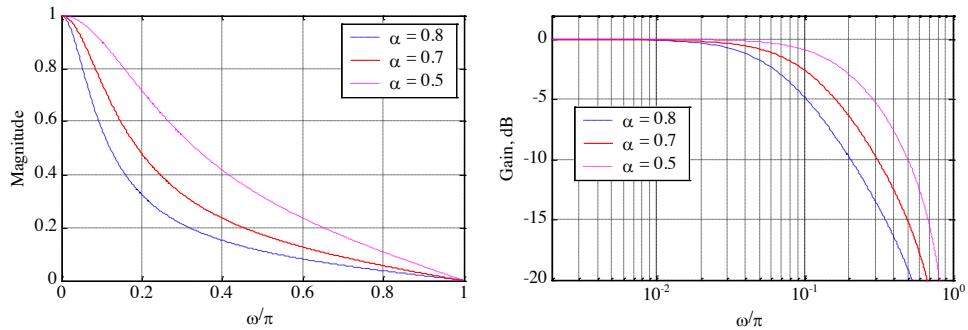


Figure 6.9: Magnitude (left) and gain (right) responses of the IIR lowpass filter.

The squared magnitude function is given by:

$$|H_{LP}(e^{j\omega})|^2 = \frac{(1-\alpha)^2(1+\cos\omega)}{2(1+\alpha^2-2\alpha\cos\omega)} \quad (6.70)$$

The derivative of $|H_{LP}(e^{j\omega})|^2$ with respect to ω is given by:

$$\frac{d|H_{LP}(e^{j\omega})|^2}{d\omega} = \frac{-(1-\alpha)^2(1+\alpha^2+2\alpha)\sin\omega}{2(1+\alpha^2-2\alpha\cos\omega)^2} \quad (6.71)$$

The derivative in Eq. 6.71 is less or equal than zero in the range $0 \leq \omega \leq \pi$, verifying again the monotonically decreasing behaviour of the magnitude function. To determine the **3-dB cutoff frequency** we set:

$$|H_{LP}(e^{j\omega_c})|^2 = \frac{1}{2} \quad (6.72)$$

in the expression for the square magnitude function, resulting in:

$$\frac{(1-\alpha)^2(1+\cos\omega_c)}{2(1+\alpha^2-2\alpha\cos\omega_c)} = \frac{1}{2} \quad (6.73)$$

which, when solved, yields:

$$\cos\omega_c = \frac{2\alpha}{1+\alpha^2} \quad (6.74)$$

The above quadratic equation can be solved for α yielding two solutions. The one resulting in a stable transfer function $H_{LP}(z)$ is given by:

$$\alpha = \frac{1-\sin\omega_c}{\cos\omega_c} \quad (6.75)$$

It follows from:

$$|H_{LP}(e^{j\omega})|^2 = \frac{(1-\alpha)^2(1+\cos\omega_c)}{2(1+\alpha^2-2\alpha\cos\omega_c)} \quad (6.76)$$

that $H_{LP}(z)$ is a BR function for $|\alpha| < 1$.

6.2.4 Highpass IIR digital filters

Example of highpass IIR digital filter

A **first-order causal highpass IIR digital filter** has a transfer function given by:

$$H_{HP}(z) = \frac{1+\alpha}{2} \frac{1-z^{-1}}{1-\alpha z^{-1}} \quad (6.77)$$

where $|\alpha| < 1$ in order to have stability. The transfer function in Eq. 6.77 has a zero at $z = 1$, i.e. at $\omega = 0$, which is in the stopband.

Its **3-dB cutoff frequency** ω_c is given by:

$$\alpha = \frac{1 - \sin \omega_c}{\cos \omega_c} \quad (6.78)$$

which is the same as that of $H_{LP}(z)$. Magnitude and gain responses of $H_{HP}(z)$ are showed in Figure 6.10. $H_{HP}(z)$ is a BR function for $|\alpha| < 1$.

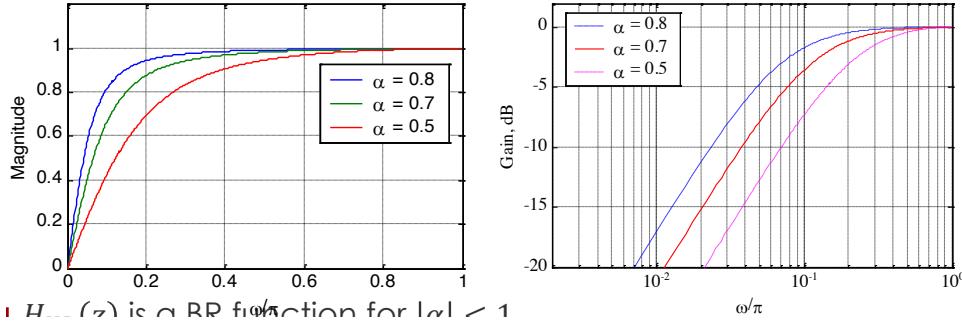


Figure 6.10: Magnitude (left) and gain (right) responses of the IIR highpass filter.

Example 60: Design of a filter

We design a first-order highpass digital filter with a 3-dB cutoff frequency of 0.8π . Now, we consider that:

$$\begin{aligned}\sin(\omega_c) &= \sin(0.8\pi) = 0.587785 \\ \cos(\omega_c) &= \cos(0.8\pi) = -0.80902\end{aligned}$$

Therefore:

$$\alpha = \frac{1 - \sin \omega_c}{\cos \omega_c} = -0.5095245 \quad (6.79)$$

Therefore:

$$H_{HP}(z) = \frac{1 + \alpha}{2} \frac{1 - z^{-1}}{1 - \alpha z^{-1}} = 0.245238 \left(\frac{1 - z^{-1}}{1 + 0.5095245 z^{-1}} \right) \quad (6.80)$$

3-dB cutoff frequency

Example of filter design

6.2.5 Bandpass IIR digital filters

A **second-order bandpass digital transfer function** is given by:

$$H_{BP}(z) = \frac{1 - \alpha}{2} \left(\frac{1 - z^{-2}}{1 - \beta(1 + \alpha)z^{-1} + \alpha z^{-2}} \right) \quad (6.81)$$

Example of bandpass IIR digital filter

Its squared magnitude function is:

$$|H_{BP}(e^{j\omega})|^2 = \frac{(1 - \alpha)^2(1 - \cos(2\omega))}{2[1 + \beta^2(1 + \alpha)^2 + \alpha^2 - 2\beta(1 + \alpha)^2 \cos \omega + 2\alpha \cos(2\alpha)]} \quad (6.82)$$

$|H_{BP}(e^{j\omega})|^2$ goes to zero at $\omega = 0$ and $\omega = \pi$. It assumes a maximum value of 1 at $\omega = \omega_0$, called the **center frequency of the bandpass filter**, where:

$$\omega_0 = \cos^{-1}(\beta) \quad (6.83)$$

Center frequency of the bandpass filter

3-dB cutoff frequencies and 3-dB bandwidth

The frequencies ω_{c_1} and ω_{c_2} where $|H_{BP}(e^{j\omega})|^2$ becomes $\frac{1}{2}$ are called the **3-dB cutoff frequencies**. The difference between the two cutoff frequencies, assuming $\omega_{c_2} > \omega_{c_1}$ is called the **3-dB bandwidth** and is given by:

$$B_w = \omega_{c_2} - \omega_{c_1} = \cos^{-1} \left(\frac{2\alpha}{1 + \alpha^2} \right) \quad (6.84)$$

The transfer function $H_{BP}(e^{j\omega})$ is a BR function if $|\alpha| < 1$ and $|\beta| < 1$. Some plots of $|H_{BP}(e^{j\omega})|$ are showed in Figure 6.11.

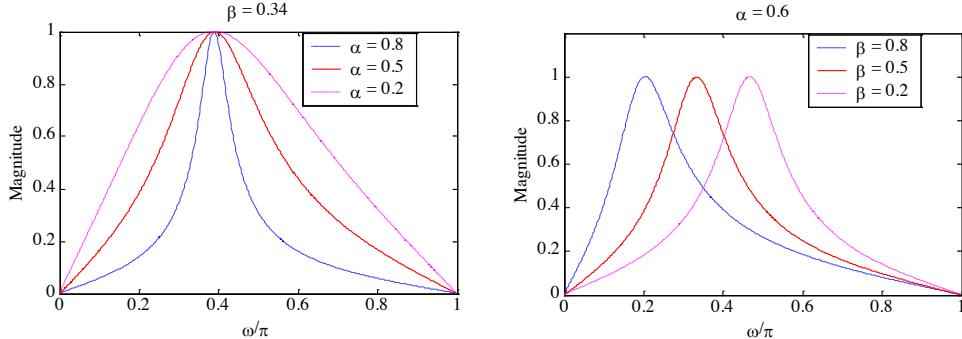


Figure 6.11: Magnitude responses of the IIR bandpass filter.

Example 61: Design of a filter

Example of filter design

We design a second-order bandpass digital filter with center frequency at 0.4π and a 3-dB bandwidth of 0.1π . We have:

$$\beta = \cos(\omega_0) = \cos(0.4\pi) = 0.309017 \quad (6.85)$$

and:

$$\frac{2\alpha}{1 + \alpha^2} = \cos(B_w) = \cos(0.1\pi) = 0.9510565 \quad (6.86)$$

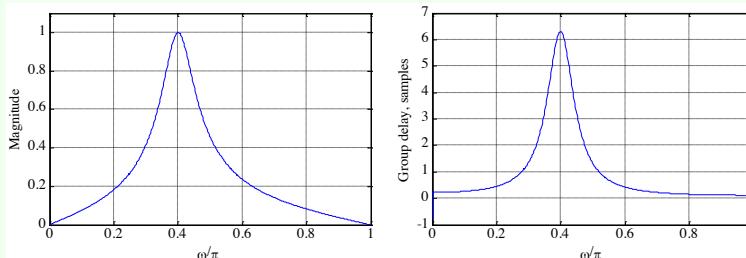
The solution of Eq. 6.86 yields $\alpha = 1.376382$ and $\alpha = 0.72654253$. The corresponding transfer functions are:

$$H'_{BP}(z) = -0.18819 \frac{1 - z^{-2}}{1 - 0.7343424z^{-1} + 1.37638z^{-2}} \quad (6.87)$$

$$H''_{BP}(z) = 0.13673 \frac{1 - z^{-2}}{1 - 0.533531z^{-1} + 0.72654253z^{-2}} \quad (6.88)$$

The poles of $H'_{BP}(z)$ are at $z = 0.3671712 \pm j1.11425636$ and they have a magnitude greater than 1. Thus, the poles of $H'_{BP}(z)$ are outside the unit circle making the transfer function unstable. On the other hand, the poles of $H''_{BP}(z)$ are at $z = 0.2667655 \pm j0.8095546$ and they have a magnitude of 0.8523746. Hence, $H''_{BP}(z)$ is BIBO stable.

Below, the plots of the magnitude function and the group delay of $H''_{BP}(z)$ are showed.



6.2.6 Bandstop IIR digital filters

A second-order bandstop digital filter has a transfer function given by:

$$H_{BS}(z) = \frac{1+\alpha}{2} \left(\frac{1-2\beta z^{-1}+z^{-2}}{1-\beta(1+\alpha)z^{-1}+\alpha z^{-2}} \right) \quad (6.89)$$

The transfer function $H_{BS}(z)$ is a BR function if $|\alpha| < 1$ and $|\beta| < 1$. Its magnitude response is showed in Figure 6.12.

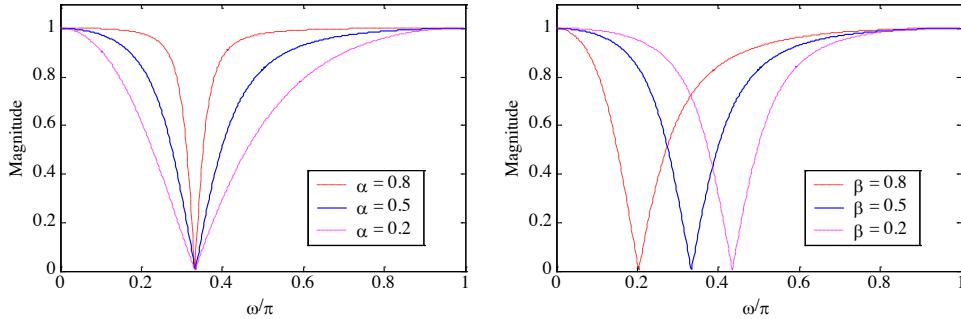


Figure 6.12: Magnitude responses of the IIR bandstop filter.

Here, the magnitude function takes the maximum value of 1 at $\omega = 0$ and $\omega = \pi$. It goes to 0 at $\omega = \omega_0$, where ω_0 , called the **notch frequency**, is given by:

$$\omega_0 = \cos^{-1}(\beta) \quad (6.90)$$

The digital transfer function $H_{BS}(z)$ is more commonly called a **notch filter**.

The frequencies ω_{c1} and ω_{c2} where $|H_{BP}(e^{j\omega})|^2$ becomes $\frac{1}{2}$ are called the **3-dB cutoff frequencies**. The difference between them, assuming $\omega_{c2} > \omega_{c1}$, is called the **3-dB notch bandwidth** and it is given by:

$$B_w = \omega_{c2} - \omega_{c1} = \cos^{-1} \left(\frac{2\alpha}{1+\alpha^2} \right) \quad (6.91)$$

Example of
bandstop IIR
digital filter

Notch frequency
and notch filters

3-dB cutoff
frequencies and
3-dB notch
bandwidth

6.2.7 Higher-Order IIR digital filters

By cascading the simple digital filters discussed so far, we can implement digital filters with sharper magnitude responses. For example, let us consider a cascade of k first-order lowpass sections characterized by the transfer function:

$$H_{LP}(z) = \frac{1-\alpha}{2} \frac{1+z^{-1}}{1-\alpha z^{-1}} \quad (6.92)$$

Example of
higher-order IIR
digital filter

The overall structure has a transfer function given by:

$$G_{LP}(z) = \left(\frac{1-\alpha}{2} \cdot \frac{1+z^{-1}}{1-\alpha z^{-1}} \right)^k \quad (6.93)$$

The corresponding squared-magnitude function is given by:

$$|G_{LP}(e^{j\omega})|^2 = \left[\frac{(1-\alpha)^2(1+\cos\omega)}{2(1+\alpha^2-2\alpha\cos\omega)} \right]^k \quad (6.94)$$

To determine the relation between its 3-dB cutoff frequency ω_c and the parameter α , we set:

$$\left[\frac{(1-\alpha)^2(1+\cos\omega)}{2(1+\alpha^2-2\alpha\cos\omega)} \right]^k = \frac{1}{2} \quad (6.95)$$

which when solved for α , yields for a stable $G_{LP}(z)$:

$$\alpha = \frac{1 + (1 - C) \cos \omega_c - \sin \omega_c \sqrt{2C - C^2}}{1 - C + \cos \omega_c} \quad (6.96)$$

where:

$$C = 2^{\frac{k-1}{k}} \quad (6.97)$$

It should be noted that, for $k = 1$, the expression given in Eq. 6.96 reduces to:

$$\alpha = \frac{1 - \sin \omega_c}{\cos \omega_c} \quad (6.98)$$

Example 62: Design of a filter

Example of filter design

We design a lowpass filter with a 3-dB cutoff frequency at $\omega_c = 0.4\pi$ using a single first-order section and a cascade of 4 first-order sections, and we compare their gain responses.

For the single first-order lowpass filter we have:

$$\alpha = \frac{1 + \sin \omega_c}{\cos \omega_c} = \frac{1 + \sin(0.4\pi)}{\cos(0.4\pi)} \quad (6.99)$$

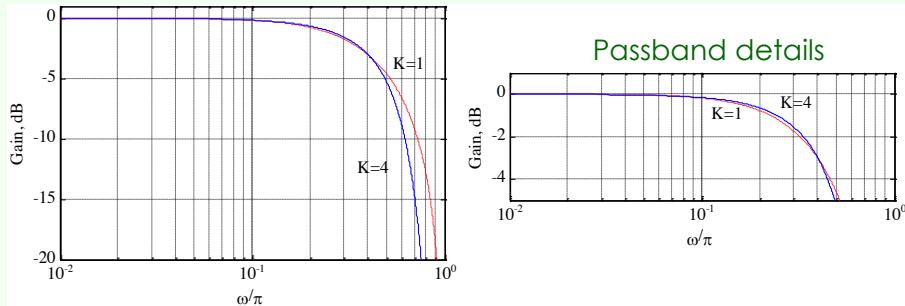
For the cascade of 4 first-order sections, we substitute $k = 4$ and get:

$$C = 2^{\frac{k-1}{k}} = 2^{\frac{4-1}{4}} = 1.6818 \quad (6.100)$$

Next, we compute:

$$\begin{aligned} \alpha &= \frac{1 + (1 - C) \cos \omega_c - \sin \omega_c \sqrt{2C - C^2}}{1 - C + \cos \omega_c} \\ &= \frac{1 + (1 - 1.6818) \cos(0.4\pi) - \sin(0.4\pi) \sqrt{2(1.6818) - (1.6818)^2}}{1 - 1.6818 + \cos(0.4\pi)} \\ &= -0.251 \end{aligned} \quad (6.101)$$

The gain responses of the two filters are showed in the plots below. As can be seen, cascading has resulted in a sharper roll-off in the gain response



Lecture 20.
Thursday 10th
December, 2020.

Linear-phase FIR transfer functions

6.3 Linear-phase FIR transfer functions

It is impossible to design an IIR transfer function with an exact linear-phase. However, it is always possible to design a FIR transfer function with an exact linear-phase response. We now develop the forms of the linear-phase FIR transfer function $H(z)$

with real impulse response $h[n]$. Let:

$$H(z) = \sum_{n=0}^N h[n]z^{-n} \quad (6.102)$$

If $H(z)$ is to have a linear-phase, its frequency response must be of the form:

$$H(e^{j\omega}) = e^{j(c\omega+\beta)} \check{H}(\omega) \quad (6.103)$$

where c and β are constants, and $\check{H}(\omega)$, called **amplitude response** or **zero-phase response**, is a real function of ω . For a real impulse response, the magnitude response $|H(e^{j\omega})|$ is an even function of ω , i.e.:

$$|H(e^{j\omega})| = |H(e^{-j\omega})| \quad (6.104)$$

Since $|H(e^{j\omega})| = |\check{H}(\omega)|$, the amplitude response is then either an even function or an odd function of ω , i.e.:

$$\check{H}(-\omega) = \pm \check{H}(\omega) \quad (6.105)$$

The frequency response satisfies the relation:

$$H(e^{j\omega}) = H^*(e^{-j\omega}) \quad (6.106)$$

or, equivalently, the relation:

$$e^{j(c\omega+\beta)} \check{H}(\omega) = e^{-j(c(-\omega)+\beta)} \check{H}(\omega) \quad (6.107)$$

If $\check{H}(\omega)$ is an even function, then the relation in Eq. 6.107 leads to:

$$e^{j\beta} = e^{-j\beta} \quad (6.108)$$

implying that either $\beta = 0$ or $\beta = \pi$. From:

$$H(e^{j\omega}) = e^{j(c\omega+\beta)} \check{H}(\omega) \quad (6.109)$$

we have:

$$\check{H}(\omega) = e^{-j(c\omega+\beta)} H(e^{j\omega}) \quad (6.110)$$

Substituting the value of β in Eq. 6.110, we get:

$$\check{H}(\omega) = \pm e^{-jc\omega} H(e^{j\omega}) = \pm \sum_{n=0}^N h[n] e^{-j\omega(c+n)} \quad (6.111)$$

Replacing ω with $-\omega$ in Eq. 6.111, we get:

$$\check{H}(-\omega) = \pm \sum_{\ell=0}^N h[\ell] e^{j\omega(c+\ell)} \quad (6.112)$$

Making a change of variable $\ell = N - n$, we rewrite Eq. 6.112 as:

$$\check{H}(-\omega) = \pm \sum_{n=0}^N h[N-n] e^{j\omega(c+N-n)} \quad (6.113)$$

Now, as $\check{H}(\omega) = \check{H}(-\omega)$, we have:

$$h[n] e^{-j\omega(c+n)} = h[N-n] e^{j\omega(c+N-n)} \quad (6.114)$$

Amplitude or zero-phase response

Eq. 6.114 leads to the condition:

$$h[n] = h[N - n] \quad 0 \leq n \leq N \quad (6.115)$$

with $c = -\frac{N}{2}$. Thus, the FIR filter with an even amplitude response will have a linear phase if it has a symmetric impulse response. If $\check{H}(\omega)$ is an odd function of ω , then from:

$$e^{j(c\omega+\beta)} \check{H}(\omega) = e^{-j(-c\omega+\beta)} \check{H}(-\omega) \quad (6.116)$$

we get $e^{j\beta} = -e^{-j\beta}$ as $\check{H}(-\omega) = -\check{H}(\omega)$. Eq. 6.116 is satisfied if $\beta = \pm\frac{\pi}{2}$. Then $H(e^{j\omega}) = e^{j(c\omega+\beta)}$ reduces to:

$$H(e^{j\omega}) = j e^{j c \omega} \check{H}(\omega) \quad (6.117)$$

Eq. 6.117 can be rewritten as:

$$\check{H}(\omega) = -j e^{-j c \omega} H(e^{j\omega}) = -j \sum_{m=0}^N h[m] e^{-j\omega(c+m)} \quad (6.118)$$

Again, as $\check{H}(\omega) = \check{H}(-\omega)$, from Eq. 6.118 we get:

$$\check{H}(-\omega) = j \sum_{\ell=0}^N h[\ell] e^{j\omega(c+\ell)} \quad (6.119)$$

Making a change of variable $\ell = N - n$, we rewrite Eq. 6.119 as:

$$\check{H}(-\omega) = j \sum_{n=0}^N h[N - n] e^{j\omega(c+N-n)} \quad (6.120)$$

Equating the RHS of Eq. 6.120 with the RHS of Eq. 6.118, we arrive at the **condition for linear phase** as:

$$h[n] = h[N - n] \quad 0 \leq n \leq N \quad (6.121)$$

with $c = -\frac{N}{2}$. Therefore, a FIR filter with an odd amplitude response will have linear-phase response if it has an antisymmetric impulse response.

Since the length of the impulse response can be either even or odd, we can define **four types of linear-phase FIR transfer functions**. In particular, for an anti-symmetric FIR filter of odd length, namely N even, $h[\frac{N}{2}] = 0$. We examine in the following discussion each of the four cases, sketched in Figure 6.13.

6.3.1 Symmetric impulse response with odd length

In this case, the degree N is even. In the following discussion we assume also $N = 8$ for simplicity. Therefore, the transfer function $H(z)$ is given by:

$$H(z) = h[0] + h[1]z^{-1} + h[2]z^{-2} + h[3]z^{-3} + h[4]z^{-4} + h[5]z^{-5} + h[6]z^{-6} + h[7]z^{-7} + h[8]z^{-8} \quad (6.122)$$

Because of symmetry, we have:

$$h[0] = h[8] \quad (6.123)$$

$$h[1] = h[7] \quad (6.124)$$

$$h[2] = h[6] \quad (6.125)$$

$$h[3] = h[5] \quad (6.126)$$

Condition for linear phase

Classification on the length of the impulse response

Case of symmetric impulse response with odd length

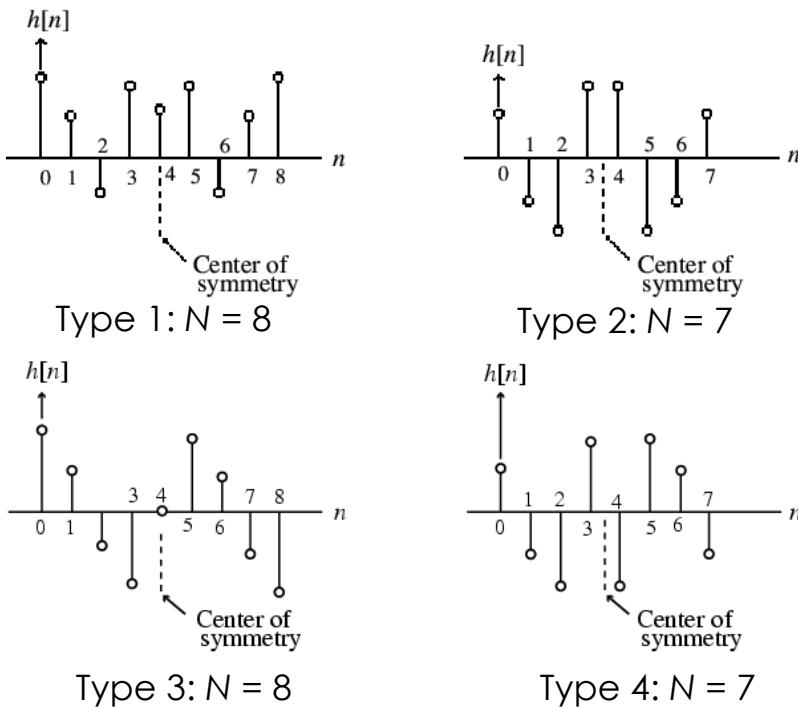


Figure 6.13: The four types of linear-phase FIR transfer functions.

Thus, we can write:

$$\begin{aligned} H(z) &= h[0](1 + z^{-8}) + h[1](z^{-1} + z^{-7}) + h[2](z^{-2} + z^{-6}) + h[3](z^{-3} + z^{-5}) + h[4]z^{-4} \\ &= z^{-4}\{h[0](z^4 + z^{-4}) + h[1](z^3 + z^{-3}) + h[2](z^2 + z^{-2}) + h[3](z + z^{-1}) + h[4]\} \end{aligned} \quad (6.127)$$

The corresponding frequency response is then given by:

Frequency response

$$H(e^{j\omega}) = e^{-j4\omega}\{2h[0]\cos(4\omega) + 2h[1]\cos(3\omega) + 2h[2]\cos(2\omega) + 2h[3]\cos(\omega) + h[4]\} \quad (6.128)$$

The quantity inside the braces is a real function of ω and can assume positive or negative values in the range $0 \leq |\omega| \leq \pi$. The phase function is given by:

Phase function

$$\theta(\omega) = -4\omega + \beta \quad (6.129)$$

where β is either 0 or π , and hence, it is a linear function of ω . The group delay is given by:

Group delay

$$\tau(\omega) = -\frac{d\theta(\omega)}{d\omega} = 4 \quad (6.130)$$

indicating a constant group delay of 4 samples.

In the general case for **Type 1 FIR filters**, the frequency response is of the form:

General case

$$H(e^{j\omega}) = e^{-jN\frac{\omega}{2}} \check{H}(\omega) \quad (6.131)$$

where the amplitude response $\check{H}(\omega)$, also called the zero-phase response, is of the form:

$$\check{H}(\omega) = h\left[\frac{N}{2}\right] + 2 \sum_{n=1}^{\frac{N}{2}} h\left[\frac{N}{2} - n\right] \cos(\omega n) \quad (6.132)$$

Example 63: Symmetric impulse response with odd length

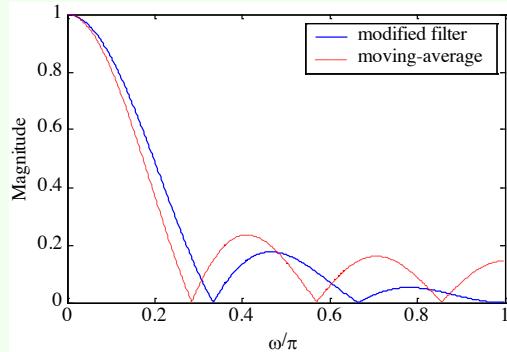
Example of symmetric impulse response with odd length

We consider:

$$H_0(z) = \frac{1}{6} \left[\frac{1}{2} + z^{-1} + z^{-3} + z^{-4} + z^{-5} + \frac{1}{2} z^{-6} \right] \quad (6.133)$$

which is seen to be a slightly modified version of a length-7 moving-average FIR filter.

This transfer function has a symmetric impulse response and therefore a linear phase response. A plot of the magnitude response of $H_0(z)$ along with that of the 7-point moving-average filter is showed below.



Note the improved magnitude response obtained by simply changing the first and the last impulse response coefficients of a moving-average (MA) filter. It can be showed that we can express:

$$H_0(z) = \frac{1}{2}(1 + z^{-1}) \cdot \frac{1}{6}(1 + z^{-1} + z^{-2} + z^{-3} + z^{-4} + z^{-5}) \quad (6.134)$$

which is seen to be a cascade of a 2-point MA filter with a 6-point MA filter. Thus, $H_0(z)$ has a double zero at $z = -1$, i.e. $\omega = \pi$.

6.3.2 Symmetric impulse response with even length

Case of symmetric impulse response with even length

In this case, the degree N is odd and we assume $N = 7$ for simplicity for the following discussion. Therefore, the transfer function is of the form:

$$H(z) = h[0] + h[1]z^{-1} + h[2]z^{-2} + h[3]z^{-3} + h[4]z^{-4} + h[5]z^{-5} + h[6]z^{-6} + h[7]z^{-7} \quad (6.135)$$

Making use of the symmetry of the impulse response coefficients, the transfer function can be rewritten as:

$$\begin{aligned} H(z) &= h[0](1 + z^{-7}) + h[1](z^{-1} + z^{-6}) + h[2](z^{-2} + z^{-5}) + h[3](z^{-3} + z^{-4}) \\ &= z^{-\frac{7}{2}} \left\{ h[0](z^{\frac{7}{2}} + z^{-\frac{7}{2}}) + h[1](z^{\frac{5}{2}} + z^{-\frac{5}{2}}) + h[2](z^{\frac{3}{2}} + z^{-\frac{3}{2}}) + h[3](z^{\frac{1}{2}} + z^{-\frac{1}{2}}) \right\} \end{aligned} \quad (6.136)$$

Frequency response

The corresponding frequency response is given by:

$$H(e^{j\omega}) = e^{-j\frac{7\omega}{2}} \left\{ 2h[0] \cos\left(\frac{7\omega}{2}\right) + 2h[1] \cos\left(\frac{5\omega}{2}\right) + 2h[2] \cos\left(\frac{3\omega}{2}\right) + 2h[3] \cos\left(\frac{\omega}{2}\right) \right\} \quad (6.137)$$

As before, the quantity inside the braces is a real function of ω and can assume positive or negative values in the range $0 \leq |\omega| \leq \pi$. Here, the phase function is given by:

$$\theta(\omega) = -\frac{7}{2}\omega + \beta \quad (6.138)$$

where β is either 0 or π . As a result, the phase is also a linear function of ω and the corresponding group delay is:

$$\tau(\omega) = \frac{7}{2} \quad (6.139)$$

indicating a group delay of $\frac{7}{2}$ samples.

The expression for the frequency response in the general case for **Type 2 FIR filters** is of the form:

$$H(e^{j\omega}) = e^{-jN\frac{\omega}{2}} \check{H}(\omega) \quad (6.140)$$

where the amplitude response is given by:

$$\check{H}(\omega) = 2 \sum_{n=1}^{\frac{N+1}{2}} h\left[\frac{N+1}{2} - n\right] \cos\left(\omega\left(n - \frac{1}{2}\right)\right) \quad (6.141)$$

6.3.3 Antisymmetric impulse response with odd length

In this case, the degree N is even and we assume $N = 8$ for simplicity for the following discussion. Therefore, applying the symmetry condition we get:

$$H(z) = z^{-4} \{ h[0](z^4 - z^{-4}) + h[1](z^3 - z^{-3}) + h[2](z^2 - z^{-2}) + h[3](z - z^{-1}) \} \quad (6.142)$$

The corresponding frequency response is given by:

$$H(e^{j\omega}) = e^{-j4\omega} e^{j\frac{\pi}{2}} \{ 2h[0] \sin(4\omega) + 2h[1] \sin(3\omega) + 2h[2] \sin(2\omega) + 2h[3] \sin(\omega) \} \quad (6.143)$$

It also exhibits a linear phase response given by:

$$\theta(\omega) = -4\omega + \frac{\pi}{2} + \beta \quad (6.144)$$

where β is either 0 or π . The group delay here is:

$$\tau(\omega) = 4 \quad (6.145)$$

indicating a constant group delay of 4 samples.

The expression for the frequency response in the general case for **Type 3 FIR filters** is of the form:

$$H(e^{j\omega}) = e^{-jN\frac{\omega}{2}} \check{H}(\omega) \quad (6.146)$$

where the amplitude response is given by:

$$\check{H}(\omega) = 2 \sum_{n=1}^{\frac{N}{2}} h\left[\frac{N}{2} - n\right] \sin(\omega n) \quad (6.147)$$

Phase function

Group delay

General case

Case of antisymmetric impulse response with odd length

Frequency response

Phase function

Group delay

General case

*Case of
antisymmetric
impulse response
with even length*

6.3.4 Antisymmetric impulse response with even length

In this case, the degree N is even and we assume $N = 7$ for simplicity for the following discussion. Therefore, applying the symmetry condition we get:

$$H(z) = z^{\frac{7}{2}} \left\{ h[0](z^{\frac{7}{2}} - z^{-\frac{7}{2}}) + h[1](z^{\frac{5}{2}} - z^{-\frac{5}{2}}) + h[2](z^{\frac{3}{2}} - z^{-\frac{3}{2}}) + h[3](z^{\frac{1}{2}} - z^{-\frac{1}{2}}) \right\} \quad (6.148)$$

Frequency response

The corresponding frequency response is given by:

$$H(e^{j\omega}) = e^{-j\frac{7\omega}{2}} e^{j\frac{\pi}{2}} \left\{ 2h[0] \sin\left(\frac{7\omega}{2}\right) + 2h[1] \sin\left(\frac{5\omega}{2}\right) + 2h[2] \sin\left(\frac{3\omega}{2}\right) + 2h[3] \sin\left(\frac{\omega}{2}\right) \right\} \quad (6.149)$$

Phase function

It again exhibits a linear phase response given by:

$$\theta(\omega) = -\frac{7}{2}\omega + \frac{\pi}{2} + \beta \quad (6.150)$$

Group delay

where β is either 0 or π . The group delay is constant and is given by:

$$\tau(\omega) = \frac{7}{2} \quad (6.151)$$

General case

The expression for the frequency response in the general case for **Type 4 FIR filters** is of the form:

$$H(e^{j\omega}) = e^{-jN\frac{\omega}{2}} \check{H}(\omega) \quad (6.152)$$

where the amplitude response is given by:

$$\check{H}(\omega) = 2 \sum_{n=1}^{\frac{N+1}{2}} h\left[\frac{N+1}{2} - n\right] \sin\left(\omega\left(n - \frac{1}{2}\right)\right) \quad (6.153)$$

6.3.5 General form of frequency response

*General form of
frequency response*

In each of the four types of linear-phase FIR filters, the **frequency response** is of the form:

$$H(e^{j\omega}) = e^{-jN\frac{\omega}{2}} e^{j\beta} \check{H}(\omega) \quad (6.154)$$

The amplitude response $\check{H}(\omega)$ for each type can become negative over certain frequency ranges, typically in the stopband.

Example 64: General form of frequency response

We consider the causal Type 1 FIR transfer function:

$$H_1(z) = -1 + 2z^{-1} - 3z^{-2} + 6z^{-3} - 3z^{-4} + 2z^{-5} - z^{-6} \quad (6.155)$$

Its amplitude and phase responses are given by:

$$\check{H}_1(\omega) = 6 - 6 \cos(\omega) + 4 \cos(2\omega) - 2 \cos(3\omega) \quad (6.156)$$

$$\theta_1(\omega) = -3\omega \quad (6.157)$$

Next, we consider the causal Type 1 FIR transfer function:

$$H_2(z) = 1 - 2z^{-1} + 3z^{-2} - 6z^{-3} + 3z^{-4} - 2z^{-5} + z^{-6} \quad (6.158)$$

Its amplitude and phase responses are given by:

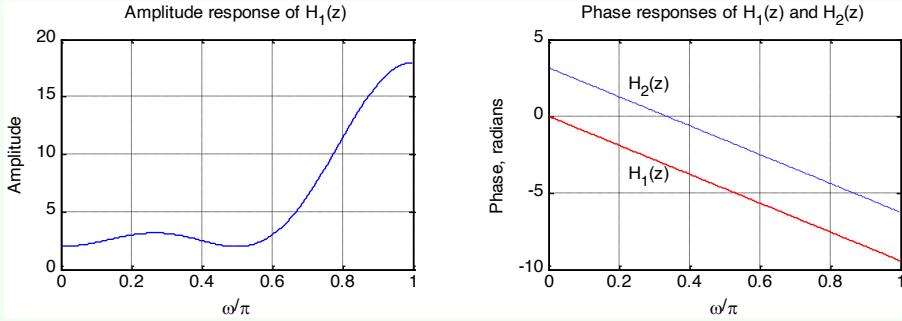
$$\check{H}_2(\omega) = -\check{H}_1(\omega) \quad (6.159)$$

$$\theta_2(\omega) = -3\omega + \pi \quad (6.160)$$

Note that:

$$|H_1(e^{j\omega})| = |H_2(e^{j\omega})| \quad (6.161)$$

Hence, $H_1(z)$ and $H_2(z)$ have identical magnitude responses but phase responses differing by π , as showed in the figure below.



Example 65: General form of frequency response

We consider the causal Type 1 FIR transfer function:

$$H_3(z) = 1 - 2z^{-1} + 3z^{-2} - 3z^{-4} + 2z^{-5} - z^{-6} \quad (6.162)$$

Its amplitude and phase responses are given by:

$$\check{H}_3(\omega) = -6 \sin(\omega) + 4 \sin(2\omega) + 2 \sin(3\omega) \quad (6.163)$$

$$\theta_3(\omega) = -3\omega + \frac{\pi}{2} \quad (6.164)$$

Next, we consider the causal Type 1 FIR transfer function:

$$H_4(z) = -1 + 2z^{-1} - 3z^{-2} + 3z^{-4} - 2z^{-5} + z^{-6} \quad (6.165)$$

Its amplitude and phase responses are given by:

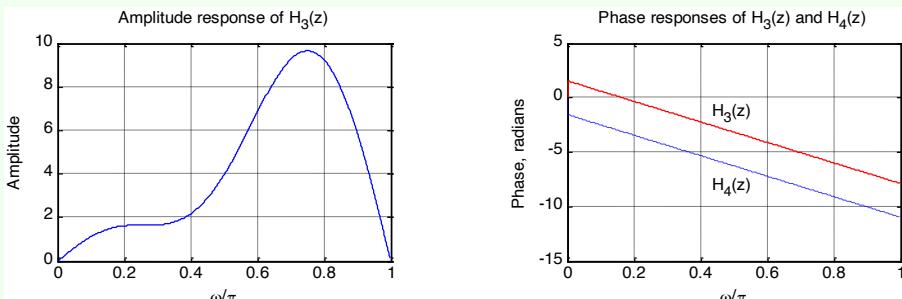
$$\check{H}_4(\omega) = -\check{H}_3(\omega) \quad (6.166)$$

$$\theta_4(\omega) = -3\omega - \frac{\pi}{2} \quad (6.167)$$

Note that:

$$|H_3(e^{j\omega})| = |H_4(e^{j\omega})| \quad (6.168)$$

Hence, $H_3(z)$ and $H_4(z)$ have identical magnitude responses but phase responses differing by π , as showed in the figure below.



General form of magnitude and phase responses

Now, in general, the **magnitude and phase responses** of the linear-phase FIR are given by:

$$|H(e^{j\omega})| = |\check{H}(\omega)| \quad (6.169)$$

$$\theta(\omega) = \begin{cases} -\frac{N\omega}{2} + \beta & \check{H}(\omega) \geq 0 \\ -\frac{N\omega}{2} + \beta - \pi & \check{H}(\omega) < 0 \end{cases} \quad (6.170)$$

General form of group delay

The **group delay** in each case is:

$$\tau(\omega) = \frac{N}{2} \quad (6.171)$$

Note that, even though the group delay is constant, since in general $|H(e^{j\omega})|$ is not a constant, the output waveform is not a replica of the input waveform.

A FIR filter with a frequency response that is a real function of ω is often called a zero-phase filter. Such a filter must have a non-causal impulse response. In fact, it needs to have a purely real-valued frequency response, and, consequently, it must have an impulse response that is even with respect to the time index $n = 0$. Therefore, it must be non-causal.

6.3.6 Zero locations

Zero locations

In this context, let us consider first a FIR filter with a symmetric impulse response:

$$h[n] = h[N - n] \quad (6.172)$$

Its transfer function can be written as:

$$H(z) = \sum_{n=0}^N h[n]z^{-n} = \sum_{n=0}^N h[N - n]z^{-n} \quad (6.173)$$

By making a change of variable $m = N - n$, we can rewrite:

$$H(z) = \sum_{n=0}^N h[N - n]z^{-n} = \sum_{m=0}^N h[m]z^{-N+m} = z^{-N} \underbrace{\sum_{m=0}^N h[m]z^m}_{H(z^{-1})} \quad (6.174)$$

Hence, for a FIR filter with a symmetric impulse response of length $N + 1$ we have:

$$H(z) = z^{-N} H(z^{-1}) \quad (6.175)$$

Mirror image polynomial

A real-coefficient polynomial $H(z)$ satisfying the above condition is called a **mirror-image polynomial (MIP)**.

Now, let us consider first an FIR filter with an antisymmetric impulse response:

$$h[n] = -h[N - n] \quad (6.176)$$

Its transfer function can be written as:

$$H(z) = \sum_{n=0}^N h[n]z^{-n} = -\sum_{n=0}^N h[N - n]z^{-n} \quad (6.177)$$

By making a change of variable $m = N - n$, we can write:

$$H(z) = -\sum_{n=0}^N h[N-n]z^{-n} = -\sum_{m=0}^N h[m]z^{-N+m} = -z^{-N}H(z^{-1}) \quad (6.178)$$

Hence, the transfer function $H(z)$ of an FIR filter with an antisymmetric impulse response satisfies the condition:

$$H(z) = z^{-N}H(z^{-1}) \quad (6.179)$$

A real-coefficient polynomial $H(z)$ satisfying the above condition is called **antimirror-image polynomial (AIP)**.

Now, it follows from the relation $H(z) = \pm z^{-N}H(z^{-1})$ that if $z = \xi_0$ is a zero of $H(z)$, so is $z = \frac{1}{\xi_0}$. Moreover, for an FIR filter with a real impulse response, the zeros of $H(z)$ occur in complex conjugate pairs. Hence, a zero at $z = \xi_0$ is associated with a zero at $z = \xi_0^*$. Thus, a complex zero that is not on the unit circle is associated with a set of 4 zeros given by:

$$z = re^{\pm j\varphi}, \quad \frac{1}{r}e^{\pm j\varphi} \quad (6.180)$$

A zero on the unit circle appear as a pair:

$$z = e^{\pm j\varphi} \quad (6.181)$$

as its reciprocal is also its complex conjugate. Since a zero at $z = \pm 1$ is its own reciprocal, it can appear only singly.

Now, a Type 2 FIR filter satisfies:

$$H(z) = z^{-N}H(z^{-1}) \quad (6.182)$$

with degree N odd. Hence, $H(-1) = (-1)^{-N}H(-1) = -H(-1)$, implying $H(-1) = 0$, i.e., $H(z)$ must have a zero at $z = -1$.

Likewise, a Type 3 or 4 FIR filter satisfies:

$$H(z) = -z^{-N}H(z^{-1}) \quad (6.183)$$

Thus:

$$H(1) = (-1)^{-N}H(1) = -H(1) \quad (6.184)$$

implying that $H(z)$ must have a zero at $z = 1$. On the other hand, only the Type 3 FIR filter is restricted to have a zero at $z = -1$ since here the degree N is even and hence:

$$H(-1) = -(-1)^{-N}H(-1) = -H(-1) \quad (6.185)$$

Typical zero locations are showed in Figure 6.14.

So, to summarize:

- **Type 1 FIR filter:** either an even number or no zeros at $z = 1$ and $z = -1$;
- **Type 2 FIR filter:** either an even number or no zeros at $z = 1$ and an odd number of zeros at $z = -1$;
- **Type 3 FIR filter:** an odd number of zeros at $z = 1$ and $z = -1$;
- **Type 4 FIR filter:** an odd number of zeros at $z = 1$ and either an even number or no zeros at $z = -1$.

antimirror image polynomial

Zeros (not) on the unit circle

Summary on zero locations depending on the type of filter

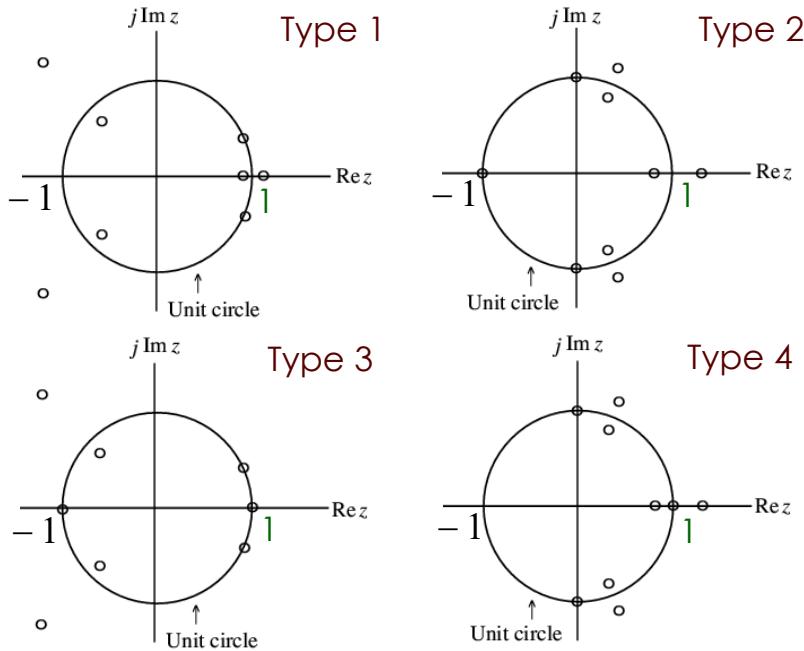


Figure 6.14: Typical zero locations.

Limitations due to zero locations

The presence of zeros at $z = \pm 1$ leads to the following limitations on the use of these linear-phase transfer functions for designing frequency-selective filters:

- a Type 2 FIR filter cannot be used to design a highpass filter since it always has a zero;
- a Type 3 FIR filter has zeros at both $z = 1$ and $z = -1$, and hence cannot be used to design either a lowpass or a highpass or a bandstop filter;
- a Type 4 FIR filter is not appropriate to design lowpass and bandstop filters due to the presence of a zero at $z = 1$;
- a Type 1 FIR filter has no such restrictions and can be used to design almost any type of filter.

Lecture 21.

Tuesday 15th

December, 2020.

6.4 Comb filters

Comb filters

The simple filters discussed so far are characterized either by a single passband and/or a single stopband. There are applications where these kind of stuff is not sufficient, so filters with multiple passbands and stopbands are required. The so called **comb filters** are an example of such instruments.

A comb filter is an LTI digital filter such that:

- its output to a scaled sum of input digital signals is equal to the scaled sum of the outputs to every one of these input signals (i.e., the filter satisfies the **superposition principle**);
- for any input signal that has a given delay, the output undergoes the same delay as the input.

Features in general form

In its most general form, a comb filter has a frequency response that is a periodic function of ω , with a period $\frac{2\pi}{L}$, where L is a positive integer. Moreover, if $H(z)$ is a filter with a single passband and/or a single stopband, a comb filter can be easily

generated from it by replacing each delay in its realization with L delays, resulting in a structure with a transfer function given by:

$$G(z) = H(z^L) \quad (6.186)$$

In particular, if $|H(e^{j\omega})|$ exhibits a peak at ω_p , then $|G(e^{j\omega})|$ will exhibit L peaks at $\frac{\omega_p k}{L}$, with $0 \leq k \leq L - 1$, in the frequency range $0 \leq \omega \leq 2\pi$. Likewise, $|H(e^{j\omega})|$ has a notch at ω_0 , then $|G(e^{j\omega})|$ will have L notches at $\frac{\omega_0 k}{L}$, with $0 \leq k \leq L - 1$, in the frequency range $0 \leq \omega \leq 2\pi$.

Another important thing to remark is that a comb filter can be generated from either an FIR or an IIR prototype filter.

Examples of comb filters

Example 66: Comb filters from FIR/IIR prototype filter

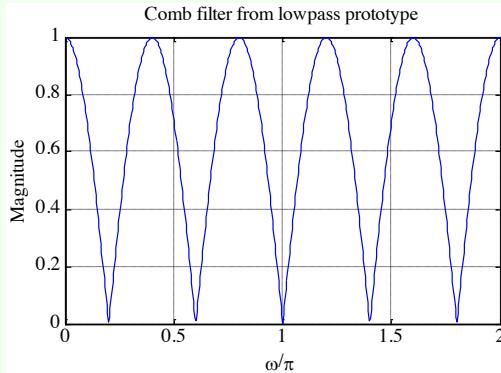
The comb filter generated from the prototype highpass FIR filter:

$$H_0(z) = \frac{1}{2}(1 + z^{-1}) \quad (6.187)$$

has a transfer function:

$$G_0(z) = H_0(z^L) = \frac{1}{2}(1 + z^{-L}) \quad (6.188)$$

$|G_0(e^{j\omega})|$ has L notches at $\omega = \frac{(2k+1)\pi}{L}$ and L peaks at $\omega = \frac{2\pi k}{L}$, with $0 \leq k \leq L - 1$, in the frequency range $0 \leq \omega < 2\pi$.



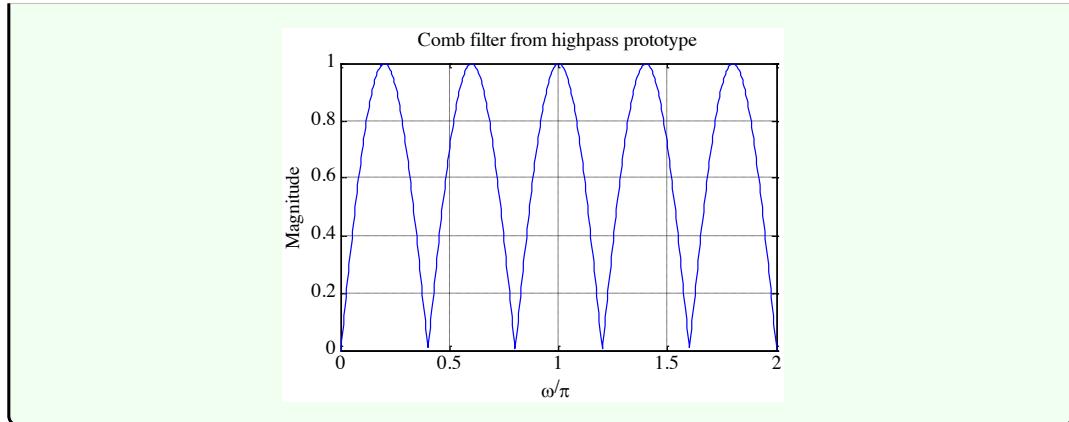
On the other hand, the comb filter generated from the prototype highpass FIR filter:

$$H_1(z) = \frac{1}{2}(1 - z^{-1}) \quad (6.189)$$

has a transfer function:

$$G_0(z) = H_1(z^L) = \frac{1}{2}(1 - z^{-L}) \quad (6.190)$$

$|G_0(e^{j\omega})|$ has L peaks at $\omega = \frac{(2k+1)\pi}{L}$ and L notches at $\omega = \frac{2\pi k}{L}$, with $0 \leq k \leq L - 1$, in the frequency range $0 \leq \omega < 2\pi$.



Depending on the application, comb filters with other types of periodic magnitude responses can be easily generated by appropriately choosing the prototype average filter.

Example 67: Comb filter from M -point moving average

The M -point moving average filter:

$$H(z) = \frac{1 - z^{-M}}{M(1 - z^{-1})} \quad (6.191)$$

can be used as a prototype. This filter has a peak magnitude at $\omega = 0$, and $M - 1$ notches at $\omega = \frac{2\pi\ell}{M}$, with $1 \leq \ell \leq M - 1$. The corresponding comb filter will have a transfer function:

$$G(z) = \frac{1 - z^{-LM}}{M(1 - z^{-L})} \quad (6.192)$$

whose magnitude has L peaks at $\omega = \frac{2\pi k}{LM}$, with $1 \leq k \leq L(M - 1)$. By choosing L and M appropriately, peaks and notches can be created at desired locations.

6.5 Digital filter structures

Digital filter structures

Now, before going on with the discussion, let us return to basic concepts of an LTI discrete-time system. Let us consider:

$$y[n] = \sum_{k=-\infty}^{\infty} x[k]h[n-k] = \sum_{k=-\infty}^{\infty} x[n-k]h[n] \quad (6.193)$$

The formula in Eq. 6.193 expresses the n^{th} output sample as a convolution sum of the input with the impulse response of the system. This can be considered as the characterization of the system and, therefore, the convolution sum can in principle be used to implement the system. It involves additions, multiplications and delays, which are simple operations. In particular, the input-output relation involves a finite sum of products:

$$y[n] = - \sum_{k=1}^N d_k y[n-k] + \sum_{k=0}^M p_k x[n-k] \quad (6.194)$$

On the other hand, an FIR system can be implemented using the convolution sum, which is a finite sum of products:

$$y[n] = \sum_{k=0}^N h[k]x[n-k] \quad (6.195)$$

Now, the actual implementation of an LTI digital filter can be either in software or hardware form, depending on the application. So, in either case, the signal variables and the filter coefficients cannot be represented with infinite precision.

A direct implementation based on either the difference equation or the finite convolution sum may not provide satisfactory performance due to the finite precision arithmetic. Thus, it is of practical interest to develop alternate realizations and choose the structure that provides satisfactory performances under this framework. In this context, a **structural representation** using interconnected basic building blocks is the first step in the hardware or software implementation of an LTI digital filter. The structural representation provides the key relations between some pertinent internal variables with the input and output that in turn provides the key to the implementation.

Structural representation

6.5.1 Block diagram representation

Block diagram representation of computational algorithms

We have seen that, in time domain, the input-output relations of an LTI digital filter is given by the convolution sum in Eq. 6.193 or by the linear constant coefficient difference Eq. 6.194. For the implementation, the input-output relationship must be described by a **valid computational algorithm**. To illustrate what we are meaning, let us consider the causal first-order LTI digital filter showed in Figure 6.15.

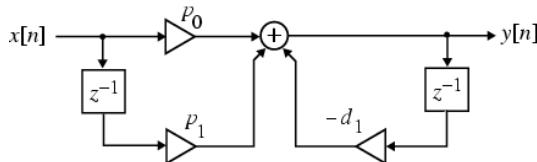


Figure 6.15: Block diagram representation of a causal first-order LTI digital filter.

This filter is described by the difference equation:

$$y[n] = -d_1y[n-1] + p_0x[n] + p_1x[n-1] \quad (6.196)$$

Using Eq. 6.196, we can compute $y[n]$ for $n \geq 0$ knowing the initial condition $y[-1]$ and the input $x[n]$ for $n \geq -1$:

$$y[0] = -d_1y[-1] + p_0x[0] + p_1x[-1] \quad (6.197)$$

$$y[1] = -d_1y[0] + p_0x[1] + p_1x[0] \quad (6.198)$$

$$y[2] = -d_1y[1] + p_0x[2] + p_1x[1] \quad (6.199)$$

$\dots = \dots$

Example with a causal first-order LTI digital filter

and so on and so forth. Each step of the calculation requires the knowledge of the previously calculated value of the output sample (delayed value of the output), the present value of the input sample and the previous one (delayed value of the input). As a result, the first-order difference equation can be interpreted as a valid computational algorithm.

The computational algorithm of an LTI digital filter can be conveniently represented in **block diagram form** using the **basic building blocks** showed in Figure 6.16. There are several advantages when adopting this representation. In fact, it is:

Basic building blocks

Advantages of block diagram representation

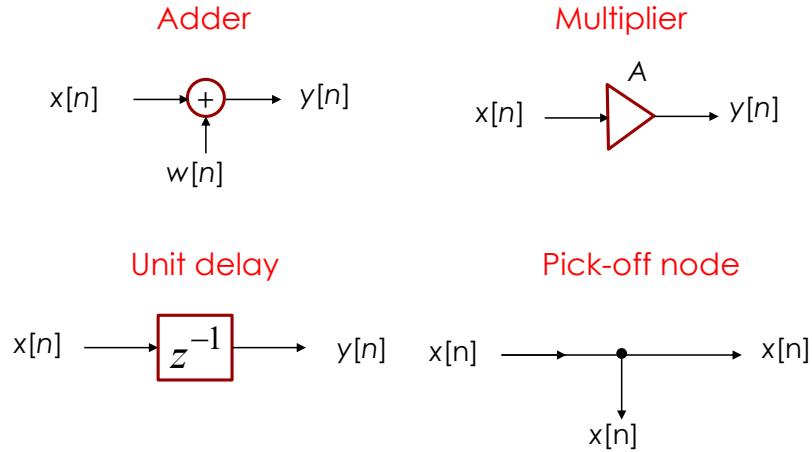


Figure 6.16: Basic building blocks of a block diagram.

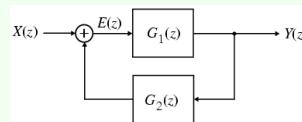
- easy to write down the computational algorithm by inspection;
- easy to analyze the block diagram to determine the explicit relation between output and input;
- easy to manipulate a block diagram to derive other “equivalent” block diagrams yielding different computational algorithms;
- easy to determine the hardware requirements;
- easier to develop block diagram representations from the transfer function directly.

Analysis block diagrams

The **analysis of block diagrams** is carried out by writing down the expressions for the output signals of each adder as a sum of its input signals, and developing a set of equations relating the filter input and output signals in terms of all internal signals. Eliminating the unwanted internal variables, then, results in the expression for the output signal as a function of the input signal and the filter parameters, that are the multiplier coefficients.

Example 68: Analysis of block diagrams

We consider the **single-loop feedback structure** showed below.



The output $E(z)$ of the adder is:

$$E(z) = X(z) + G_2(z)Y(z) \quad (6.200)$$

However, from the scheme we can see that:

$$Y(z) = G_1(z)E(z) \quad (6.201)$$

Eliminating $E(z)$ from Eqs. 6.200 and 6.201, we arrive at:

$$[1 - G_1(z)G_2(z)]Y(z) = G_1(z)X(z) \quad (6.202)$$

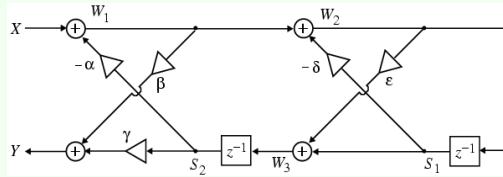
which leads to:

$$H(z) = \frac{Y(z)}{X(z)} = \frac{G_1(z)}{1 - G_1(z)G_2(z)} \quad (6.203)$$

Example with a single-loop feedback structure

Example 69: Analysis of block diagrams

We analyze the **cascaded lattice structure** showed in the scheme below, where the z -dependence of signal variables is not showed for brevity.



Example with a cascaded lattice structure

The output signals of the four adders are given by:

$$W_1 = X - \alpha S_2 \quad (6.204)$$

$$W_2 = W_1 - \delta S_1 \quad (6.205)$$

$$W_3 = S_1 + \varepsilon W_2 \quad (6.206)$$

$$Y = \beta W_1 + \gamma S_2 \quad (6.207)$$

Moreover, from the scheme we can observe:

$$S_2 = z^{-1} W_3 \quad (6.208)$$

$$S_1 = z^{-1} W_2 \quad (6.209)$$

Substituting the last two relations in the first four equations, we get:

$$W_1 = X - \alpha z^{-1} W_3 \quad (6.210)$$

$$W_2 = W_1 - \delta z^{-1} W_2 \quad (6.211)$$

$$W_3 = z^{-1} W_2 + \varepsilon W_2 \quad (6.212)$$

$$Y = \beta W_1 + \gamma z^{-1} W_3 \quad (6.213)$$

Therefore, we get:

$$W_2 = \frac{W_1}{1 + \delta z^{-1}} \quad (6.214)$$

$$W_3 = (\varepsilon + z^{-1}) W_2 \quad (6.215)$$

Combining the last two equations we get:

$$W_3 = \frac{\varepsilon + z^{-1}}{1 + \delta z^{-1}} W_1 \quad (6.216)$$

Substituting Eq. 6.216 the ones for W_1 and Y , we finally arrive at:

$$H(z) = \frac{Y}{X} = \frac{\beta + (\beta\delta + \gamma\varepsilon)z^{-1} + \gamma z^{-2}}{1 + (\delta + \alpha\varepsilon)z^{-1} + \alpha z^{-2}} \quad (6.217)$$

6.5.2 The delay-free loop problem

For the physical realizability of the digital filter structure, it is necessary that the block representation contains no delay-free loops. To illustrate the **delay-free loop problem**, we consider the structure in Figure 6.17.

Necessary condition for physical realizability of the structure

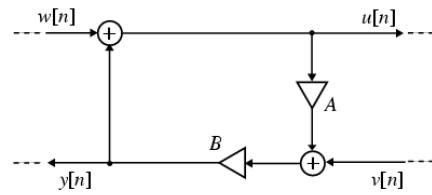


Figure 6.17: Example of structure with a delay loop.

From the analysis of this structure, we get:

$$u[n] = w[n] + y[n] \quad (6.218)$$

$$y[n] = B(v[n] + Au[n]) \quad (6.219)$$

which, when combined, results in:

$$y[n] = B(v[n] + A(w[n] + y[n])) \quad (6.220)$$

The determination of the current value of $y[n]$ requires the knowledge of the same value. However, this is physically impossible to achieve due to the finite time required to carry out all arithmetic operations on a digital machine.

Solving the delay-free loop problem

There exist methods to detect the presence of delay-free loops in an arbitrary structure, along with methods to locate and remove these loops without the overall input-output relation. One possibility of removal is achieved by replacing the portion of the overall structure containing the delay-free loops by an equivalent realization with no delay-free loops. To illustrate the process, let us consider again Eqs. 6.218 and 6.219, characterizing the structure in Figure 6.17. Substituting Eq. 6.219 into Eq. 6.218, we arrive at:

$$u[n] = w[n] + Bv[n] + ABu[n] \quad (6.221)$$

Solving Eq. 6.221, we get:

$$u[n] = \frac{1}{1 - AB}(w[n] + Bv[n]) \quad (6.222)$$

By reordering Eq. 6.218, we get:

$$y[n] = u[n] - w[n] \quad (6.223)$$

Therefore, we have obtained a delay-free loop realization based on Eqs. 6.222 and 6.223, and the scheme of the new structure is showed in Figure 6.18.

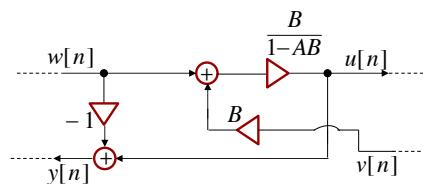


Figure 6.18: Structure of Figure 6.17 after removal of delay-free loops.

6.5.3 Canonic and non-canonic structures

Canonic and non-canonic structures

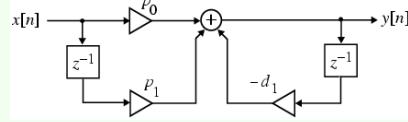
A digital filter structure is said to be **canonic** if the number of delays in the block diagram representation is equal to the order of the transfer function. Otherwise, it is a **non-canonic structure**.

Example 70: Canonic and non-canonic structures

The structure characterized by:

$$y[n] = -d_1 y[n-1] + p_0 x[n] + p_1 x[n-1] \quad (6.224)$$

and showed in the scheme below is non-canonic, as it employs two delays to realize a first-order difference equation.



6.5.4 Equivalent structures

Another important topic in digital filter design is the equivalence. Two digital filter structures are defined to be **equivalent** if they have the same transfer function. In the following discussion we describe several methods for the generation of equivalent structures. However, we state now that a fairly simple way to do it from a given realization is via the **transpose operation**, which consists in:

- reversing all the paths;
- replacing pick-off nodes by adders, and viceversa;
- interchanging the input and output nodes.

Let us consider an example of this technique. This is showed in Figure 6.19 along with the redrawn transposed structure.

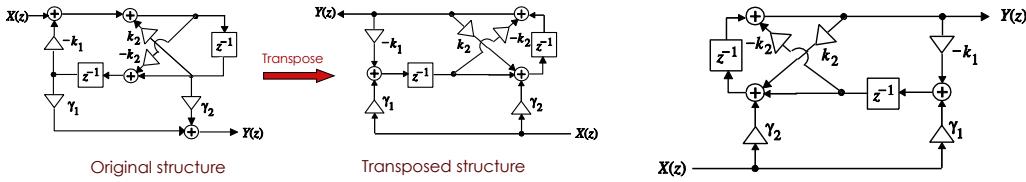


Figure 6.19: Example of the application of the transpose operation (left) and redrawn transposed structure (right).

We have to remark that all the other methods for developing equivalent structures are based on a specific algorithm for each structure. Moreover, there are literally an infinite number of equivalent structures realizing the same transfer function. Thus, it is impossible to develop all the equivalent realizations and in this course we will restrict our attention to a discussion of some commonly used structures.

Under infinite precision arithmetic, any given realization of a digital filter behaves identically to any other equivalent structure. However, in practice, due to the finite wordlength limitations, a specific realization behaves totally differently from its other equivalent realizations. Hence, it is important to choose a structure that has the least quantization effects when implemented. One way to arrive at such a structure is to determine a large number of equivalent structures, analyze the finite wordlength effects in each case, and select the one showing the least effects. In certain cases, it is possible to develop a structure that by construction has the least quantization effects. Here, we review some simple realizations that in many applications are quite adequate.

Lecture 22.
Thursday 17th
December, 2020.

Equivalent
structures

Transpose
operation

Equivalent
structures and
finite precision
arithmetics

6.6 Basic FIR digital filter structures

A **causal FIR filter of order N** is characterized by a transfer function $H(z)$ given by:

$$H(z) = \sum_{n=0}^N h[n]z^{-n} \quad (6.225)$$

which is a polynomial in z^{-1} . In the time-domain the input-output relation of the FIR filter in Eq. 6.225 is given by:

$$y[n] = \sum_{n=0}^N h[k]x[n-k] \quad (6.226)$$

6.6.1 Direct form FIR filter structures

It is important to remark that an FIR filter of order N is characterized by $N + 1$ coefficients and, in general, requires $N + 1$ multipliers and N two-input adders. Structures in which the multiplier coefficients are precisely the coefficients of the transfer function are called **direct form structures**.

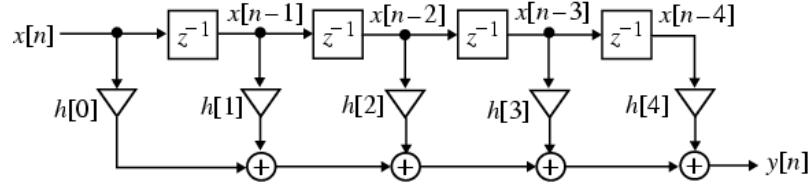


Figure 6.20: Example of FIR filter structure.

In particular, an analysis of the structure with $N = 4$, as showed in Figure 6.20, yields:

$$y[n] = h[0]x[n] + h[1]x[n - 1] + h[2]x[n - 2] + h[3]x[n - 3] + h[4]x[n - 4] \quad (6.227)$$

which is precisely of the form of the convolution sum description. The direct form structure in Figure 6.20 is also known as a **transversal filter**. The transpose of this direct form structure is showed in Figure 6.21. Both direct form structures are canonic with respect to the delays.

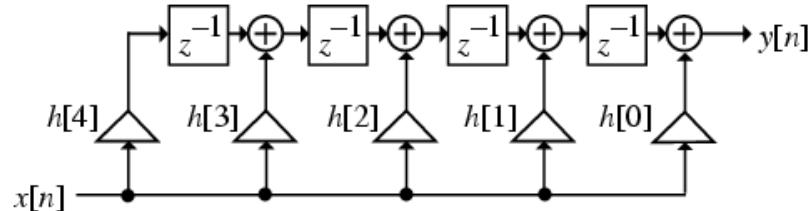


Figure 6.21: Transpose of the direct form structure in Figure 6.20.

6.6.2 Cascade form FIR filter structures

A higher-order FIR transfer function can also be realized as a **cascade** of second-order FIR sections and possibly a first-order section. To this end, we express $H(z)$ as:

$$H(z) = h[0] \prod_{i=1}^k (1 + \beta_{1,k}z^{-1} + \beta_{2,k}z^{-2}) \quad (6.228)$$

where $k = \frac{N}{2}$, if N is even, or $k = \frac{N+1}{2}$, if N is odd, with $\beta_{2k} = 0$. A cascade realization for $N = 6$ is showed in Figure 6.22. Each second-order section in the structure can also be realized in the transposed direct form.

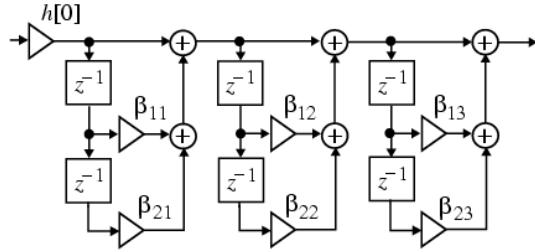


Figure 6.22: Example of cascade realization for $N = 6$ of an FIR filter.

6.6.3 Polyphase FIR structures

The **polyphase decomposition** of $H(z)$ leads to a parallel form structure. To illustrate this approach, let us consider a causal FIR transfer function $H(z)$ with $N = 8$:

$$H(z) = \sum_{\ell=0}^8 h[\ell]z^{-\ell} \quad (6.229)$$

$H(z)$ can be expressed as a sum of two terms, with one term containing the even-indexed coefficients and the other containing the odd-indexed ones:

$$\begin{aligned} H(z) &= \sum_{\ell=0}^4 h[2\ell]z^{-2\ell} + \sum_{\ell=1}^4 h[2\ell-1]z^{-(2\ell-1)} \\ &= \sum_{\ell=0}^4 h[2\ell]z^{-2\ell} + z^{-1} \sum_{\ell=0}^3 h[2\ell+1]z^{-2\ell} \end{aligned} \quad (6.230)$$

By using the notation:

$$E_0(z) = h[0] + h[2]z^{-1} + h[4]z^{-2} + h[6]z^{-3} + h[8]z^{-4} \quad (6.231)$$

$$E_1(z) = h[1] + h[3]z^{-1} + h[5]z^{-2} + h[7]z^{-3} \quad (6.232)$$

we can express $H(z)$ as:

$$H(z) = E_0(z^2) + z^{-1}E_1(z^2) \quad (6.233)$$

The decomposition in Eq. 6.233 is more commonly known as the **2-branch polyphase decomposition**. A realization of $H(z)$ based on the latter is thus showed in Figure 6.23.

Polyphase decomposition and FIR structures

2-branch polyphase decomposition

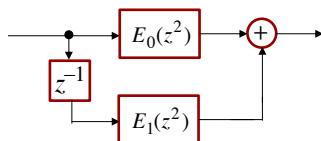


Figure 6.23: Example of structure of 2-branch polyphase decomposition.

In a similar manner, by grouping the terms in the original expression for $H(z)$, we can re-express it in the form:

$$H(z) = E_0(z^3) + z^{-1}E_1(z^3) + z^{-2}E_2(z^3) \quad (6.234)$$

where now:

$$E_0(z) = h[0] + h[3]z^{-1} + h[6]z^{-2} \quad (6.235)$$

$$E_1(z) = h[1] + h[4]z^{-1} + h[7]z^{-2} \quad (6.236)$$

$$E_2(z) = h[2] + h[5]z^{-1} + h[8]z^{-2} \quad (6.237)$$

3-branch polyphase decomposition

The decomposition of $H(z)$ in Eq. 6.234 is more commonly known as the **3-branch polyphase decomposition**, and a realization of $H(z)$ based on the latter is thus showed in Figure 6.24.

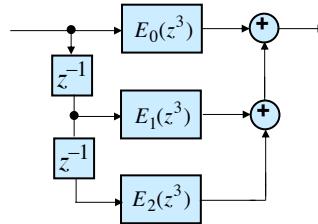


Figure 6.24: Example of structure of 3-branch polyphase decomposition.

General case of polyphase decomposition

In the general case, an **L-branch polyphase decomposition** of an FIR transfer function of order N is of the form:

$$H(z) = \sum_{m=0}^{L-1} z^{-m} E_m(z^L) \quad (6.238)$$

where:

$$E_m(z) = \sum_{n=0}^{\lfloor \frac{N+1}{L} \rfloor} h[Ln+m]z^{-m} \quad (6.239)$$

with $h[n] = 0$ for $n > N$.

We have to remark that the subfilters $E_m(z^L)$ in the polyphase realization of an FIR transfer function are also FIR filters and they can be realized using any methods described so far. However, to obtain a canonic realization of the overall structure, the delays in all the subfilters must be shared. For example, a canonic realization of a length-9 FIR transfer function is obtained using **delay sharing** is showed in Figure 6.25.

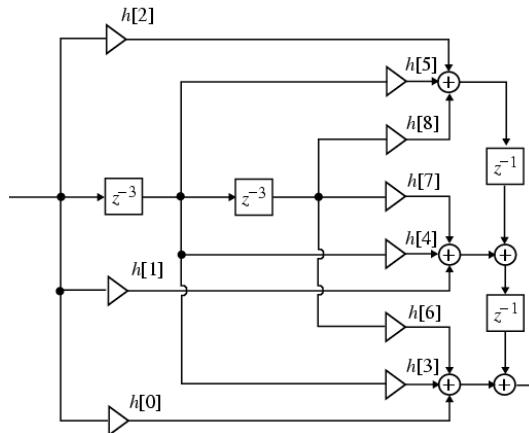


Figure 6.25: Example of the structure of a length-9 FIR transfer function with delay sharing.

6.6.4 Linear-phase FIR structures

The symmetry (or antisymmetry) property of a linear-phase FIR filter can be exploited to reduce the number of multipliers into almost half of that in the direct form implementations. For example, we consider a length-7 Type 1 FIR transfer function with a symmetric impulse response:

$$H(z) = h[0] + h[1]z^{-1} + h[2]z^{-2} + h[3]z^{-3} + h[2]z^{-4} + h[1]z^{-5} + h[0]z^{-6} \quad (6.240)$$

If we rewrite $H(z)$ in the form:

$$H(z) = h[0](1 + z^{-6}) + h[1](z^{-1} + z^{-5}) + h[2](z^{-2} + z^{-4}) + h[3]z^{-3} \quad (6.241)$$

we obtain the realization showed in Figure 6.26.

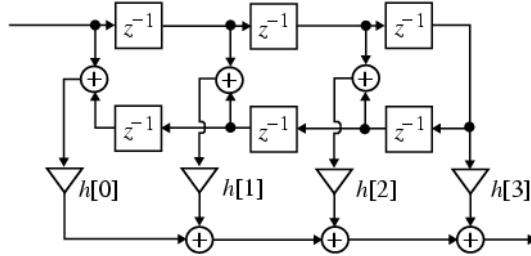


Figure 6.26: Example of the structure of a length-7 Type 1 FIR transfer function with a symmetric impulse response.

A similar decomposition can be applied to a Type 2 FIR transfer function. For example, let us consider a length-8 Type 2 FIR transfer function expressed as:

$$H(z) = h[0](1 + z^{-7}) + h[1](z^{-1} + z^{-6}) + h[2](z^{-2} + z^{-5}) + h[3](z^{-3} + z^{-4}) \quad (6.242)$$

The corresponding realization is showed in Figure 6.27.

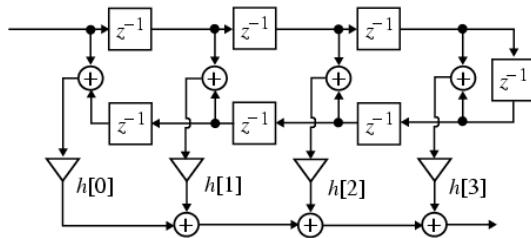


Figure 6.27: Example of the structure of a length-8 Type 2 FIR transfer function with a symmetric impulse response.

From the previous structures, we observe that:

- the Type 1 linear-phase structure for a length-7 FIR filter requires 4 multipliers, whereas a direct form realization requires 7 of them;
- the Type 2 linear-phase structure for a length-8 FIR filter requires 4 multipliers, whereas a direct form realization requires 8 of them.

Similar savings occur in the realization of Type 3 and 4 linear-phase FIR filters, with antisymmetric impulse responses.

Symmetry properties for structure simplification

Tapped delay line**6.6.5 Tapped delay line**

A **tapped delay line (TDL)** is a delay line with at least one “**tap**”. The ladder extracts a signal output from somewhere within the delay line, optionally scales it, and usually sums with other taps to form an output signal. A tap may be **interpolating** or **non-interpolating**. The ladder extracts the signal at some fixed integer delay relative to the input.

In this context, in some applications such as musical and sound processing, FIR filter structures of the form in Figure 6.28 are employed.

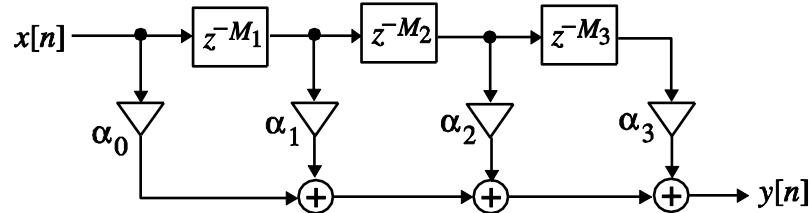


Figure 6.28: Example of structure of a tapped delay line.

the structure consists of a chain of $M_1 + M_2 + M_3$ unit delays with taps at the input, at the end of the first M_1 delays, at the end of the next M_2 delays, and at the output. Signals at these taps are then multiplied by constants α_0 , α_1 , α_2 and α_3 , then they are added to form the output. The direct form FIR structure in Figure 6.29 is seen to be a special case of the tapped delay line, where there is a tap after each unit delay.

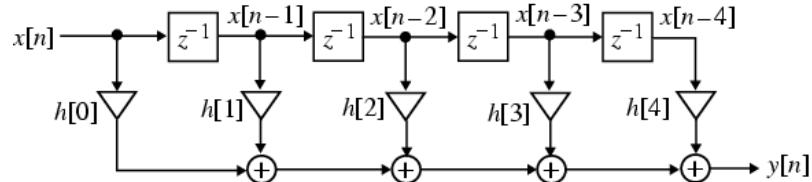


Figure 6.29: Special case of the tapped delay line in Figure 6.28, with a tap after each unit delay.

Lecture 23.
Tuesday 22nd
December, 2020.

*Basic IIR digital
filter structures*

*Issues of finite
precision of
coefficients*

Issues of overflow

6.7 Basic IIR digital filter structures

IIR filters allow both zeros and poles, while FIR allow zeros only. Therefore, IIR filters can be more **selective** for a given filter order. IIR are also called **recursive filters**, since their output depends on past inputs and past output samples. From this point we can understand that their design is not guaranteed to be stable and, in addition, they can be particularly sensitive to coefficient quantization.

The finite precision of the coefficients can lead to several issues:

- in order to be unconditionally stable and causal, all the system poles must be inside the unit circle ($|z| < 1$). The coefficient roundoff may inadvertently move a pole outside the unit circle;
- finite coefficient precision “quantizes” the pole locations. This may change the frequency response from the ideal case even if still stable.

These are not the only issues to care about. In fact, we can also have **overflow** issues:

- the gain from input to storage nodes in the filter may exceed unity. This can cause filter state to be saturated (clipped), resulting in distortion;

- typically, we must scale down, namely attenuate, the input signal, then scale up, namely amplify, by an equal amount on the output.

Now, the causal IIR digital filters we are concerned with in this course are characterized by a real rational transfer function of z^{-1} or, equivalently, by a constant coefficient difference equation. From the ladder representation, it can be seen that the realization of the causal IIR digital filters requires some form of **feedback**.

Another concept to keep in mind is that an N^{th} order IIR digital transfer function is characterized by $2N + 1$ unique coefficients, and in general, it requires $2N + 1$ multipliers and $2N$ two-input adders for the implementation.

6.7.1 Direct form IIR filter structures

The **direct form IIR filter structures** are such that the multiplier coefficients are precisely the coefficients of the transfer function. For example, let us consider for simplicity a 3rd-order IIR filter with a transfer function:

$$H(z) = \frac{P(z)}{D(z)} = \frac{p_0 + p_1 z^{-1} + p_2 z^{-2} + p_3 z^{-3}}{1 + d_1 z^{-1} + d_2 z^{-2} + d_3 z^{-3}} \quad (6.243)$$

We can implement $H(z)$ as a cascade of two filter sections, as showed in Figure 6.30, where:

$$H_1(z) = \frac{W(z)}{X(z)} = P(z) = p_0 + p_1 z^{-1} + p_2 z^{-2} + p_3 z^{-3} \quad (6.244)$$

$$H_2(z) = \frac{Y(z)}{W(z)} = \frac{1}{D(z)} = \frac{1}{1 + d_1 z^{-1} + d_2 z^{-2} + d_3 z^{-3}} \quad (6.245)$$

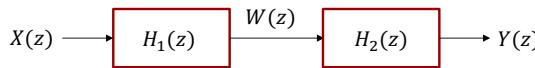


Figure 6.30: IIR filter implementation as a cascade of two filter sections.

The filter section $H_1(z)$ can be seen to be an FIR filter and it can be realized as showed in Figure 6.31 and characterized by:

$$w[n] = p_0 x[n] + p_1 x[n - 1] + p_2 x[n - 2] + p_3 x[n - 3] \quad (6.246)$$

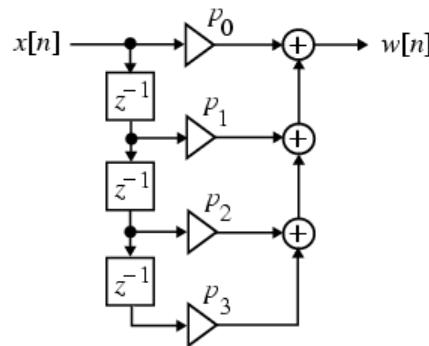


Figure 6.31: Filter section $H_1(z)$ of Figure 6.30.

On the other hand, the time domain representation of $H_2(z)$ is given by:

$$y[n] = w[n] - d_1 y[n - 1] - d_2 y[n - 2] - d_3 y[n - 3] \quad (6.247)$$

Direct form IIR filter structure

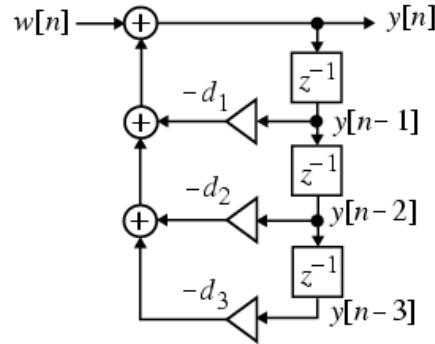


Figure 6.32: Filter section $H_2(z)$ of Figure 6.30.

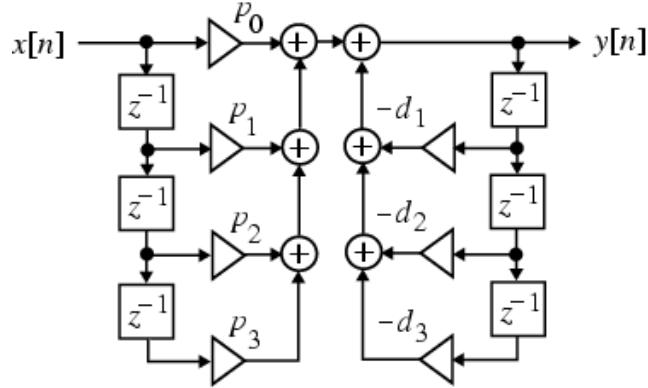


Figure 6.33: Cascade of the two structures realizing $H_1(z)$ and $H_2(z)$.

Direct form I structure

Direct form I_\dagger structure

The realization of $H_2(z)$ follows from Eq. 6.247 and it is showed in Figure 6.32.

A cascade of the two structures realizing $H_1(z)$ and $H_2(z)$ leads to the realization of $H(z)$, showed in Figure 6.33 and known as the **direct form I structure**.

Note that the direct form I structure is non-canonic as it employs 6 delays to realize a 3rd-order transfer function. A transpose of this type of form structure is showed in Figure 6.33 and it is called **direct form I_\dagger structure**.

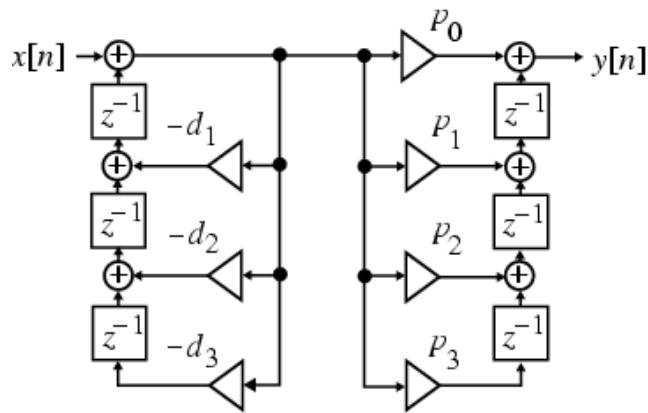


Figure 6.34: Transpose of the direct form I structure in Figure 6.33.

Direct form II and II_\dagger structures

Various other non-canonic direct form structures can be derived by simple block diagram manipulations, as showed in Figure 6.35.

Now, we consider the direct form structure in Figure 6.36. The signal variables at the nodes ① and ① are the same, and hence the two twop delays can be shared. Likewise, the signal variables at the nodes ② and ② are the same, permitting the sharing of the middle two delays. Following the same argument, the bottom two

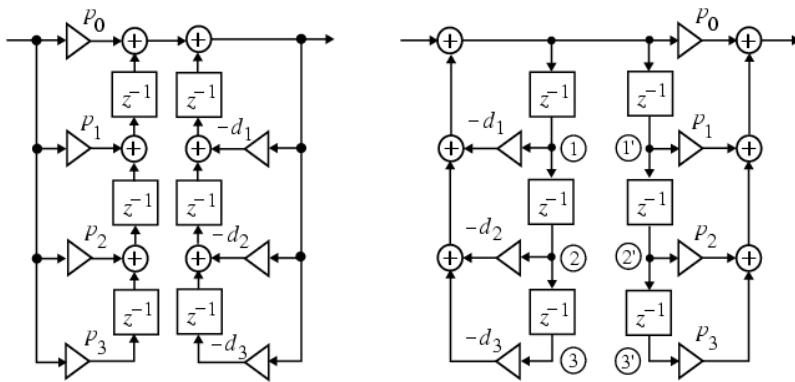


Figure 6.35: Other non-canonic direct form structures.

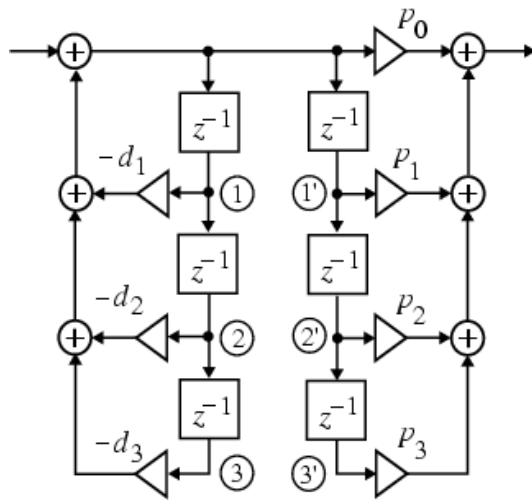
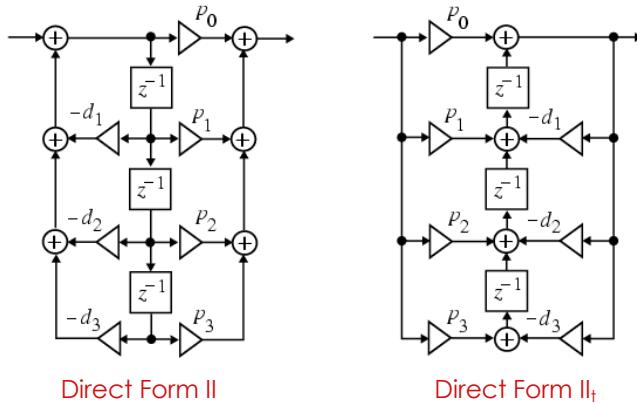


Figure 6.36: Example of direct form structure.

delays can be shared. The sharing of all the delays reduces the total number of delays to 3, resulting in a canonic realization, showed in Figure 6.37 along with its transpose structure. The direct form realizations of an N^{th} -order IIR transfer function should now be evident.

Figure 6.37: Example of direct form II and II_t structures.

6.7.2 Cascade form IIR filter structures

By expressing the numerator and the denominator polynomials of the transfer function as a product of polynomials of lower degree, a digital filter can be realized as a

Cascade form IIR filter structures

cascade of low-order filter sections.

We consider for example:

$$H(z) = \frac{P(z)}{D(z)} = \frac{P_1(z)P_2(z)P_3(z)}{D_1(z)D_2(z)D_3(z)} \quad (6.248)$$

Examples of cascade realizations of this filter, obtained by different pole-zero pairings and by different ordering of sections, are showed respectively in Figures 6.38 and 6.39.

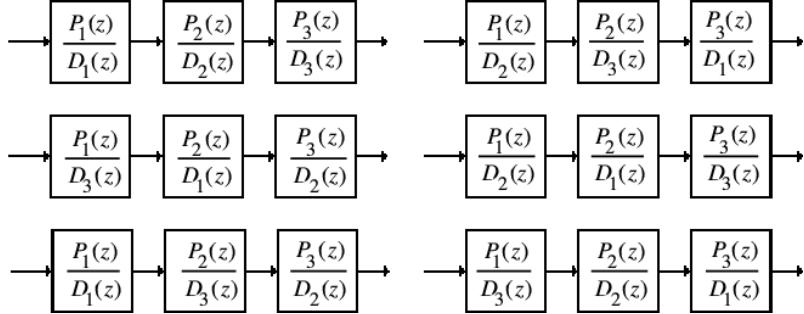


Figure 6.38: Cascade realizations of $H(z)$ obtained by different pole-zero pairings.

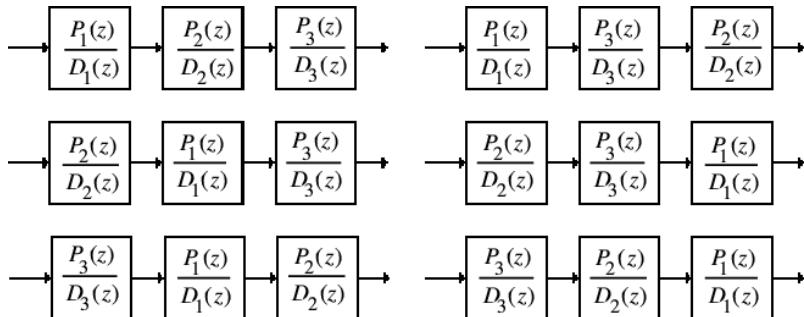


Figure 6.39: Cascade realizations of $H(z)$ obtained by different ordering of sections.

There are altogether a total of 36 different cascade realizations of Eq. 6.248, based on pole-zero pairings and ordering. Due to the finite wordlength effects, each of them behaves differently from the others.

Usually, the polynomials are factored into a product of 1st-order and 2nd-order polynomials. In this case, $H(z)$ is expressed as:

$$H(z) = p_0 \prod_k \left(\frac{1 + \beta_{1,k}z^{-1} + \beta_{2,k}z^{-2}}{1 + \alpha_{1,k}z^{-1} + \alpha_{2,k}z^{-2}} \right) \quad (6.249)$$

In Eq. 6.249, for a first-order factor, we have $\alpha_{2,k} = \beta_{2,k} = 0$.

If we take as example the 3rd-order transfer function:

$$H(z) = p_0 \left(\frac{1 + \beta_{1,1}z^{-1}}{1 + \alpha_{1,1}z^{-1}} \right) \cdot \left(\frac{1 + \beta_{1,2}z^{-1} + \beta_{2,2}z^{-2}}{1 + \alpha_{1,2}z^{-1} + \alpha_{2,2}z^{-2}} \right) \quad (6.250)$$

One possible realization is showed in Figure 6.40.

Let us consider another example. We compare the direct form II and the cascade form realizations of:

$$H(z) = \frac{0.44z^{-1} + 0.362z^{-2} + 0.02z^{-3}}{1 + 0.4z^{-1} + 0 - 18z^{-2} - 0.2z^{-3}} \quad (6.251)$$

By factorizing the numerator and the denominator, we obtain:

$$H(z) = \left(\frac{0.44 + 0.362z^{-1} + 0.02z^{-2}}{1 + 0.8z^{-1} + 0.5z^{-2}} \right) \cdot \left(\frac{z^{-1}}{1 - 0.4z^{-1}} \right) \quad (6.252)$$

The diagrams of both realizations are showed in Figure 6.41.

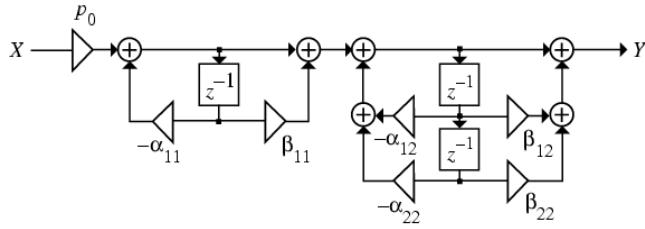
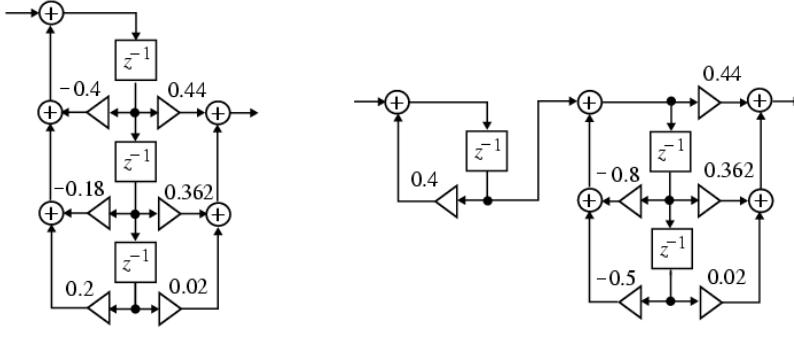


Figure 6.40: Possible realization of Eq. 6.250.



Direct form II

Cascade form

Figure 6.41: Diagrams of direct form II (left) and cascade form (right) realizations of Eq. 6.251.

6.7.3 Parallel form IIR filter structures

A partial-fraction expansion of the transfer function in z^{-1} leads to the **parallel form I structure**. Assuming simple poles, the transfer function $H(z)$ can be expressed as:

$$H(z) = \gamma_0 + \sum_k \left(\frac{\gamma_{0,k} + \gamma_{1,k}z^{-1}}{1 + \alpha_{1,k}z^{-1} + \alpha_{2,k}z^{-2}} \right) \quad (6.253)$$

In Eq. 6.253, for a real pole we have $\alpha_{2,k} = \gamma_{1,k} = 0$.

A direct partial-fraction expansion of the transfer function z leads to the **parallel form II structure**. Assuming simple poles, the transfer function $H(z)$ can be expressed as:

$$H(z) = \delta_0 + \sum_k \left(\frac{\delta_{1,k}z^{-1} + \delta_{2,k}z^{-2}}{1 + \alpha_{1,k}z^{-1} + \alpha_{2,k}z^{-2}} \right) \quad (6.254)$$

In Eq. 6.254, for a real pole we have $\alpha_{2,k} = \delta_{2,k} = 0$.

The two basic parallel realizations of a 3rd-order IIR transfer function are showed in Figure 6.42.

Now, let us consider the following example. A partial-fraction expansion in z^{-1} of Eq. 6.251 yields:

$$H(z) = -0.1 + \frac{0.6}{1 - 0.4z^{-1}} + \frac{-0.5 - 0.2z^{-1}}{1 + 0.8z^{-1} + 0.5z^{-2}} \quad (6.255)$$

The corresponding parallel form I realization is showed in Figure 6.43. Likewise, a partial fraction expansion of $H(z)$ in z yields:

$$H(z) = \frac{0.24z^{-1}}{1 - 0.4z^{-1}} + \frac{0.2z^{-1} + 0.25z^{-2}}{1 + 0.8z^{-1} + 0.5z^{-2}} \quad (6.256)$$

and the corresponding parallel form II realization is showed in Figure 6.43 as well.

Parallel form I structure

Parallel form II structure

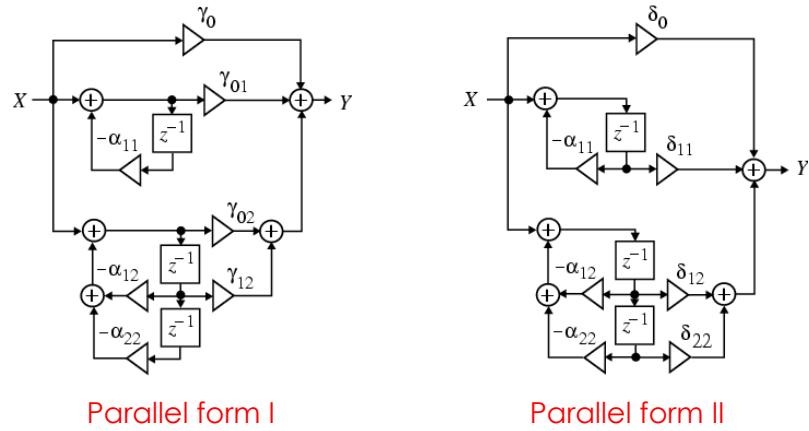


Figure 6.42: Diagrams of parallel form I (left) and form II (right) realizations of a 3rd-order IIR transfer function.

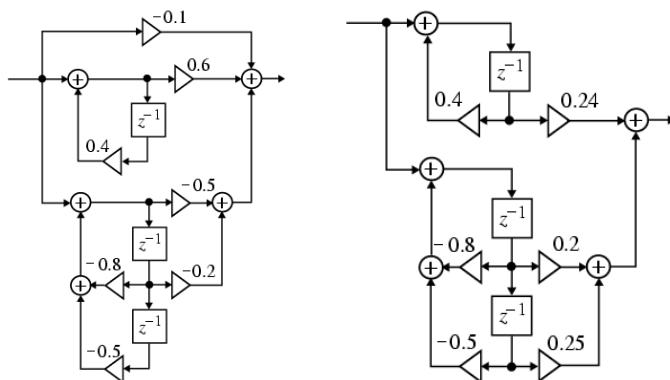


Figure 6.43: Parallel form I (left) and form II (right) realizations of Eq. 6.251.

Bibliography

- [1] *Signal Processing for Communications*, P. Prandoni and M. Vetterli, https://www.sp4comm.org/docs/sp4comm_corrected.pdf
- [2] *Digital Signal Processing*, A. V. Oppenheim and R. W. Schafer, Prentice Hall
- [3] *Digital Signal Processing: Digital Signal Processing: A Computer-Based Approach*, S. K. Mitra