1. Fundamentals

1.1 Preamble

Living in a Sampled World

- **Sampling Theorem**

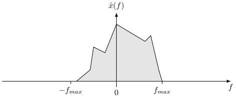
  $x(t)$ analog signal with maximum frequency f_{max}.
  $x[n] = x(nT_s)$ samples obtained from $x(t)$ by sampling at frequency $f_s = 1/T_s$.

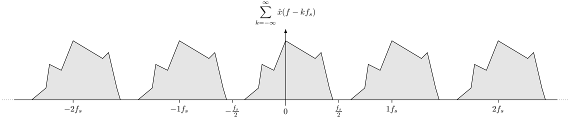
  The "must" condition in order to have $x[n]$ carrying the same information as $x(t)$ is:

  $$f_s > 2 f_{\text{max}} \, .$$

  If f_s is given and $f_s < 2 f_{\text{max}}$, then $x(t)$ needs to be low passed, so that the maximum frequency $f_{\text{LP}_{\text{max}}}$ of the low passed analog signal $x_{\text{LP}}(t)$ satisfies $f_s > 2 f_{\text{LP}_{\text{max}}}$.
  Better lose the saddle than the horse!

- **Spectrum and Replicas**

<table>
<tr><th>signal</th><th>F.T.</th><th>spectrum</th></tr>
<tr><td>$x(t)$</td><td>$\hat{x}(f)$</td><td>


</td></tr>
<tr><td>$x[n]$</td><td>$\sum_{k=-\infty}^{\infty} \hat{x}(f - k f_s)$</td><td>


</td></tr>
</table>

Living in a Sampled World

- **Time & Frequency Domains**

  This course deals (mostly) with **discrete time signals** or, more generally, **countable data**

  $$x[0], x[1], \ldots, x[N-1]\,,$$

  where, in practical scenarios, N is finite.

  - Time domain samples are commonly indexed with integers;
  - Frequency domain samples are indexed with integers, *i.e.*, **sample indexes**, or expressed in **normalized frequencies** or in **radians**.

Living in a Sampled World

- **Sample Indexes in the Frequency Domain**

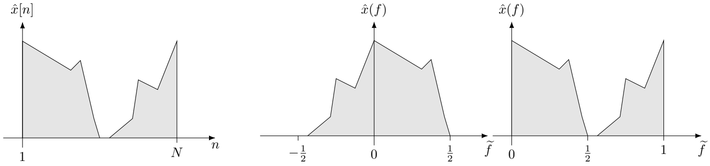
  When dealing with a finite sequence of samples $x[0], \ldots, x[N-1]$, the transformation into the frequency domain leads to yet another sequence of samples $\hat{x}[0], \ldots, \hat{x}[N-1]$. Recalling that the spectrum of a sampled signal is periodic, the samples $\hat{x}[0], \ldots, \hat{x}[N-1]$ represent the period on the positive frequencies

- **Normalized Frequencies**

  The spectrum of a samples signal $x[n]$ is f_s-periodic. The period (where all the information is contained) can be represented in the interval $(-\frac{f_s}{2}, \frac{f_s}{2}]$ or equivalently $[0, f_s)$.

  Frequencies can be normalized with respect to the sampling frequency f_s, therefore obtaining a spectrum period in the interval $(-\frac{1}{2}, \frac{1}{2}]$ or equivalently $[0, 1)$.

  When working with samples, the sampling frequency information is not necessarily available!

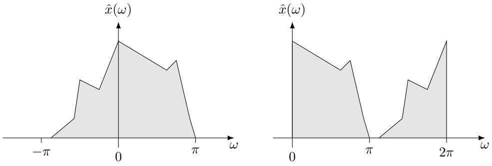


Living in a Sampled World

- **Radians**

  The quantity $\omega = 2\pi f$ is defined as angular frequency or radians per unit of time. In a discrete world, it also can be normalized.

  With an abuse of notation, the normalized radians per unit of time are defined as $\omega = 2\pi \frac{f}{f_s}$, and the correspond to radians $\in [0, 2\pi)$.



1.2 Linear Time Invariant Systems

Linear Time Invariant Systems

- **Impulse Response** $h[n]$

Output of a system when the input is the discrete time impulse $\delta[n] = \begin{cases} 1 & \text{if } n = 0 \\ 0 & \text{if } n \neq 0 \end{cases}$

- **Linear Time Invariant System - LTI**

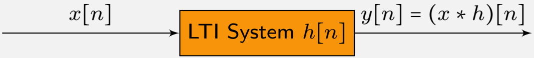
$$x_1[n] \to y_1[n]\,, \quad x_2[n] \to y_2[n] \Rightarrow \alpha x_1[n] + \beta x_2[n] \to \alpha y_1[n] + \beta y_2[n]\,,$$

and

$$x[n] \to y[n] \Rightarrow x[n-m] \to y[n-m]\,.$$

Property

The impulse response h of a linear time invariant system gives a complete description of the input-output relation of the system, and $y[n] = (x * h)[n] = \sum_{k \in \mathbb{Z}} x[n-k]h[k]$.



Linear Time Invariant Systems

- **Proof of the statement** $y[n] = (x * h)[n]$

  To prove that the output $y[n]$ to an input $x[n]$ of an LTI system with impulse response $h[n]$ is given by $y[n] = (x * h)[n]$ we proceed as follows

  - Write the input $x[n]$ as
  $$x[n] = \sum_{k=-\infty}^{+\infty} x[k]\delta[n-k]\,.$$
  - By linearity, the system output to the above sum $\sum_{k=-\infty}^{+\infty} x[k]\delta[n-k]$ is given by the sum of the system outputs to each individual terms $x[k]\delta[n-k]$.
  - By linearity and time invariance, the system output to each individual term $x[k]\delta[n-k]$ of the sum is given by $x[n]h[n-k]$.
  - Finally, combining the previous two statements, the system output $y[n]$ to the input $x[n]$ is given by
  $$y[n] = \sum_{k=-\infty}^{+\infty} x[n]h[n-k] = (x * h)[n]\,.$$

Linear Time Invariant Systems

- **Stability Bounded-Input Bounded-Output - BIBO**

  A system is BIBO stable if a bounded input signal produces a bounded output signal.

Property

A LTI system is BIBO stable if and only if its impulse response h is summable.

$$\text{LTI system BIBO stable} \iff \sum_{n=-\infty}^{\infty} |h[n]| < \infty.$$

- **Causal Systems**

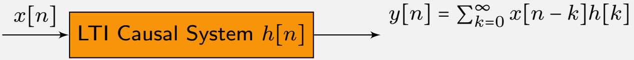
  No output before an input stimulus is received

Property

For a LTI system

$$\text{Causality} \iff h[n] = 0, \quad \forall n < 0\,.$$

Therefore $y[n] = (x * h)[n] = \sum_{k=0}^{\infty} x[n-k]h[k]$.



Linear Time Invariant Systems

- **Norms**

  Used to "measure" distances/difference between functions or in general non singleton elements (data samples, impulse responses)

  - ℓ_1-**norm** $\|h[n]\|_1 = \sum_{n\in\mathbb{Z}} |h[n]|$ (L_1-norm $\|f(t)\|_1 = \int_{\mathbb{R}} |f(t)|dt$)
  - ℓ_2-**norm** $\|h[n]\|_2 = \left(\sum_{n\in\mathbb{Z}} |h[n]|^2\right)^{1/2}$ (L_2-norm $\|f(t)\|_2 = \int_{\mathbb{R}} |f(t)|^2 dt$)
  - ℓ_∞-**norm** $\|h[n]\|_\infty = \max_{n\in\mathbb{Z}} |h[n]|$

- **Norms & Stability**

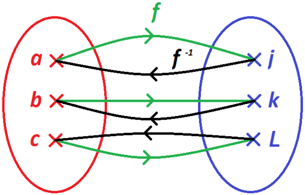
  BIBO stability can be rewritten as

$$\|x\|_\infty < \infty \Rightarrow \|y\|_\infty < \infty \iff \|h\|_1 < \infty.$$

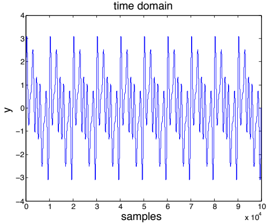
1.3 Transforms

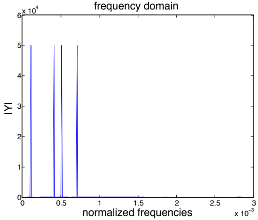
Transforms

- **Motivation**
  - Move to a space where certain computations are much easier to perform



  - Get another view of the data





Transforms

- **The Fourier Series - FS**
  - Fourier Coefficients

    $f(t)$ T-periodic function with $\int_0^T |f(t)|dt < \infty$ or $\int_0^T |f(t)|^2 dt < \infty$. The coefficients

    $$C_n = \frac{1}{T}\int_0^T f(t)e^{-i2\pi \frac{n}{T} t} dt \quad n \in \mathbb{Z},$$

    are called the **Fourier coefficients** of f, also noted as $C_n(f)$.

  - Fourier Series Expansion

    Suppose $\sum_{n\in\mathbb{Z}} |C_n(f)| < \infty$ or $\sum_{n\in\mathbb{Z}} |C_n(f)|^2 < \infty$. Then

    $$f(t) = \sum_{n\in\mathbb{Z}} C_n(f)e^{i2\pi \frac{n}{T} t}.$$

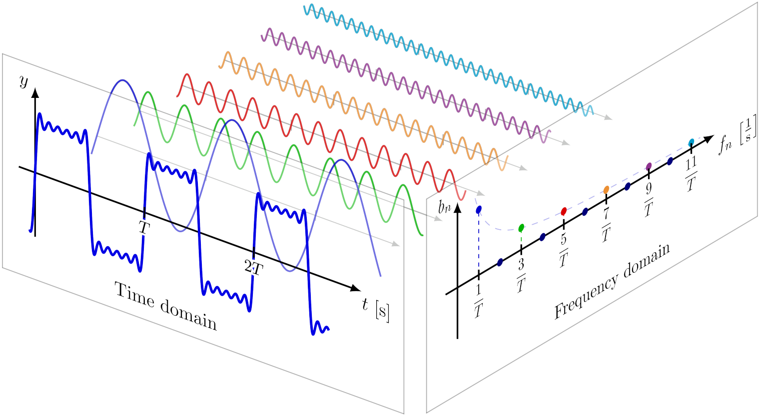
    The above expression is called **Fourier series expansion** of f.

Remark
Notice that if $\int_0^T |f(t)|^2 dt < \infty$ then $\sum_{n\in\mathbb{Z}} |C_n(f)|^2 < \infty$, and viceversa.

The same does not hold for $\int_0^T |f(t)| dt < \infty$ or $\sum_{n\in\mathbb{Z}} |C_n(f)| < \infty$.

Transforms

- **The Fourier Series - FS: Interpretation**



Transforms

- **The Fourier transform - FT**
  - Fourier Transform

    $f(t)$ with $\|f(t)\|_1 < +\infty$ or $\|f(t)\|_2 < +\infty$ (otherwise FT **not** defined)

$$\widehat{f}(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} \, dt$$

  - Inverse Fourier Transform

    $\widehat{f}(\omega)$ with $\|\widehat{f}(\omega)\|_1 < +\infty$ or $\|\widehat{f}(\omega)\|_2 < +\infty$ (otherwise iFT **not** defined)

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{f}(\omega) e^{j\omega t} \, d\omega$$

Remark

Notice that if $\|f(t)\|_2 < +\infty$ then $\|\widehat{f}(\omega)\|_2 < +\infty$, and viceversa.

The same does not hold for $\|f(t)\|_1 < +\infty$ or $\|\widehat{f}(\omega)\|_1 < +\infty$.

Transforms

- **Discrete Time Fourier Transform - DTFT**
  - Discrete time Fourier transform

    $x[n]$ with $\|x[n]\|_1 < +\infty$ or $\|x[n]\|_2 < +\infty$

    $$\widehat{x}(\omega) = \sum_{n \in \mathbb{Z}} x[n] e^{-j\omega n} \quad \omega \in [0, 2\pi]\,.$$

> Remark
>
> If the DTFT of a certain sequence exists, then it is a 2π-periodic function. This suggests that we can define a series expansion for a periodic function $\widehat{x}(e^{j\omega})$.

  - Inverse discrete time Fourier Transform

    $\widehat{x}(e^{j\omega})$ with $\|\widehat{x}(\omega)\|_1 < +\infty$ or $\|\widehat{x}(\omega)\|_2 < +\infty$

    $$x[n] = \frac{1}{2\pi} \int_0^{2\pi} \widehat{x}(\omega) e^{j\omega n} d\omega\,.$$

Transforms

- **Discrete Time Fourier Series - DTFS (aka Discret Fourier Transform - DFT)**

  N samples $x[1], x[2], \dots, x[N]$, with $\|x[n]\|_1 < +\infty$

  - Transform: $\hat{x}[k] = \sum_{n=1}^{N} x[n] e^{-j2\pi \frac{(n-1)(k-1)}{N}}, \quad k = 1, \dots, N;$
  - Inverse Transform: $x[n] = \frac{1}{N} \sum_{k=1}^{N} \hat{x}[k] e^{j2\pi \frac{(n-1)(k-1)}{N}}, \quad n = 1, \dots, N.$

Transforms

- **Discrete Fourier Transform - Meaning**

The formula of the inverse DFT $x[n] = \frac{1}{N}\sum_{k=1}^{N}\hat{x}[k]e^{j2\pi\frac{(n-1)(k-1)}{N}}$, $n = 1 \ldots N$, shows us that the signal can be written as the **weighted** sum of **complex sinusoids**.

– The complex sinusoids are $e^{j2\pi\frac{(n-1)(k-1)}{N}}$, $\quad n,k = 1,\ldots,N$;

– The weights are $\hat{x}[k] = \sum_{n=1}^{N} x[n]e^{-j2\pi\frac{(n-1)(k-1)}{N}}$, $\quad k = 1,\ldots,N$.

That is

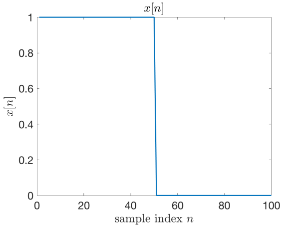
$$x[n] = \frac{1}{N}\left(\hat{x}[1] + \hat{x}[2]e^{j2\pi\frac{(n-1)}{N}} + \hat{x}[3]e^{j2\pi\frac{(n-1)2}{N}} + \ldots + \hat{x}[N]e^{j2\pi\frac{(n-1)(N-1)}{N}}\right).$$

Transforms

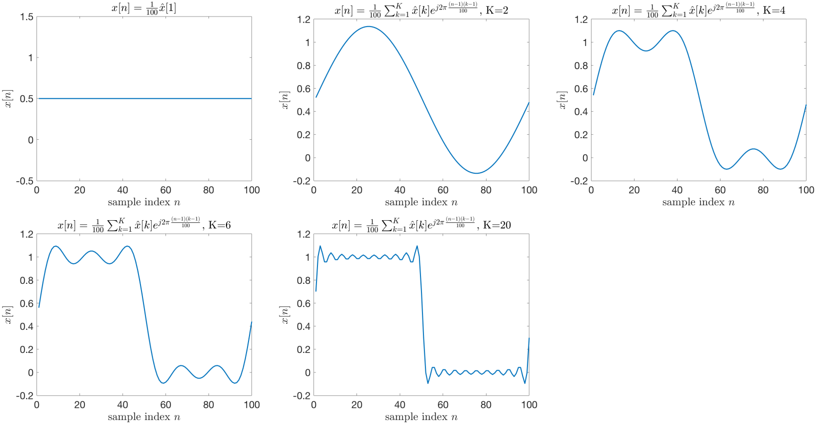
- **Discrete Fourier Transform - Meaning**

  $x[n]$, $n = 1, \ldots, 100$, with decomposition as sinusoids:

  $$x[n] = \frac{1}{100}\sum_{k=1}^{100} \hat{x}[k] e^{j2\pi\frac{(n-1)(k-1)}{100}}, \ n = 1 \ldots 100.$$



  Partial sum of sinusoids:



Transforms

Discrete Fourier Transform - Matrix Form

N samples $x[1], x[2], \dots, x[N]$

- Transform: $\hat{x}[k] = \sum_{n=1}^{N} x[n] e^{-j2\pi \frac{(n-1)(k-1)}{N}}$, $k = 1, \dots, N$;
- Inverse Transform: $x[n] = \frac{1}{N} \sum_{k=1}^{N} \hat{x}[k] e^{j2\pi \frac{(n-1)(k-1)}{N}}$, $n = 1, \dots, N$.

Let $W_N = e^{-j2\pi \frac{1}{N}}$ and define the matrix $\mathbf{W}_N = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & W_N & \dots & W_N^{N-1} \\ \vdots & \vdots & & \vdots \\ 1 & W_N^{N-1} & \dots & W_N^{(N-1)(N-1)} \end{bmatrix}$,

- Transform:
  $\hat{x}[k] = \sum_{n=1}^{N} x[n] W_N^{(n-1)(k-1)}$, $k = 1, \dots, N$, and $\begin{bmatrix} \hat{x}[1] \\ \vdots \\ \hat{x}[N] \end{bmatrix} = \mathbf{W} \begin{bmatrix} x[1] \\ \vdots \\ x[N] \end{bmatrix}$
- Inverse Transform:
  $x[n] = \frac{1}{N} \sum_{k=1}^{N} \hat{x}[k] W_N^{-(n-1)(k-1)}$, $n = 1, \dots, N$, and $\begin{bmatrix} x[1] \\ \vdots \\ x[N] \end{bmatrix} = \frac{1}{N} \mathbf{W}^* \begin{bmatrix} \hat{x}[1] \\ \vdots \\ \hat{x}[N] \end{bmatrix}$.

Transforms

- **Discrete Fourier Transform & Discrete Time Fourier Transform**

  Consider a signal $x[n]$ with finite number of samples, *i.e.*, $N < \infty$ samples $x[1], x[2], \ldots, x[N]$. The corresponding DTFT reads

$$\widehat{x}(\omega) = \sum_{n=1}^{N} x[n] e^{-j\omega(n-1)} \,.$$

> Remark
> Notice that the index of the exponential is $(n-1)$ since we consider $x[1]$ as the initial sample (1 is therefore the "origin" of the indexes).

When the number of samples are finite, the DFT can be easily obtained by discretizing the DTFT. Recalling that $\omega \in [0, 2\pi)$, by setting $\omega = \frac{2\pi(k-1)}{N}$ for $k = 1, \ldots, N$, we obtain the DTF.

Transforms

- **Fourier Transformation's Properties: Parseval**

  The Fourier transform conserves the square norm.

  - Continuous time Fourier Transform (FT)

$$\int_{-\infty}^{+\infty} f^*(t)g(t)\,dt = \frac{1}{2\pi}\int_{-\infty}^{+\infty} F^*(\omega)G(\omega)\,d\omega$$

  - Discrete time Fourier Transform (DTFT)

$$\sum_{n=-\infty}^{\infty} x^*[n]y[n] = \frac{1}{2\pi}\int_{0}^{2\pi} \widehat{x}^*(\omega)\widehat{y}(\omega)\,d\omega$$

  - Discrete time Fourier Series (DTFS)

$$\sum_{n=0}^{N-1} x^*[n]y[n] = \frac{1}{N}\sum_{n=0}^{N-1} \widehat{x}^*[n]\widehat{y}[n].$$

Remark

$1/2\pi$ and $1/N$ come from the definition of Fourier transform we have used. Other definitions exist, where a normalization is introduced to remove these constants.

Transforms

- **Fourier Transformation's Properties: Duality**

support (time)	$\Leftrightarrow$	transform support (frequency)
finite	$\Leftrightarrow$	infinite
infinite	$\Leftrightarrow$	finite
discrete	$\Leftrightarrow$	periodic
periodic	$\Leftrightarrow$	discrete
periodic discrete	$\Leftrightarrow$	periodic discrete

> **Remark**
> When using the DFT both the discrete time signal and the corresponding discrete frequency transform are considered periodic.

- **On the Domain of the Fourier Transformation**

  We commonly consider Fourier transformations to be applications from the time domain to the frequency domain. In such a framework, we have a clear interpretation of the transformation. Nevertheless, from a mathematical point of view, as far as the condition of existence are satisfied, Fourier transformations can be applied to any space.

Transforms

- **z-Transform**

  $h[n]$, $n \in \mathbb{Z}$ not necessarily absolutely summable. Then, the two-sided (or bilateral) z-transform is defined as

$$H(z) = \sum_{n=-\infty}^{+\infty} h[n] z^{-n} \qquad z \in \mathbb{C}.$$

- **z-Transform: Region of convergence**

  The z transform exists if the sum converges

$$\sum_{n=-\infty}^{+\infty} |h[n] r^{-n}| < +\infty \,,$$

  The **region of convergence** is defined by

$$\text{ROC}_f \triangleq \left\{ z : \sum_{n=-\infty}^{+\infty} \left| h[n] z^{-n} \right| < \infty \right\}.$$

  Notice that, by expressing the z variable with magnitude and phase $z = re^{j\omega}$, the region of convergence depends only on the magnitude r of the z variable.

Transforms

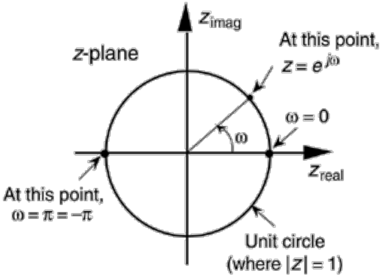
- **z-Transform & DTFT**

  The z-Transform is the extension from the unitary circle to the complex plane of the DTFT. Indeed $z = e^{j2\pi \frac{n}{N}}$ is the restriction of a complex variable to the circle of radius $\left|e^{j2\pi \frac{n}{N}}\right| = 1$.

  By expressing the z variable with magnitude and phase $z = re^{j\omega}$ then z is on a circle of radius r and

  $$H(re^{j\omega}) = \sum_{n=-\infty}^{+\infty} h[n] r^{-n} e^{-j\omega n}.$$

  For $r = 1$, *i.e.*, $z = e^{j\omega}$, we obtain the DTFT of $h[n]$.



1.4 LTI Causal Systems & Transforms

LTI Causal Systems & Transforms

- **The Perfect Analysis & Design Tool**

  The z-Transform and the Fourier Transforms (in particular DTFT & DFT) are the ideal tools to analyze and design systems.

  The Fourier Transform (DTFT) of the impulse response $h[n]$ is called **Transfer Function** of the system, and it enables to

  - Define the effect of the system on the signal, and in particular define the system as a filter (low pass, band pass, high pass);
  - Characterize particular systems.

  The z-Transform of the impulse response $h[n]$ enables to

  - Characterize the system in terms of poles and zeros;
  - Analyze its stability;
  - Write the system input-output relation as a finite difference equation and characterize its realizability.

LTI Causal Systems & Transforms

- **Fourier Transform: Eigenfunctions of LTI Systems**

  Consider the output of a LTI system $h[n]$

  $$y[n] = \sum_k h[k]x[n-k] = \sum_k h[n-k]x[k]\,.$$

  When $x[n] = e^{j\omega_0 n}$

  $$y[n] = \sum_k h[k]e^{j\omega_0(n-k)} = e^{j\omega_0 n}\sum_k h[k]e^{-j\omega_0 k} = e^{j\omega_0 n}H(e^{j\omega_0}).$$

  The output corresponding to a complex exponential is a complex exponential.

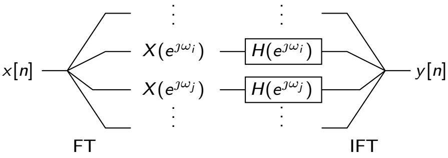
  By analogy with eigenvectors and eigenvalues of a linear function, we say that the **complex exponentials are eigenfunctions of linear time invariant system**.

LTI Causal Systems & Transforms

- **Fourier Transform: Eigenfunctions of LTI Systems**

$$y[n] = \sum_k h[k]e^{j\omega_0(n-k)} = e^{j\omega_0 n}\sum_k h[k]e^{-j\omega_0 k} = e^{j\omega_0 n}H(e^{j\omega_0}).$$

$H(e^{j\omega_0})$ represents the eigenvalue of the system and it is the Fourier transform of $h[n]$ evaluated for $\omega = \omega_0$. By linearity we can decompose the input and output signals as a sum of exponentials.

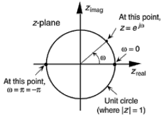


$$Y(e^{j\omega_0}) = H(e^{j\omega_0})X(e^{j\omega_0})$$

Fourier decomposition of LTI filtering (convolution) decomposes the input into eigenfunctions $e^{j\omega}$, which are weighted by the eigenvalues $H(e^{j\omega})$.

The DTFT of the

LTI Causal Systems & Transforms



- **z-transform: Fractional $H(z)$ & Realizable Systems**

  A LTI causal system $h[n]$ is realizable if it can be implemented using a finite number of operations per sample.

  A priori, if the impulse response $h[n]$ has an infinite number of values $h[0], h[1], \ldots$, the system is not realizable. Nevertheless, if the input-output equation of the system can be written as a linear constant-coefficient difference equation,

$$y[n] + \sum_{k=1}^{M} a_k y[n-k] = \sum_{k=0}^{N} b_k x[n-k] , \quad N, M < \infty ,$$

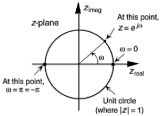
  the system is realizable. This is equivalent of saying that the z-transform has a **rational form**

$$H(z) = \frac{B(z)}{A(z)} = \frac{\sum_{k=0}^{N} b_k z^{-k}}{1 + \sum_{k=1}^{M} a_k z^{-k}} , \quad N, M < \infty .$$

  Notice that when the impulse response $h[n]$ has a finite number of values $h[0], h[1], \ldots, h[N]$ ($h[k] = 0 \ \forall k > N$), we have $A(z) = 1$.

  A realizable system (*i.e.*, with rational z-transform as above) with $h[n]$ having an infinite number of values is called a **Infinite Impulse Response - IIR** system, while in the case of a finite number of values ($A(z) = 1$), it is called a **Finite Impulse Response - FIR** system.

LTI Causal Systems & Transforms



- **z-transform: Fractional $H(z)$, Realizable Systems, Poles & Zeros**

  Consider a fractional $H(z)$ corresponding to a realizable system (numerator and denominator are finite polynomials in z^{-1}).

$$H(z) = \frac{B(z)}{A(z)} = \frac{\sum_{k=0}^{N} b_k z^{-k}}{1 + \sum_{k=1}^{M} a_k z^{-k}}, \quad N, M < \infty .$$

  Let's write the numerator and denominator as polynomials in z

$$H(z) = \frac{b_0 + b_1 z^{-1} + \ldots + b_N z^{-N}}{1 + a_1 z^{-1} + \ldots + a_M z^{-M}} = \frac{z^{-N}\left(b_0 z^N + b_1 z^{N-1} + \ldots + b_N\right)}{z^{-M}\left(z^M + a_1 z^{M-1} + \ldots + a_M\right)}$$

  If $N > M$, then

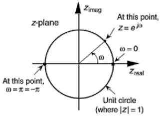
$$H(z) = \frac{z^{-N}\left(b_0 z^N + b_1 z^{N-1} + \ldots + b_N\right)}{z^{-M}\left(z^M + a_1 z^{M-1} + \ldots + a_M\right)} = \frac{\left(b_0 z^N + b_1 z^{N-1} + \ldots + b_N\right)}{z^{N-M}\left(z^M + a_1 z^{M-1} + \ldots + a_M\right)} = \frac{\tilde{B}(z)}{\tilde{A}(z)} .$$

  If $M > N$, then

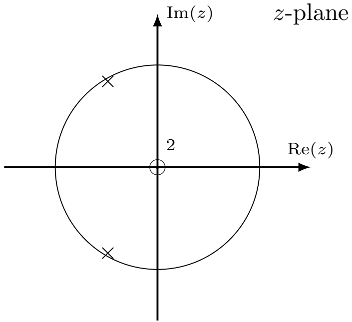
$$H(z) = \frac{z^{-N}\left(b_0 z^N + b_1 z^{N-1} + \ldots + b_N\right)}{z^{-M}\left(z^M + a_1 z^{M-1} + \ldots + a_M\right)} = \frac{z^{M-N}\left(b_0 z^N + b_1 z^{N-1} + \ldots + b_N\right)}{\left(z^M + a_1 z^{M-1} + \ldots + a_M\right)} = \frac{\tilde{B}(z)}{\tilde{A}(z)} .$$

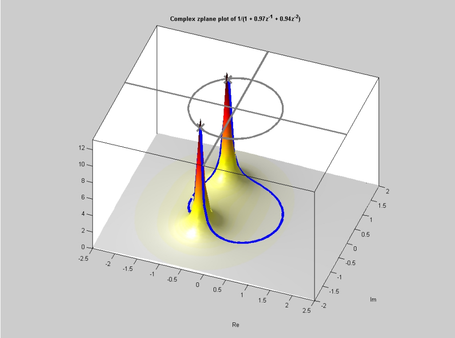
  - The zeros of $\tilde{B}(z)$, *i.e.* z such that $\tilde{B}(z) = 0$ are called **zeros** of $H(z)$.
  - The zeros of $\tilde{A}(z)$, *i.e.* z such that $\tilde{A}(z) = 0$ are called **poles** of $H(z)$.

LTI Causal Systems & Transforms



- z-**transform: Fractional** $H(z)$**, Realizable Systems, Poles & Zeros**





3D image from DSP First

LTI Causal Systems & Transforms

- **Numerical Exercise #1.1: z-Transform & System Design**

  By choosing the values (or positions) of poles and zeros we can design a LTI system, *i.e.*, we can define the transfer function, that is, the effect of the system on the input signal.

  The Matlab applicaiton `filterDesigner` provides an handy tool for system design based on poles and zeros positioning.

LTI Causal Systems & Transforms

- **z-Transform ROC and BIBO stability**

  Let $h[n]$ be the impulse response of a LTI causal system. The ROC of the corresponding z-Transform $H(z)$ includes the unit circle if and only if $\|h[n]\|_1 < \infty$.

  Therefore, the ROC of $H(z)$ contains the unit circle then the system is BIBO stable.

  Indeed, BIBO stability implies the existence of the DTFT of the impulse response , and therefore, the existence of the z-Transform on the unit circle!

LTI Causal Systems & Transforms

- **Minimum Phase System**

  A system with impulse response $h[n]$ is called **minimum phase** if it is stable, causal and all zeros are inside or on the unit circle. If all the zeros are only inside the unit circle, we say that the sequence is **strictly** minimum phase.

  A systems with strictly minimum phase impulse response have always a stable, causal and minimum phase inverse system.

- **Zero-Phase System**

  A zero-phase system is such that the its transfer function has phase equal to zero for any frequency, *i.e.*,

$$H(e^{j\omega}) = H^*(e^{j\omega}), \quad \text{or in terms of } z\text{-transform} \quad H(z) = H^*(\frac{1}{z^*}).$$

- **Spectral Factorization**

  If $H(z)$ is zero-phase then it can be decomposed as

$$H(z) = A(z)A(z^{-1}).$$

  where $A(z)$ and $A(z^{-1})$ have conjugate pairs of zeros and poles. We can arbitrarily choose to assign the zeros strictly inside the unit circle to $A(z)$ obtaining a **minimum phase spectral factor**.

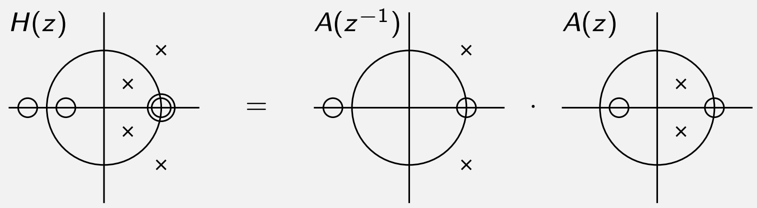
LTI Causal Systems & Transforms

- **Spectral Factorization (Example)**

  Consider a z-Transform $H(z)$ with poles and zeros given below.

  One can easily check that it is zero phase by observing that zeros (and poles) are at reciprocal positions and zeros on the unit circle have even multiplicity.

  The minimum phase spectral factor $A(z)$ is obtained by taking the poles and zeros inside the unit circle and half of the zeros on the unit circle.



LTI Causal Systems & Transforms

- **z Variable as Delay Operator**

  Consider the output of an LTI causal system $h[n]$

$$y[n] = (h * x)[n] = \sum_{k=0}^{\infty} h[k]x[n-k] = h[0]x[n] + h[1]x[n-1] + h[2]x[n-2] + \dots .$$

  By interpreting z^{-1} as a **delay operator**, *i.e.*, $x[n]z^{-1} = x[n-1]$, we have

$$\begin{aligned} y[n] &= h[0]x[n] + h[1]x[n-1] + h[1]x[n-2] + \dots \\ &= h[0]x[n] + h[1]x[n]z^{-1} + h[2]x[n]z^{-2} + \dots \\ &= x[n]\left(h[0] + h[1]z^{-1} + h[2]z^{-2} + \dots\right) \\ &= x[n]\sum_{k=0}^{\infty} h[k]z^{-k} = x[n]H(z) . \end{aligned}$$

  where $H(z)$ has to be treated as an operator.

Remark
Be careful that $x[n]H(z)$ is not to be interpreted as the product of a time domain signal and a z-transform, but as the operator $H(z)$ applied on the time domain signal $x[n]$.

1.5 Elements of Probability & Discrete Time Stochastic Processes

Elements of Probability

- **Random Variables**

  A quantity (*e.g.* a physical measurement), that randomly takes values over a certain set $\mathcal{A}$ ($\mathcal{A} = \mathbb{R}$ real r.v., $\mathcal{A} = \mathbb{Z}$ integer valued r.v., $\mathcal{A} = \{0, 1\}$ binary r.v.).

- **Law of a Random Variable (Cumulative Distribution Function)**

  $F_X(a) = \mathrm{P}(X \leq a)$ describes the probabilistic behavior of a r.v. X.

  – For continuous value random variables:
  We can define the **probability density function** $f_X(a) = \frac{dF_X(a)}{da}$. Then

  $$\mathrm{P}(X \in (a, b]) = \int_a^b f_X(x)\, dx\,, \quad a, b \in \mathcal{A} \subseteq \mathbb{R}$$

  and in particular $F_X(a) = \mathrm{P}(X \leq a) = \int_{-\infty}^{a} f_X(x)\, dx\,.$

  – For discrete value random variables:
  We can compute the **probability** of taking a particular value as $\mathrm{P}(X = a)$. Then

  $$F_X(k) = \mathrm{P}(X \leq k) = \sum_{n=-\infty}^{k} \mathrm{P}(X = n)\,, \quad k \in \mathcal{A} \subseteq \mathbb{Z}$$

Remark

If X is a continuous value random variable that has a probability density function $f_X(a)$, *i.e.*, $F_X(a)$ is differentiable, then $\mathrm{P}(X = a) = 0$

Elements of Probability

- **Moments of a Random Variable**

  Moment of order k of X: $\mathrm{E}\left[X^k\right]$ where we suppose $\mathrm{E}\left[|X|^k\right] < \infty$.
  - First order moment $\mathrm{E}[X]$ is the expected value
  - Second order moment is related to the variance $\mathrm{Var}(X) = \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2$.
  - Covariance of two random variables X, and Y:

  $$\mathrm{cov}(X, Y) = \mathrm{E}\left[XY^*\right] - \mathrm{E}[X]\,\mathrm{E}[Y]^*$$

- **Independence of Two Random Variables**

  X and Y are said to be independent if and only if their joint cumulative distribution is equal to the product of the respective cumulative distributions, *i.e.*,

  $$F_{XY}(a, b) = \mathrm{P}(X \le a \text{ and } Y \le b) = \mathrm{P}(X \le a)\,\mathrm{P}(Y \le b) = F_X(a)\,F_Y(b)\ .$$

  Independence **implies** $\mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y]$ and $\mathrm{cov}(X, Y) = 0$.

> Remark
> The latter is a necessary condition of independence but, in the general case, it is **not** a sufficient condition. It is a necessary and sufficient condition only if X and Y are jointly Gaussian distributed (try to prove this assertion as an exercise).

Discrete Time Stochastic Processes

- **Definition**

  A discrete time stochastic process is a sequence of random variables $\{X[n]\}_{n\in\mathbb{Z}}$, *i.e.*, $X[1], \ldots, X[n], \ldots$, where the index n accounts for temporal dependency.

- **Law**

  The law of a discrete time stochastic processes is the joint cumulative distribution functions of $(X[k_1], \ldots, X[k_n])$, for every $(k_1, \ldots, k_n)$ and every $n \in \mathbb{N}$, *i.e.*,

$$F_{X[k_1]\ldots X[k_n]}(a_1, \ldots, a_n) = \mathrm{P}\left(X[k_1] \leq a_1, \ldots, \ldots X[k_n] \leq a_n\right) .$$

Discrete Time Stochastic Processes

- **Correlation and Covariance**

  The correlation and the covariance are second order moments:

  – Correlation of the stochastic process at two time indexes

  $$R(k,l) = \mathrm{E}\Big[X[k]\,X^*[l]\Big];$$

  – Covariance of the stochastic process at two time indexes

  $$\Gamma(k,l) = \mathrm{cov}\Big(X[k], X[l]\Big) = \mathrm{E}\Big[X[k]\,X^*[l]\Big] - \mathrm{E}\Big[X[k]\Big]\mathrm{E}\Big[X[l]\Big]^*.$$

Remark

As the name says, the correlation (and similarly, the covariance), provides a measurement of the inter-dependency of the process at different time instants.

It is therefore the key tool for analyzing patterns, and therefore for statistical signal processing!

Discrete Time Stochastic Processes

- **Stationarity**

A discrete time stochastic process $\{X[n]\}_{n\in\mathbb{Z}}$ is said to be stationary if and only if

$$\mathrm{P}\left(X[k_1]\leq a_1,\ldots,\ldots X[k_n]\leq a_n\right)=\mathrm{P}\left(X[k_1+l]\leq a_1,\ldots,\ldots X[k_n+l]\leq a_n\right),$$

for all $(k_1,\ldots,k_n)$ and for all $l,n\in\mathbb{N}$, *i.e.*,

$$F_{X[k_1]\ldots X[k_n]}\left(a_1,\ldots,a_n\right)=F_{X[k_1+l]\ldots X[k_n+l]}\left(a_1,\ldots,a_n\right),$$

for all $(k_1,\ldots,k_n)$, and for all $l,n\in\mathbb{N}$.

Remark

Consequences of stationarity (necessary but not sufficient)

- $\mathrm{E}\left[X[n]^k\right]$ = constant (independent of n);
- $\mathrm{E}\left[X[k]\,X^*[l]\right]=R(k-l)$ and $\mathrm{cov}\left(X[k],X[l]\right)=\Gamma(k-l)$, for every $k,l\in\mathbb{Z}$.

Discrete Time Stochastic Processes

- **Wide Sense Stationarity - WSS**

  A discrete time stochastic process $\{X[n]\}_{n\in\mathbb{Z}}$ is said to be wide sense stationary if and only if

  - $\mathrm{E}[X[n]] = m$ constant, independent of n;
  - $\mathrm{E}\Big[X[k]\,X^*[l]\Big] = R(k-l)$, for every $k, l \in \mathbb{Z}$,*i.e.*, the correlation depends only on the time difference. Consequently, $\mathrm{cov}\Big(X[k], X[l]\Big) = \Gamma(k-l)$.
  - $\mathrm{Var}(X[n]) = \sigma^2$ constant, independent of n, with $\sigma^2 < \infty$.

Discrete Time Stochastic Processes

- **Power Spectral Density of w.s.s. processes**

  Let $X[n]$ be a wide sense stationary process with correlation $R_X(k)$. Assume that $R_X(k)$ is absolutely summable. Then, the power spectral density $S_X(\omega)$ of $X[n]$ is defined as

$$S_X(\omega) = \sum_{k=-\infty}^{\infty} R_X(k)\, e^{-i\omega k}\,.$$

  The power spectral density corresponds to an average of the square Fourier transforms of the possible realizations of the process and it gives insight on the distribution of the (average) energy of the process.

Remark

- The power spectral density does not represent the wide sense stochastic process, but only its second order characteristics. Indeed, while a deterministic signal can be obtained from its Fourier transform, it is not generally possible to characterize the process, *i.e.*, to obtain its law, given its power spectral density;
- The power spectral density is the distribution of the energy on the frequency domain of the correlation, that is of the key tool for pattern analysis.

Discrete Time Stochastic Processes

- **Filtering in the "Frequency" domain: The Fundamental Filtering Formula**

  Let $X[n]$ be a wide sense stationary process, with summable correlation $R_X(k)$, and power spectral density $S_X(\omega)$.

  Call $Y[n]$ the stochastic process obtained by filtering $X[n]$ with a stable filter h (BIBO/ℓ^1), *i.e.*,

  $$Y[n] = \sum_{k=-\infty}^{\infty} h_{n-k} X[k].$$

  Then

  - $Y[n]$ is a wide sense stationary process;
  - $\mathrm{E}[Y] = \mathrm{E}[X] \sum_{k=-\infty}^{\infty} h_k$;
  - $R_Y(k)$ is absolutely summable;
  - $Y[n]$ has power spectral density

    $$S_Y(\omega) = |H(\omega)|^2 S_X(\omega),$$

    where $H(\omega)$ the discrete Fourier transform of the filter h.

1.6 Elements of Hilbert Spaces

Elements of Hilbert Spaces

- **What is a Hilbert Space?**

  A Hilbert space is nothing but a collection of elements with very nice properties, over which we can define two functions:

  - An **inner product** $< \cdot, \cdot >$ between two elements.
    The inner product is related to the notion of projection of one element onto the other.
    When the inner product between two elements is zero, they are said to be **orthogonal**;

  - A **norm** $\|\cdot\|$ of an element.
    The norm is related to the notion of energy.
    The norm of the difference between two elements represents the **distance** between them, that is, the **energy of the error** between them.

Elements of Hilbert Spaces

- **The Euclidean Space as Hilbert Space (Example)**

  Call $\vec{a}_x$, $\vec{a}_y$, $\vec{a}_z$, the unit vectors of the Euclidean space $\mathbb{R}^3(x, y, z)$.

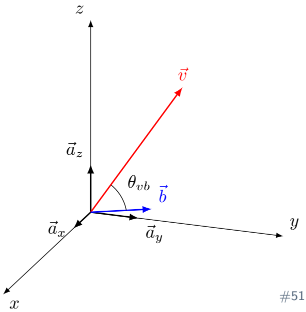
  Given a vector $\vec{v} = v_x\vec{a}_x + v_y\vec{a}_y + v_z\vec{a}_z$

  – The norm is the vector length, *i.e.*, $\|\vec{v}\| = \left(v_x^2 + v_y^2 + v_z^2\right)^{\frac{1}{2}}$

  – The inner product between the vector $\vec{v} = v_x\vec{a}_x + v_y\vec{a}_y + v_z\vec{a}_z$ and the vector $\vec{b} = b_x\vec{a}_x + b_y\vec{a}_y + b_z\vec{a}_z$ is their scalar product, *e.g.*,

$$< \vec{v}, \vec{b} >= \|\vec{v}\| \left\|\vec{b}\right\| \cos\theta_{vb} = v_x b_x + v_y b_y + v_z b_z \ ,$$

  where θ_{vb} is the angle between the two vectors $\vec{v}$ and $\vec{b}$.



Elements of Hilbert Spaces

- **Random Variables with Finite Variance as Hilbert Space (Example)**

  Random variables with finite variance form a Hilbert space.

  The collection of random variables X such that $\mathrm{Var}(X) < \infty$ is denoted as $\mathrm{L}^2(P)$.

  - The inner product is
$$\langle A, B\rangle_{L^2(P)} = \mathrm{E}\left[AB^*\right]$$
  - The norm is
$$\|A\|_{L^2(P)} = \mathrm{E}\left[|A|^2\right].$$

- **Discrete Time Stochastic Processes as Hilbert Space (Example)**

  W.s.s. processes form a Hilbert space.

  This is just a particular case of the space of random variables with finite variance: For a w.s.s. process $X[n]$ we have, by definition, $\mathrm{E}\left[|X[n]|^2\right] < \infty$.

Elements of Hilbert Spaces

- **The Projection Theorem**

  Let E and S be two Hilbert spaces such that $S \subset E$. Then, for every element v of E it exists a unique element b of S such that

  - b is the element of S that minimizes the norm of the difference $v - b$, that is

  $$b = \arg\min_{c \in S} \|v - c\| \; ;$$

  - The difference $v - b$ is orthogonal to any element in S, that is

  $$\langle v - b, c \rangle = 0 \,, \quad \forall \; c \in S \,,$$

  The element b is called the **orthogonal projection of v onto S**.

Elements of Hilbert Spaces

- **The Projection Theorem in the Euclidean Space $\mathbb{R}^3(x,y,z)$ (Example)**

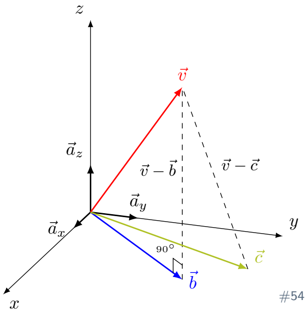
  Call $\vec{b}$ the orthogonal projection of the vector $\vec{v}$, defined in the space $E = \mathbb{R}^3(x,y,z)$, onto the plane $S = \mathbb{R}^2(x,y)$ $(S \subset E)$.

  In particular:

  - $\vec{b}$ is the vector in the plane that minimizes the norm of the difference $\vec{v} - \vec{b}$

  $$\left\|\vec{v} - \vec{b}\right\| \leq \left\|\vec{v} - \vec{c}\right\| \;\; \forall \;\; \vec{c} \in \mathbb{R}^2(x,y)\,;$$

  - The difference $\vec{v} - \vec{b}$ is orthogonal to any vector in the plane $S = \mathbb{R}^2(x,y)$.



Elements of Hilbert Spaces

- **The Projection Theorem in the Space of w.s.s. Processes**

  Let E and S be two spaces of w.s.s. signals such that $S \subset E \subseteq \mathrm{L}^2(P)$. Then, for every w.s.s. signal $X[n]$ in E it exists a unique w.s.s. signal $Y[n]$ in S such that

  – $Y[n]$ is the signal in S minimizing the norm of $X[n] - Y[n]$, that is

$$Y[n] = \underset{U[n]\in S}{\arg\min}\ \mathrm{E}\left[|X[n] - U[n]|^2\right];$$

  – The difference $X[n] - Y[n]$ is orthogonal to any element in S, that is

$$\mathrm{E}\left[(X[n] - Y[n])U[n]^*\right] = 0, \quad \forall\ U[n] \in S,$$

  The w.s.s. signal $Y[n]$ is the **orthogonal projection** of $X[n]$ onto S.

Elements of Hilbert Spaces

- **The Projection Theorem in the Space of w.s.s. Processes**

  Consider the particular case of an AR process (of order 3) in canonical form

$$X[n] = -p_1 X[n-1] - p_2 X[n-2] - p_3 X[n-3] + W[n]\,,$$

where $X[n]$ is, by definition, w.s.s., $\mathrm{E}\,[W[n]] = 0$, $\mathrm{E}\,[X[k]W[n]^*] = 0$ for every $k \leq n-1$.
Call E the space generated by $X[k]$, for all $k \leq n$, and by $W[n]$.
Call S the space generated by $X[k]$, for all $k \leq n-1$, where $S \subset E$.

Then the orthogonal projection of $X[n] \in E$ onto S is given by

$$Y[n] = -p_1 X[n-1] - p_2 X[n-2] - p_3 X[n-3]\,.$$

Indeed, for every vector in S, that is for every $X[k]$ with $k \leq n-1$,

$$\begin{aligned}< X[n] - Y[n], X[k] > &= < X[n] - (-p_1 X[n-1] - p_2 X[n-2] - p_3 X[n-3]), X[k] > \\ &= < W[n], X[k] > = \mathrm{cov}\,(W[n], X[k]) = 0\,.\end{aligned}$$

Elements of Hilbert Spaces

- **Intuitive Property**

  Let v be a element of E that we can write as

  $$v = b + c$$

  where $b \in S$ and c is orthogonal to any element in S. Then, the orthogonal projection of v onto S is straightforwardly given by b.

Elements of Hilbert Spaces

- **Intuitive Property in the Euclidean Space (Example)**

  Consider a vector
  $$\vec{v} = v_x \vec{a}_x + v_y \vec{a}_y + v_z \vec{a}_z \, .$$

  Call $\vec{b} = v_x \vec{a}_x + v_y \vec{a}_y \in S = \mathbb{R}^2(x,y)$ and $\vec{c} = v_z \vec{a}_z$, the latter being orthogonal to any vector in $S = \mathbb{R}^2(x,y)$.

  Therefore, the orthogonal projection of $\vec{v}$ onto $S = \mathbb{R}^2(x,y)$ is given by
  $$\vec{b} = v_x \vec{a}_x + v_y \vec{a}_y \, .$$

Elements of Hilbert Spaces

- **Intuitive Property in the Space of w.s.s. Processes (Example)**

  Let $X[n]$ be a element of E that we can write as $X[n] = U[n] + V[n]$, where $U[n] \in S$ and $V[n]$ is orthogonal to any element in S. Then, the orthogonal projection of $X[n]$ onto S is straightforwardly given by $U[n]$.

  In the specific case of AR process (of order 3) in canonical form

  $$X[n] = -p_1 X[n-1] - p_2 X[n-2] - p_3 X[n-3] + W[n],$$

  we have $U[n] = -p_1 X[n-1] - p_2 X[n-2] - p_3 X[n-3] \in S$ and $V[n] = W[n]$ orthogonal any vector in S.

  Therefore the orthogonal projection of $X[n]$ onto S gives

  $$Y[n] = U[n] = -p_1 X[n-1] - p_2 X[n-2] - p_3 X[n-3].$$

Elements of Hilbert Spaces

- **Intuitive Property in the Space of w.s.s. Processes:**
  Optimal Predictor as a Projection onto the Past (Example)

  Let $X[n]$ be a w.s.s. stochastic process.

  Call S=space formed by the linear combinations of $X[n-1],\ldots,X[n-N]$, and E=space formed by the linear combinations of $X[n+k],\ldots,X[n],\ldots,X[n-N]$.

  To predict $X[n+k]\in E$ we use a linear combination of its past

  $$a_1 X[n-1]+\ldots+a_N X[n-N].$$

  Notice that, clearly, such a linear combination belongs to S.

  Among all the possible linear combinations we choose as optimal the one that minimizes the mean square error (MSE)

  $$\mathrm{E}\left[\left|X[n+k]-(a_1 X[n-1]+\ldots+a_N X[n-N])\right|^2\right].$$

  Gosh! Do I now really have to minimize the MSE by computiting all the derivatives w.r.t. a_n and set them to zero?

Elements of Hilbert Spaces

- **Intuitive Property in the Space of w.s.s. Processes:**
  Optimal Predictor as a Projection onto the Past (Example)

  No! ... no effort with the Projection Theorem!

  Recall that to predict $X[n+k] \in E$ we search for the linear combination of its past $a_1 X[n-1] + \ldots + a_N X[n-N]$ that minimizes the MSE

$$\mathrm{E}\left[\left| X[n+k] - (a_1 X[n-1] + \ldots + a_N X[n-N]) \right|^2 \right].$$

  In Summary: we want to estimate an element of E with the element $Y[n]$ of S that minimizes the MSE $\mathrm{E}\left[|X[n+k] - Y[n]|^2 \right]$.

  By the projection theorem
  - $Y[n]$ is the orthogonal projection of $X[n+k]$ onto S;
  - $Y[n]$ is such that $\mathrm{E}\left[(X[n+k] - Y[n]) U[n]^* \right] = 0$, for all $U[n] \in S$.

  Therefore instead of computing annoying derivatives we just have to solve the linear system

$$\mathrm{E}\left[(X[n+k] - Y[n]) X[n-l]^* \right] = 0, \quad l = 1, \ldots, N.$$

1.7 Empirical Statistics

Empirical Statistics

- **Why do we need empirical statistics?**

  To compute the moments of a w.s.s. process $\{X[n]\}_{n\in\mathbb{Z}}$ it necessary to know the law F_X of the process.

  In practice, we observe realizations of the stochastic process, say $x[1],\ldots,x[N]$, but we do not know F_X.

  Estimators for the moments of the stochastic process based on the observed realization $x[1],\ldots,x[N]$ = **empirical statistics**.

Empirical Statistics

- **Empirical Mean, Correlation, and Variance**
  - Empirical mean $\widehat{m}$

$$\widehat{m}\,(x[1],\ldots,x[N]) = \frac{1}{N}\sum_{n=1}^{N} x[n]$$

  - Empirical variance $\widehat{\sigma}^2$

$$\widehat{\sigma}^2\,(x[1],\ldots,x[N]) = \frac{1}{N}\sum_{n=1}^{N} x[n]^2 - \widehat{m}^2$$

  - Empirical correlation

$$\widehat{R}_X(k)\,(x[1],\ldots,x[N]) = \alpha \sum_{n=1}^{N-k} x[n+k]x[n]^*\,, \quad k = 0,\ldots,N-1\,,$$

where we either have $\alpha = \frac{1}{N-k}$ or $\alpha = \frac{1}{N}$.

Empirical Statistics

- **Empirical Correlation: Meaning**

  The correlation consists in multiplying each sample with its neighbor at k-positions, and then compute the average of such multiplications.

  For $k = 0$, the correlation corresponds to the power of the signal $\widehat{R}_X(0) = \dfrac{1}{N}\sum_{n=1}^{N}|x[n]|^2$.

  For $k \neq 0$ the correlation gives a measure of the average dependency (similarity) between each sample its neighbor at k-positions.

  The measure of the dependency is particularly true for zero mean signals, in which case a zero correlation indicates independency between samples.

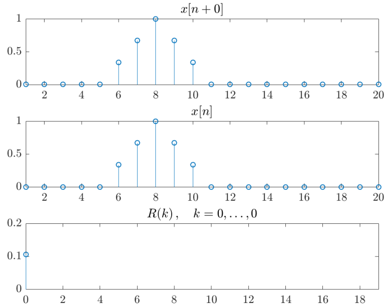
  When the mean is not zero, the correlation value also accounts for the square of the mean.

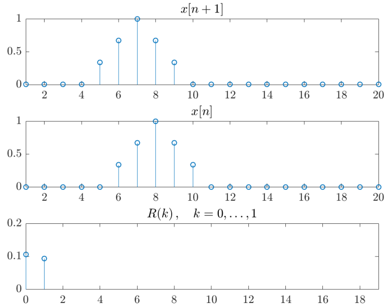
Empirical Statistics

Empirical Correlation: A Visual Example

$$\widehat{R}_X(0) = \alpha \sum_{n=1}^{N} |x[n]|^2 \,,$$

$$\widehat{R}_X(1) = \alpha \sum_{n=1}^{N-1} x[n+1]x^*[n] \,,$$

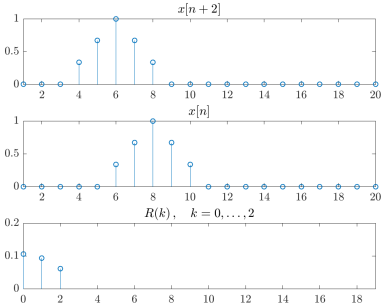


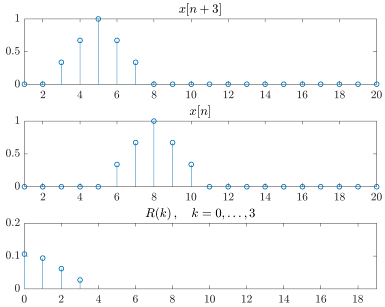


Empirical Statistics

▶ Empirical Correlation: A Visual Example

$$\widehat{R}_X(2) = \alpha \sum_{n=1}^{N-2} x[n+2]x^*[n]\,, \qquad \widehat{R}_X(3) = \alpha \sum_{n=1}^{N-3} x[n+3]x^*[n]\,,$$



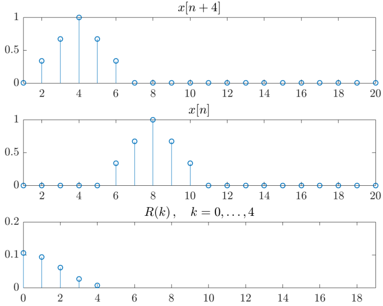


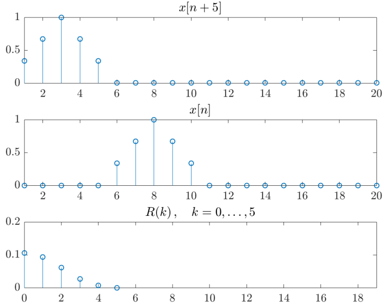
Empirical Statistics

Empirical Correlation: A Visual Example

$$\widehat{R}_X(4) = \alpha \sum_{n=1}^{N-4} x[n+4]x^*[n]\,,$$

$$\widehat{R}_X(5) = \alpha \sum_{n=1}^{N-5} x[n+5]x^*[n]\,,$$





Empirical Statistics

- **Bias and Variance of Empirical Statistics**

  For each moment different estimator can be proposed, each presenting its own advantages.

  Let $\widehat{S}\,(x[1], \ldots, x[N])$ be the empirical statistics for a certain probabilistic moment S.

  In order to evaluate how good the empirical statistics is, we analyze its **bias**

$$\mathrm{E}\left[\widehat{S}\,(X[1], \ldots, X[N])\right] - S\,,$$

  and **variance**

$$\mathrm{Var}\left(\widehat{S}\,(X[1], \ldots, X[N]) - S\right)\,.$$

  A good empirical statistics has zero bias, *i.e.*, is **unbiased**, and has a (low) variance that decreases as the number of realizations increases.

  However, it is not possible in general to obtain such a configuration and often we have to chose a tradeoff between low bias and low variance.

Empirical Statistics

- **Empirical Correlation: Unbiased**

The unbiased version of the correlation is defined for a normalizing parameter $\alpha = \dfrac{1}{N - |k|}$

$$\widehat{R}_X^{\text{NB}}(k) = \frac{1}{N - |k|} \sum_{n=1}^{N-|k|} x[n+k]x^*[n]\,,$$

At lag k, the correlation $\widehat{R}_X^{\text{NB}}(k)$ is based on the sum of $N - |k|$ multiplications, averaged over $\frac{1}{N-|k|}$.

- **Pros:**
  With such a normalizing parameter we have the standard averaging formula for $N - |k|$ elements. For instance, if $x[n] = c$ is constant, the correlation is always equal to c^2 (all the samples are similar).

- **Cons:**
  At extreme lags, that is when k is close to $N - 1$, the average is based on very few elements, making it inaccurate. For instance

$$\widehat{R}_X^{\text{NB}}(N-2) = \frac{1}{2}\left(x[N-1]x^*[1] + x[N]x^*[2]\right)\,, \quad \text{and} \quad \widehat{R}_X^{\text{NB}}(N-1) = x[N]x^*[1]\,,$$

Assignement
Check that indeed such an estimator is unbiased.

Empirical Statistics

- **Numerical Exercise: Empirical Correlation Unbiased**

  We want to compute the unbiased correlation of a zero mean Gaussian white noise, *i.e.*, a signal whose samples are independent.

  – Generate $N = 1000$ samples of a zero mean Gaussian white noise (Python Numpy command `x = numpy.random.normal(0,1,N)` or the Matlab command `x=randn(1,N)`);

  – Compute the unbiased correlation (Python Numpy command `R=numpy.correlate(x,x, mode='full')` to be scaled, or Matlab command `RxxNB=xcorr(x,'unbiased')`;

  – Plot the correlation

  – What do you remark?

Empirical Statistics

- **Empirical Correlation: Biased**

  The Biased version of the correlation is defined for a normalizing parameter $\alpha = \frac{1}{N}$

  $$\widehat{R}_X^{\mathsf{B}}(k) = \frac{1}{N} \sum_{n=1}^{N-|k|} x[n+k]x^*[n] \,.$$

  At lag k, the correlation $\widehat{R}_X^{\mathsf{B}}(k)$ is based on the sum of $N - |k|$ multiplications, averaged over $\frac{1}{N}$.

  - **Pros:**
    With such a normalizing parameter $N - |k|$ elements are averaged over N (and not over $N - |k|$!). The averaging for extreme lags, that is when k is close to $N - 1$, is therefore highly penalized (few elements averaged over N). But at extreme lags the averaging is more inaccurate, so, finally, the constant factor $1/N$ reduces its influence.

  - **Cons:**
    When the signal is deterministic or constant, the constant factor $1/N$ linearly reduces the amplitude of the correlation, that is, it introduces a bias.

Empirical Statistics

- **Numerical Exercise: Empirical Correlation Biased**

  Let's repeat the previous numerical exercise by computing the biased correlation of a zero mean Gaussian white noise.

  – Generate $N = 1000$ samples of a zero mean Gaussian white noise (Python Numpy command `x = numpy.random.normal(0,1,N)` or the Matlab command `x=randn(1,N)`);

  – Compute the biased correlation (Python Numpy command `R=numpy.correlate(x,x, mode='full')` to be scaled by N, or Matlab command `RxxNB=xcorr(x,'biased')`;

  – Plot the correlation

  – What do you remark now?

Empirical Statistics

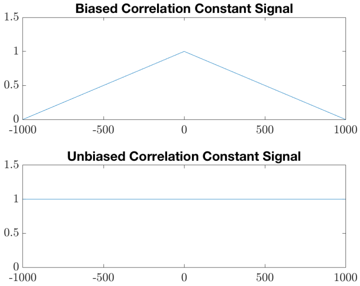
- **Empirical Correlation: Biased vs Unbiased**

Based on the previous examples, one might think that the biased correlation is the optimal choice. Well not every time!

In the definition of the correlation we mentioned that it accounts for the mean of the signal. If the mean is not zero, the constant term $1/N$ will reduce the influence of the mean as k increases.

Take the extreme case where the signal is constant, for instance $x[n] = 1$, $n = 1, \ldots, 1000$.

All the samples are similar (they are the same!), therefore, each sample will be equally correlated with the others. If we compute the biased correlation we get a linearly decreasing function as k increases. On the other hand, the unbiased correlation looks as expected: A constant!



Empirical Statistics

- **Numerical Exercise: Biased vs Unbiased Correlation**

  Let's apply the correlation to a non zero mean signal (different than a constant!)

  - Generate $N = 1000$ samples of a uniform white noise (in the interval $[0, 1)$);
  - Compute the mean;
  - Compute the biased correlation;
  - Plot the biased correlation. What do you observe?
  - Compute the unbiased correlation;
  - Plot the biased correlation. What do you observe?