

Signal Processing

Authors: G. Chardon , R. Combes, J. Fiorina,
E. Lahalle, J. Picheral, P. Rodriguez,
C. Soussen, G. Valmorbida, A. Wautier

CentraleSupélec, March 2025

CONTENTS

Summary	7
Notations	9
1 MATHEMATICAL PREREQUISITES	11
1.1 Finite dimensional vector spaces	11
1.1.1 Definition	11
1.1.2 Linear transformation	12
1.1.3 Norm, inner product	12
1.2 Banach spaces	13
1.2.1 ℓ^p spaces	14
1.2.2 L^p spaces	14
1.3 Hilbert spaces	14
1.3.1 Fourier series	15
1.4 Distributions	16
1.4.1 Tempered distributions	19
1.4.2 Compactly supported distributions	19
1.5 Random variables and processes	20
1.5.1 Brief reminders of statistics	23
2 LINEAR TIME INVARIANT SYSTEMS	25
2.1 Signal concept	25
2.1.1 Definitions	25
2.1.2 Energy representations	26
2.2 Convolution	26
2.2.1 Continuous case: convolutions of functions	27
2.2.2 Discrete case	28
2.2.3 Signals modeled by distributions	28
2.2.4 Properties of the convolution product	30
2.3 Concept of system	30
2.3.1 Definition	30
2.3.2 Properties of systems	31
2.4 Linear convolution filters	32
2.4.1 Definition	33
2.4.2 Finite impulse response filters (FIR)	34
2.4.3 Infinite impulse response filters (IIR)	36
2.4.4 Difference equations	36
2.5 Eigenfunctions	37
2.6 Conclusion	38
3 FREQUENCY DOMAIN REPRESENTATION	39
3.1 Continuous time Fourier transform	40
3.1.1 Fourier transform of signals in L^1	40
3.1.2 Fourier transform of L^2 signals	42
3.1.3 Fourier transform in $\mathcal{S}'$	43

3.1.4	Some classical examples	46
3.2	Fourier series representation	48
3.2.1	Signals with limited support	48
3.2.2	Periodic signals	49
3.3	Discrete-time Fourier transform	49
3.3.1	Fourier transform of ℓ^1 signals	50
3.3.2	Fourier transform of ℓ^2 signals	51
3.3.3	Some classical results	52
3.4	Discrete Fourier transform	53
3.5	Application to linear filtering	54
3.5.1	Filtering in the frequency domain	54
3.5.2	Synthesis of FIR filters using the DFT	57
3.6	Spectral analysis	59
3.6.1	Time windowing of continuous time signals	59
3.6.2	Spectral resolution	61
3.6.3	Time windowing of discrete time signals	62
3.7	Conclusion	63
4	SAMPLING AND RECONSTRUCTION	67
4.1	Time sampling	67
4.2	Reconstruction	70
4.2.1	Band limited signals in $[-B, B]$	70
4.2.2	Shannon's theorem	70
4.2.3	Proof of Shannon's theorem	72
4.2.4	Band limited signals in $[A, B]$	74
4.3	Sampling in practice	75
4.3.1	Anti-aliasing filter	75
4.3.2	Practical reconstruction	75
4.4	Conclusion	76
5	SIGNALS MODELED BY RANDOM PROCESSES	77
5.1	Definition and characterization of random processes	77
5.1.1	Definition	77
5.1.2	Finite-dimensional distributions	79
5.1.3	Gaussian process	80
5.1.4	Mean and autocorrelation	80
5.2	Stationary processes	81
5.2.1	Strict-sense stationarity	82
5.2.2	Wide-sense stationarity	82
5.2.3	Power spectral density	83
5.2.4	White noise	86
5.2.5	Linear filtering of random processes	88
5.2.6	Autoregressive model	90
5.3	Conclusion	92
6	ESTIMATION BASED ON RANDOM MODELS	93
6.1	Estimation of the mean of a random process	94
6.2	Estimation of the autocorrelation function	95
6.3	Estimation of the PSD: periodogram	98

6.3.1	Definition	98
6.3.2	Statistical properties of the periodogram	100
6.3.3	Variants of the periodogram method	101
6.4	Parametric estimation of the PSD	104
6.4.1	Principle	104
6.4.2	Case of a signal modeled by an AR process	104
6.5	Optimal linear estimator	106
6.5.1	Definition and properties	107
6.5.2	Prediction of a random process	108
6.5.3	Wiener filter for denoising	110
6.6	Conclusion	113
	Bibliography	116

SUMMARY

CONTEXT AND ISSUES

The digital world produces large volumes of signals of all kinds (audio, image, video, physical measurements) in different domains such as healthcare, telecommunication, industry or environment. Processing the information from these data is fundamental for medical diagnosis, data compression, noise removal from an audio signal, etc. This course introduces methods to address various problems such as filtering, transmission, denoising of signals or spectral analysis.

ORGANIZATION

This handout introduces basic methods based on mathematical concepts such as Hilbert spaces, distributions, and random vectors. These concepts are revisited in chapter 1. Chapter 2 introduces the notions of signals and systems, in particular linear convolution filters. Chapter 3 studies Fourier transforms, an essential tool in signal processing. Fourier transforms are introduced for continuous and discrete time signals and used to analyze the frequency content of signals (spectral analysis) and to characterize linear filters in the frequency domain. The sampling theory introduced in chapter 4 provides conditions under which the information contained in a signal is preserved after time discretization. Shannon's theorem is the fundamental result ensuring exact reconstruction of a band-limited signal. Finally, chapters 5 and 6 introduce a probabilistic framework to characterize the statistical properties of signals, seen as random processes. The analysis of stationary signals is based on the auto-correlation function (time domain) and the power spectral density (frequency domain), formally introduced in chapter 5. Chapter 6 discusses how to estimate the auto-correlation and spectral density from experimental data, with applications to denoising, filtering, data prediction, and spectral analysis.

NOTATIONS

VECTORS

$\mathbf{x} \in \mathbb{C}^n$	vector
$x[k]$ or x_k	k -th coordinate of the vector

FREQUENCY REPRESENTATIONS

$(\mathcal{F}x)(f)$ or $X(f)$	Fourier transform
$(\mathcal{F}x)(\nu)$ or $X(\nu)$	discrete time Fourier transform
$\mathbf{X} \in \mathbb{C}^n$	discrete Fourier transform of a vector $\mathbf{x} \in \mathbb{C}^n$

RANDOM VARIABLES

X_k	random variable at time k
-------	-----------------------------

INNER PRODUCTS AND NORMS

$(x, y)_{L^2}$	inner product in $L^2(\mathbb{R})$, equal to $\int_{\mathbb{R}} x y^* d\lambda$
$(x, y)_{L^2}$	inner product in $L^2(0, T)$, equal to $\frac{1}{T} \int_{[0, T]} x y^* d\lambda$
(x, y)	inner product in $\ell^2(\mathbb{Z})$, equal to $\sum_{k \in \mathbb{Z}} x[k] y^*[k]$
$\ \cdot\ _{L^p}$	L^p norm, with $\ x\ _{L^p}^p = \int_{\mathbb{R}} x(t) ^p dt$
$\ \cdot\ _p$	ℓ^p norm, with $\ \mathbf{x}\ _p^p = \sum_k x[k] ^p$

DISTRIBUTIONS

$\langle T, \varphi \rangle$	evaluation of distribution T on a test function φ
------------------------------	---

DIRAC VS. KRONECKER

δ_a	Dirac distribution centered at point a : $\langle \delta_a, \varphi \rangle = \varphi(a)$
$\delta = \delta_0$	centered Dirac distribution: $\langle \delta, \varphi \rangle = \varphi(0)$
$\mathbb{I}_T$	Dirac comb: $\mathbb{I}_T = \sum_{k \in \mathbb{Z}} \delta_{kT}$
$\delta[k]$	Kronecker function: $\delta[k] = 1$ if $k = 0$, and 0 otherwise

COMPLEX EXPONENTIALS

$e_f(t) = \exp(2i\pi ft)$	complex exponential of frequency f
$e_n(t) = \exp\left(2i\pi \frac{nt}{T}\right)$	complex exponential used for Fourier series [simplified notation for $e_{n/T}$]
$e_\nu[k] = \exp(2i\pi \nu k)$	discrete complex exponential of frequency ν
$\mathbf{e}_n \in \mathbb{C}^N$	Discretized version of function $t \mapsto \exp(2i\pi nt)$ on $[0, 1[$: $\mathbf{e}_n = \left\{ \exp(2i\pi \frac{nk}{N}), k = 0, \dots, N-1 \right\}$

MATHEMATICAL PREREQUISITES

The purpose of this chapter is to lay the mathematical foundations of this course: functional spaces, basic operations on signals, reminders of probability, and to provide links with the issues covered in the signal processing course. We will call “signal” a function, a distribution, or a digital sequence, mostly time-dependent (audio signal, telecommunications), space-dependent (image, geographical data), or of both (video).

1.1 FINITE DIMENSIONAL VECTOR SPACES

1.1.1 *Definition*

A vector space over a scalar field $\mathbb{K}$ (here, $\mathbb{R}$ or $\mathbb{C}$) is a set E equipped with addition (operator $+$ from $E \times E$ to E) and multiplication (operator $\cdot$ from $\mathbb{K} \times E$ to E). These operators are such that $(E, +)$ is a group, that associativity and distributivity properties are satisfied, and that the neutral element of $\cdot$ is the neutral element of the multiplication in $\mathbb{K}$.

For finite-dimensional vector spaces, there exists a basis spanning the vector space. A basis is a family $\mathcal{B}$ of vectors $(\mathbf{e}_i)_{1 \leq i \leq N}$ such that any vector $\mathbf{x}$ of E is written in a unique way as a linear combination of the basis elements:

$$\mathbf{x} = \sum_{i=1}^N x_i \mathbf{e}_i.$$

The weights x_i are referred to as the coordinates of $\mathbf{x}$ in basis $\mathcal{B}$. All basis of the same vector space contain the same number N of vectors. We call N the dimension of the vector space.

APPLICATION IN SIGNAL PROCESSING : Signals living in a finite dimension space will be seen as elements of a vector space. For an audio signal, we can think of amplification or attenuation of sound (multiplication with a scalar), and mixing two sounds (addition).

1.1.2 Linear transformation

We call linear transformation a mapping L from E to F such that for all $\mathbf{u}$ and $\mathbf{v}$ in E and λ in $\mathbb{K}$,

$$\begin{aligned} L(\mathbf{u} + \mathbf{v}) &= L(\mathbf{u}) + L(\mathbf{v}) \\ L(\lambda \mathbf{u}) &= \lambda L(\mathbf{u}). \end{aligned}$$

The linear property allows us to simplify this description. Given bases $\mathcal{B}_E$ of E and $\mathcal{B}_F$ of F , one can show that the coefficients of $\mathbf{y} = L(\mathbf{x})$ read:

$$y_n = \sum_{m=1}^N a_{nm} x_m \quad (1)$$

where N is the dimension of E , and scalars $a_{nm} \in \mathbb{K}$ depend on L , $\mathcal{B}_E$ and $\mathcal{B}_F$. Denoting $\mathbf{A}$ is the 2D matrix containing a_{nm} for all n, m , (1) rereads:

$$\mathbf{y} = \mathbf{A}\mathbf{x}. \quad (2)$$

$\mathbf{A}$ will be referred to as the matrix of transformation L in bases $\mathcal{B}_E$ and $\mathcal{B}_F$. One can thus represent L by the related matrix $\mathbf{A}$, and compute the output $L(\mathbf{x})$ on a computer as the matrix-vector product (2).

More, some transformations from E to E are diagonalizable. Computing $\mathbf{y} = \mathbf{A}\mathbf{x}$ boils down to multiplying the coefficients of $\mathbf{x}$ in a well-chosen basis (of eigenvectors, such that $L(\mathbf{u}_n) = \lambda_n \mathbf{u}_n$):

$$y_n = \lambda_n x_n.$$

This description allows for more simple interpretation and calculation of the action of L on vectors in E .

APPLICATION IN SIGNAL PROCESSING : in signal processing, we will deal with a specific family of linear transformations (in finite dimension or not), called *linear filters*, that are translation invariant. In the same way that we describe a linear application by a matrix, a filter will be described by its impulse response. Instead of matrix products, we will use convolution. The Fourier transform will play the role of diagonalization.

1.1.3 Norm, inner product

A norm $N(\cdot)$ on E is a function from E to $\mathbb{R}_+$ satisfying the properties:

- separability: $N(\mathbf{x}) = 0 \Rightarrow \mathbf{x} = 0$,
- homogeneity: $N(\lambda \mathbf{x}) = |\lambda| N(\mathbf{x})$,
- triangular inequality: $N(\mathbf{x} + \mathbf{y}) \leq N(\mathbf{x}) + N(\mathbf{y})$.

A vector space endowed with a norm is called a normed vector space.

A norm makes it possible to define a topology, and therefore to speak of convergence of a sequence of vectors $\mathbf{x}_n$ approximating a vector $\mathbf{x}$. For finite dimensional spaces on $\mathbb{R}$ or $\mathbb{C}$, all norms are equivalent and therefore define the same topology: the notion of convergence does not depend on the chosen norm. The following norms are often encountered:

- $\|\mathbf{x}\|_1 = \sum_{i=1}^N |x_i|$
- $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^N |x_i|^2}$
- $\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq N} |x_i|$.

A finite dimensional space endowed with an inner product is said to be *pre-Hilbertian*. An inner product (Hermitian, if $\mathbb{K} = \mathbb{C}$) is a bilinear (sesquilinear) application verifying the following properties:

- symmetry: $(\mathbf{x}, \mathbf{y}) = (\mathbf{y}, \mathbf{x})^*$
- positivity: $(\mathbf{x}, \mathbf{x}) \in \mathbb{R}_+$
- definiteness: $(\mathbf{x}, \mathbf{x}) = 0 \Leftrightarrow \mathbf{x} = 0$.

An inner product induces a norm $\|\mathbf{x}\| = \sqrt{(\mathbf{x}, \mathbf{x})}$. Specifically, the inner product $(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^N x_i y_i^*$ induces the norm $\|\cdot\|_2$. The Cauchy-Schwarz equality states that:

$$|(\mathbf{x}, \mathbf{y})| \leq \|\mathbf{x}\| \|\mathbf{y}\|. \quad (3)$$

One can then define orthonormal bases, *i.e.*, bases $(\mathbf{e}_i)_{1 \leq i \leq N}$ such that $(\mathbf{e}_i, \mathbf{e}_j) = 1$ if $i = j$, and 0 otherwise. For orthonormal bases, the decomposition of $\mathbf{x}$ in terms of basis vectors $\mathbf{e}_i$ reads:

$$\mathbf{x} = \sum_{i=1}^N x_i \mathbf{e}_i \quad (4)$$

with weights

$$x_i = (\mathbf{x}, \mathbf{e}_i). \quad (5)$$

1.2 BANACH SPACES

Any space of finite dimension admits a basis. We can therefore always represent a vector by its coordinates, and all norms defined on a given vector space are equivalent. Moreover, any finite-dimensional vector space on $\mathbb{R}$ or $\mathbb{C}$ is complete.

The situation is far more complex in infinite dimension. We call Banach space a complete and normed vector space. We will focus our attention on two types of Banach spaces.

1.2.1 ℓ^p spaces

The ℓ^p space ($p \geq 1$) is the space of sequences $(x[n])_{n \in \mathbb{Z}}$ such that $\sum_{n \in \mathbb{Z}} |x[n]|^p < \infty$. Specifically,

- ℓ^1 is the space of integrable sequences, $\|x\|_1 = \sum_{n \in \mathbb{Z}} |x[n]|$
- ℓ^2 is the space of finite energy sequences, $\|x\|_2^2 = \sum_{n \in \mathbb{Z}} |x[n]|^2$
- ℓ^∞ is the space of bounded sequences, $\|x\|_\infty = \sup_{n \in \mathbb{Z}} |x[n]|$.

We denote $\delta[n]$ the sequence defined by

$$\delta[n] = \begin{cases} 1 & \text{if } n = 0, \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

δ is called the **Kronecker function** or **Kronecker impulse**.

Remark. Notation δ will also refer to the Dirac distribution. Unlike the Dirac distribution, the Kronecker function $\delta[n]$ is a sequence defined on $\mathbb{Z}$.

1.2.2 L^p spaces

Similarly, the spaces $L^p(I)$ (with I an interval of $\mathbb{R}$ and $1 \leq p \leq \infty$) are Banach spaces. We first define $\mathcal{L}^p(I)$, the set of measurable functions x defined on I such that $\int_I |x|^p d\lambda < \infty$. The space $L^p(I)$ is then the set of equivalence classes of $\mathcal{L}^p(I)$ by the relation "equal almost everywhere". We will specifically focus on:

- integrable functions $L^1(\mathbb{R})$, $\|x\|_{L^1} = \int_{\mathbb{R}} |x| d\lambda$,
- finite energy functions $L^2(\mathbb{R})$, $\|x\|_{L^2}^2 = \int_{\mathbb{R}} |x|^2 d\lambda$,
- essentially bounded functions $L^\infty(\mathbb{R})$, $\|x\|_{L^\infty} = \sup |x|$.

Note that these spaces do not take into account the regularity of functions (i.e., the functions in L^p have no reason to be continuous). Also, there is no systematic inclusion properties (except when I is bounded, in which case $L^p(I) \subset L^q(I)$ for $p > q$).

APPLICATION IN SIGNAL PROCESSING: ℓ^p spaces will be used to model discrete-time signals while L^p spaces will refer to continuous time signals. When $p = 2$ (see the next section), we will refer to finite energy signals, where the energy is $\|\cdot\|_2^2$.

1.3 HILBERT SPACES

The particular case of $L^2(I)$ deserves specific attention. In the case where the interval I is not bounded, we define the inner product in $L^2(I)$ by:

$$(x, y)_{L^2} = \int_{\mathbb{R}} xy^* d\lambda. \quad (7)$$

When I is bounded, the inner product is normalized by the length of I^1 . Thus, the inner product on $L^2([0, T])$ is defined by:

$$(x, y)_{L^2} = \frac{1}{T} \int_{[0, T]} xy^* d\lambda. \quad (8)$$

The inner product obviously satisfies the Cauchy-Schwarz inequality.

$L^2(I)$ is a Hilbert space, *i.e.*, a Banach space with an inner product. Moreover, it is separable (there is a countable subset dense in L^2). So, there exists a Hilbert basis $(e_i)_{i \in \mathbb{Z}}$: any function $x \in L^2(I)$ can be written as:

$$x = \sum_{i \in \mathbb{Z}} c_i e_i \quad (9)$$

with convergence in L^2 norm. Moreover, the family $(e_i)_{i \in \mathbb{Z}}$ is orthonormal. The coefficients c_i are then given by:

$$c_i = (x, e_i)_{L^2}. \quad (10)$$

The Riesz representation theorem is an important property of Hilbert spaces: a continuous linear form ℓ on a Hilbert space H can always be written

$$\forall x \in H, \ell(x) = (x, x_\ell)_H$$

where $x_\ell \in H$. We can therefore identify H with its dual.

We can also define the Sobolev spaces H^m , whose functions x are such that x and its derivatives up to order m (in the sense of distributions) are in L^2 . These spaces take into account the regularity of functions. For the special case of H^1 , the inner product is:

$$(x, y)_{H^1} = (x, y)_{L^2} + (x', y')_{L^2}.$$

APPLICATION IN SIGNAL PROCESSING : the construction of bases of Hilbert spaces such as $L^2(\mathbb{R})$ is very important since it is linked to the decomposition of continuous time signals. We can mention discrete wavelet bases or local cosine bases, and for signals with limited bandwidth, the bases of cardinal signs, allowing to represent a sampled signal without loss of information.

1.3.1 Fourier series

The space $L^2([0, T])$ of integrable square functions defined on $[0, T]$ is a separable Hilbert space. A classical orthonormal basis of $L^2([0, T])$ is the Fourier basis defined by:

$$e_n(t) = \exp\left(i2\pi \frac{nt}{T}\right), \quad n \in \mathbb{Z}. \quad (11)$$

¹ for reasons that will become clearer in paragraph 1.3.1.

The decomposition reads:

$$x = \sum_{n \in \mathbb{Z}} c_n e_n \quad (12)$$

$$x(t) = \sum_{n \in \mathbb{Z}} c_n \exp\left(i2\pi \frac{nt}{T}\right) \quad (13)$$

with weights given by:

$$\begin{aligned} c_n &= (x, e_n)_{L^2} \\ &= \frac{1}{T} \int_{[0, T]} x(t) \exp\left(-i2\pi \frac{nt}{T}\right) dt \end{aligned} \quad (14)$$

APPLICATION IN SIGNAL PROCESSING: Fourier series allow to decompose a periodic signal (speech signal, musical note, etc.) into sums of harmonics. The fact that a function belongs to a Sobolev space H^m guarantees the fast decrease of coefficients $|c_n|$.

1.4 DISTRIBUTIONS

As mentioned above, in a Hilbert space H (for example L^2), one can identify the vectors of H with continuous linear forms over H . This is especially true for finite dimension spaces. On the other hand, for general L^p spaces ($p \neq 2$), this is no longer true. For $p > 1$, the dual of L^p is isomorphic to L^q with $1/p + 1/q = 1$. For L^p spaces defined on a bounded interval (for example, $[0, 1]$), we notice that for $p > 2$, the dual L^p is not equal to L^p and is even strictly larger (indeed, it is L^q , with $q = p/(p-1) < p$). We can therefore make the dual grow by limiting the starting space.

This idea is pushed to the extreme by the distributions. Let $\mathcal{D}'(\mathbb{R})$ denote the distribution space, that is, the set of continuous linear forms on the space $\mathcal{D}(\mathbb{R})$ of compactly supported functions C^∞ , called test functions. The starting space being here very restricted, its dual is absolutely gigantic.

The space $\mathcal{D}'(\mathbb{R})$ includes regular distributions associated with locally integrable functions (and therefore, all functions of L^p , for $1 \leq p \leq \infty$). In this case, the evaluation of a test function φ by a regular distribution T_x (related to function x) is given by

$$\langle T_x, \varphi \rangle = \int_{\mathbb{R}} x \varphi d\lambda. \quad (15)$$

However, not all distributions are associated with functions. In particular, the Dirac distribution δ

$$\langle \delta, \varphi \rangle = \varphi(0) \quad (16)$$

cannot be written as the integral of φ against a function. More generally, we define δ_a by

$$\langle \delta_a, \varphi \rangle = \varphi(a). \quad (17)$$

We have in particular

$$\delta_0 = \delta. \quad (18)$$

Warning: do not confuse the Dirac distribution δ with the Kronecker function $\delta[k]$ defined in (6). The latter is a numerical sequence.

We also define the Dirac comb as the sum of Dirac distributions spaced by a :

$$\mathbb{L}_a = \sum_{n \in \mathbb{Z}} \delta_{na}. \quad (19)$$

Equality and convergence in the sense of distributions are defined by:

$$\begin{aligned} T = S &\Leftrightarrow \forall \varphi \in \mathcal{D}(\mathbb{R}), \langle T, \varphi \rangle = \langle S, \varphi \rangle, \\ T_n \rightarrow T &\Leftrightarrow \forall \varphi \in \mathcal{D}(\mathbb{R}), \langle T_n, \varphi \rangle \rightarrow \langle T, \varphi \rangle. \end{aligned}$$

Every distribution is differentiable. Its derivative is still a distribution, defined by

$$\langle T', \varphi \rangle = -\langle T, \varphi' \rangle, \quad (20)$$

which extends the integration by parts when T is a sufficiently regular function. In particular, the Heaviside function $x = \mathbb{1}_{\mathbb{R}^+}$ is not differentiable in the classical sense, but its derivative in the sense of distributions is the Dirac distribution δ (note: the value of $\mathbb{1}_{\mathbb{R}^+}$ in zero is, from the point of view of the distributions, of no importance). This result can be extended as follows: let $I = [a, b]$ be an interval of $\mathbb{R}$ and x a differentiable function. The derivative of $x\mathbb{1}_I$ is given by

$$(x\mathbb{1}_I)' = x'\mathbb{1}_I + x(a)\delta_a - x(b)\delta_b. \quad (21)$$

The discontinuities of the function lead to Dirac distributions in the derivative, with amplitudes equal to those of the jumps.

An affine change of variables $\chi(x) = ax + b$ is defined as follows:

$$\langle T \circ \chi, \varphi \rangle = \frac{1}{|a|} \langle T, \varphi \circ \chi^{-1} \rangle. \quad (22)$$

Note that when T is a regular distribution (associated to some locally integrable function f), (22) identifies with the formula obtained by calculating the integral of $(f \circ \chi)\varphi$ and changing variables within this integral.

Specifically, for the Dirac distribution and $\chi(x) = ax - b$, we get:

$$\delta_0 \circ \chi = \frac{1}{|a|} \delta_{b/a}. \quad (23)$$

APPLICATION IN SIGNAL PROCESSING : Distributions are useful for modeling linear time invariant systems, defined later in the document, by convolutions. In particular, the neutral element of the convolution is the distribution δ . The δ_a operator acts as a delay on a signal, with a shift in time equal to a . We will sometimes use the notation $\delta(t - a)$ instead of δ_a , to make this delay more explicit. We should keep in mind that a distribution is

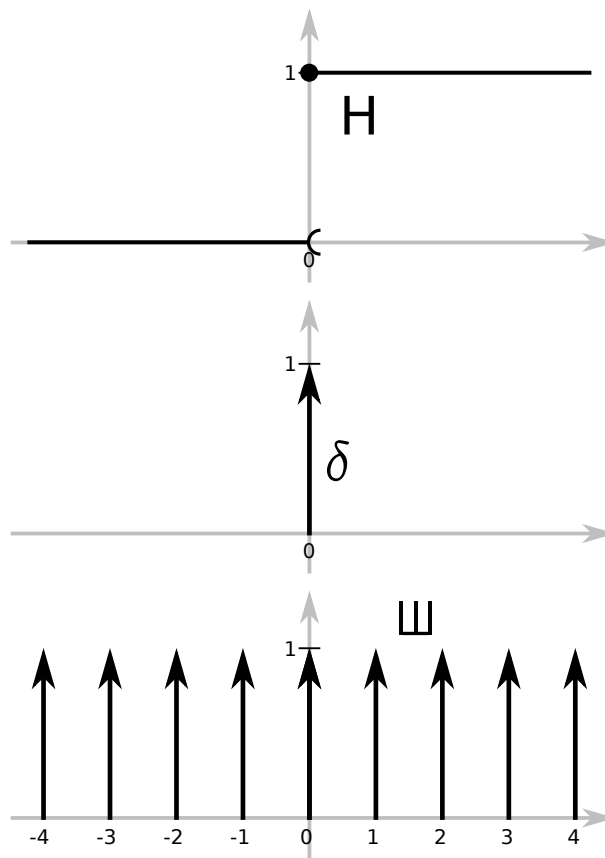


Figure 1: Heaviside function $H = \mathbb{1}_{\mathbb{R}_+}$, Dirac distribution $\delta = \delta_0$ and Dirac comb $\mathbb{W}_a$ for $a = 1$.

not a function of t , and that this notation is nothing but a convenient abuse. Note that the Dirac distribution can be seen as the limit, in the sense of distributions, of a sequence of impulses of integral 1 and supports more and more concentrated around zero. The Dirac comb will be important for the study of the sampling operation (transition from a continuous time signal to a discrete time signal).

1.4.1 Tempered distributions

We define the Schwartz space $\mathcal{S}$ as the set of C^∞ functions φ on $\mathbb{R}$ such that

$$\forall \alpha, \beta \in \mathbb{N}, \sup_{t \in \mathbb{R}} |t^\alpha \varphi^{(\beta)}(t)| < \infty, \quad (24)$$

that is, the derivatives $\varphi^{(\beta)}(t)$ decrease faster than any negative power of $|t|$. The members of its dual $\mathcal{S}'$ are called tempered distributions. The test functions being in the Schwartz space ($\mathcal{D} \subset \mathcal{S}$), the tempered distributions are also distributions ($\mathcal{S}' \subset \mathcal{D}'$).

Some examples of tempered distributions include:

- locally integrable functions with at most polynomial growth,
- Dirac distributions,
- Dirac combs.

1.4.2 Compactly supported distributions

Before defining distributions with compact support, we recall that the support of a function, denoted $\text{supp}(\varphi)$, is defined as the closure of the set of values x for which $\varphi(x) \neq 0$.

Definition 1.1. A distribution $T \in \mathcal{D}'(\mathbb{R})$ is said to be zero on a open set $\mathcal{O}$ of $\mathbb{R}$ if for all $\varphi \in \mathcal{D}(\mathbb{R})$ such that $\text{supp}(\varphi) \subset \mathcal{O}$, $\langle T, \varphi \rangle = 0$.

For example, one can easily verify that δ is zero on $\mathbb{R}^*$ and that for all $x \in L^1(\mathbb{R})$, “ T_x is zero” on $\mathcal{O}$ is equivalent to $x|_{\mathcal{O}} = 0$ almost everywhere.

We now extend the concept of support to distributions.

Definition 1.2 (Distribution support). The support of a distribution $T \in \mathcal{D}'(\mathbb{R})$ is the complement of the largest open where T is zero.

Example 1.1.

- If $a_i \in \mathbb{R}$ and $\alpha_i \in \mathbb{C}$, $\text{supp}(\sum_{i=1}^k \alpha_i \delta_{a_i}) = \{a_1, \dots, a_k\}$.
- If $x \in L^1(\mathbb{R})$, $\text{supp}(T_x) = \text{supp}(x)$.

Definition 1.3 (Distributions with compact support). *We denote $\mathcal{E}'(\mathbb{R})$ the subspace of $\mathcal{D}'(\mathbb{R})$ formed of distributions with compact support.*

$\mathcal{E}'(\mathbb{R})$ contains the Dirac distributions δ_a , their derivatives $\delta_a^{(k)}$ for all $k \in \mathbb{N}$, and the set of functions $x \in L^1(\mathbb{R})$ with compact support (seen as regular distributions). $\mathcal{E}'(\mathbb{R})$ contains the space of test functions: $\mathcal{D}(\mathbb{R}) \subset \mathcal{E}'(\mathbb{R})$.

We also have the following nesting property:

$$\mathcal{E}'(\mathbb{R}) \subset \mathcal{S}'(\mathbb{R}) \subset \mathcal{D}'(\mathbb{R}). \quad (25)$$

APPLICATION IN SIGNAL PROCESSING: The tempered distribution space is the most general framework for defining a Fourier transform. They allow to calculate the Fourier transform of a Dirac distribution or a Dirac comb, but also a complex exponential $e(t) = \exp(i2\pi f_0 t)$, which are neither in L^1 nor in L^2 .

1.5 RANDOM VARIABLES AND PROCESSES

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. A random variable X is a measurable application from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B})$ where $\mathcal{B}$ refers to the Borel sigma-algebra. A complex random variable X_c can be decomposed as $X_c = X_r + iX_i$, where X_r and X_i are real random variables.

The law of the random variable X is the measure $\mathbf{P}_X$ on $\mathbb{R}$ such that $\mathbf{P}_X(A) = \mathbf{P}(X^{-1}(A))$. The expectation of X is defined by:

$$\begin{aligned} \mathbb{E}(X) &= \int_{\Omega} X(\omega) d\mathbf{P}(\omega) \\ &= \int_{\mathbb{R}} x d\mathbf{P}_X(x). \end{aligned} \quad (26)$$

Its variance $\text{Var}(X)$ reads

$$\begin{aligned} \text{Var}(X) &= \int_{\Omega} |X(\omega) - \mathbb{E}(X)|^2 d\mathbf{P}(\omega) \\ &= \int_{\mathbb{R}} |x - \mathbb{E}(X)|^2 d\mathbf{P}_X(x). \end{aligned} \quad (27)$$

Moreover, a random variable is fully characterized by its cumulative function

$$F_X(x) = \mathbf{P}(X \leq x) \quad (28)$$

and its characteristic function

$$\varphi_X(t) = \mathbb{E}(\exp(itX)). \quad (29)$$

If the law $\mathbf{P}_X(x)$ is absolutely continuous with respect to the Lebesgue measure (i.e., $d\mathbf{P}_X = f_X d\lambda$), we say that X admits the probability density

function f_X . A very important case of density function is the normal distribution $\mathcal{N}(\mu, \sigma^2)$, with mean μ and variance σ^2 , defined by:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \quad (30)$$

Random vectors are random variables of Ω in $\mathbb{R}^N$. Their probability law is a measure on $\mathbb{R}^N$. For a pair of random variables (X, Y) , we define the covariance by:

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))^*) \\ &= \int_{\mathbb{R}^2} (x - \mathbb{E}(X))(y - \mathbb{E}(Y))^* d\mathbf{P}_{XY}(x, y), \end{aligned} \quad (31)$$

denoting $\mathbf{X}^*$ the Hermitian conjugate of $\mathbf{X}$ (i.e., the conjugate transpose matrix). The mean of a random vector $\mathbf{X}$ is $\mu_{\mathbf{X}} = \mathbb{E}(\mathbf{X})$ and the covariance matrix reads:

$$\Sigma_{\mathbf{X}} = \mathbb{E}((\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{X} - \mu_{\mathbf{X}})^*), \quad (32)$$

whose coefficients are equal to $\Sigma_{\mathbf{X}, nm} = \text{Cov}(X_n, X_m)$. The covariance matrix is always Hermitian positive semi-definite. Indeed, for any vector $\mathbf{a} = (a_j)_{1 \leq j \leq N}$, $\mathbf{a}^* \Sigma_{\mathbf{X}} \mathbf{a} = \text{Var}(\langle \mathbf{a}, \mathbf{X} \rangle) \geq 0$.

By definition, a Gaussian random vector is such that any linear combination of its components follows a normal law. A real Gaussian random vector $\mathbf{X}$ is fully defined by its mean μ and covariance matrix Σ . Its characteristic function reads:

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \exp\left(i\mu_{\mathbf{X}}^t \mathbf{t} - \frac{1}{2} \mathbf{t}^t \Sigma \mathbf{t}\right). \quad (33)$$

Moreover, if Σ is invertible, the random vector admits a density given by:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{N/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{x} - \mu_{\mathbf{X}})^t \Sigma^{-1} (\mathbf{x} - \mu_{\mathbf{X}})\right). \quad (34)$$

Fig. 2 gives an example of a Gaussian vector. Note that it is not sufficient that the marginal distributions of the components of the vector are Gaussian random variables for the vector to be Gaussian. This is illustrated by the example of fig. 3.

A *discrete time random process* $(X_n(\omega))$ is a sequence of random variables for $n \in \mathbb{Z}$. This concept generalizes that of random vectors. A process is said to be Gaussian if any vector extracted from the process is Gaussian. Realizations of the process $(X_n(\omega))$, i.e., values of the process obtained for given ω 's, are also called *trajectories*. Fig. 4 shows realizations of a simple random process, where each X_k takes values -1 and 1, and where $X_{k+1} = X_k$ with probability 0.9, and $X_{k+1} = -X_k$ with probability 0.1.

APPLICATION IN SIGNAL PROCESSING: many signals can be modeled by random processes, in the sense that it is useful to think of them as realizations of a random process. This type of modeling is, for example, very useful for unvoiced speech signals (whisper, consonants such as /ch/, /s/, /f/), for

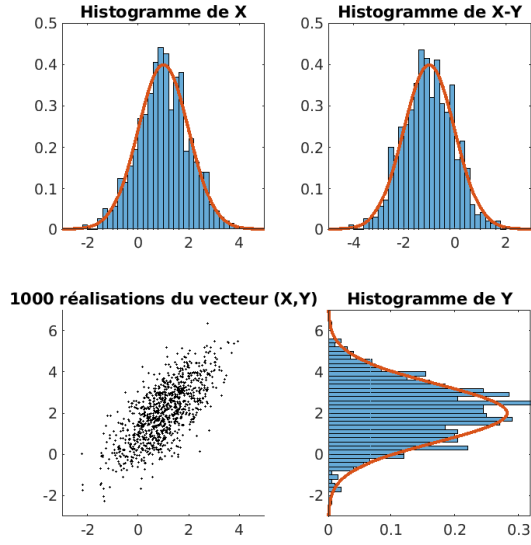


Figure 2: Realizations of a Gaussian vector (X, Y) of mean $\mu = (1, 2)$ and covariance matrix $\Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$. The laws of any linear combination X and Y are normal.

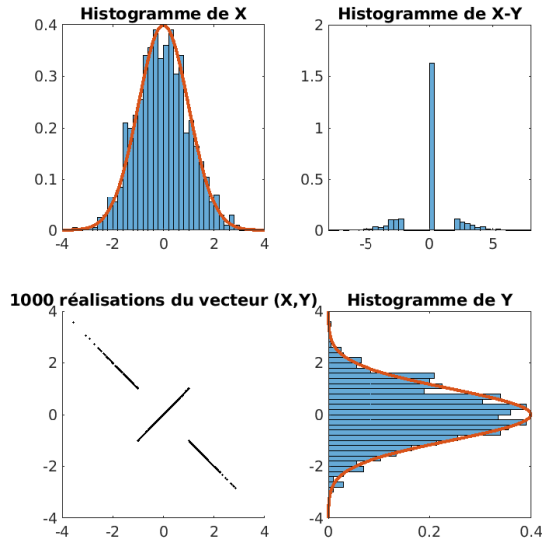


Figure 3: Realizations of a random vector (X, Y) , where $Y = X\mathbb{1}_{|X|<1} - X\mathbb{1}_{|X|\geq 1}$ and X is a Gaussian variable of law $\mathcal{N}(0, 1)$. The marginal laws of X and Y are Gaussian, but $X - Y$ does not follow a normal distribution. (X, Y) is therefore not a Gaussian vector.

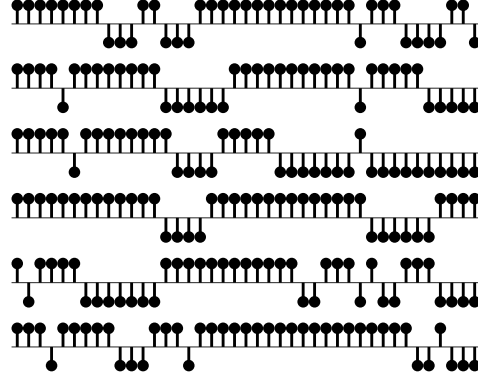


Figure 4: Six trajectories of a random process with values -1 or 1 .

which it is more informative to give the characteristics of the random process that produced them, rather than the values of their specific trajectories, taking into account the variability between trajectories. In signal processing, we will manipulate the autocorrelation of a signal (between two instants n and m), equal to $\mathbb{E}(X_n X_m^*)$, which coincides with the element $\Sigma_{X,nm}$ when $\mathbf{X}$ is centered.

1.5.1 Brief reminders of statistics

The purpose of statistics is to estimate characteristics of a random variable (mean, variance, parameters of a family of probability distributions, etc.) from realizations of this random variable. For example, from a finite number of dice rolls, we seek to estimate the probability of each face of the dice.

Here is a simple example of estimation. The empirical mean $\hat{\mu}_N$ is an estimator of the mean μ of a random variable X . With (mutually) independent and identically distributed (*i.i.d.*) data X_n for $1 \leq n \leq N$,

$$\hat{\mu}_N = \frac{1}{N} \sum_{n=1}^N X_n. \quad (35)$$

This estimator is also a random variable. We can calculate its expectation and variance:

$$\mathbb{E}(\hat{\mu}_N) = \mathbb{E}(X) \quad (36)$$

$$\text{Var}(\hat{\mu}_N) = \frac{1}{N^2} \sum_{n=1}^N \text{Var}(X_n) \quad (37)$$

$$= \frac{1}{N} \text{Var}(X). \quad (38)$$

This estimator is unbiased (its expectation is equal to that of X , which we seek to estimate), and consistent (its variance tends to 0 when the number of data points tends to infinity).

APPLICATION IN SIGNAL PROCESSING: as seen before, it is common to think of a signal as the realization of a random process. With some additional assumptions (stationarity, ergodicity), we will see that it is possible to estimate the parameters of a random process from a unique realization. For example, one can estimate the parameters of a speech generative model from a single recording. The problem of estimating the frequency of a sinusoid in noise will also be tackled within random process framework.

 LINEAR TIME INVARIANT SYSTEMS

This chapter studies linear and time invariant filters, that are fundamental in signal processing. We first give basic definitions of continuous-time and discrete-time signals. Then, we define the convolution product of two signals and exhibit its main mathematical properties. We study the properties of linear convolution filters, and show that the complex exponential signals are eigenfunctions of any linear filter. As such, they characterize the filter in the frequency domain.

2.1 SIGNAL CONCEPT

2.1.1 Definitions

A deterministic signal, or more simply a signal, is a function of one or more time or space variables, with values in $\mathbb{R}$ or $\mathbb{C}$. There are two major classes of signals: continuous-time and discrete-time signals.

A **continuous (time) signal**, called *analog signal*, is a function of a continuous variable $t \in \mathbb{R}$. We will denote it

$$\begin{aligned} x : \mathbb{R} &\rightarrow \mathbb{C} \\ t &\mapsto x(t). \end{aligned}$$

Variable t can represent time, but also space ($t = (x, y)$ for images), wavelength, etc.

A **discrete (time) signal** is a sequence

$$x = \{x[k], k \in \mathbb{Z}\}.$$

Very often, a discrete signal is obtained by *sampling* the values of a continuous time signal at different times $t_k, k \in \mathbb{Z}$. We thus have:

$$x[k] = x_c(t_k), \forall k \in \mathbb{Z},$$

denoting x_c the continuous signal associated with the discrete signal x . For this reason, discrete signals are also called sampled signals. Note that a **digital signal** is a discrete and quantized signal, i.e., it can only take a finite number of values. *Quantization* is the operation $q = Q(x)$ which transforms a signal with continuous values ($x[k] \in \mathbb{R}$) into a signal with discrete values. For example, if q is $\mathbb{N}$ -valued, $q[k]$ is the nearest integer to $x[k]$.

Deterministic signals will be seen as mathematical objects living in functional spaces like $L^p(\mathbb{R})$ (continuous time) or $\ell^p(\mathbb{Z})$ (discrete time). In contrast to deterministic signals, we define *stochastic signals* as random processes. They will be studied in chapters 5 and 6.

2.1.2 Energy representations

Let x be a continuous-time deterministic signal. We define when they exist, the following quantities:

- Mean power $P_x = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{[-T, T]} |x(t)|^2 dt$
- Energy $E_x = \int_{\mathbb{R}} |x(t)|^2 dt = \|x\|_{L^2}^2$.

A signal of finite energy therefore satisfies $x \in L^2(\mathbb{R})$ and $P_x = 0$. Similar definitions apply in the discrete case, with $E_x = \|x\|_2^2$.

We define the cross-correlation between two finite energy signals by:

$$\forall \tau \in \mathbb{R}, \gamma_{xy}(\tau) = \int_{\mathbb{R}} x(t) y^*(t - \tau) dt \quad (\text{continuous case}) \quad (39)$$

$$\forall n \in \mathbb{Z}, \gamma_{xy}[n] = \sum_k x[k] y^*[k - n] \quad (\text{discrete case}) \quad (40)$$

The cross-correlation is equal to the inner product between x and the delayed version of y :

$$\begin{aligned} \gamma_{xy}(\tau) &= (x, y(\cdot - \tau)) & (\text{continuous case}) \\ \gamma_{xy}[n] &= (x, y[\cdot - n]) & (\text{discrete case}) \end{aligned}$$

The Signal-to-Noise Ratio (SNR) quantifies the level of disturbance of a signal during its transmission in a communication network. Denoting by u the noiseless signal and $x = u + b$ the noisy signal (assuming an additive noise b), the SNR is defined by:

$$\text{SNR} = \left(\frac{P_u}{P_b} \right) \quad \text{or} \quad \text{SNR} = \left(\frac{E_u}{E_b} \right) \quad (41)$$

for finite power and finite energy signals, respectively. The SNR is the most often expressed in dB:

$$\text{SNR}_{\text{dB}} = 10 \log_{10}(\text{SNR}). \quad (42)$$

2.2 CONVOLUTION

Convolution is a fundamental operation in signal processing. We first define it in L^p spaces, and then for discrete signals (ℓ^p spaces) and for distributions.

2.2.1 Continuous case: convolutions of functions

Definition 2.1 (Convolution of two functions). Consider two functions $x : \mathbb{R} \rightarrow \mathbb{C}$ and $y : \mathbb{R} \rightarrow \mathbb{C}$. The convolution product of x and y , denoted $x * y$, is the function, if defined, defined by:

$$\forall t \in \mathbb{R}, (x * y)(t) = \int_{\mathbb{R}} x(\theta)y(t - \theta)d\theta \quad (43)$$

$$= \int_{\mathbb{R}} x(t - \theta)y(\theta)d\theta. \quad (44)$$

$(x * y)(t)$ is sometimes denoted $x * y(t)$. In this handout, we favor the writing $(x * y)(t)$ to clearly mark that $(x * y)(t)$ depends on all values of $y(\theta)$.

Note that the convolution is not always defined. For example, if $x = y = \mathbb{1}_{\mathbb{R}}$, then $\forall t, (x * y)(t) = \infty$. Furthermore, the convolution has a regularizing power (see CIP course). For instance, for discontinuous functions $x = y = \mathbb{1}_{[-\frac{T}{2}, \frac{T}{2}]}$, where $T > 0$, $x * y$ is the "triangle" (continuous) function of width $2T$ and amplitude T :

$$(x * y)(t) = (T - |t|) \mathbb{1}_{[-T, T]}(t). \quad (45)$$

Convolution is defined everywhere (possibly almost everywhere) depending on the space in which the x and y signals live.

Theorem 2.1. The convolution of two functions in L^p and L^q is defined in the following cases:

x	y	$x * y$	defined
L^1	L^1	L^1	almost everywhere
L^1	L^2	L^2	almost everywhere
L^1	L^∞	L^∞	everywhere
L^2	L^2	L^∞	everywhere

The norm of $x * y$ is bounded by the product of the norms of x and y .

The proof is elementary and left as an exercise.

Going further...

Theorem 2.1 is generalized using Young's inequality. Let $p, q, r \in [1, \infty]$ with

$$\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}.$$

The convolution of $x \in L^p$ and $y \in L^q$ is in L^r . Moreover,

$$\|x * y\|_r \leq \|x\|_p \|y\|_q.$$

Remark. Equation (43) is closely linked to the cross-correlation equation, see (39). For x and $y \in L^2(\mathbb{R})$, $\gamma_{xy}(\tau)$ is equal to the inner product between x and the delayed signal $y(\cdot - \tau)$. By denoting $\bar{y}(t) = y^*(-t)$, we have

$$\gamma_{xy} = x * \bar{y}. \quad (46)$$

2.2.2 Discrete case

Definition 2.2 (Convolution in the discrete case). Let $x = \{x[k], k \in \mathbb{Z}\}$ and $y = \{y[k], k \in \mathbb{Z}\}$. The convolution product of x and y is the sequence $x * y$, if defined, defined by:

$$\forall n \in \mathbb{Z}, (x * y)[n] = \sum_{k \in \mathbb{Z}} x[k] y[n - k] \quad (47)$$

$$= \sum_{k \in \mathbb{Z}} x[n - k] y[k]. \quad (48)$$

$(x * y)[n]$ is sometimes denoted $x * y[n]$. We prefer writing $(x * y)[n]$.

The existence conditions are similar to the continuous case; we can write the convolution of two signals in $\ell^1(\mathbb{Z})$, of two signals in $\ell^2(\mathbb{Z})$, of a signal $\ell^1(\mathbb{Z})$ with a signal in $\ell^\infty(\mathbb{Z})$, etc.

2.2.3 Signals modeled by distributions

It is necessary to extend convolution to distributions to be able to define some filters like delay filters (55) $y = \delta_\tau * x$. The reader is redirected to chapter 1 for a definition of distribution spaces on which we are working.

Convolution of a C^∞ function and a distribution

The convolution of a $C^\infty(\mathbb{R})$ function with a distribution is defined as a direct extension of (44).

Definition 2.3. Let $x \in C^\infty(\mathbb{R})$ and $Y \in \mathcal{D}'(\mathbb{R})$. Assume that one of the following assumptions is satisfied:

1. $x \in \mathcal{D}(\mathbb{R})$ and $Y \in \mathcal{D}'(\mathbb{R})$;
2. $x \in \mathcal{S}(\mathbb{R})$ and $Y \in \mathcal{S}'(\mathbb{R})$;
3. $x \in C^\infty(\mathbb{R})$ and $Y \in \mathcal{E}'(\mathbb{R})$.

The function $x * Y$ is defined by:

$$\forall t \in \mathbb{R}, (x * Y)(t) = \langle Y, x(t - \cdot) \rangle \quad (49)$$

where $x(t - \cdot)$ denotes the function $\theta \mapsto x(t - \theta)$.

Theorem 2.2. Under the previous assumptions, $x * Y \in C^\infty(\mathbb{R})$. Moreover, for all $k \in \mathbb{N}$, $(x * Y)^{(k)} = x^{(k)} * Y = x * Y^{(k)}$.

Example 2.1. For $Y = \delta_a$ and $x \in C^\infty(\mathbb{R})$, (49) implies that:

$$\forall t \in \mathbb{R}, (x * \delta_a)(t) = x(t - a). \quad (50)$$

Going further...

Definition 2.4 (Convolution of distributions). Let $X \in \mathcal{E}'(\mathbb{R})$ and $Y \in \mathcal{D}'(\mathbb{R})$. The convolution product $X * Y$ is the distribution defined by:

$$\forall \phi \in \mathcal{D}(\mathbb{R}), \langle X * Y, \phi \rangle = \langle X, \theta \mapsto \langle Y, \phi(\cdot + \theta) \rangle \rangle \quad (51)$$

$$= \langle Y, \theta \mapsto \langle X, \phi(\cdot + \theta) \rangle \rangle \quad (52)$$

This definition has a direct link with the definition (43)-(44) of the convolution of two functions x and y . Let $x \in L^1(\mathbb{R})$ and $y \in L^1(\mathbb{R})$. According to Theorem 2.1, $x * y \in L^1(\mathbb{R})$. The related regular distribution reads:

$$\begin{aligned} \forall \phi \in \mathcal{D}(\mathbb{R}), \langle T_{x*y}, \phi \rangle &= \int_{\mathbb{R}} (x * y)(t) \phi(t) dt \\ &= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} x(\theta) y(t - \theta) d\theta \right] \phi(t) dt \\ &= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} y(t - \theta) \phi(t) dt \right] x(\theta) d\theta \\ &= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} y(u) \phi(u + \theta) du \right] x(\theta) d\theta \\ &= \int_{\mathbb{R}} x(\theta) \langle T_y, \phi(\cdot + \theta) \rangle d\theta \end{aligned} \quad (53)$$

using Fubini's theorem to invert both integrals. (53) clearly identifies with (51). Similarly, one can establish a direct link between (52) and (44).

Remark. The convolution product of two distributions is not always defined. Due to the nesting property $\mathcal{E}' \subset \mathcal{S}' \subset \mathcal{D}'$, the above definition (with $X \in \mathcal{E}'$, $Y \in \mathcal{D}'$) covers the following cases:

- $X \in \mathcal{D}$, $Y \in \mathcal{D}'$;
- X and $Y \in \mathcal{E}'$. Here, we have $X * Y \in \mathcal{E}'$;
- $X \in \mathcal{E}'$ and $Y \in \mathcal{S}'$. Here, we have $X * Y \in \mathcal{S}'$.

Example 2.2.

- For $X = \delta_a \in \mathcal{E}'(\mathbb{R})$, we have for $Y \in \mathcal{D}'(\mathbb{R})$,

$$\forall \phi \in \mathcal{D}(\mathbb{R}), \langle \delta_a * Y, \phi \rangle = \langle Y, \phi(\cdot + a) \rangle.$$

For $a = 0$, this yields $\delta * Y = Y$.

- In the specific case where $Y = y \in C^0(\mathbb{R})$, we have:

$$\langle \delta_a * y, \phi \rangle = \int_{\mathbb{R}} y(t) \phi(t + a) dt = \int_{\mathbb{R}} y(t - a) \phi(t) dt.$$

We have therefore (in the distributions sense) $(\delta_a * y)(t) = y(t - a)$ which extends (50) to the convolution of δ_a with a continuous function.

2.2.4 Properties of the convolution product

The properties are stated whatever the nature of the signals x , y and z provided that the manipulated quantities are well defined.

- Commutativity: $x * y = y * x$
- Associativity: $x * (y * z) = (x * y) * z$
- Distributivity: $x * (y + z) = (x * y) + (x * z)$
- Neutral element (continuous case) is the Dirac distribution $\delta_0 = \delta$:
 $\delta * y = y$.
- Neutral element (discrete case) is the Kronecker function $\delta[k]$:
 $\delta * y = y$.
- Translation: $x(t - \tau) = (x * \delta_\tau)(t)$
- Differentiation (possibly in the sense of distributions): $\frac{d}{dt}(x * y) = \frac{dx}{dt} * y = x * \frac{dy}{dt}$
- Support (possibly in the sense of distributions): $\text{supp}(x * y) \subset \text{supp}(x) + \text{supp}(y) = \{t_1 + t_2 : t_1 \in \text{supp}(x), t_2 \in \text{supp}(y)\}$.
- Cross-correlation: $\gamma_{xy} = x * \bar{y}$ with $\bar{y}(t) = y^*(-t)$.

2.3 CONCEPT OF SYSTEM

2.3.1 Definition

A system is a mapping $y = S(x)$ from the space of input signals x to that of output signals y . The spaces considered are typically L^p spaces for continuous signals and ℓ^p spaces for discrete signals. Schematic representation of systems is based on the "black box" diagram of Fig. 5.

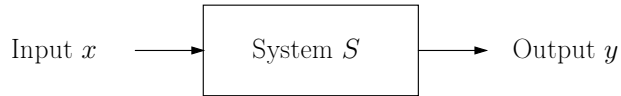


Figure 5: Graphical representation of a system.

A system models transformations undergone by a signal, including the alteration of a physical signal (audio signal, electromagnetic wave) during its transmission through a propagation medium, or the improvement of the signal content after a signal processing task (denoising of a sound record, deconvolution to remove blur in an image). Hereafter, we only consider single input - single output systems.

2.3.2 Properties of systems

Definition 2.5 (Homogeneity). *A system is said to be homogeneous in input/output, if the input and output signals are of the same nature (continuous or discrete).*

Definition 2.6 (Linearity). *A system S is linear if for all linear combinations*

$$x = a_1 x_1 + a_2 x_2, \quad a_1, a_2 \in \mathbb{R},$$

we have

$$S(x) = a_1 S(x_1) + a_2 S(x_2).$$

In other words, the output signal $S(x)$ reads as the weighted sum of signals $S(x_1)$ and $S(x_2)$ with the same weights a_1 and a_2 as in the weighted sum $x = a_1 x_1 + a_2 x_2$.

Remark. *Physical systems are very rarely linear. A model S is an equation that represents the expected behavior of a phenomenon according to the laws of physics. We often perform a linearization around a so-called operating point (or equilibrium point) of the system to get a linear model.*

The following definitions are given for signals continuous time signals. Similar definitions apply to discrete time signals. We recall that $(\delta_\tau * x)(t) = x(t - \tau)$ represents a time shift of $x(t)$, also called translation. For $\tau > 0$, the signal $\delta_\tau * x$ is a delayed version of $x(t)$.

Definition 2.7 (Translation invariance). *A system S is invariant by translation if for any input x and for all $\tau \in \mathbb{R}$,*

$$S(\delta_\tau * x) = (S(x)) * \delta_\tau. \quad (54)$$

In other words, the output of the filter associated with the delayed signal coincides with the delayed version of the output signal $S(x)$, with the same delay.

Definition 2.8 (Homogeneous linear filter). *A linear filter is a system that satisfies both properties of linearity and translation invariance.*

The convolution $y = h * x$ is an important case of linear filtering, where h (called impulse response) characterizes the filter. We can check from the properties of the convolution product (see section 2.2.4), that this system is linear (obvious) and translation invariant, because the output associated to a delayed signal reads:

$$h * (\delta_\tau * x) = (h * x) * \delta_\tau.$$

Two remarkable examples are obtained when $h = \delta_\tau$ (time-delay system) and $h = \mathbb{1}_{\mathbb{R}_+}$ (integrator).

Example 2.3.

- *Time-delay:* $y = \delta_\tau * x$ or $y = \delta[\cdot - n] * x$:

$$y(t) = x(t - \tau), \quad \text{or} \quad y[k] = x[k - n]. \quad (55)$$

- *Integrator:* $y = \mathbb{1}_{\mathbb{R}_+} * x$ or $y = \mathbb{1}_{\mathbb{N}} * x$:

$$y(t) = \int_{-\infty}^t x(\tau) d\tau, \quad \text{or} \quad y[k] = \sum_{n=-\infty}^k x[n] \quad (56)$$

Definition 2.9 (Causal system). *A system is causal if for any time moment t_0 , the output signal at time t_0 only depends on the values of the input signal for times $t \leq t_0$.*

Example 2.4. *The integrator defined in (56) is a causal system. On the other hand, the calculation of the average in a centered interval*

$$y(t) = \frac{1}{2T} \int_{t-T}^{t+T} x(\theta) d\theta \quad (57)$$

is not causal, because it is necessary to know the value of the input $x(\theta)$ for $\theta > t$ to calculate $y(t)$.

Causality seems natural for physical systems (the cause occurs before the effect). However, one may consider non-causal systems in the case of offline processing. For example, when an entire signal has been recorded or stored in a buffer memory, one can perform *a posteriori* calculations such as (57) on a window centered around time t .

The stability of a system can be characterized in various ways. We retain the definition in the sense of Bounded-Input, Bounded-Output (BIBO).

Definition 2.10 (Bounded-Input, Bounded Output Stability). *A system S is BIBO stable if for any input $x \in L^\infty(\mathbb{R})$, $S(x) \in L^\infty(\mathbb{R})$.*

2.4 LINEAR CONVOLUTION FILTERS

In section 2.3, a linear filter was defined as a linear time invariant (LTI) system. Hereafter, the systems are assumed homogeneous in input/output. Furthermore, we focus on filters defined by a convolution.

2.4.1 Definition

Definition 2.11 (Impulse response). *Let S be a linear filter whose input-output relationship involves a convolution:*

$$y = S(x) = h * x, \quad (58)$$

*(provided that $h * x$ is well defined). The signal h is called impulse response of the filter. It is possibly a distribution.*

The Dirac distribution $\delta = \delta_0$ (continuous time signals) or the Kronecker function $\delta[k]$ (discrete time signals) being the neutral elements of the convolution, we have

$$h = S(\delta). \quad (59)$$

Hence the name impulse response: h is the response of the filter to a pulse input (Dirac or Kronecker depending on the context).

Interpretation of the calculation of the convolution product

Take the example of a discrete-time system and a pulse input ($x[k] = \delta[k]$). The output $y[k] = h[k]$ corresponds to the impulse response, as shown in Fig. 6.

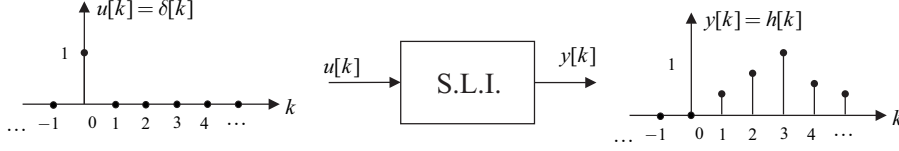


Figure 6: Impulse response of a LTI system.

Now consider an arbitrary input $x[k]$. Given the impulse response, the output

$$y[k] = (h * x)[k] = \sum_{n \in \mathbb{Z}} x[n] h[k - n]$$

is obtained as a linear combination of the delayed versions of h (the signals $k \mapsto h[k - n]$) weighted by the amplitudes of the pulses $x[n]$. Since the convolution product is commutative, we also have

$$y[k] = \sum_{n \in \mathbb{Z}} h[n] x[k - n].$$

$h[n]$ can thus also be interpreted as a weighting sequence: the term $h[n]$ weights the value of the input passed n times ago (i.e., $x[k - n]$) to contribute to the calculation of y for time k .

Stability

The BIBO stability of the filter is guaranteed if

$$h \in L^1(\mathbb{R}) \quad (\text{continuous case})$$

$$h \in \ell^1(\mathbb{Z}) \quad (\text{discrete case})$$

Indeed, by Theorem 2.1, these conditions imply that if x is bounded ($x \in L^\infty(\mathbb{R})$, respectively $\ell^\infty(\mathbb{Z})$), then $h * x$ is bounded.

Causality

By writing the definition of the convolution product (here for continuous time signals):

$$y(t) = (h * x)(t) = \int_{\mathbb{R}} h(\theta)x(t - \theta)d\theta$$

it appears that the filter is causal if and only if $h(t) = 0$ for $t < 0$ (respectively $h[k] = 0$ for $k < 0$), i.e.,

$$\text{supp}(h) \subset \mathbb{R}_+ \quad (\text{continuous case})$$

$$\text{supp}(h) \subset \mathbb{N} \quad (\text{discrete case})$$

The input-output relationship is then written

$$y(t) = \int_{\mathbb{R}_+} h(\theta)x(t - \theta) d\theta = \int_{]-\infty, t]} h(t - \theta)x(\theta) d\theta.$$

and rereads

$$y(t) = \int_{[0, t]} h(\theta)x(t - \theta) d\theta = \int_{[0, t]} h(t - \theta)x(\theta) d\theta.$$

if $x(t) = 0$ for $t < 0$. Similar expressions are obtained for discrete time signals.

Remark. For a stable discrete-time system, $h[n]$ tends to 0 when $|n|$ tends to infinity. In a nutshell, the values of the input signal at times very far from time k , have little impact on the value of the output for time k .

2.4.2 Finite impulse response filters (FIR)

Definition 2.12. A FIR (**Finite Impulse Response**) filter is a filter whose impulse response has a bounded (time limited) support.

Example 2.5 (Derivative filter). The derivative filter computes the difference between two consecutive points. The output y is:

$$y[k] = x[k] - x[k - 1]. \quad (60)$$

Identifying this equation with

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

we get the impulse response:

$$h[n] = \begin{cases} 1 & \text{if } n = 0, \\ -1 & \text{if } n = 1, \\ 0 & \text{elsewhere.} \end{cases}$$

The support of the impulse response is $\{0, 1\}$. It is bounded, therefore h is an FIR filter.

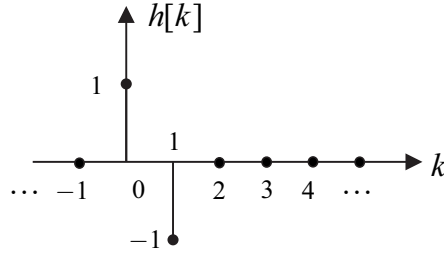


Figure 7: Impulse response of a derivative filter.

Example 2.6 (Moving average filter). This filter calculates the average value of the signal over a window of $2K + 1$ points centered on the current point:

$$y[k] = \frac{1}{2K + 1} \sum_{n=-K}^K x[k-n]. \quad (61)$$

The filter is not causal because $y[k]$ depends on inputs at times later than k . By identifying (61) with $(h * x)[k]$, we get:

$$h[n] = \begin{cases} 1 & \text{if } |n| \leq K, \\ 0 & \text{elsewhere.} \end{cases}$$

The support of the impulse response is $\{-K, \dots, K\}$. It is bounded, therefore h is an FIR filter.

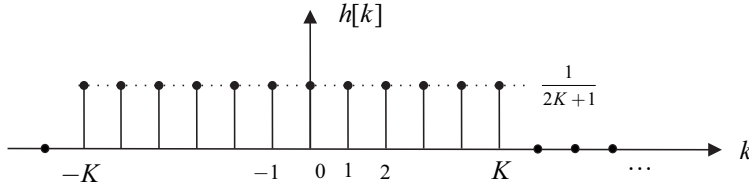


Figure 8: Impulse response of a moving average filter.

2.4.3 Infinite impulse response filters (IIR)

Definition 2.13. An IIR (**Infinite Impulse Response**) filter is a filter whose impulse response has an unbounded (time unlimited) support.

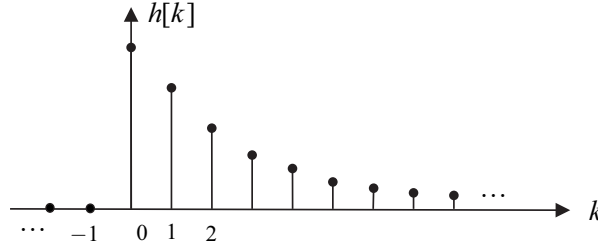


Figure 9: Impulse response of a causal IIR filter.

Example 2.7. Causal exponential smoothing Fig. 9 shows the impulse response of a filter for which $h[n] = \alpha^n$ with $0 < \alpha < 1$. The input-output relationship reads:

$$y[k] = \sum_{n=0}^{\infty} \alpha^n x[k-n]. \quad (62)$$

2.4.4 Difference equations

The input/output relationship of a filter can be described by a convolution, but also by a recursive equation between the outputs at different times. Consider an ordinary differential equation:

$$y(t) - \alpha_1 y'(t) - \dots - \alpha_p y^{(p)}(t) = \beta_0 x(t) + \dots + \beta_m x^{(m)}(t), \quad t \in \mathbb{R}_+$$

(with p initial conditions at $t = 0$). Its time discretization using the finite difference method yields an equation of the form:

$$y[k] = \sum_{i=1}^p a_i y[k-i] + \sum_{j=0}^m b_j x[k-j] \quad (63)$$

where the parameters a_i and b_j do not depend on k . This equation is called **finite difference equation** of order p .

Since the impulse response h identifies with the output of the filter for input $x[k] = \delta[k]$, it comes

$$\forall k > m, \quad h[k] = \sum_{i=1}^p a_i h[k-i] \quad (64)$$

$$\forall k \in \{0, \dots, m\}, \quad h[k] = \sum_{i=1}^p a_i h[k-i] + b_k. \quad (65)$$

Assuming the system is causal, these equations allow one to deduce h by solving a linear system.

2.5 EIGENFUNCTIONS

In finite dimension, the eigenvalues and eigenvectors of a linear application provide essential information on its action on a vector. The following theorem states that all stable linear invariant systems admit the complex exponential signals as eigenfunctions, i.e., the output signal yielded by the filter for these functions is proportional to the input signal.

Theorem 2.3. Consider a stable filter of impulse response h . The output signal $y = h * e_f$, i.e., the response to the LTI system to the complex exponential

$$e_f(t) = \exp(2i\pi ft)$$

is proportional to e_f :

$$y(t) = H(f) e_f(t).$$

The weight is called frequency response. It is given by the Fourier transform of h :

$$H(f) = \int_{\mathbb{R}} h(t) \exp(-2i\pi ft) dt \quad (66)$$

Theorem 2.4. Consider a stable discrete-time filter of impulse response h . The output signal $y = h * e_v$, i.e., the response to the LTI system to the discrete-time complex exponential

$$e_v[k] = \exp(2i\pi vk)$$

is proportional to e_v :

$$y[k] = H(v) e_v[k].$$

The weight is given by the discrete-time Fourier transform of h :

$$H(v) = \sum_{k \in \mathbb{Z}} h[k] \exp(-2i\pi vk) \quad (67)$$

Proof. (Continuous case) Let $\tau \in \mathbb{R}$. Note that $e_f * \delta_{-\tau} = \exp(i2\pi f\tau) e_f$. By temporal invariance and linearity of S , we therefore have:

$$S(e_f) * \delta_{-\tau} = S(e_f * \delta_{-\tau}) = \exp(i2\pi f\tau) S(e_f).$$

It comes that

$$\forall \tau, S(e_f)(0 + \tau) = \exp(i2\pi f\tau) S(e_f)(0),$$

that is, $S(e_f) = S(e_f)(0) e_f$. So e_f is an eigenfunction of S .

In the case where S admits an impulse response $h \in L^1$, the eigenvalue related to e_f is given by:

$$S(e_f)(0) = (e_f * h)(0) = \int_{\mathbb{R}} h(t) \exp(-2i\pi ft) dt.$$

□

Note that in the discrete case, the signals e_ν and $e_{\nu+n}$ are identical for $n \in \mathbb{Z}$, and that $H(\nu)$ is periodic with period 1.

Going further...

In the theoretical viewpoint, the frequency response of a filter can be measured by applying the filter to a complex exponential, and measuring the factor by which it is amplified or attenuated, and phase shifted. This is impossible in real life since complex exponentials have infinite support. However, in the case of stable and causal filters whose impulse response h is L^1 , the response to a complex exponential for positive times ($e_f \mathbb{1}_{\mathbb{R}_+}$) tends to the exponential response e_f when t tends to infinity. Indeed, the difference between the two responses is given by:

$$S(e_f)(t) - S(e_f \mathbb{1}_{\mathbb{R}_+})(t) = S(e_f \mathbb{1}_{\mathbb{R}_-})(t) = \int_{\mathbb{R}} e_f(t - \tau) h(\tau) \mathbb{1}_{[t, +\infty[}(\tau) d\tau$$

which tends to 0 when $t \rightarrow +\infty$ by dominated convergence.

We can therefore measure the frequency response by introducing a complex exponential from $t = 0$ in the system, and waiting for extinction of the transitional regime.

2.6 CONCLUSION

The links between linear time invariant systems and the Fourier transform are of paramount importance in signal processing. They justify the frequency analysis of systems (for example, plotting Bode diagrams, which is nothing but determining the eigenvalues of a system). Indeed, knowing the frequency response $H(f)$ of a system, one can reconstruct its impulse response using the inverse Fourier transform, and therefore predict its action on any input signal. Chapter 3 recalls important properties of the Fourier transform, that are useful for analyzing linear time invariant systems.

For continuous time systems, the link between impulse response and frequency response was given here only when the impulse response is an integrable function. Let us stress that many filters admit an impulse response that is not integrable, especially those given by a distribution (think of the elementary case of time delay systems, whose impulse response is a translated Dirac distribution). The extension of the Fourier transform to tempered distributions is also covered in chapter 3, allowing frequency analysis of these systems.

FREQUENCY DOMAIN REPRESENTATION

Fourier transforms are linear transforms applying to continuous and discrete time signals. They make it possible to represent signals in the frequency domain. For example, the content of audio signals can be easily interpreted when signals are expressed in the frequency domain. The frequency representation also simplifies many signal processing tasks, since convolution in the time domain leads to a simple elementwise product in the frequency domain. Frequency analysis consists of finding a decomposition of the form

$$x(t) = \int_{\mathbb{R}} X(f) \exp(2i\pi ft) df \quad (68)$$

which can be interpreted as a decomposition of a signal x on the space of complex exponentials $e_f : t \mapsto \exp(2i\pi ft)$. (68) can be seen as an extension of Fourier series representation (well-known in the case of periodic signals) to non-periodic signals and distributions.

Chapter organization

We will successively introduce the four Fourier transforms:

- The continuous-time Fourier transform, for a continuous time signal $x(t)$ of infinite support.
- The Fourier series representation, for a continuous time signal of compact support (or for a periodic signal).
- The discrete-time Fourier transform, for a discrete time signal $x[k]$ with infinite support.
- The discrete Fourier transform, for a discrete-time signal of finite support, represented as a vector $\mathbf{x} \in \mathbb{C}^N$.

These transforms will be applied to linear filtering and spectral analysis.

3.1 CONTINUOUS TIME FOURIER TRANSFORM

The Fourier transform (FT) of a signal

$$\begin{aligned} x : \mathbb{R} &\rightarrow \mathbb{C} \\ t &\mapsto x(t) \end{aligned}$$

is the function $\mathcal{F}x : \mathbb{R} \rightarrow \mathbb{C}$ defined¹ by:

$$\forall f \in \mathbb{R}, \mathcal{F}x(f) = \int_{\mathbb{R}} e^{-2i\pi ft} x(t) dt. \quad (69)$$

Similarly, the conjugate Fourier transform of a signal x is the function $\mathcal{F}^*x : \mathbb{R} \rightarrow \mathbb{C}$ defined by:

$$\forall f \in \mathbb{R}, \mathcal{F}^*x(f) = \int_{\mathbb{R}} e^{2i\pi ft} x(t) dt. \quad (70)$$

The modulus $|\mathcal{F}x(f)|$ is most often used to give a graphical representation of the Fourier transform, called **spectrum** of signal x . However, two signals can have the same spectrum and be significantly different in the time domain.

Notations. $\mathcal{F}x(f)$ should be understood as $(\mathcal{F}x)(f)$. When it is necessary to make explicit the dependence of x with respect to t , we will write $\mathcal{F}\{x(t)\}(f)$. We will very often use the short notation $X := \mathcal{F}x$. Thus,

$$\forall f \in \mathbb{R}, X(f) = \int_{\mathbb{R}} e^{-2i\pi ft} x(t) dt. \quad (71)$$

Remark. When x is not in $L^1(\mathbb{R})$, integrals (69) and (70) are possibly divergent. These definitions are no more valid. We will see though that the FT definition can be generalized to a large class of signals.

We start by defining the Fourier transform in $L^1(\mathbb{R})$, before addressing the extensions to $L^2(\mathbb{R})$ and to tempered distributions, in $\mathcal{S}'(\mathbb{R})$. The reader is redirected to the handout of CIP for proofs of the first theorems.

3.1.1 Fourier transform of signals in L^1

The definition of FT is the simplest for $x \in L^1(\mathbb{R})$, because the integral (69) is defined at any point f . Note that for $f = 0$, we have:

$$X(0) = \int_{\mathbb{R}} x(t) dt. \quad (72)$$

Note also that the FT of an even function ($x(-t) = x(t)$) is an even function ($X(-f) = X(f)$). Similarly, the FT of an odd function is an odd function. Moreover, if x real-valued, then X has Hermitian symmetry: $\forall f, X(-f) = X^*(f)$.

¹ The alternative definition is sometimes found: $X(\omega) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(-i\omega t) x(t) dt$.

Theorem 3.1. Let $x \in L^1$. Its Fourier transform satisfies:

- $X \in L^\infty$,
- $X \in C^0$,
- (Riemann-Lebesgue) $\lim_{|f| \rightarrow \infty} X(f) = 0$,
- If $x \in L^1$ and $\text{supp}(x)$ is bounded, then $X \in C^\infty(\mathbb{R})$.

According to Theorem 3.1, the FT of a signal in L^1 is continuous and bounded. However, X is not necessarily in L^1 , as shown in the example below.

Example 3.1 (Rectangular window). Let $x = \mathbb{1}_{[-\frac{T}{2}, \frac{T}{2}]}$. We calculate

$$X(f) = \int_{[-\frac{T}{2}, \frac{T}{2}]} e^{-2i\pi ft} dt = T \text{sinc}_c(T\pi f) \quad (73)$$

where the sinc function is defined by

$$\text{sinc}_c(\theta) := \begin{cases} \frac{\sin(\theta)}{\theta} & \text{if } \theta \neq 0 \\ 1 & \text{if } \theta = 0 \end{cases} \quad (74)$$

The FT of the even rectangular window is real valued, but not in $L^1(\mathbb{R})$. It vanishes at $f = \frac{k}{T}$, $k \in \mathbb{Z}^*$. $X(0) = T$ is equal to its integral, see (72).

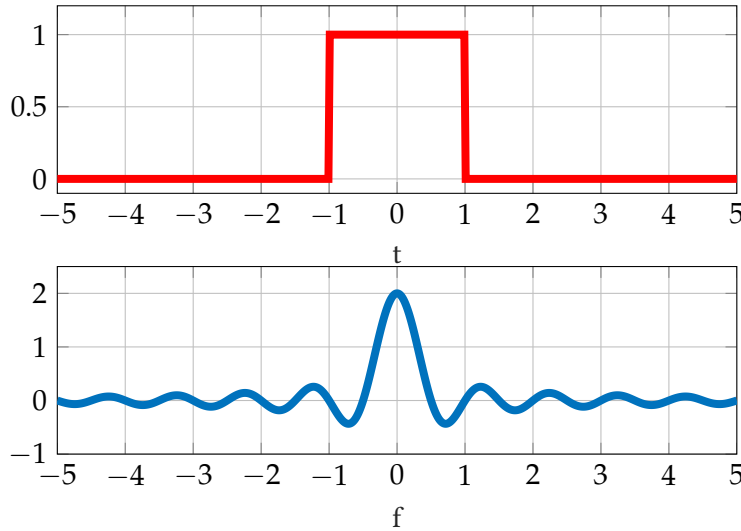


Figure 10: Rectangular window of width 2 and its FT. The FT is a sinc vanishing when f is multiple of $1/2$ (and non zero).

Table 1: Fourier transform properties.

	$z(t)$	$Z(f)$
Linearity	$(\alpha x + \beta y)(t)$	$\alpha X(f) + \beta Y(f)$
Time-dilatation / Contraction	$x(\alpha t)$	$\frac{1}{ \alpha } X\left(\frac{f}{\alpha}\right)$
Duality of FT	$(\mathcal{F}X)(t)$	$X(-f)$
Product	$x(t)y(t)$	$(X * Y)(f)$
Convolution	$(x * y)(t)$	$X(f)Y(f)$
Delay	$x(t - \tau_0)$	$e^{-2i\pi\tau_0 f} X(f)$
Modulation	$e^{2i\pi f_0 t} x(t)$	$X(f - f_0)$
Differentiation	$x^{(n)}(t)$	$(2i\pi f)^n X(f)$
Product by t^n	$t^n x(t)$	$\mathcal{F}x^{(n)}(f) / (-2i\pi)^n$

Properties

The main properties of the FT are listed in Table 1 and proved in appendix. It is worth noting that the FT of a time-dilated signal is the contracted version of the FT of the input signal, with the same expansion/contraction factor $(\alpha, 1/\alpha)$. It should also be noted that the FT of a convolution product is equal to the entrywise product of FTs. Conversely, the FT of a product of functions is equal to the convolution product of their FTs. An important case of application of this result is the following: the FT of a delayed signal is equal to the modulated version of the FT of the starting signal (and vice versa). These results are deduced from those on the FT of the convolution product and the entrywise product, noting that $x(t - \tau_0) = (\delta_{\tau_0} * x)(t)$. The FTs of δ_{τ_0} and $e_{f_0}(t) = e^{2i\pi f_0 t}$ are given in Table 2.

A fundamental property is that the conjugate Fourier transform $\mathcal{F}^*$ is the inverse Fourier transform operator.

Theorem 3.2. *if $x \in L^1(\mathbb{R})$ and $\mathcal{F}x \in L^1(\mathbb{R})$, then we have $\mathcal{F}^* \mathcal{F}x(t) = x(t)$ almost everywhere.*

The inverse Fourier transform is therefore written:

$$x(t) = \int_{\mathbb{R}} e^{2i\pi f t} X(f) df. \quad (75)$$

3.1.2 Fourier transform of L^2 signals

The Fourier transform extends to an isometry from $L^2(\mathbb{R})$ into $L^2(\mathbb{R})$.

Definition 3.1. Let $x \in L^2(\mathbb{R})$. We define $x_n = x\mathbb{1}_{[-n,n]}$ for $n \in \mathbb{Z}$. The FT of x is the limit in L^2 :

$$X = \lim_{n \rightarrow \infty}^{L^2} X_n \quad (76)$$

of the Fourier transforms X_n of functions x_n (seen as L^1 functions). The conjugate Fourier transform extends similarly.

The reader is redirected to the CIP course for more details. The main properties are listed below.

Theorem 3.3 (Plancherel-Parseval). The FT is an isometry from $L^2(\mathbb{R})$ into $L^2(\mathbb{R})$. It preserves the norm and the inner product:

- $\forall x \in L^2(\mathbb{R}), \|X\| = \|x\|.$
- $\forall x, y \in L^2(\mathbb{R}), (X, Y) = (x, y).$

The first property can be interpreted as the conservation of energy ($E_x = \|x\|^2 = E_X$) whatever the representation, time or frequency, chosen for the signal.

Theorem 3.4 (Inverse Fourier Transform). For all $x \in L^2(\mathbb{R})$, $\mathcal{F}\mathcal{F}^*x = \mathcal{F}^*\mathcal{F}x = x$ almost everywhere.

3.1.3 Fourier transform in $\mathcal{S}'$

The FT properties on the Schwartz space $\mathcal{S}(\mathbb{R})$ (see definition (24)) are key ingredients towards the generalization of FT to distributions. First, let us note that $\mathcal{S}(\mathbb{R}) \subset L^2(\mathbb{R}) \cap L^1(\mathbb{R})$. Moreover, unlike $\mathcal{D}(\mathbb{R})$, $\mathcal{S}(\mathbb{R})$ is stable by Fourier transform (cf. handout of the CIP course).

Lemma 3.1. $\forall \varphi \in \mathcal{S}(\mathbb{R}), \mathcal{F}\varphi \in \mathcal{S}(\mathbb{R}).$

This is the reason why we consider the space of tempered distributions $T \in \mathcal{S}'(\mathbb{R})$ to generalize the FT.

Definition 3.2. Let $T \in \mathcal{S}'(\mathbb{R})$. The Fourier transform of T is the tempered distribution $\mathcal{F}T \in \mathcal{S}'(\mathbb{R})$ defined by:

$$\forall \varphi \in \mathcal{S}(\mathbb{R}), \langle \mathcal{F}\{T\}, \varphi \rangle = \langle T, \mathcal{F}\varphi \rangle. \quad (77)$$

The space of distributions with compact support $\mathcal{E}'(\mathbb{R})$ being included in $\mathcal{S}'(\mathbb{R})$, the previous definition is still valid for $T \in \mathcal{E}'(\mathbb{R})$.

Remark. (77) is a generalization of the following lemma to distributions.

Lemma 3.2. if x and $\varphi \in \mathcal{S}(\mathbb{R})$, then $\mathcal{F}x \varphi$ and $\varphi \mathcal{F}x \in L^1(\mathbb{R})$, and

$$\int_{\mathbb{R}} \mathcal{F}x \varphi = \int_{\mathbb{R}} x \mathcal{F}\varphi.$$

$\langle \mathcal{F}\{T\}, \varphi \rangle$ in (77) plays the role of $\int_{\mathbb{R}} \mathcal{F}x \varphi$ here.

Let us now apply the definition to the calculation of the FT of the Dirac distribution.

Example 3.2 (FT of a Dirac). Let $f_0 \in \mathbb{R}$. For $\varphi \in \mathcal{S}(\mathbb{R})$,

$$\langle \mathcal{F}\delta_{f_0}, \varphi \rangle = \langle \delta_{f_0}, \mathcal{F}\varphi \rangle = (\mathcal{F}\varphi)(f_0) = \int_{\mathbb{R}} e^{-2i\pi f_0 t} \varphi(t) dt = \langle e_{-f_0}, \varphi \rangle.$$

We thus have

$$\mathcal{F}\delta_{f_0} = e_{-f_0}. \quad (78)$$

Specifically, for $f_0 = 0$,

$$\mathcal{F}\delta = \mathbb{1}_{\mathbb{R}}. \quad (79)$$

Theorem 3.5 (Inverse Fourier Transform). The FT is an isomorphism from $\mathcal{S}'(\mathbb{R})$ into itself, and the inverse Fourier transform $\mathcal{F}^{-1} = \mathcal{F}^*$ is the distribution defined by:

$$\forall \varphi \in \mathcal{S}(\mathbb{R}), \langle \mathcal{F}^*\{T\}, \varphi \rangle = \langle T, \mathcal{F}^*\varphi \rangle.$$

Example 3.3 (FT of complex exponentials). Let $f_0 \in \mathbb{R}$. Repeating the calculation in the previous example with the inverse FT, we find $\mathcal{F}^*\delta_{f_0} = e_{f_0}$, hence

$$\mathcal{F}e_{f_0} = \mathcal{F}\mathcal{F}^*\delta_{f_0} = \delta_{f_0}. \quad (80)$$

It follows that the FT of the distributions associated with the trigonometric signals are given by

$$\mathcal{F}\{\cos(2\pi f_0 t)\} = \mathcal{F}\left\{\frac{e_{f_0} + e_{-f_0}}{2}\right\} = \frac{\delta_{-f_0} + \delta_{f_0}}{2}$$

$$\mathcal{F}\{\sin(2\pi f_0 t)\} = \mathcal{F}\left\{\frac{e_{f_0} - e_{-f_0}}{2i}\right\} = \frac{\delta_{-f_0} - \delta_{f_0}}{2i}.$$

When $f_0 = 0$, we get:

$$\mathcal{F}\mathbb{1}_{\mathbb{R}} = \delta_0 = \delta.$$

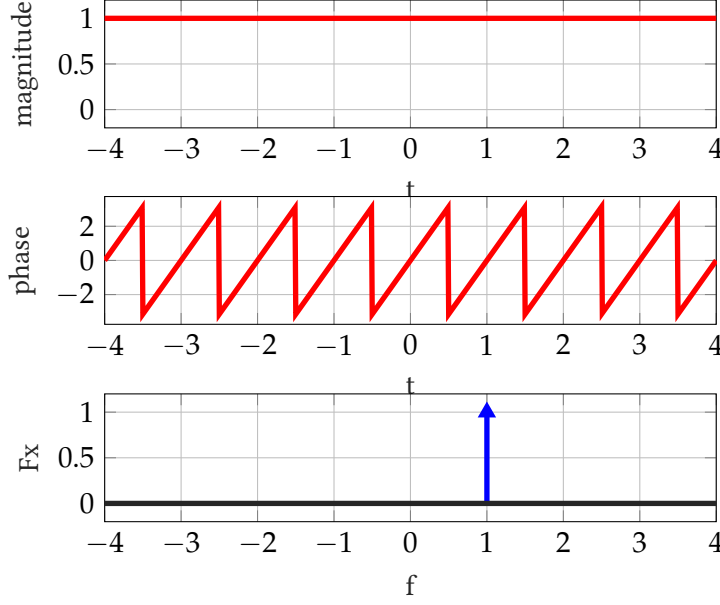


Figure 11: Function $e_{f_0}(t) = \exp(2i\pi f_0 t)$ for $f_0 = 1$: time variation of the amplitude and the phase, plot of its FT.

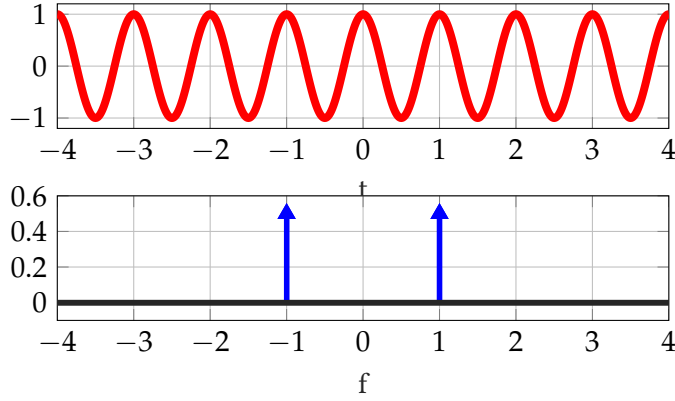


Figure 12: Function $\cos(2\pi f_0 t)$ for $f_0 = 1$: time and frequency representations.

The calculation of the FT of a Dirac comb is based on the following lemma.

Lemma 3.3 (Poisson summation formula). *For $\varphi \in \mathcal{S}(\mathbb{R})$ and $T > 0$, we have*

$$\frac{1}{T} \sum_{k \in \mathbb{Z}} \varphi\left(\frac{k}{T}\right) = \sum_{k \in \mathbb{Z}} \mathcal{F}\varphi(kT). \quad (81)$$

Proof. The proof can be found in the CIP booklet with weaker assumptions:

1. $\varphi \in L^1(\mathbb{R}) \cap \mathcal{C}^0(\mathbb{R})$,
2. $\exists M > 0, \exists \alpha > 1 : \forall x \in \mathbb{R}, |\varphi(x)|(1 + |x|^\alpha) \leq M$,
3. $\sum_{k \in \mathbb{Z}} |\mathcal{F}\varphi(kT)| < \infty$.

The assumptions are all met for $\varphi \in \mathcal{S}(\mathbb{R})$, since:

1. $\mathcal{S}(\mathbb{R}) \subset L^1(\mathbb{R}) \cap \mathcal{C}^\infty(\mathbb{R})$,
2. φ is rapidly decreasing, see (24),
3. according to Lemma 3.1, $\mathcal{F}\varphi \in \mathcal{S}(\mathbb{R})$. $\mathcal{F}\varphi$ is thus rapidly decreasing.

□

Theorem 3.6 (FT of a Dirac comb). *Let $T > 0$. Then the Dirac comb $\mathbb{W}_T$ satisfies $\mathbb{W}_T \in \mathcal{S}'(\mathbb{R})$. Moreover, we have:*

$$\mathcal{F}\mathbb{W}_T = \frac{1}{T}\mathbb{W}_{\frac{1}{T}} \quad (82)$$

Proof. By applying (81) to the test functions φ , we have:

$$\begin{aligned} \forall \varphi \in \mathcal{S}(\mathbb{R}), \quad \frac{1}{T} \langle \mathbb{W}_{\frac{1}{T}}, \varphi \rangle &= \frac{1}{T} \sum_{k \in \mathbb{Z}} \varphi\left(\frac{k}{T}\right) \\ &= \sum_{k \in \mathbb{Z}} (\mathcal{F}\varphi)(kT) \\ &= \langle \mathbb{W}_T, \mathcal{F}\varphi \rangle \\ &= \langle \mathcal{F}\mathbb{W}_T, \varphi \rangle. \end{aligned}$$

hence the equality (82). □

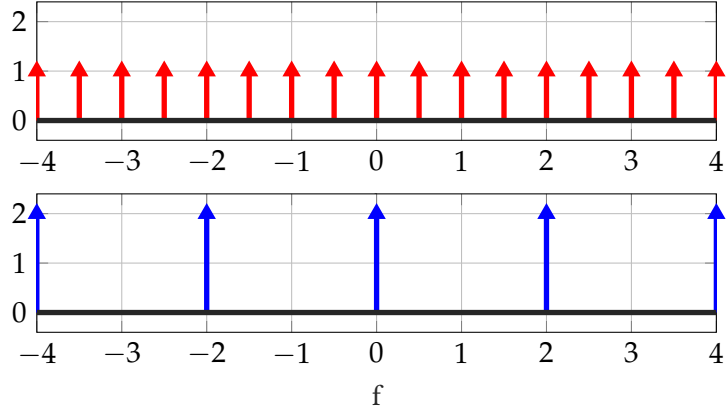


Figure 13: Dirac comb: time representation $\mathbb{W}_T$ and frequency representation $\frac{1}{T}\mathbb{W}_{\frac{1}{T}}$ for $T = \frac{1}{2}$.

3.1.4 Some classical examples

The most classical results are summarized in Table 2. We now detail two classical calculations.

The first is the FT of a triangle function. Let

$$\triangle_{[-T,T]}(t) = \left(1 - \frac{|t|}{T}\right) \mathbb{1}_{[-T,T]}(t) \quad (83)$$

Table 2: Classical Fourier transform identities. Notice the symmetry of the expressions (lines 1-2, 3-4, 5-6, 7-8) due to the duality property of the FT: $(\mathcal{F}\mathcal{F}x)(t) = x(-t)$.

x	$x \in \dots$	X
$\mathbb{1}_{[-T,T]}$	$L^1(\mathbb{R})$	$f \mapsto 2T \sin_c(2\pi T f)$
$t \mapsto \sin_c(2\pi f_0 t)$	$L^2(\mathbb{R})$	$\frac{1}{2f_0} \mathbb{1}_{[-f_0, f_0]}$
$\mathbb{1}_{\mathbb{R}}$	$\mathcal{S}'(\mathbb{R})$	δ
δ	$\mathcal{S}'(\mathbb{R})$	$\mathbb{1}_{\mathbb{R}}$
$t \mapsto \exp(2i\pi f_0 t)$	$\mathcal{S}'(\mathbb{R})$	δ_{f_0}
δ_{τ_0}	$\mathcal{S}'(\mathbb{R})$	$f \mapsto \exp(-2i\pi \tau_0 f)$
$t \mapsto \cos(2\pi f_0 t)$	$\mathcal{S}'(\mathbb{R})$	$\frac{\delta_{-f_0} + \delta_{f_0}}{2}$
$t \mapsto \sin(2\pi f_0 t)$	$\mathcal{S}'(\mathbb{R})$	$\frac{\delta_{f_0} - \delta_{-f_0}}{2i}$
$\mathbb{W}_T$	$\mathcal{S}'(\mathbb{R})$	$\frac{1}{T} \mathbb{W}_{\frac{1}{T}}$
$t \mapsto \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{t^2}{2\sigma^2})$	$L^1(\mathbb{R})$	$f \mapsto \exp(-2\sigma^2\pi^2 f^2)$

be the "triangle" function centered at 0, with amplitude 1 and support $[-T, T]$. Since

$$\triangle_{[-T,T]} = \frac{1}{T} \left(\mathbb{1}_{[-\frac{T}{2}, \frac{T}{2}]} * \mathbb{1}_{[-\frac{T}{2}, \frac{T}{2}]} \right) \quad (84)$$

and the FT of a rectangular function is known, see (73), the property on the FT of the convolution product yields:

$$\mathcal{F}\triangle_{[-T,T]}(f) = T \sin_c^2(T\pi f). \quad (85)$$

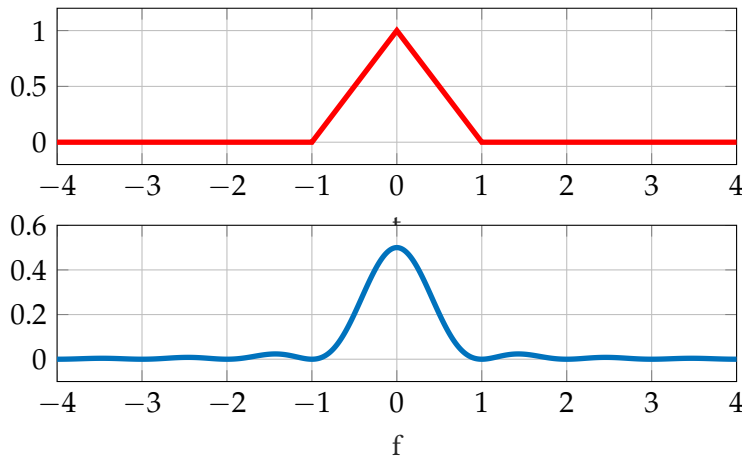


Figure 14: $\triangle_{[-T,T]}$ function for $T = 1$: time and frequency representations.

The second calculation is the FT of a cardinal sine. We set $s(t) = \text{sinc}(\pi t)$. According to the duality property of the Fourier transform (see table 2) and since the indicator function $\mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]}$ is even,

$$\mathcal{F}\mathcal{F}\mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]} = \mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]}.$$

Since $\mathcal{F}\mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]} = s$, the previous equation rereads $\mathcal{F}s = \mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]}$, i.e.,

$$\mathcal{F}\{\text{sinc}(\pi t)\}(f) = \mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]}(f). \quad (86)$$

Applying the dilatation / contraction property with $\alpha = 2f_0$, this result generalizes as follows (for $f_0 > 0$):

$$\begin{aligned} \mathcal{F}\{\text{sinc}(2\pi f_0 t)\}(f) &= \frac{1}{2f_0} \mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]} \left(\frac{f}{2f_0} \right) \\ &= \frac{1}{2f_0} \mathbb{1}_{[-f_0, f_0]}(f). \end{aligned} \quad (87)$$

3.2 FOURIER SERIES REPRESENTATION

3.2.1 Signals with limited support

The Fourier series representation theorem is stated for signals with limited support in the interval $[0, T]$. The space $L^2([0, T])$ is endowed with the inner product $(x, y) = \frac{1}{T} \int_{[0, T]} xy^* d\lambda$.

Theorem 3.7. Any signal $x \in L^2([0, T])$ can be decomposed in the Hilbert basis of $L^2([0, T])$ formed by functions:

$$\begin{aligned} e_k : [0, T] &\rightarrow \mathbb{C} \\ t &\mapsto \exp\left(2i\pi \frac{kt}{T}\right). \end{aligned}$$

The decomposition reads:

$$x = \sum_{k \in \mathbb{Z}} c_k e_k \quad (88)$$

where

$$c_k = (x, e_k) = \frac{1}{T} \int_0^T \exp\left(-2i\pi \frac{kt}{T}\right) x(t) dt. \quad (89)$$

We also have Parseval's identity:

$$\|x\|_{L^2(0, T)}^2 = \frac{1}{T} \int_{[0, T]} |x(t)|^2 dt = \sum_{k \in \mathbb{Z}} |c_k|^2. \quad (90)$$

The series (88) converges in L^2 . Under certain assumptions, we have point-wise and uniform convergence properties.

Theorem 3.8 (Dirichlet). *If x is piecewise $\mathcal{C}^1$, then (88) is valid at any point t where x is continuous. if x is furthermore continuous, convergence of the Fourier series to x is uniform.*

3.2.2 Periodic signals

The Fourier transform of periodic signals is defined in the sense of distributions. Consider a T -periodic signal x and its pattern $x_T := x\mathbb{1}_{[0,T]}$ over one period. We can write $\forall t, x(t) = \sum_{k \in \mathbb{Z}} x_T(t - kT)$, or equivalently,

$$x = x_T * \mathbb{I}_T. \quad (91)$$

Assume that $x_T \in L^1(0, T)$, so that $X_T \in \mathcal{C}^\infty(\mathbb{R})$. According to Theorem 3.6, the Fourier transform of x reads:

$$X = \mathcal{F}\{x_T * \mathbb{I}_T\} = X_T \frac{1}{T} \mathbb{I}_{\frac{1}{T}}.$$

Thus we have

$$X = \sum_{k \in \mathbb{Z}} c_k \delta_{\frac{k}{T}} \quad (92)$$

where

$$\begin{aligned} c_k &= \frac{1}{T} X_T \left(\frac{k}{T} \right) \\ &= \frac{1}{T} \int_0^T \exp \left(-2i\pi \frac{kt}{T} \right) x(t) dt \end{aligned} \quad (93)$$

are equal to the Fourier coefficients of x_T defined in (89).

Note that a periodic function has a spectrum formed of equally spaced impulses according to (92). Conversely, we show in the next section that the Fourier transform of a discrete time signal is periodic.

3.3 DISCRETE-TIME FOURIER TRANSFORM

The Fourier transform of a discrete-time signal (DTFT) $x : \mathbb{Z} \rightarrow \mathbb{C}$ is defined by

$$\forall \nu \in \mathbb{R}, \mathcal{F}x(\nu) = X(\nu) = \sum_{k \in \mathbb{Z}} x[k] e^{-2i\pi \nu k} \quad (94)$$

whenever this series is convergent (especially when $x \in \ell^1(\mathbb{Z})$). When the series is divergent, definition (94) is no more valid. This said, the DTFT can be defined for broader ranges of signals (in $\ell^2(\mathbb{Z})$, for periodic signals) using the distribution framework.

As for the continuous case, we write $X := \mathcal{F}x$, and use the notation $\mathcal{F}\{x[k]\}(\nu)$ when the dependence of x on k needs to be explicit.

Remark (normalized frequency, high and low frequencies). *The discrete-time Fourier transform is a periodic signal*

$$\begin{aligned} X : \mathbb{R} &\rightarrow \mathbb{C} \\ \nu &\mapsto X(\nu) \end{aligned}$$

of period 1. The variable ν is called normalized frequency.

$X(\nu)$ being 1-periodic, the low frequencies correspond to $\nu \approx 0$, but also to $\nu \approx \ell$ for $\ell \in \mathbb{Z}$. High frequencies correspond to $\nu \approx \frac{1}{2} + \ell$, $\ell \in \mathbb{Z}$.

It is common practice to display a graphical representation of $|X(\nu)|$ over one period, for $\nu \in [0, 1]$ (with the highest frequency $\nu = \frac{1}{2}$ in the center of the figure, and the lowest frequencies $\nu = 0$ and 1 on the left and right edges) or for $\nu \in [-\frac{1}{2}, \frac{1}{2}]$ (with the 0 frequency in the center of the figure, and high frequencies $\nu \approx \pm \frac{1}{2}$ in the left and right edges).

3.3.1 Fourier transform of ℓ^1 signals

In the case of $\ell^1(\mathbb{Z})$ signals, the DTFT shares the same properties as the FT for the L^1 signals.

Theorem 3.9. Let $x \in \ell^1$. The DTFT satisfies:

- $X \in L^\infty$;
- $X \in \mathcal{C}^0$.

We also have, from the definition, that

$$X(0) = \sum_{k \in \mathbb{Z}} x[k]. \quad (95)$$

Example 3.4 (Rectangular window). Consider the (discrete) rectangular window of width $2N + 1$ given by

$$\mathbb{1}_{[-N, N]}[k] = \sum_{n=-N}^N \delta[k - n]$$

The DTFT writes:

$$\mathcal{F} \left\{ \mathbb{1}_{[-N, N]} \right\} (\nu) = \sum_{k=-N}^N e^{-2i\pi k\nu} = \begin{cases} \frac{\sin(\pi(2N+1)\nu)}{\sin(\pi\nu)} & \text{if } \nu \neq 0 \\ 2N + 1 & \text{if } \nu = 0. \end{cases}$$

Note that the sinc function, obtained in the continuous case (see (73)) is now replaced here by a ratio of sines.

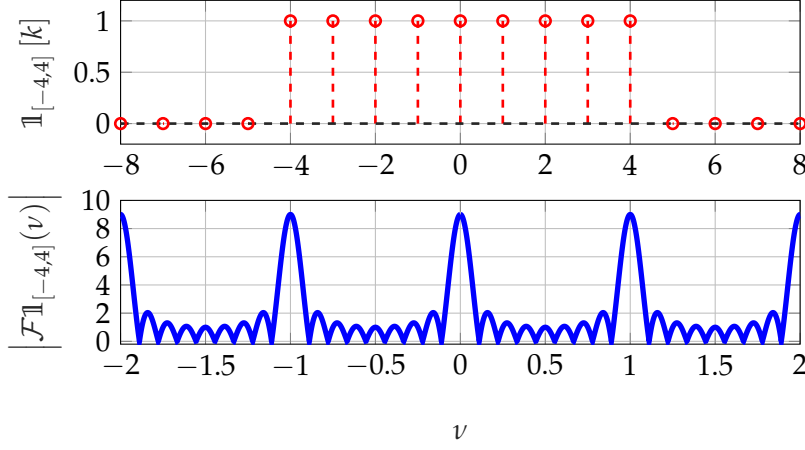


Figure 15: Discrete rectangular window $\mathbb{1}_{[-N,N]}$: time and frequency representations (modulus). As for any discrete time signal, the DTFT is 1-periodic.

3.3.2 Fourier transform of ℓ^2 signals

Theorem 3.10. *The DTFT operator $\mathcal{F}$ is a (bijective) isometry from $\ell^2(\mathbb{Z})$ into $L^2([0, 1])$, which leads to the conservation of the inner product and the energy (Parseval's equality):*

$$\begin{aligned} \forall x, y \in \ell_2(\mathbb{Z}), \quad \sum_{k \in \mathbb{Z}} x[k] y^*[k] &= \int_{[0,1]} X(\nu) Y^*(\nu) \, d\nu \\ \forall x \in \ell_2(\mathbb{Z}), \quad \sum_{k \in \mathbb{Z}} |x[k]|^2 &= \int_{[0,1]} |X(\nu)|^2 \, d\nu. \end{aligned}$$

The inverse transform is given by

$$x[k] = \int_{[0,1[} X(\nu) e^{2i\pi k\nu} \, d\nu \quad (96)$$

Remark. (94) is nothing but the Fourier series representation of the (1-periodic) function $X = \mathcal{F}x$. The coefficients of this decomposition are expressed as the inner products of X with the complex exponential signals, see (96). X being 1-periodic, its energy is defined as the integral of $|X(\nu)|^2$ over one period, such as $[0, 1]$.

The main properties of the DTFT are summarized in Table 3. Similar to the FT of continuous signals, the DTFT of a convolution product is the entrywise product of the DTFTs of each discrete signal. Conversely, the DTFT of a product is the **circular** convolution product of the DTFTs. X and Y being 1-periodic functions, the circular convolution $\circledast$ is defined by:

$$\forall \nu \in \mathbb{R}, \quad (X \circledast Y)(\nu) = \int_{[0,1[} X(\theta) Y(\nu - \theta) \, d\theta. \quad (97)$$

Table 3: Properties of the Discrete Time Fourier Transform.

	$z[k]$	$Z(\nu)$
Linearity	$\alpha x[k] + \beta y[k]$	$\alpha X(\nu) + \beta Y(\nu)$
Duality	$\mathcal{F}\{X(\nu)\}[k]$	$X(-\nu)$
Delay	$x[k - k_0]$	$e^{-2i\pi\nu k_0} X(\nu)$
Modulation	$e^{2i\pi\nu_0 k} x[k]$	$X(\nu - \nu_0)$
Entrywise product	$x[k] y[k]$	$(X \otimes Y)(\nu)$
Convolution	$(x * y)[k]$	$X(\nu) Y(\nu)$

3.3.3 Some classical results

Table 4 summarizes classical DTFT expressions. Some signals such as the complex exponentials are neither in ℓ^1 nor in ℓ^2 , but in ℓ^∞ . The definition of their DTFT is (very) subtle and will not be further elaborated here. The interested reader is referred to the book [5]. One should notice that **the DTFT of a discrete sinusoid is the sum of two Dirac combs**, and that the DTFT of a rectangular function is not a cardinal sine function, but a ratio of sine functions.

Table 4: DTFT of classical discrete time signals. The DTFT signals are always 1-periodic.

$x[k]$	$x \in \dots$	X
$\delta[k]$ (Kronecker function)	$\ell^1(\mathbb{Z})$	$\mathbb{1}_{\mathbb{R}}$
$\delta[k - k_0]$	$\ell^1(\mathbb{Z})$	$\nu \mapsto \exp(-2i\pi\nu k_0)$
$\mathbb{1}_{[-N,N]}[k] = \sum_{n=-N}^N \delta[k - n]$	$\ell^1(\mathbb{Z})$	$\nu \mapsto \frac{\sin(\pi(2N+1)\nu)}{\sin(\pi\nu)}$
$\mathbb{1}_{\mathbb{Z}}[k] = \sum_{n=-\infty}^{\infty} \delta[k - n]$	$\ell^\infty(\mathbb{Z})$	$\mathbb{1}_{\mathbb{Z}}$ (Dirac comb)
$\exp(2i\pi\nu_0 k)$	$\ell^\infty(\mathbb{Z})$	$\sum_{\ell \in \mathbb{Z}} \delta_{\nu_0 + \ell}$ (Dirac comb)
$\cos(2\pi\nu_0 k)$	$\ell^\infty(\mathbb{Z})$	$\frac{1}{2} \sum_{\ell \in \mathbb{Z}} (\delta_{-\nu_0 + \ell} + \delta_{\nu_0 + \ell})$
$\sin(2\pi\nu_0 k)$	$\ell^\infty(\mathbb{Z})$	$\frac{1}{2i} \sum_{\ell \in \mathbb{Z}} (\delta_{\nu_0 + \ell} - \delta_{-\nu_0 + \ell})$

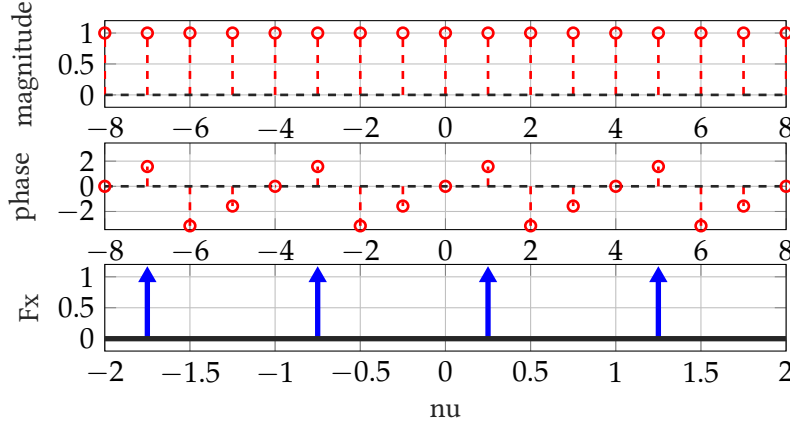


Figure 16: Complex exponential $\exp(2i\pi\nu_0 k)$: time representation (magnitude and phase) and frequency representation for $\nu_0 = 0.25$. The DTFT is 1-periodic.

Remark (Calculation method). For signals in ℓ^1 , the DTFT is calculated by formula (94). For signals not being in ℓ^1 , one may combine the classical results of Table 4 with the properties stated in Table 3: DTFT of delayed signals, of modulated signals, of a linear combination, etc.

3.4 DISCRETE FOURIER TRANSFORM

The discrete Fourier transform (not to be confused with the DTFT) is a very useful tool for digital signal processing from a practical point of view. Unlike DTFT, it applies to a discrete **signal of finite length**, represented by a vector $\mathbf{x} = \{x_0, \dots, x_{N-1}\} \in \mathbb{C}^N$.

Definition 3.3 (Discrete Fourier Transform). The **Discrete Fourier Transform** (DFT) is the linear application

$$\begin{aligned} \mathbb{C}^N &\rightarrow \mathbb{C}^N \\ \mathbf{x} &\mapsto \mathbf{X} = \mathbf{F}_N \mathbf{x} \end{aligned}$$

where $\mathbf{F}_N \in \mathbb{C}^{N \times N}$ is the matrix defined by

$$\mathbf{F}_N[n, k] = e^{-2i\pi \frac{nk}{N}} \quad (98)$$

for n and $k \in \{0, \dots, N-1\}$.

The DFT makes it possible to define a frequency representation X_n , $n \in \{0, \dots, N-1\}$ of a signal $\mathbf{x}$ with

$$\begin{aligned} X_n &= \sum_{k=0}^{N-1} \mathbf{F}_N[n, k] x_k \\ &= \sum_{k=0}^{N-1} e^{-2i\pi \frac{kn}{N}} x_k. \end{aligned} \quad (99)$$

Note that $\mathbf{F}_N$ is a symmetric Vandermonde matrix, whose inverse is equal (within a multiplicative factor) to its conjugate.

Theorem 3.11 (Inverse DFT). *The inverse transform of the DFT is the linear operation $\mathbf{x} = \mathbf{F}_N^{-1} \mathbf{X}$, where*

$$\mathbf{F}_N^{-1} = \frac{1}{N} \mathbf{F}_N^*. \quad (100)$$

Thus, the inverse DFT reads:

$$x_k = \frac{1}{N} \sum_{n=0}^{N-1} e^{2i\pi \frac{nk}{N}} X_n \quad (101)$$

for $k \in \{0, \dots, N-1\}$.

Remark. (101) can be interpreted as the representation of $\mathbf{x}$ in the orthogonal basis formed by the complex exponential signals of frequency $\frac{n}{N}$, denoted $\mathbf{e}_n = \{\exp(2i\pi \frac{nk}{N}), k = 0, \dots, N-1\}$:

$$\mathbf{x} = \sum_{n=0}^{N-1} \left(\frac{1}{N} X_n \right) \mathbf{e}_n. \quad (102)$$

Parseval's theorem guarantees energy conservation in the time and frequency domains (up to a multiplying factor).

Theorem 3.12 (Parseval). *For $\mathbf{x} \in \mathbb{C}^N$, we have $\|\mathbf{F}_N \mathbf{x}\|_2^2 = N \|\mathbf{x}\|_2^2$.*

Proof.

$$\|\mathbf{F}_N \mathbf{x}\|_2^2 = (\mathbf{F}_N \mathbf{x})^* (\mathbf{F}_N \mathbf{x}) = \mathbf{x}^* \mathbf{F}_N^* \mathbf{F}_N \mathbf{x} = \mathbf{x}^* (N\mathbf{I}) \mathbf{x} = N \|\mathbf{x}\|_2^2.$$

□

Note that matrix-vector product calculations such as $\mathbf{F}_N \mathbf{x}$ and $\mathbf{F}_N^{-1} \mathbf{X}$ usually cost $\mathcal{O}(N^2)$ elementary operations (additions and multiplications). However, there are fast algorithms (called FFT, for *Fast Fourier Transform*) allowing to perform these calculations in $\mathcal{O}(N \log(N))$ operations, by exploiting the particular structure of matrix $\mathbf{F}_N$.

3.5 APPLICATION TO LINEAR FILTERING

3.5.1 Filtering in the frequency domain

Consider a linear convolution filter of impulse response h . The input-output relationship reads $y(t) = (h * x)(t)$ or $y[k] = (h * x)[k]$. This relationship is simplified in the frequency domain because the FT (DTFT) of a convolution product is equal to the entrywise product of the FT (DTFT).

Continuous time filters

Definition 3.4 (Frequency response). *The input-output relationship of a filter is written in the frequency domain in the form:*

$$\forall f \in \mathbb{R}, Y(f) = H(f) X(f) \quad (103)$$

where $H(f) = (\mathcal{F}h)(f)$ is called **frequency response** or transfer function.

Any complex exponential e_f is an eigenfunction of a filter (see section 2.5). The related output reads:

$$y = H(f) e_f. \quad (104)$$

The modulus and phase of $H(f)$ can be seen as the gain in magnitude and the phase shift undergone by the signal e_f when passing through the filter.

Example 3.5. *An ideal low-pass filter is characterized by the frequency response $H(f) = \mathbb{1}_{[-f_c, f_c]}(f)$, where f_c refers to the cutoff frequency. This filter eliminates the high frequencies corresponding to the discontinuities (jumps in signal values) and the noise contained in the signal. So, the filter provides a smooth form of the input signal. According to Table 2, its impulse response reads:*

$$h(t) = 2f_c \operatorname{sinc}(2\pi f_c t).$$

The filter is neither causal nor with finite impulse response (hence the name "ideal"). In practice, windowing techniques are used to design a causal and stable filter whose frequency response is a good approximation of $\mathbb{1}_{[-f_c, f_c]}$.

Discrete time filters

Definition 3.5. *The input-output relationship of a discrete-time filter can be written in the frequency domain as:*

$$\forall \nu, Y(\nu) = H(\nu) X(\nu) \quad (105)$$

where $H(\nu) = (\mathcal{F}h)(\nu)$ is the frequency response. The DTFT being periodic, they are calculated over an interval of length 1 (generally, for $\nu \in [0, 1[$ or $\nu \in [-\frac{1}{2}, \frac{1}{2}[$).

Definition 3.6. A digital filter is called:

- high-pass if high frequencies are preserved:

$$H(\nu) \neq 0 \text{ for } \nu_c \leq |\nu| \leq \frac{1}{2}$$

and low frequencies are eliminated:

$$H(\nu) \approx 0 \text{ for } |\nu| \leq \nu_c,$$

denoting $\nu_c \in (0, \frac{1}{2})$ the cutoff frequency;

- low-pass if low frequencies are preserved ($H(0) \neq 0$) and high frequencies are eliminated ($H(\pm\frac{1}{2}) = 0$);
- band-pass if frequencies within a certain range are preserved ($H(\nu) \neq 0 \iff |\nu| \in [\nu_1, \nu_2]$ with $0 < \nu_1 < \nu_2 < \frac{1}{2}$);
- band-stop if frequencies within a certain interval are eliminated ($H(\nu) \approx 0 \iff |\nu| \in [\nu_1, \nu_2]$ with $0 < \nu_1 < \nu_2 < \frac{1}{2}$).

Example 3.6. The derivative filter is defined by $y[n] = x[n] - x[n-1]$. Its impulse response reads:

$$h[k] = \delta[k] - \delta[k-1] \quad (106)$$

(see section 2.4.2). The support of h is therefore equal to $\{0, 1\}$. Let us compute the frequency response of the filter:

$$\begin{aligned} H(\nu) &= h[0] + h[1] \exp(-2i\pi\nu) \\ &= 1 - \exp(-2i\pi\nu) \\ &= 2i \exp(-i\pi\nu) \sin(\pi\nu). \end{aligned} \quad (107)$$

h is a high-pass filter because $H(0) = 0$ and the maximum values of $|H(\nu)|$ are reached for $|\nu| = 1/2$.

Example 3.7. The moving average filter is defined by

$$y[n] = \frac{1}{2L+1} \sum_{k=-L}^L x[n-k].$$

We can show that its frequency response reads:

$$H(\nu) = \frac{\sin(2\pi(2L+1)\nu)}{(2L+1) \sin(\pi\nu)}$$

It is a low pass filter.

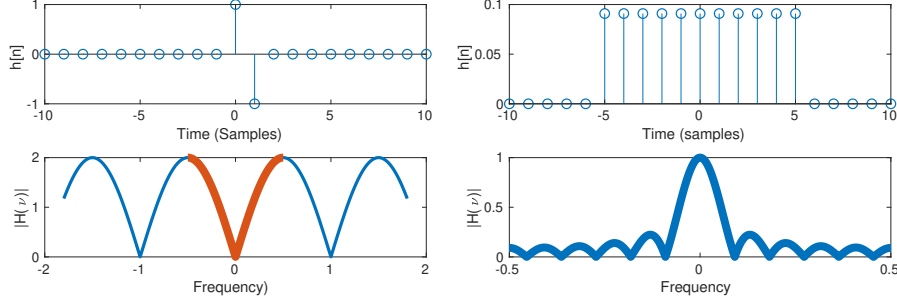


Figure 17: Impulse and frequency responses of the derivative (left) and moving average filters (right). These filters are low pass and high pass, respectively.

Difference equations

Some IIR filters are characterized by difference equations such as:

$$\forall k, y[k] = \sum_{\ell=1}^p a_{\ell} y[k - \ell] + x[k] \quad (108)$$

where p is the order of the filter. Then, the following signals are equal:

$$y = \sum_{\ell=1}^p a_{\ell} y[\cdot - \ell] + x.$$

The impulse response of the filter being of infinite support, it is hard to evaluate. On the other hand, the frequency response $H(\nu)$ can be simply calculated. We apply the property of the DTFT on delayed signals (see Table 3):

$$\forall \nu, Y(\nu) = \sum_{\ell=1}^p a_{\ell} e^{-2i\pi\ell\nu} Y(\nu) + X(\nu). \quad (109)$$

Identifying this expression with (105), we get:

$$\forall \nu, H(\nu) = \frac{1}{1 - \sum_{\ell=1}^p a_{\ell} e^{-2i\pi\ell\nu}}. \quad (110)$$

3.5.2 Synthesis of FIR filters using the DFT

Consider a causal FIR filter, and let p be the size of the impulse response. The input-output relationship reads:

$$y[n] = \sum_{k=0}^{p-1} h_k x[n - k].$$

We use the DFT to calculate the frequency response of the filter, $(\mathcal{F}h)(\nu)$, for specific ν -values.

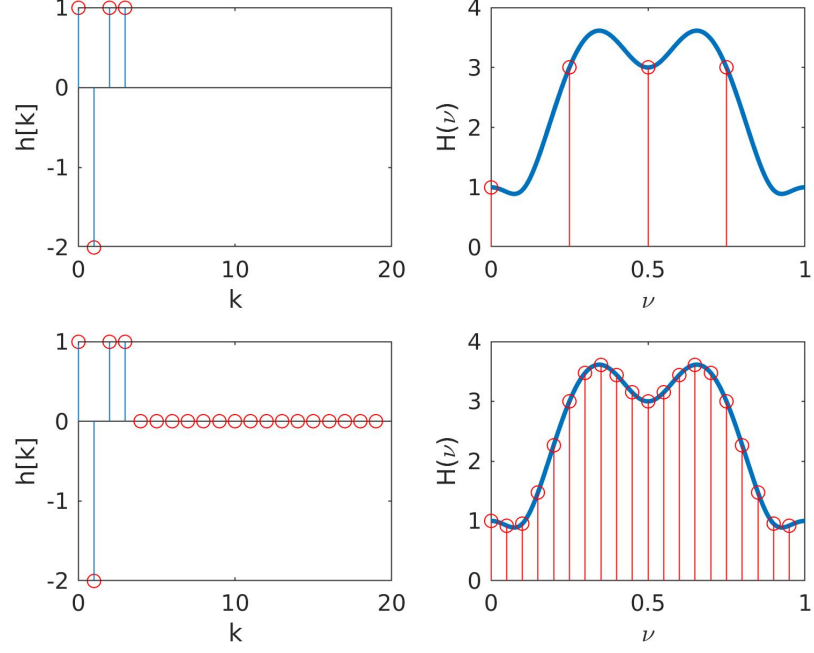


Figure 18: Calculation of the frequency response of a FIR filter. Right: the DTFT $(\mathcal{F}h)(\nu)$ is shown in continuous line. Its DFT-evaluation (for 4 and 20 frequency points, respectively) is in red. Zero-padding (bottom figures) improves the frequency representation.

The DFT operates on the vector $\mathbf{h} = \{h_k, k = 0, \dots, p-1\} \in \mathbb{C}^p$. The DFT output vector $\mathbf{H} = \mathbf{F}_p \mathbf{h}$ (see (99)) coincides with the evaluation of $(\mathcal{F}h)(\nu)$ for frequencies that are multiple of $\frac{1}{p}$:

$$\forall n \in \{0, \dots, p-1\}, H_n = (\mathcal{F}h)\left(\frac{n}{p}\right) \quad (111)$$

So, $\mathcal{F}h$ is calculated on a frequency grid whose step (equal to the spacing between two points on this grid) is $\frac{1}{p}$.

The **zero-padding** technique consists of artificially adding zeros in vector $\mathbf{h}$ in order to improve the frequency representation. Let us denote $\mathbf{h}' = \{\mathbf{h}, \mathbf{0}_{M-p}\} \in \mathbb{C}^M$ the vector formed by adding $M-p$ zeros to the end of the vector $\mathbf{h}$ (with $M > p$). The DFT $\mathbf{H}' = \mathbf{F}_M \mathbf{h}'$ evaluates the frequency response on the grid of step $\frac{1}{M}$:

$$\forall m \in \{0, \dots, M-1\}, H'_m = (\mathcal{F}h)\left(\frac{m}{M}\right). \quad (112)$$

The effect of zero padding is illustrated in Fig. 18: the number of frequency points considered is much larger using zero padding. This yields an obvious accuracy improvement in the DFT plot (in red).

Once the frequency response of the filter has been calculated, we can calculate the frequency representation of the output $y = h * x$ associated with a given input x , using the formula $Y(\nu) = (\mathcal{F}h(\nu)) X(\nu)$. Care should

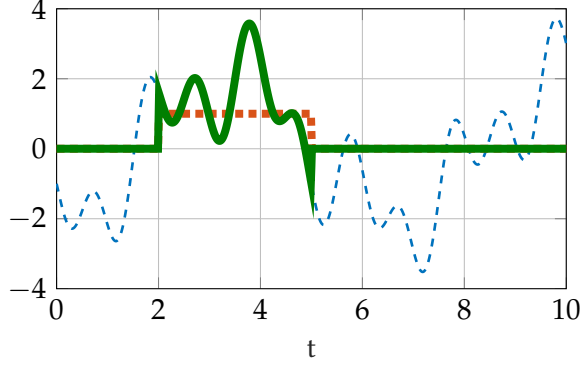


Figure 19: Rectangular windowing: $x(t) = s(t) \mathbb{1}_{[a,b]}(t)$. The observed signal x is shown in green.

be taken to calculate the DTFT $X(\nu)$ on the same grid as the grid used to calculate $(\mathcal{F}h)(\nu)$, i.e., a grid of step $\frac{1}{M}$ according to (112).

3.6 SPECTRAL ANALYSIS

Spectral analysis consists of decomposing a signal as a sum of sinusoids. This is a classical problem encountered in various fields ranging from sound analysis to radar source localisation, and analysis of the light emitted by stars in astrophysics. The Fourier transform is the natural tool to carry out such decomposition.

3.6.1 Time windowing of continuous time signals

In practice, the signal $x(t)$ that we want to analyze has been observed for a finite duration, for example for $t \in [-\frac{T}{2}, \frac{T}{2}]$. We can therefore model it as the product of a stationary ideal signal $s(t)$, corresponding to an infinite horizon, by a rectangular window:

$$x(t) = s(t) \mathbb{1}_{[-\frac{T}{2}, \frac{T}{2}]}(t). \quad (113)$$

From the properties of the FT, we get:

$$\begin{aligned} X &= S * \mathcal{F} \mathbb{1}_{[-\frac{T}{2}, \frac{T}{2}]} \\ &= S * \{T \operatorname{sinc}(\pi T f)\}. \end{aligned} \quad (114)$$

For instance, consider the sine signal $s(t) = \cos(2\pi f_0 t)$, whose FT is $S = \frac{1}{2}(\delta_{f_0} + \delta_{-f_0})$. (114) rereads:

$$\forall f, X(f) = \frac{T}{2} (\operatorname{sinc}(\pi T (f - f_0)) + \operatorname{sinc}(\pi T (f + f_0))). \quad (115)$$

Graphically, the Dirac impulses at $f = \pm f_0$ are replaced by cardinal sines, still centered at $\pm f_0$. Their main and side lobes are shown in Figure 20.

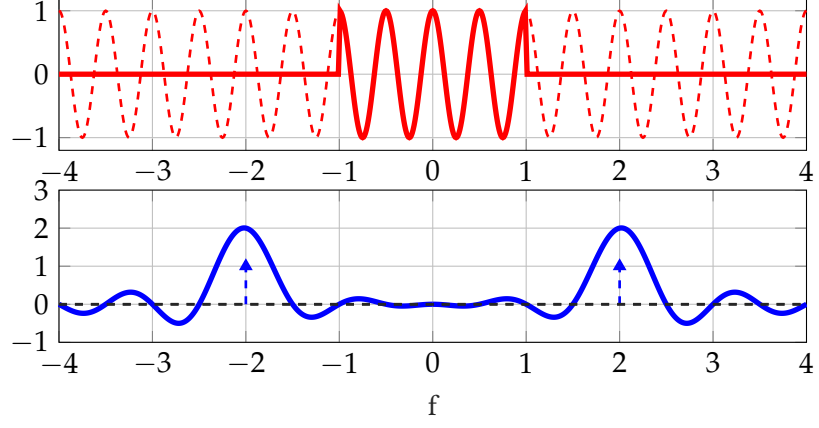


Figure 20: Sinusoid with a rectangular window: time and frequency representations.

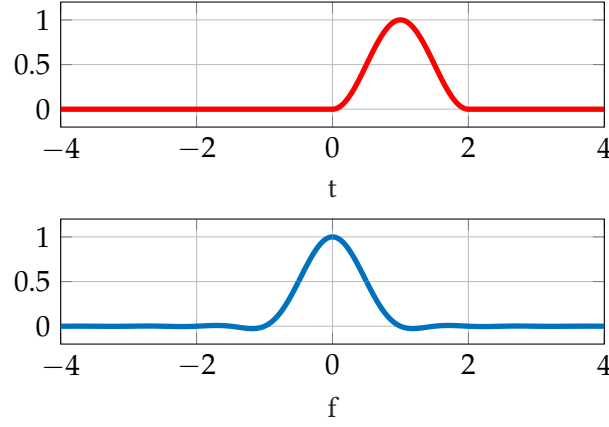


Figure 21: Hann window: time and frequency representations (modulus).

In the case of signals whose content are richer than a simple sinusoid, the presence of multiple side lobes and their superposition yield a difficulty in distinguishing close frequencies.

The effects of rectangular windowing can be somehow reduced by replacing (113) with:

$$\forall t \in \mathbb{R}, x(t) = s(t) w_{[-\frac{T}{2}, \frac{T}{2}]}(t) \quad (116)$$

where $w_{[-\frac{T}{2}, \frac{T}{2}]}(t)$ is the triangular window $\Delta_{[-T/2, T/2]}$ defined in (83), or an alternative window of specific shape like the Hann window:

$$w_{[-\frac{T}{2}, \frac{T}{2}]}(t) = \cos^2\left(\frac{\pi t}{T}\right) \mathbb{1}_{[-\frac{T}{2}, \frac{T}{2}]}(t) \quad (117)$$

shown in Fig. 21. These windows are of width T and yield a fast decay of side lobes (in $\frac{1}{f^2}$ or $\frac{1}{f^3}$, respectively). On the other hand, the width of the main lobe is large ($\frac{4}{T}$), see Fig. 22. In comparison, when using a rectangular window, the width of the main lobe more narrow (width of $\frac{2}{T}$) and the magnitude of side lobes decreases more slowly (in $\frac{1}{f}$).

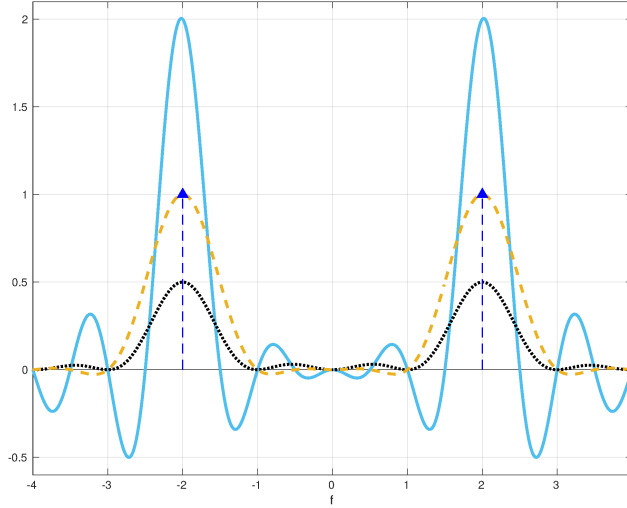


Figure 22: Windowing effects on a sinusoid: rectangular (continuous line), triangular (black dotted line), Hann (yellow dotted line) windows.

3.6.2 Spectral resolution

The concept of spectral resolution is related to the ability to distinguish close sinusoids, or more precisely, complex exponentials at close frequencies. The spectral resolution is influenced by the truncation window considered.

Definition 3.7 (Spectral resolution). *Let s_1 and s_2 be two complex exponentials of same magnitude, of respective frequencies f_1 and f_2 , defined over an observation window of duration T . The spectral resolution is the minimum frequency gap for which s_1 and s_2 can be distinguished by Fourier transform, regardless of the phase shift.*

For a rectangular window, the spectral resolution is equal to $\frac{2}{T}$, i.e., both complex exponentials can be distinguished as long as

$$|f_1 - f_2| \geq \frac{2}{T}. \quad (118)$$

For a mixture of more than two exponentials, the rule (118) is applied to the minimum difference between the frequencies of two of the exponentials.

The result (118) illustrates that one can distinguish more easily two complex exponentials of close frequencies when the observation time is long enough. The formula (118) results from the expression of the Fourier transform of the rectangular window: $T \sin_c(\pi T f)$, see (114). The width of the main lobe of this signal is equal to $\frac{2}{T}$ because $\sin_c(\pi T f)$ vanishes for $f = \pm \frac{1}{T}$.

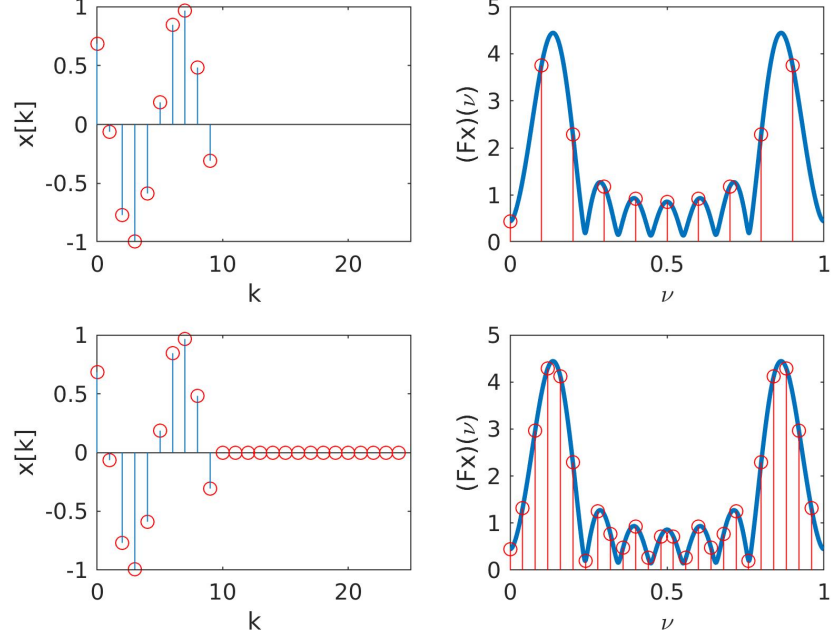


Figure 23: Frequency representations of a discrete signal of length $N = 10$. Right, the DTFT $X(\nu)$ is plotted in continuous line. Its evaluation by DFT is in red. The computation is realized for N frequency points (top figures). For a computation with M frequency points ($M > N$), we use zero-padding by adding $M - N$ zeros to the signal $x[k]$ ($M = 25$ bottom figures).

3.6.3 Time windowing of discrete time signals

For any discrete signal $s[k]$, the principle of windowing remains valid. The windowed signal is expressed by:

$$\forall k \in \mathbb{Z}, x[k] = s[k] w_{\{0, \dots, N-1\}}[k] \quad (119)$$

where $w_{\{0, \dots, N-1\}}$ is a discrete version of the previous windows (rectangular, triangular or Hann) of length N .

To compute the DTFT of x on a computer, we only keep the points k for which $w_{\{0, \dots, N-1\}}[k] \neq 0$. We therefore have a vector $\mathbf{x} = \{x_k, k = 0, \dots, N-1\}$ of length N , denoted $x_k = x[k]$. The zero-padding technique introduced in section 3.5.2 consists of artificially adding $M - N$ zeros to this vector, for $M > N$. The DFT of the augmented vector $\mathbf{x}$ provides an evaluation $\mathbf{X} \in \mathbb{C}^M$ of $\mathcal{F}x(\nu)$ on the frequency grid

$$\nu_m = \frac{m}{M}, \quad m = 0, \dots, M-1$$

of step $\frac{1}{M}$. Let us denote $\mathbf{X} = \text{TFD}(\mathbf{x})$. By locating the position, magnitude and phase of the local maximizers of $|X_m|$, we can retrieve the parameters of the sinusoids contained in the input signal $s[k]$.

The benefit of zero-padding is illustrated in Fig. 23, when $M = 2N$. There is a noticeable gain in accuracy while locating the frequency of the discrete sinusoid.

3.7 CONCLUSION

The Fourier transform is a fundamental tool to analyze stationary signals. Many signals are not stationary. More sophisticated tools are used such as the short-term Fourier transform or the wavelet transform. Finally, the FT plays a major role in understanding time-sampling of continuous signals. This operation is necessary for transmission, storage and reconstruction of time signals and will be detailed in the next chapter.

APPENDIX

The FT properties will be exhibited for L^1 functions.

Linearity. Let $z = \alpha x + \beta y$.

$$\begin{aligned} Z(f) &= \int_{\mathbb{R}} e^{-2i\pi ft} (\alpha x(t) + \beta y(t)) dt \\ &= \alpha \int_{\mathbb{R}} e^{-2i\pi ft} x(t) dt + \beta \int_{\mathbb{R}} e^{-2i\pi ft} y(t) dt \\ &= \alpha X(f) + \beta Y(f) \end{aligned}$$

Time dilatation. Let $z(t) = x(\alpha t)$.

$$\begin{aligned} Z(f) &= \int_{\mathbb{R}} e^{-2i\pi ft} x(\alpha t) dt \\ &= \int_{\mathbb{R}} e^{-2i\pi f \frac{\theta}{\alpha}} x(\theta) \frac{1}{|\alpha|} d\theta \\ &= \frac{1}{|\alpha|} X\left(\frac{f}{\alpha}\right). \end{aligned}$$

Duality of the FT. The inverse Fourier transform is given by

$$x(t) = \int_{\mathbb{R}} e^{2i\pi t\eta} X(\eta) d\eta.$$

Hence, we have

$$x(-t) = \int_{\mathbb{R}} e^{-2i\pi t\eta} X(\eta) d\eta = (\mathcal{F}X)(t)$$

Convolution product. Let $z = x * y$.

$$\begin{aligned}
 Z(f) &= \int_{\mathbb{R}} e^{-2i\pi f t} \int_{\mathbb{R}} x(\theta) y(t - \theta) d\theta dt \\
 &= \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-2i\pi f t} x(\theta) y(t - \theta) d\theta dt \\
 &= \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-2i\pi f(t-\theta)} e^{-2i\pi f \theta} x(\theta) y(t - \theta) d\theta dt \\
 &= \int_{\mathbb{R}} e^{-2i\pi f \theta} x(\theta) \int_{\mathbb{R}} e^{-2i\pi f(t-\theta)} y(t - \theta) dt d\theta \\
 &= \int_{\mathbb{R}} e^{-2i\pi f \theta} x(\theta) \int_{\mathbb{R}} e^{-2i\pi f \tau} y(\tau) d\tau d\theta \\
 &= \int_{\mathbb{R}} e^{-2i\pi f \theta} x(\theta) d\theta \int_{\mathbb{R}} e^{-2i\pi f \tau} y(\tau) d\tau \\
 &= X(f) Y(f).
 \end{aligned}$$

Entrywise product. Let $z = xy$.

$$\begin{aligned}
 Z(f) &= \int_{\mathbb{R}} e^{-2i\pi f t} x(t) y(t) dt \\
 &= \int_{\mathbb{R}} e^{-2i\pi f t} \left(\int_{\mathbb{R}} e^{2i\pi t \eta} \mathcal{F}x(\eta) d\eta \right) y(t) dt \\
 &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} e^{-2i\pi(f-\eta)t} y(t) dt \right) \mathcal{F}x(\eta) d\eta \\
 &= \int_{\mathbb{R}} \mathcal{F}y(f - \eta) \mathcal{F}x(\eta) d\eta \\
 &= (X * Y)(f) (\mathcal{F}y * \mathcal{F}x)(f)
 \end{aligned}$$

Time-delay. Let $z(t) = x(t - \tau)$. We can calculate:

$$\begin{aligned}
 Z(f) &= \int_{\mathbb{R}} e^{-2i\pi f t} x(t - \tau) dt \\
 &= \int_{\mathbb{R}} e^{-2i\pi f(\theta + \tau)} x(\theta) d\theta \\
 &= e^{-2i\pi f \tau} \int_{\mathbb{R}} e^{-2i\pi f \theta} x(\theta) d\theta \\
 &= e^{-2i\pi f \tau} X(f).
 \end{aligned}$$

We can find the same result more directly, by noting that $z = \delta_{\tau} * x$, and exploiting the property on the FT of a convolution product:

$$Z(f) = (\mathcal{F}\delta_{\tau})(f) X(f) = e^{-2i\pi f \tau} X(f)$$

Modulation / Frequency Translation.

$$\begin{aligned}
 \mathcal{F}(e_{f_0} x)(f) &= \int_{\mathbb{R}} e^{-2i\pi f t} \left(e^{2i\pi f_0 t} x(t) \right) dt \\
 &= \int_{\mathbb{R}} e^{-2i\pi(f-f_0)t} x(t) dt \\
 &= X(f - f_0).
 \end{aligned}$$

We can find the same result more directly by exploiting the property on the FT of an entrywise product:

$$\mathcal{F}(e_{f_0} x) = (\mathcal{F}e_{f_0}) * X = \delta_{f_0} * X.$$

Derivative. The FT is defined if $\lim_{t \rightarrow \infty} x(t) = 0$. An integration by parts leads to:

$$\begin{aligned} \mathcal{F}\{x'\}(f) &= \int_{\mathbb{R}} e^{-2i\pi ft} x'(t) dt \\ &= \lim_{a \rightarrow \infty} \left(e^{-2i\pi fa} x(a) - e^{2i\pi fa} x(-a) \right) + \int_{\mathbb{R}} (2i\pi f e^{-2i\pi ft}) x(t) dt \\ &= (2i\pi f) X(f). \end{aligned}$$

By induction, we conclude that

$$\mathcal{F}\{x^{(n)}\}(f) = (2i\pi f)^n X(f).$$

Multiplication by t^n . The derivative of the FT is

$$\begin{aligned} X'(f) &= \frac{d}{df} \int_{\mathbb{R}} e^{-2i\pi ft} x(t) dt \\ &= (-2i\pi) \frac{d}{df} \int_{\mathbb{R}} e^{-2i\pi ft} tx(t) dt \\ &= (-2i\pi) \mathcal{F}\{tx(t)\}(f) \end{aligned}$$

Hence we have $\mathcal{F}\{tx(t)\}(f) = \frac{1}{(-2i\pi)} X'(f)$. The property

$$\mathcal{F}\{t^n x(t)\}(f) = \frac{1}{(-2i\pi)^n} X^{(n)}(f)$$

is shown by induction.

SAMPLING AND RECONSTRUCTION

Digital processing of a signal coming from the real world, requires the signal to be represented in discrete form, *i.e.*, a sequence of numbers (real or complex), that can be stored in the memory of a computer.

The concept of time discretization of a (continuous time) signal has already been introduced in the mathematics course:

- in CIP, the existence of orthogonal bases of separable Hilbert spaces was established. For example, there is an isomorphism between functions in $L^2([0, 1])$ and the sequences in $\ell^2(\mathbb{Z})$ given by Fourier series.
- in PDE, discretization by finite elements or finite differences was addressed, where the regularity of functions (in the C^k sense or within Sobolev spaces) allows one to establish convergence of the discretization to the target continuous function.

In this course, the discretization of signals is addressed from the sampling point of view, which targets the following question: under what conditions is it possible to reconstruct a continuous time signal $x(t)$ from its values x_n at certain times t_n ?

In the case of uniform sampling, we will see that a signal can be perfectly reconstructed from its samples when its Fourier transform has compact support, provided that the sampling frequency is large enough.

4.1 TIME SAMPLING

Definition 4.1 (Time sampling). *Sampling is the construction of a discrete-time signal $x_d[n]$ from a continuous time signal $x(t)$. We will limit ourselves to uniform sampling, that is, given a sampling period $T_s > 0$,*

$$x_d[n] = x(nT_s). \quad (120)$$

We also define the sampling frequency $f_s = 1/T_s$.

Obviously, the signal x has to be continuous, otherwise equation (120) does not make sense. We assume that $x(t)$ is such that its Fourier transform is in $L^1(\mathbb{R})$. We know indeed that the inverse Fourier transform of an L^1 function is continuous and bounded. So, $x(t)$ is continuous and bounded. Consequently, the discrete signal x_d lays in $\ell^\infty(\mathbb{Z})$.

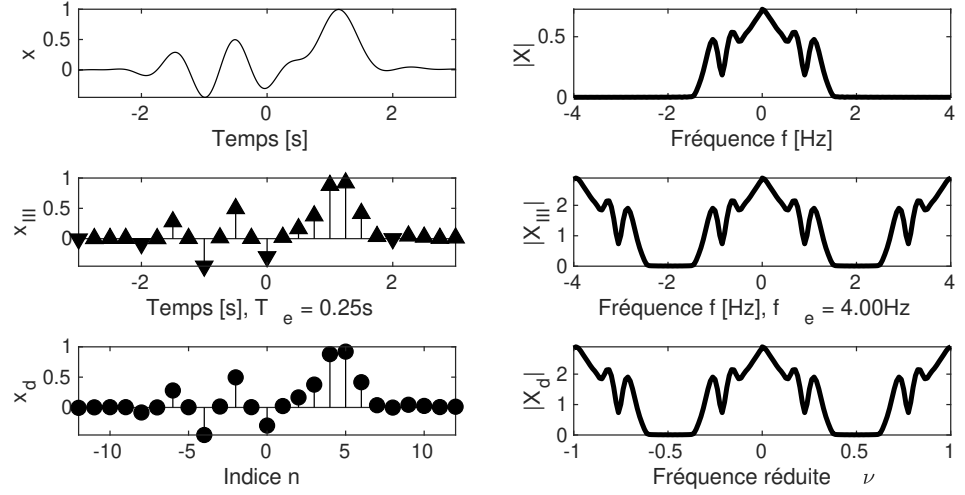


Figure 24: Sampling at $f_s = 4$ Hz. Left: signals x , $x_{||}$ and x_d . Right: their spectra. The FT $X_{||}(f)$ is f_s -periodic. It is equal to $X(f)$ on $[-f_s/2, f_s/2]$ up to a multiplicative factor. $X_d(v)$ is 1-periodic. The graphs of $|X_{||}(f)|$ and $|X_d(v)|$ are identical up to a scaling factor.

We further define the signal

$$x_{||} = \sum_{n \in \mathbb{Z}} x(nT_s) \delta_{nT_s} \quad (121)$$

which can be rewritten as the product of x and the Dirac comb $\mathbb{I}_{T_s}$:

$$x_{||} = \mathbb{I}_{T_s} x. \quad (122)$$

The signal $x_{||}$ makes it possible to link the continuous time signal $x(t)$ and the discrete-time signal $x_d[n]$ since

$$x_{||} = \sum_{n \in \mathbb{Z}} x_d[n] \delta_{nT_s}. \quad (123)$$

The left column of figure 24 displays a representation of these three signals.

The following theorem relates the FT of x , $x_{||}$ and x_d :

Theorem 4.1. Let $x(t)$ be a signal whose Fourier transform $X(f)$ is in L^1 . Define $x_{||}$ and x_d as above. The Fourier transform $X_{||}$ of $x_{||}$ is given by:

$$X_{||}(f) = f_s \sum_{n \in \mathbb{Z}} X(f - n f_s) \quad (124)$$

and the discrete-time Fourier transform of x_d is given by:

$$X_d(v) = f_s \sum_{n \in \mathbb{Z}} X(v f_s - n f_s). \quad (125)$$

The series converge in the L^1 sense on the intervals $]-\frac{f_s}{2}, \frac{f_s}{2}[$ and $]-\frac{1}{2}, \frac{1}{2}[$, respectively.

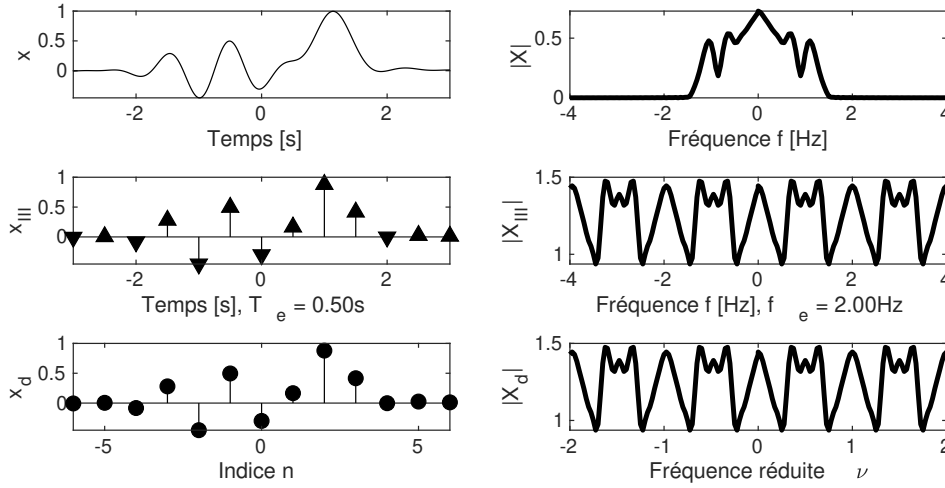


Figure 25: Sampling at $f_s = 2$ Hz. Left: signals x , x_{III} and x_d . Right: their spectra. In the spectrum $|X_{\text{III}}(f)|$, the copies of the spectrum $|X(f)|$ overlap (spectral aliasing).

Proof. X_{III} is the FT of the product of III_{T_s} and x . It is therefore equal to the convolution of the FT of both signals. According to theorem 3.6 on the FT of a Dirac comb,

$$X_{\text{III}}(f) = \left(\frac{1}{T_s} \text{III}_{\frac{1}{T_s}} * X \right)(f) = f_s \sum_{n \in \mathbb{Z}} X(f - n f_s).$$

In $L^1\left(-\frac{f_s}{2}, \frac{f_s}{2}\right)$, the sequence $N \rightarrow \sum_{n=-N}^N X(f - n f_s)$ is Cauchy because of the dominated convergence theorem. Since $L^1\left(-\frac{f_s}{2}, \frac{f_s}{2}\right)$ is completeness, the sequence is convergent.

The second identity of the theorem is obtained by calculating the Fourier transform of (123):

$$X_{\text{III}}(f) = \sum_{n \in \mathbb{Z}} x_d[n] e^{-i2\pi f n T_s} = X_d(f T_s),$$

denoting $\nu = f T_s = f / f_s$. $\square$

We will keep in mind that **the FT of a signal sampled at frequency f_s is f_s -periodic with the periodization pattern $X(f)$ replicated every f_s . Conversely, the FT of a periodic signal is a line spectrum**, see section 3.2.

Remark (Complex exponential signals). *Theorem 4.1 can only be applied when the Fourier transform X lays in L^1 . One can, however, derive generalized expressions when X is a distribution with compact support. Consider the complex exponential signal $x(t) = \exp(2i\pi f_0 t)$. Then, $X = \delta_{f_0} \notin L^1$. Since $x_d[n] = x(n T_s) = \exp(2i\pi \frac{f_0}{f_s} n)$, we calculate that (see Table 4):*

$$X_d = \sum_{n \in \mathbb{Z}} \delta_{\frac{f_0}{f_s} - n}. \quad (126)$$

This result shares similarities with (125) since the spectral line at $f = f_0$ in signal X induces a spectral line at $\frac{f_0}{f_s} - n$ for all $n \in \mathbb{Z}$ in signal X_d . Not though that the weighting factor f_s in (125) does not appear anymore.

Interpretation of Theorem 4.1

Sampling in time domain is equivalent to periodizing the spectrum of the signal. For arbitrary signals, it is therefore impossible to reconstruct a signal from its samples, since many copies of the spectrum can overlap, thus preventing reconstruction of the original spectrum. This is visible in Fig. 25.

However, when the replicated versions of the spectrum (with shifts every f_s) have disjoint supports as in Fig. 24, it is possible to reconstruct the spectrum of the original signal. This observation is formally stated hereafter.

4.2 RECONSTRUCTION

Hereafter, we show that for any band limited signal in L^2 , if the signal is sampled at a sufficiently small period, then it can be exactly reconstructed.

4.2.1 Band limited signals in $[-B, B]$

Definition 4.2 (Band-limited signal). *A band-limited signal is such that the support of its Fourier transform is a subset of $[-B, B]$ for some finite B . We will denote by PW_B the set of L^2 signals having a limited band in $[-B, B]$. PW_B is therefore the pre-image by the Fourier transform of the set of L^2 functions with support in $[-B, B]$.*

For example, the function defined by $x(t) = \text{sin}_c(\pi t)$, where we recall the definition of sin_c given in (74) :

$$\text{sin}_c(\theta) := \begin{cases} \frac{\text{sin}(\theta)}{\theta} & \text{si } \theta \neq 0 \\ 1 & \text{si } \theta = 0 \end{cases}$$

satisfies $x \in PW_{\frac{1}{2}}$ because its FT is $X = \mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]}$. Using time dilation - contraction properties of the FT, the function defined by $x_B(t) = \text{sin}_c(2\pi Bt)$ with $B > 0$ satisfies $x_B \in PW_B$ since $X_B = \frac{1}{2B} \mathbb{1}_{[-B, B]}$.

Functions in PW_B satisfy the assumptions of Theorem 4.1. Indeed, the FT of a function $x \in L^2$ is in L^2 , moreover any function $X \in L^2$ having a compact support is in L^1 .

4.2.2 Shannon's theorem

Shannon's theorem, also called the **sampling theorem**, is a fundamental theorem guaranteeing that information contained in a band-limited signal $x(t)$ is conserved after time-sampling if the sampling period is small enough.

Theorem 4.2 (Shannon's theorem). *Let $x \in PW_B$ and $T_s = 1/(2B)$. For all $t \in \mathbb{R}$, we have:*

$$\begin{aligned} x(t) &= \sum_{n \in \mathbb{Z}} x(nT_s) \operatorname{sinc} \left(\pi \frac{t - nT_s}{T_s} \right) \\ &= \sum_{n \in \mathbb{Z}} x_d[n] \operatorname{sinc} \left(\pi \frac{t - nT_s}{T_s} \right) \end{aligned} \quad (127)$$

Remark. Obviously, $x \in PW_B$ implies that $x \in PW_{B'}$ for $B' \geq B$. So, the theorem remains valid for any $T_s \leq \frac{1}{2B}$.

Shannon's criterion for a band-limited signal

For L^2 functions that are band-limited in $[-B, B]$, it is sufficient to sample at frequency

$$f_s \geq 2B \quad (\text{Shannon's criterion})$$

to perfectly reconstruct the signal $x(t)$ from its samples $x_d[n]$. Except very specific situations (e.g., holes in the spectrum), Shannon's criterion is also a necessary condition: sampling at frequency f_s without error is possible for signals whose having a limited band in $[-f_s/2, f_s/2]$ only.

In the example of Fig. 24, $B = 1.5$ Hz and $f_s = 4$ Hz. Shannon's criterion is met ($f_s \geq 2B$). It is not met on the example of Fig. 25, where $B = 1.5$ Hz and $f_s = 2$ Hz ($f_s < 2B$), hence the spectral aliasing phenomenon.

Interpretation of Shannon's theorem

Equation (127) shows that $x(t)$ can be interpolated from its samples $x_d[n]$. (127) can be interpreted as the convolution product of $x_{\sqcup}$ and $\operatorname{sinc}(\pi t/f_s)$. The FT of this convolution product is equal to the elementwise product of the FT of $x_{\sqcup}$ and the rectangular window $\mathbb{1}_{[-B, B]}$ (FT of the cardinal sine). Under the condition that $x \in PW_B$, one can notice that in the sum (124),

- For $n = 0$, the function $f \mapsto X(f)$ has a support in $[-B, B]$.
- For $n \neq 0$, the function $f \mapsto X(f - nf_s)$ has a support in $[nf_s - B, nf_s + B]$. This support does not intersect $[-B, B]$.

The reconstruction formula (127) can therefore be interpreted as a low pass filter that cuts out high order frequency content (for $|f| \geq B$) introduced by discretization.

Figures 26 and 27 display two reconstructed signals when the Shannon criterion is satisfied or not.

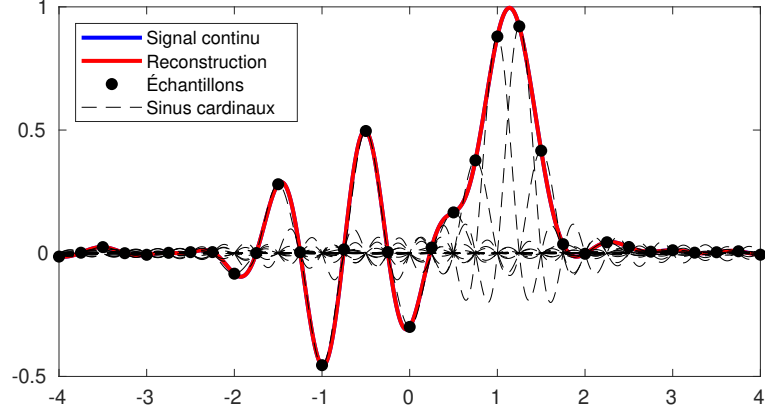


Figure 26: Case where Shannon's criterion is satisfied. Reconstruction of signal $x(t)$ using the sum (127).

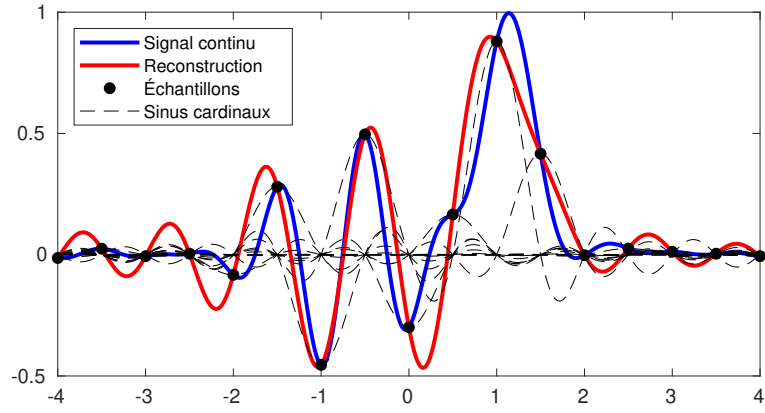


Figure 27: Case where the Shannon's criterion is not satisfied. Reconstruction of signal $x(t)$ using the sum (127).

4.2.3 Proof of Shannon's theorem

The theorem can be interpreted as the decomposition of x in an orthogonal basis of PW_B . The theorem is proved in two steps:

1. Construction of an orthogonal basis of PW_B ensuring convergence of (127) in L^2 .
2. Pointwise convergence analysis of the series.

Orthogonal basis of PW_B

Lemma 4.1. *The family of functions s_n defined by*

$$s_n(t) = \text{sin}_c \left(\pi \frac{t - nT_s}{T_s} \right), \quad n \in \mathbb{Z} \quad (128)$$

is an orthogonal basis of PW_B . The coefficients x_n of a function $x \in PW_B$ in this basis are given by:

$$x_n = x(nT_s), \quad (129)$$

and

$$x = \sum_{n \in \mathbb{Z}} x_n s_n \quad (130)$$

in the L^2 sense, where $T_s = \frac{1}{2B}$.

Proof. We have $s_0(t) = \text{sinc}(\pi \frac{t}{T_s})$. Since $s_n(t) = s_0(t - nT_s)$, we get

$$S_n(f) = S_0(f) e^{-2i\pi f n T_s}.$$

Furthermore, according to (87), the FT of s_0 reads:

$$S_0(f) = \frac{1}{2B} \mathbb{1}_{[-B, B]}(f).$$

The Fourier basis of $L^2([-B, B])$ is given by:

$$S_n(f) = \frac{1}{2B} \exp\left(-2i\pi \frac{nf}{2B}\right) \mathbb{1}_{[-B, B]}(f).$$

The Fourier transform being an isometry, the inverse Fourier transform of signals S_n yield a basis of PW_B . That is, the set of signals $\{s_n, n \in \mathbb{Z}\}$ is a basis of PW_B . According to Parseval's identity,

$$\|s_n\|_{L^2}^2 = \|S_n\|_{L^2}^2 = \int_{\mathbb{R}} |S_n(f)|^2 = \frac{1}{2B}.$$

Moreover, the weights X_n of $X(f)$ in the basis $\{S_n\}$ of L^2 identify with the weights x_n of x in the basis $\{s_n\}$ of PW_B . These weights are given by:

$$X_n = \frac{(X, S_n)}{\|S_n\|^2} = \int_{[-B, B]} X(f) \exp\left(2i\pi \frac{nf}{2B}\right) df$$

where $X = \sum_{n \in \mathbb{Z}} X_n s_n$ in the L^2 sense. By interpreting the integral as the inverse Fourier transform of $X(f)$ evaluated at $t = \frac{n}{2B} = nT_s$, we find

$$x_n = X_n = x(nT_s).$$

□

Convergence of the series

We now prove pointwise convergence of the series. The family s_n being an orthogonal basis, we get

$$\|x\|_{L^2}^2 = \sum_{n \in \mathbb{Z}} |x_n|^2 \|s_n\|_{L^2}^2 = \frac{1}{2B} \sum_{n \in \mathbb{Z}} |x_n|^2, \quad (131)$$

which shows that the sequence $(x_n)_{n \in \mathbb{Z}}$ is in $\ell^2(\mathbb{Z})$. Moreover, for all t , the sequence $(s_n(t))_{n \in \mathbb{Z}}$ is square integrable. By interpreting (127) as the inner product of $(x_n)_{n \in \mathbb{Z}}$ and $(s_n(t))_{n \in \mathbb{Z}}$, we obtain the convergence of this series using the Cauchy-Schwarz inequality.

We also show that the sum of the series is continuous. On the interval $t \in [-T_s, T_s]$:

- for all t , the sequence $(x_n s_n(t))_{n \in \mathbb{Z}}$ is in ℓ^1 ,

- for all n , $t \mapsto x_n s_n(t)$ is continuous,
- for all t , $|x_n s_n(t)| \leq |x_n| \max(1, (\pi(|n| - 1))^{-1})$. The upper bound reads as an elementwise product of two ℓ^2 sequences, therefore it is in ℓ^1 .

Thus, the series is a continuous function on $[-T_s, T_s]$. We can repeat this proof on any interval of the form $[-T_s + kT_s, T_s + kT_s]$ to prove continuity on $\mathbb{R}$. The sum of the series being continuous, and equal to the continuous function $x(t)$ in the L^2 sense, the series converges pointwise to x .

Going further...

For signals whose Fourier transform is a distribution with support in $[-B, B]$ (which ensures continuity and at most polynomial growth), the theorem must be slightly modified. The sampling frequency $f_s = 1/T_s$ must be strictly greater than $2B$, and the signal $x(t)$ is reconstructed by the formula

$$\forall t \in \mathbb{R}, x(t) = \sum_{n \in \mathbb{Z}} x(nT_s) \phi(t - nT_s) \quad (132)$$

where ϕ is chosen such that its Fourier transform Φ satisfies:

- $\Phi \in C^\infty$,
- $\Phi(f) = 1/f_s$ for $|f| \leq B$,
- $\Phi(f) = 0$ for $|f| \geq f_s/2$.

The series (132) can be written as $x_{\square} * \phi$. Its FT reads:

$$X_{\square}(f)\Phi(f) = f_s \sum_{n \in \mathbb{Z}} X(f - nf_s) \Phi(f). \quad (133)$$

The conditions on the values of Φ and the support of $X(f)$ allow to select only the term of the series for $n = 0$, i.e., to perfectly reconstruct $X(f)$.

Moreover, regularity of Φ is necessary to guarantee pointwise convergence of (132). Indeed, the samples $x(nT_s)$ are no longer bounded, but only with at most polynomial growth. The regularity of Φ implies that ϕ decreases faster than any polynomial, which guarantees convergence of the series.

4.2.4 Band limited signals in $[A, B]$

The sampling theorem (Theorem 4.2) is stated for band-limited signals in $[-B, B]$. The so-called “modulated” signals, having a support included in a non-symmetric interval $[A, B]$, can obviously be considered as band-limited signals in $[-C, C]$, with $C = \max(|A|, |B|)$. We can then apply the sampling theorem for $f_s \geq 2C$. Hereafter, we show that the bound $f_s \geq 2C$ can be improved, allowing one to use a sampling frequency lower than $2C$.

Shannon’s criterion for a modulated signal

Let $x(t)$ be an L^2 signal having a limited band in $[A, B]$. Assuming that the central frequency $\frac{A+B}{2}$ is known, it is possible to reconstruct this signal from the sampled signal $x_d[n]$ at a frequency f_s such that:

$$f_s \geq B - A. \quad (134)$$

Indeed, the so called “baseband” (*i.e.*, unmodulated) signal is written:

$$y(t) = x(t) e^{-i2\pi \frac{A+B}{2}t} = x(t) e^{-i\pi(A+B)t}. \quad (135)$$

Using the properties of the Fourier transform (see Tab. 1 for the FT of a modulated signal), we easily show that $y(t)$ is band limited in $[-\frac{B-A}{2}, \frac{B-A}{2}]$. Shannon’s theorem applied to $y(t)$ then guarantees exact reconstruction of $y(t)$ for $f_s \geq B - A$:

$$y(t) = \sum_{n \in \mathbb{Z}} y(nT_s) \operatorname{sinc}(\pi(t - nT_s)/T_s).$$

Using (135), we get the reconstruction equation for $x(t)$:

$$x(t) = \sum_{n \in \mathbb{Z}} x(nT_s) \operatorname{sinc}(\pi(t - nT_s)/T_s) e^{i\pi(A+B)(t-nT_s)}.$$

Note that this formula involves $A + B$. Concretely, it is necessary to know the value of the central frequency $\frac{A+B}{2}$ to be able to reconstruct a signal $x(t)$.

4.3 SAMPLING IN PRACTICE

4.3.1 Anti-aliasing filter

The Shannon theorem assumes that the signal to be sampled has a spectrum with compact support. This is not always the case in practice. Even when the spectrum has a compact support, it is not always possible to use a high enough sampling frequency. Sampling is therefore always preceded by a low-pass filtering of the signal, which cuts out the frequency band above $f_s/2$. It should be noted that this filter is necessarily implemented in analog form.

4.3.2 Practical reconstruction

The reconstruction formulas (127) and (132) raise several implementation issues. Indeed, the reconstruction of x at time t requires the knowledge of all the samples $x(nT_s)$, including those located in the future. Zero-order hold is a simple technique allowing to approximate the signal to be reconstructed by a piecewise constant signal, defined by:

$$\tilde{x}(t) = \sum_{n \in \mathbb{Z}} x_d[n] \mathbb{1}_{[nT_s, (n+1)T_s[}(t). \quad (136)$$

Reconstruction at time t only requires the knowledge of the sample at the previous time nT_s , where $n = \lfloor t/T_s \rfloor$.

The zero-order hold model can be analyzed by writing $\tilde{x}$ as a convolution product:

$$\tilde{x} = \mathbb{1}_{[0, T_s[} * x_{\square}. \quad (137)$$

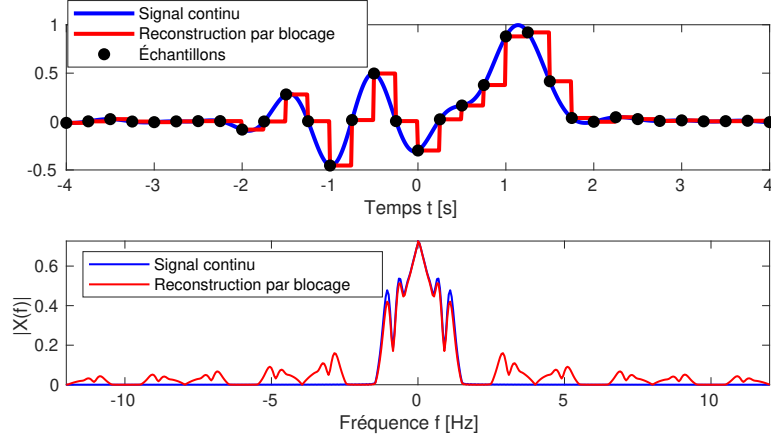


Figure 28: Zero-order hold reconstruction. The filtering of the signal in the useful band and the replicates of the attenuated spectrum are visible on the spectrum.

The FT of $\tilde{x}$ is therefore given by

$$\tilde{X}(f) = X_{\square}(f) \left(T_s e^{-i\pi f T_s} \frac{\sin \pi f T_s}{\pi f T_s} \right). \quad (138)$$

The effects of zero-order hold are:

- the introduction of a time shift of duration $T_s/2$ (term $e^{-i\pi f T_s}$),
- a filtering of the signal in the useful band ($(\sin \pi f T_s)/(\pi f T_s)$ is not constant in the useful band),
- the presence of spectral aliasing attenuated in $1/f$.

The last two effects can be limited by the use of a low-pass analog filter applied to $\tilde{x}$.

4.4 CONCLUSION

The sampling theory allows to link the Fourier transform of a continuous signal $x(t)$ and the discrete time Fourier transform of its sampled version x_d . We will first keep in mind that sampling a signal at a sampling period T_s results in periodizing its spectrum with period $f_s = 1/T_s$. Moreover, the information content of $x(t)$ is preserved whenever the Shannon criterion is satisfied, that is, when $f_s \geq 2B$ with B the maximum frequency contained in signal x . In this case, we can (theoretically) perfectly reconstruct $x(t)$ from the values of $x_d[n]$, $n \in \mathbb{Z}$. When Shannon's criterion is not satisfied, an anti-aliasing filter (low-pass) is used prior to time-sampling. Here, the signal reconstructed from the samples is an approximate version of the signal $x(t)$, containing its low-frequency components only.

SIGNALS MODELED BY RANDOM PROCESSES

The deterministic signal theory reaches its limits when dealing with signals whose values are not perfectly known (typically, one cannot use a function from $\mathbb{R}$ to $\mathbb{R}$). Random processes are powerful probabilistic models allowing to characterize the statistical properties of signals and to elaborate processing techniques exploiting these properties.

Consider a physical measurement (e.g., an audio recording) yielding a noisy signal. Clearly, one cannot provide a deterministic mathematical model that would perfectly explain the value of the noise at any time. Using random processes, it is possible to characterize the statistical properties of the noise and useful signal, thus allowing for signal denoising and reduction of the noise influence (this process will be detailed in section 6.5.3 of chapter 6).

Consider another example, in speech processing. Two sounds may seem identical from an auditory viewpoint, whereas the related associated signals have very different values. By studying the statistical properties of these signals, one can classify sounds e.g., in speech recognition or synthesize new signals that are very similar from an auditory viewpoint (speech synthesis).

5.1 DEFINITION AND CHARACTERIZATION OF RANDOM PROCESSES

Discrete time random processes are a generalization of random vectors to deal with signal models having an infinite time support. On the contrary, random vectors are defined from N scalar random variables. Therefore, they can be used to model a signal of finite dimension.

5.1.1 Definition

We will restrict ourselves to *discrete time* random processes.

Definition 5.1 (discrete time random process). A **discrete time random process**, denoted X_n , is a sequence of random variables $\{X_n(\omega)\}_{n \in \mathbb{Z}}$ defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$.

In signal processing, the index n usually corresponds to time or space.

We call **realization**¹ of the process X_n , the deterministic signal, denoted $x[n]$, induced from a realization of all random variables $\{X_n(\omega)\}_{n \in \mathbb{Z}}$. Thus, for a given ω , we have:

$$\forall n \in \mathbb{Z}, \quad x[n] = X_n(\omega).$$

It is important to make the distinction between:

- the random process, which is a sequence of random variables modeling time signals,
- the realization(s) of the process which are observed deterministic signals.

Practically, the realizations at hand are used to estimate the properties of a random process. These properties are further exploited to improve the quality of observed signals (e.g., noise suppression, signal deconvolution to remove the blur introduced by a measuring device).

Remark. We also commonly use the terms **stochastic processes** or **time series** to refer to random processes. A realization of a random process is also referred to as a **trajectory** or **observation**.

Example 5.1. Here are three examples of random processes with a representation of their realizations, see Fig. 29:

1. Sequence of i.i.d. variables $X_n = A_n$ with $A_n \sim \mathcal{N}(0, \sigma^2)$.
2. Constant process with random variable $Y_n = A$ with $A \sim \mathcal{N}(0, \sigma^2)$.
3. Modulation of amplitude by random variables
 $Z_n = A \cos(2\pi\nu_0 n) + B \sin(2\pi\nu_0 n)$ where $(A, B) \sim \mathcal{N}(\mathbf{0}, \Sigma)$.

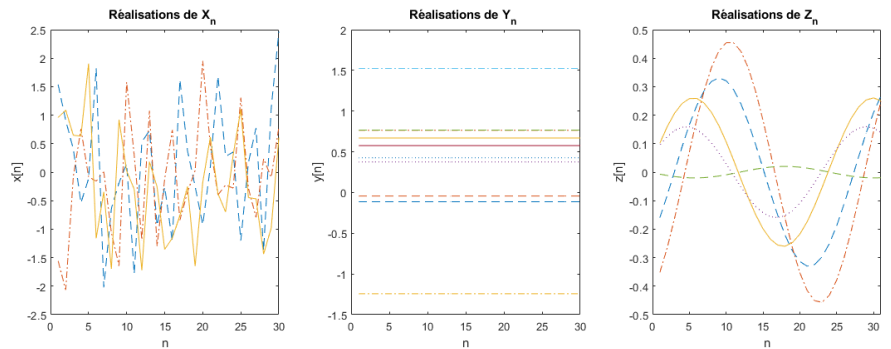


Figure 29: A few realizations of three random processes.

¹ In statistics, we talk about data rather than realizations.

Going further...

Continuous time random processes.

The most general definition of random processes is as follows.

Definition 5.2. A random process $X(\omega, t)$ is an application defined by

$$\begin{aligned} X &: \Omega \times \mathcal{T} \longrightarrow K \\ (\omega, t) &\longmapsto X(\omega, t), \end{aligned}$$

where

- $(\Omega, \mathcal{E}, P)$ is a probability space,
- K is a topological or measurable space (usually, a real or complex normed vector space),
- $\mathcal{T}$ is a set,

and such that for any $t \in \mathcal{T}$, the partial application

$$\begin{aligned} \Omega &\longrightarrow K \\ \omega &\longmapsto X(\omega, t) \end{aligned}$$

is a random variable.

t is the variable of evolution of the signal, usually time or space. For fixed ω_0 , $X(\omega_0, t)$ is a realization of the random process. It is a signal depending on t only. Depending on the set defined by $\mathcal{T}$, we obtain a discrete or continuous time process:

- $t \in \mathbb{R}$: continuous time process
- $t \in \mathbb{Z}$: discrete time process (we rather note $X(\omega, n)$ with $n \in \mathbb{Z}$)

5.1.2 Finite-dimensional distributions

The law of a random process is fully characterized by its finite-dimensional distributions for any order.

Definition 5.3 (Finite-dimensional distributions). For $N > 0$, the finite-dimensional distributions of order N are the joint probability distributions defined for any $(n_1, n_2, \dots, n_N) \in \mathbb{Z}^N$ by

$$P(X_{n_1} \in [a_1, b_1], X_{n_2} \in [a_2, b_2], \dots, X_{n_N} \in [a_N, b_N]).$$

However, this characterization is generally hard to derive and interpret. It is therefore rarely used in practice.

5.1.3 Gaussian process

Definition 5.4 (Gaussian process). A **process is Gaussian** if its finite-dimensional distributions are all Gaussian: $\forall N > 0$, $\forall (n_1, n_2, \dots, n_N) \in \mathbb{Z}^N$,

$$(X_{n_1}, X_{n_2}, \dots, X_{n_N}) \sim \mathcal{N}(\mu_{n_1, n_2, \dots, n_N}, \Sigma_{n_1, n_2, \dots, n_N})$$

or equivalently, if any (finite) weighted combination of X_n is a Gaussian random variable: $\forall N > 0$, $\forall (n_1, n_2, \dots, n_N) \in \mathbb{Z}^N$, $\forall (\alpha_1, \alpha_2, \dots, \alpha_N) \in \mathbb{R}^N$,

$$\exists \mu, \sigma : \sum_{i=1}^N \alpha_i X_{n_i} \sim \mathcal{N}(\mu, \sigma^2)$$

Notice that the fact that the marginal laws of X_n are Gaussian is not sufficient to ensure that the process X is Gaussian. One can construct a non-Gaussian process whose random variables X_n are Gaussian for all n .

5.1.4 Mean and autocorrelation

We have seen that the complete description of a random process by its finite-dimensional distribution is a rather complex task. We often limit ourselves to a partial description using the first two moments: **mean** and **autocorrelation**. For example, one can distinguish different speech phonemes (such as the /f/ and /s/ sounds) from the first two moments of an audio signal.

A second reason for focusing on the first two moments is that for a Gaussian process, they fully characterize the process. For non Gaussian processes, they only yield a partial characterization.

Definition 5.5 (Mean). We define the **mean** (or first order moment) of a random process X_n by

$$\mu_X[n] = \mathbb{E}(X_n).$$

Definition 5.6 (Autocorrelation). We define the **autocorrelation** (or second order moment) of a random process X_n by

$$r_X[n_1, n_2] = \mathbb{E}(X_{n_1} X_{n_2}^*).$$

Remark. If $\mathbb{E}(|X_n|^2) < \infty$, the random variables X_n belong to the space $L^2(\Omega, \mathcal{F}, \mathcal{P})$. The autocorrelation $r_X[n_1, n_2] = \mathbb{E}(X_{n_1} X_{n_2}^*)$ rewrites as the inner product between X_{n_1} and X_{n_2} in $L^2(\Omega, \mathcal{F}, \mathcal{P})$.

μ_X and r_X are both deterministic signals. $\mu_X[n]$ gives the mean value of the random process at time n while $r_X[n_1, n_2]$ gives the correlation between times n_1 and n_2 .

For most random processes modeling physical phenomena, $r_X[n_1, n_2]$ decreases when $|n_2 - n_1|$ grows. Intuitively, this corresponds to the fact that two samples taken at close times are more correlated.

Definition 5.7 (Cross-correlation). We define the **cross-correlation** between two processes X_n and Y_n by

$$r_{XY}[n_1, n_2] = \mathbb{E} (X_{n_1} Y_{n_2}^*).$$

Remark. The cross-correlation of a process with itself coincides with its autocorrelation, since $r_{XX}[n_1, n_2] = r_X[n_1, n_2]$.

Definition 5.8 (Uncorrelated random processes). The random processes X_n and Y_n are **uncorrelated** if

$$\forall n_1, n_2 \in \mathbb{Z}, \quad r_{XY}[n_1, n_2] - \mu_X[n_1] \mu_Y^*[n_2] = 0,$$

or equivalently,

$$\forall n_1, n_2 \in \mathbb{Z}, \quad \mathbb{E} ((X_{n_1} - \mu_X[n_1])(Y_{n_2} - \mu_Y[n_2])^*) = 0.$$

Remark. If $Z = X + Y$ with X and Y uncorrelated and centered then:

$$r_Z[n_1, n_2] = r_X[n_1, n_2] + r_Y[n_1, n_2].$$

Going further...

The centered second order moment is called **autocovariance**. It is defined as $\mathbb{E} ((X_{n_1} - \mu_X[n_1])(Y_{n_2} - \mu_Y[n_2])^*)$. The **cross-covariance** is defined likewise: $\mathbb{E} ((X_{n_1} - \mu_X[n_1])(Y_{n_2} - \mu_Y[n_2])^*)$. If X_n and Y_n are centered, their auto/cross-covariance coincides with their auto/cross-correlation.

5.2 STATIONARY PROCESSES

Stationarity is a key notion. In a nutshell, stationarity means that the statistical properties of the process do not change over time. A trivial example of non-stationary process is when the mean value $\mu_X[n]$ is varying with respect to n .

5.2.1 Strict-sense stationarity

Definition 5.9 (Strict-sense stationarity). *A random process is strictly stationary if its finite-dimensional distributions are translation invariant: for all $N > 0$ and $k \in \mathbb{Z}$, the random vectors $(X_{n_1}, \dots, X_{n_N})$ and $(X_{n_1+k}, \dots, X_{n_N+k})$ have the same distribution.*

Strict-sense stationarity may be difficult to establish and of limited interest. Therefore, we will focus on wide-sense stationarity.

5.2.2 Wide-sense stationarity

Definition 5.10 (Wide-sense stationarity). *A random process is **wide-sense stationary (WSS)** (or **second order stationary** or **weak-sense stationary**) if:*

- the mean is constant (it does not depend on time):

$$\forall n, \quad \mu_X[n] = \mu_X.$$

- it is of finite power: $\forall n, \mathbb{E}(|X_n|^2) < \infty$
- the autocorrelation depends only on the time difference between both times considered:

$$\forall n_1, n_2 \quad r_X[n_1, n_2] = \gamma_X[n_1 - n_2].$$

Note that if X_n is WSS, then $\mathbb{E}(|X_n|^2) = r_X[n, n]$ is constant, and $r_X[n + k, n] = \mathbb{E}(X_{n+k}X_n^*) = \mathbb{E}(X_nX_{n-k}^*)$. Hereafter, we will use the simplified notation:

$$\gamma_X[k] := r_X[n + k, n] = \mathbb{E}(X_{n+k}X_n^*) \quad (139)$$

for the autocorrelation function. Similarly, the cross-correlation between two jointly stationary processes is denoted by:

$$\gamma_{XY}[k] := \mathbb{E}(X_{n+k}Y_n^*). \quad (140)$$

Summarizing, two functions should be distinguished:

- the bivariate autocorrelation function (defined for all random processes): $r_X[n_1, n_2] = \mathbb{E}(X_{n_1}X_{n_2}^*)$.
- the univariate autocorrelation function (defined only for stationary processes): $\gamma_X[k] = \mathbb{E}(X_{n+k}X_n^*)$.

In the following, all processes are assumed wide-sense stationary. This implies that their second order moments are finite. We will therefore use notations γ_X and γ_{XY} .

Definition 5.11 (Variance). The **variance** of a stationary process X_n is defined by:

$$\sigma_X^2 = \mathbb{E}(|X_n - \mu_X|^2) = \mathbb{E}(|X_n|^2) - |\mu_X|^2.$$

Definition 5.12 (Power). The **power** of a stationary process X_n is the value of the autocorrelation at zero:

$$P_X = \mathbb{E}(|X_n|^2) = \gamma_X[0] = \sigma_X^2 + |\mu_X|^2.$$

For a centered process, the power is equal to the variance.

Definition 5.13 (Signal-to-noise ratio). The **Signal-to-noise ratio** quantifies the perturbation of a stationary process $X_n = U_n + W_n$, where U_n is the noiseless signal and W_n the perturbing noise. It is defined by the ratio of the powers of U_n and W_n . It is most often expressed in dB:

$$\text{SNR}_{\text{dB}} = 10 \log_{10} \left(\frac{\mathbb{E}(|U_n|^2)}{\mathbb{E}(|W_n|^2)} \right). \quad (141)$$

Property. The autocorrelation of a stationary process is finite and maximum at zero:

$$\forall k \in \mathbb{Z}, |\gamma_X[k]| \leq \gamma_X[0].$$

Proof. The result is trivial using the Cauchy-Schwarz inequality with $\gamma_X[k] = \langle X_{n+k}, X_n \rangle$ (by denoting $\langle \cdot, \cdot \rangle$ the inner product between random variables). Moreover, by definition of a wide-sense stationary process, $\gamma_X[0]$ is finite. $\square$

Property. The autocorrelation has Hermitian symmetry:

$$\forall k \in \mathbb{Z}, \gamma_X[k] = \gamma_X^*[-k]$$

Proof.

$$\gamma_X[-k] = \mathbb{E}(X_{n-k} X_n^*) = \mathbb{E}(X_{n-k}^* X_n)^* = \gamma_X^*[k].$$

$\square$

5.2.3 Power spectral density

It is important to characterize a random process in the frequency domain. For example, telecommunication signals are commonly modeled by random

processes, and one usually seeks to quantify the spectral band used by these signals in order to optimize the transmission channels. The main tool to characterize stationary random processes in the frequency domain is the Power Spectral Density (PSD).

Definition 5.14. The *Power spectral density (PSD)* of a stationary random process X_n is defined by:

$$\forall \nu \in \mathbb{R}, \quad \Gamma_X(\nu) = \lim_{N \rightarrow \infty} \mathbb{E} \left(\frac{1}{2N+1} \left| \sum_{k=-N}^N X_k e^{-i2\pi k\nu} \right|^2 \right).$$

Theorem 5.1 (Wiener-Khinchin). The power spectral density is the Fourier transform of the autocorrelation of the random process:

$$\forall \nu \in \mathbb{R}, \quad \Gamma_X(\nu) = \sum_{k \in \mathbb{Z}} \gamma_X[k] e^{-i2\pi \nu k}.$$

Proof. By definition of the PSD:

$$\Gamma_X(\nu) = \lim_{N \rightarrow \infty} \mathbb{E} \left(\frac{1}{2N+1} \left| \sum_{k=-N}^N X_k \exp(-i2\pi k\nu) \right|^2 \right). \quad (142)$$

We write:

$$\begin{aligned} \Gamma_X^N(\nu) &= \mathbb{E} \left(\frac{1}{2N+1} \left| \sum_{k=-N}^N X_k \exp(-i2\pi k\nu) \right|^2 \right) \\ &= \frac{1}{2N+1} \mathbb{E} \left(\left(\sum_{k=-N}^N X_k \exp(-i2\pi k\nu) \right)^* \left(\sum_{k=-N}^N X_k \exp(-i2\pi k\nu) \right) \right) \\ &= \frac{1}{2N+1} \mathbb{E} \left(\sum_{k=-N}^N \sum_{n=-N}^N X_k^* X_n \exp(-i2\pi(n-k)\nu) \right) \\ &= \frac{1}{2N+1} \sum_{k=-N}^N \sum_{n=-N}^N \mathbb{E}(X_k^* X_n) \exp(-i2\pi(n-k)\nu) \\ &= \frac{1}{2N+1} \sum_{k=-N}^N \sum_{n=-N}^N \gamma_X[n-k] \exp(-i2\pi(n-k)\nu) \\ &= \sum_{m=-2N}^{2N} \frac{2N+1-|m|}{2N+1} \gamma_X[m] \exp(-i2\pi m\nu) \\ &= \sum_{m \in \mathbb{Z}} \max \left(0, \frac{2N+1-|m|}{2N+1} \right) \gamma_X[m] \exp(-i2\pi m\nu) \end{aligned}$$

If $\gamma_X \in L^1$, we can bound the terms in the series by $|\gamma_X[m]|$ and apply the dominated convergence theorem to invert limit and sum. We get:

$$\Gamma_X(\nu) = \lim_{N \rightarrow \infty} \Gamma_X^N(\nu) = \sum_{m \in \mathbb{Z}} \gamma_X[m] \exp(-i2\pi m\nu).$$

Moreover, $\Gamma_X(\nu)$ is continuous.

If $\gamma_X \notin L^1$, the limit has to be calculated in the sense of distributions:

$$\lim_{N \rightarrow \infty} \langle \Gamma_X^N, \phi \rangle = \left\langle \sum_{m \in \mathbb{Z}} \gamma_X[m] \delta_m, \Phi \right\rangle$$

where ϕ is a test function and Φ denotes its Fourier transform. □

The Wiener-Khinchin theorem is frequently used to compute the power spectral density because it is simpler to calculate the Fourier transform of the autocorrelation function than applying the definition of the PSD.

Property. *The power spectral density is real and non-negative:*

$$\forall \nu, \Gamma_X(\nu) \in \mathbb{R}_+$$

This property is a direct consequence of the definition.

Remark. *If the process X_n is real valued, then its PSD is even:*

$$\forall \nu, \Gamma_X(\nu) = \Gamma_X(-\nu).$$

Indeed, if $X_n \in \mathbb{R}$, then $\gamma_X[k] \in \mathbb{R}$. Since γ_X has Hermitian symmetry, we have $\gamma_X[-k] = \gamma_X[k]$ for all $k \in \mathbb{N}$. From the properties of the Fourier transform, $\Gamma_X(\nu)$ is even.

Property. *The **power** of a stationary process is equal to the integral of the power spectral density:*

$$\gamma_X[0] = \int_{-1/2}^{1/2} \Gamma_X(\nu) d\nu.$$

Proof. According to the Wiener-Khinchin theorem, the autocorrelation is the inverse Fourier transform of the PSD:

$$\gamma_X[n] = \int_{-1/2}^{1/2} \Gamma_X(\nu) e^{i2\pi\nu n} d\nu$$

For $n = 0$, we get the result. $\square$

The name “power spectral density” is justified by the fact that integrating the PSD in the spectral domain yields the power of the random process. By integrating the PSD in a specific frequency band, we obtain the power of the process in this frequency band.

Calculation of the PSD

The usual method for analytically calculation of the PSD is as follows:

1. Compute the mean $\mu_X[n] = \mathbb{E}(X_n)$
2. Compute the autocorrelation $r_X[k_1, k_2] = \mathbb{E}(X_{k_1} X_{k_2}^*)$
3. Check that the process is WSS:
 - $\mu_X[k] = \mu_X$,
 - $r_X[k_1, k_2] = \gamma_X[k_1 - k_2]$
4. Check the consistency of your results using the properties of $\gamma_X[k]$:
 - $\gamma_X[0] \geq |\gamma_X[k]|$,

- $\gamma_X[k] = \gamma_X^*[-k]$.

5. Calculate the power spectral density: $\Gamma_X(\nu) = \mathcal{F}\gamma_X(\nu)$.

5.2.4 White noise

A white noise is a random process used to model various sources of error:

- measurement errors due to sensor imperfections (measurement noise);
- the inaccuracy of the mathematical model describing the useful signal, when this model is too rough.

The white noise then models the non-informative part of the data. In other scenarios, a white noise is used to describe the useful (informative) part of the data: when placed as input of a filter, the output signal is a random process describing the diversity of signals that can be observed in practice.

Definition 5.15 (white noise in the strict sense). *A process W_n is a **white noise in the strict sense** if the following properties are met:*

- W_n is a centered wide-sense stationary process,
- $\gamma_W[n] = \sigma^2\delta[n]$,
- W_n and W_k are independent for all $k \neq n$.

Definition 5.16 (white noise in the wide sense). *W_n is a **white noise in the wide sense** (or in the weak sense), of variance σ^2 if the following properties are met:*

- W_n is a centered wide-sense stationary process,
- $\gamma_W[n] = \sigma^2\delta[n]$.

Remark.

- A white noise being centered, its power equals its variance: $P_W = \sigma^2$.
- A Gaussian white noise is always a white noise in the strict sense because uncorrelation implies independence for Gaussian processes.

Property. *Let W_n be a white noise in the wide sense. We have:*

- W_n and W_m are **uncorrelated** for $n \neq m$.
- **The PSD of W_n is constant:** $\Gamma_W = \sigma^2\mathbb{1}_{\mathbb{R}}$.

The name “white noise” comes from this property by analogy to white light whose spectrum contains all the frequency components in the same proportions.

Convention. Throughout the rest of the handout, the short name “white noise” refers to white noise in the wide sense.

Example 5.2. Consider two random processes: a white noise and a colored noise. Fig. 30 displays a realization of these processes, their autocorrelation and their PSD.

- **Realizations:** From an empirical viewpoint, by observing a realization of each random process, we notice that for the white noise (left), the successive values seem to be uncorrelated; while for the colored noise (right), we observe a positive/negative sequence, showing a certain correlation.

We can confirm these comments from the theoretical viewpoint:

- **Autocorrelation:** the autocorrelation of the white noise is impulse-shaped, while that of the colored noise is positive valued for samples spaced by an even number of points and negative for odd intervals.

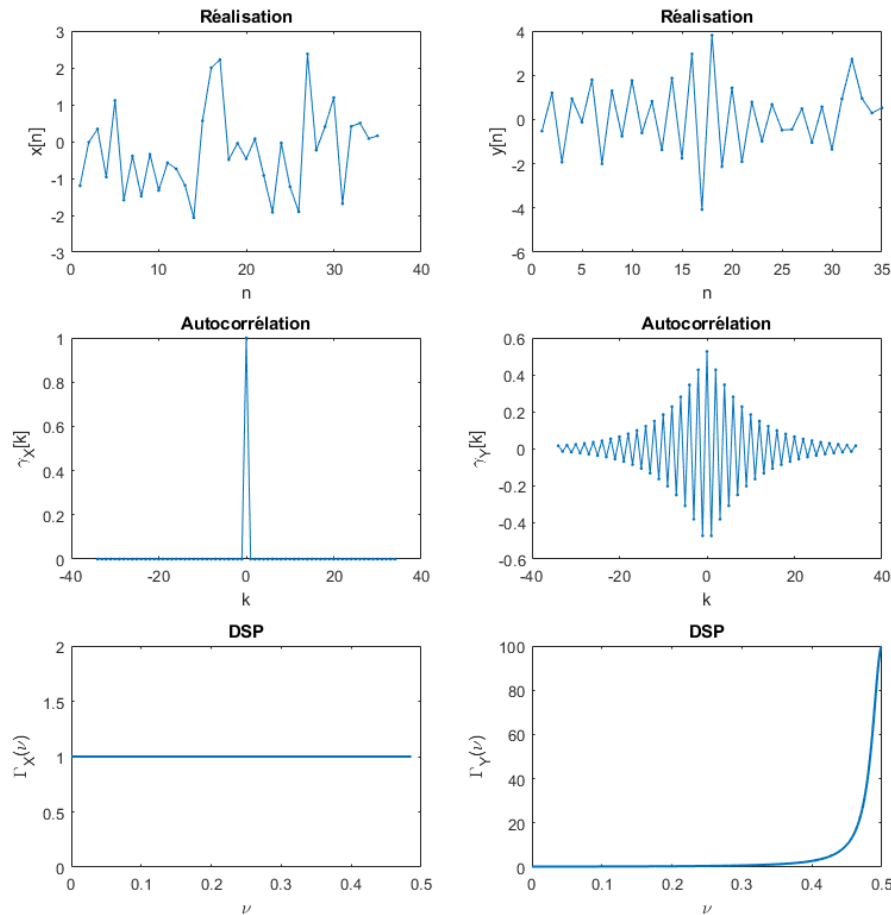


Figure 30: Comparison of two random processes: white noise (left) and colored noise (right).

- **Power Spectral Density:** the white noise PSD is constant. The colored noise PSD has high frequency content, which is consistent with the fast variations observed in the realization signal.

5.2.5 Linear filtering of random processes

Linear convolution filters are one of the main tools in signal processing. The purpose of this section is to analyze the properties of a filtered stationary random process.

Interference formula

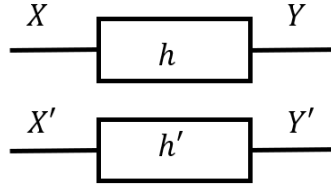


Figure 31: Linear filtering of two random processes (interference formula).

Theorem 5.2 (Interference formula). *Let:*

- X_n and X'_n be two wide-sense stationary random processes,
- h and h' be the impulse responses (ℓ^1) of two stable filters
- Y_n and Y'_n denote the output (filtered) random processes:

$$Y_n = (h * X)_n = \sum_{k \in \mathbb{Z}} h[k] X_{n-k},$$

$$Y'_n = (h' * X')_n = \sum_{k \in \mathbb{Z}} h'[k] X'_{n-k}.$$

Then:

- the filtered processes Y_n and Y'_n are stationary,
- their cross-correlation reads:

$$\gamma_{YY'} = h * \bar{h}' * \gamma_{XX'}$$

with $\bar{h}'[k] = h'^*[-k]$.

Proof.

$$\begin{aligned}
\gamma_{YY'}[k] &= \mathbb{E}(Y_{n+k} Y_n'^*) \\
&= \mathbb{E} \left(\sum_{i \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} h[i] h'^*[j] X_{n+k-i} X_{n-j}'^* \right) \\
&= \sum_{i \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} h[i] h'^*[j] \mathbb{E}(X_{n+k-i} X_{n-j}'^*) \\
&= \sum_{j \in \mathbb{Z}} h'^*[j] \sum_{i \in \mathbb{Z}} h[i] \gamma_{XX'}[j+k-i] \\
&= \sum_{j \in \mathbb{Z}} h'^*[j] g[j+k] \quad \text{with } g = h * \gamma_{XX'} \\
&= \sum_{j \in \mathbb{Z}} \bar{h}'[-j] g[j+k] \\
&= \sum_{j \in \mathbb{Z}} \bar{h}'[j] g[k-j] \\
&= (\bar{h}' * g)[k] \\
&= (\bar{h}' * h * \gamma_{XX'})[k]
\end{aligned}$$

□

This theorem (referred to as the **interference formula**) provides a relation between the second order moments of the filtered random processes and those of the input processes. The interference formula has many practical applications, some of which are detailed hereafter.

Autocorrelation and PSD of a filtered random process

Corollary 5.1. *Let*

- X_n be a wide-sense stationary process,
- h be the impulse response (ℓ^1) of a stable filter.

*The filtered process $Y_n = h * X_n$ has the following properties:*

- *the autocorrelation of Y_n reads: $\gamma_Y = h * \bar{h} * \gamma_X$,*
- *the PSD of Y_n reads: $\Gamma_Y(\nu) = |H(\nu)|^2 \Gamma_X(\nu)$.*

Proof. Apply the interference formula with $X' = X$ and $h' = h$. Then, use the Fourier transform property: $\mathcal{F}\bar{h}(\nu) = H^*(\nu)$. □

This corollary is an important result. It allows us to write the PSD (as well as the autocorrelation function) of the filtered process with respect to the input random process.

Identification of the impulse response of a linear filter

Identification of the impulse response of an unknown linear filter is a classical problem. One possible approach aims to use a white noise as input of the filter, and calculate the cross-correlation between this white noise and the related output signal.

Corollary 5.2. *Let*

- W_n be a white noise,
- h be the impulse response (ℓ^1) of a stable filter.

*The cross-correlation between the input noise W_n and the filtered process $Y_n = h * W_n$ is proportional to the impulse response of the filter:*

$$\gamma_{YW}[k] = \sigma^2 h[k].$$

Proof. Apply the interference formula with inputs $X_n := W_n$ and $X'_n := W_n$, and filters $h[n]$ and $h'[n] := \delta[n]$.

The output signals read $Y_n = (h * W)_n$ and $W_n = (h' * W)_n$.

Thus, $\gamma_{YW} = h * \bar{h}' * \gamma_{XX'} = h * \delta * \gamma_W = \sigma^2 h$. $\square$

5.2.6 Autoregressive model

Autoregressive models are commonly used to study time series in various contexts such as pollution measurements, financial data, or audio time series. They allow one to predict future data from past observations. A second reason for using AR models is that their parameters can be estimated in a simple way, as we will see in Chapter 6.

Definition

Definition 5.17. (Autoregressive model) *An autoregressive model (AR) of order p is random process obtained by filtering a white noise using a stable filter of transfer function:*

$$H(\nu) = \frac{1}{1 - \sum_{k=1}^p a_k e^{-i2\pi k\nu}} \quad \text{where } a_1, \dots, a_p \in \mathbb{R}.$$

Time domain: recurrence equation

According to section 3.5.1, the AR model can be expressed in time domain using a recurrence equation.

Theorem 5.3. *An AR process Y_n of order p reads:*

$$Y_n = \sum_{k=1}^p a_k Y_{n-k} + W_n \quad (143)$$

where W_n is a white noise.

Remark. *The impulse response of the filter is of infinite duration. It is therefore more relevant to use the recurrence equation above to describe the input-output relationship, because the latter involves a finite number of terms Y_{n-k} .*

Parameters of an AR model

An AR process of order p is fully parameterized by:

- the variance σ^2 of the generating white noise W_n ,
- the coefficients $\{a_k\}_{1 \leq k \leq p}$ of the recurrence equation.

PSD of an AR process

Since an AR process is obtained by filtering a white noise, its PSD can be written using Corollary 5.1:

$$\Gamma_Y(\nu) = \sigma^2 \left| \frac{1}{1 - \sum_{k=1}^p a_k e^{-i2\pi k\nu}} \right|^2. \quad (144)$$

Autocorrelation of an AR model

Property. The autocorrelation of an AR process satisfies:

$$\forall k \geq 0, \quad \gamma_Y[k] = \sum_{i=1}^p a_i \gamma_Y[k-i] + \sigma^2 \delta[k] \quad (145)$$

with $\gamma_Y[k] = \gamma_Y[-k]$ for $k < 0$.

Proof.

$$\begin{aligned} \gamma_Y[k] &= \mathbb{E}(Y_n Y_{n-k}) \\ &= \mathbb{E} \left(\left(\sum_{i=1}^p a_i Y_{n-i} + W_n \right) Y_{n-k} \right) \\ &= \sum_{i=1}^p a_i \mathbb{E}(Y_{n-i} Y_{n-k}) + \mathbb{E}(W_n Y_{n-k}) \\ &= \sum_{i=1}^p a_i \gamma_Y[k-i] + \sigma^2 \delta[k] \end{aligned}$$

The result $\mathbb{E}(W_n Y_{n-k}) = \sigma^2 \delta[k]$ comes from (143) and the fact that W_n is white noise ($\mathbb{E}(W_i W_j) = \sigma^2 \delta[i-j]$). Indeed, according to (143), Y_{n-k} depends only on W_i at times $i \leq n-k$. We further obtain:

- for $k > 0$, $\mathbb{E}(W_n Y_{n-k}) = 0$;
- for $k = 0$,

$$\begin{aligned} \mathbb{E}(W_n Y_{n-k}) &= \mathbb{E}(W_n Y_n) \\ &= \mathbb{E} \left(W_n \left(\sum_{i=1}^p a_i Y_{n-i} + W_n \right) \right) \\ &= \sum_{i=1}^p a_i \mathbb{E}(W_n Y_{n-i}) + \mathbb{E}(W_n^2) \\ &= 0 + \sigma^2. \end{aligned}$$

□

Equation (145) yields a system of $p + 1$ linear equations, obtained for $0 \leq k \leq p$, referred to as the **Yule-Walker** equations. One can estimate the $p + 1$ model parameters, namely $\{a_1, a_2, \dots, a_p, \sigma^2\}$ by inverting this system. This subject will be further addressed in Chapter 6.

5.3 CONCLUSION

This chapter first introduced the notion of stationary discrete-time process. A classical way to characterize a stationary process is by computing the first two moments (mean and autocorrelation function) in the time domain and the power spectral density in the frequency domain.

Another central concept is that of white noise, used either for modeling errors within observed data or to generate a random process using filtering techniques. The autocorrelation of a white noise is an impulse function at 0 while its PSD is constant for all frequencies.

At last, stationary processes such as AR models can be generated by filtering a white noise. The interference formula is a key tool to characterize the generated process in the time and frequency domains.

ESTIMATION BASED ON RANDOM MODELS

The previous chapter introduced mathematical tools to model time signals using random processes. In practice, we have one or several realizations of a random process, from which we wish to estimate statistical parameters such as the mean, autocorrelation, PSD or other model parameters. The purpose of this chapter is to define estimators, using concepts common with the statistics course.

As an example, consider speech processing, where one wants to identify the nature of elementary sounds (called phonemes) for automatic recognition purpose. Some phonemes such as /s/ or /f/ can be modeled by autoregressive (AR) random processes. In practice, we have a recording of an audio signal corresponding to a single realization of the AR process. The objective is to estimate the AR model parameters from this realization in order to identify the phoneme.

Throughout this chapter, **we will only consider discrete-time, wide-sense stationary processes**. Let us specify our main notations:

- A random process is denoted with a capital letter X_n , $n \in \mathbb{Z}$. The index n is related to the variable of evolution (e.g., time).
- The deterministic signal corresponding to a realization of this process is denoted with a lowercase letter: $x[n]$, $n \in \mathbb{Z}$.
- The estimator (or the statistic) is denoted with a hat: $\hat{\mu}$. Note that this is a random variable (or a random process).

Remark. $x[k]$ is referred to as the k -th sample (in the sense of time sampling, see Chapter 4) of signal x , that is, the value of signal x observed at time k . Similarly, X_k will be referred to as the k -th sample of the random process X_n ($n \in \mathbb{Z}$). X_k is the random variable X_n for $n = k$. Note that in statistics, a “sample” of a random variable Y refers to a realization $Y(\omega)$ from some ω . The meaning of sampling is thus completely different. To avoid any confusion, sampling will always refer to time sampling in this chapter.

6.1 ESTIMATION OF THE MEAN OF A RANDOM PROCESS

One of the difficulties of estimation in signal processing is the fact that usually, we have only one realization of the random process. In other words, it is only possible to observe one realization. If we draw a parallel with a random variable, it would be like trying to estimate its mean or variance from a single data!

This is why we will deal with ergodic signals, for which the statistical mean (over various realizations) can be replaced by the mean over the time samples for a single realization. Obviously, this approach only makes sense for stationary signals, for which the quantities to be estimated do not depend on time.

Definition 6.1 (Estimator of the mean). *The time estimator of the mean from N times is defined by:*

$$\hat{\mu}_X^N = \frac{1}{N} \sum_{n=0}^{N-1} X_n.$$

Property. *This estimator is unbiased.*

Proof. The bias is defined by $\text{bias}(\hat{\mu}_X^N) = \mathbb{E}(\hat{\mu}_X^N) - \mu_X$. It is easy to see that $\mathbb{E}(\hat{\mu}_X^N) = \frac{1}{N} \sum_{n=0}^{N-1} \mathbb{E}(X_n) = \mu_X$. So, $\text{bias}(\hat{\mu}_X^N) = 0$. $\square$

Definition 6.2 (First-order ergodicity). *The stationary process X_n is first-order ergodic if the time estimator of the mean converges in L^2 to the mean of the process:*

$$\hat{\mu}_X^N = \frac{1}{N} \sum_{n=0}^{N-1} X_n \xrightarrow[N \rightarrow \infty]{L^2} \mu_X$$

or equivalently (since the mean estimator is unbiased), if

$$\text{Var}(\hat{\mu}_X^N) \xrightarrow[N \rightarrow \infty]{} 0$$

In practice, we compute a realization of the random variable $\hat{\mu}_X^N$ from a single realization $x[n] = X_n(\omega)$ of a stationary and ergodic process X_n . This realization reads: $\hat{\mu}_X^N(\omega) = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$. The ergodic property allows us to replace the average over random realizations μ_X by a time average $\hat{\mu}_X^N(\omega)$ of the samples $x[n]$.

Example 6.1.

- A white noise is first-order ergodic. Indeed,

$$\begin{aligned}
 \text{Var}(\hat{\mu}_X^N) &= \frac{1}{N^2} \sum_{n'=0}^{N-1} \sum_{n=0}^{N-1} \mathbb{E}(X_n X_{n'}^*) \\
 &= \frac{1}{N^2} \sum_{n=0}^{N-1} \mathbb{E}(|X_n|^2) \quad \text{since } \mathbb{E}(X_n X_{n'}^*) = 0, \forall n \neq n' \\
 &= \frac{\gamma_X[0]}{N} \xrightarrow{N \rightarrow \infty} 0.
 \end{aligned}$$

- The constant random value process defined by $\forall n, X_n = A$ with $A \sim \mathcal{N}(0, \sigma^2)$ is not ergodic. Indeed,

$$\begin{aligned}
 \text{Var}(\hat{\mu}_X^N) &= \frac{1}{N^2} \sum_{n'=0}^{N-1} \sum_{n=0}^{N-1} \mathbb{E}(X_n X_{n'}^*) \\
 &= \frac{1}{N^2} \sum_{n'=0}^{N-1} \sum_{n=0}^{N-1} \mathbb{E}(A^2) \\
 &= \sigma^2
 \end{aligned}$$

does not converge to 0.

6.2 ESTIMATION OF THE AUTOCORRELATION FUNCTION

Definition 6.3 (Unbiased autocorrelation estimator). *The unbiased time estimator of the autocorrelation function is defined by:*

$$\begin{aligned}
 \hat{\gamma}_X^{N, \text{unb}}[k] &= \frac{1}{N-k} \sum_{\ell=0}^{N-k-1} X_{\ell+k} X_{\ell}^* \quad \text{for } 0 \leq k < N, \\
 \hat{\gamma}_X^{N, \text{unb}}[k] &= \left(\hat{\gamma}_X^{N, \text{unb}}[-k] \right)^* \quad \text{for } -N < k < 0.
 \end{aligned}$$

Property. *This estimator is unbiased for all k such that $|k| < N$.*

Proof. We check that $\mathbb{E}(\hat{\gamma}_X^{N, \text{unb}}[k]) = \gamma_X[k]$ for $0 \leq k < N$. This result is further extended to $k < 0$ by exploiting that functions γ_X and $\gamma_X^{N, \text{unb}}$ are Hermitian symmetric. $\square$

$\hat{\gamma}_X^{N, \text{unb}}[k]$ is an estimator of the autocorrelation function only when $|k| < N$. It is rarely used because its variance is huge when k is close to N . In the limit case $|k| = N - 1$, the average operation within the definition of $\hat{\gamma}_X^{N, \text{unb}}[k]$ only includes one term (which yields a huge variance). On the contrary, for $k = 0$, the average operation includes N terms. Therefore, the variance of $\hat{\gamma}_X^{N, \text{unb}}[0]$ is reasonably small when N is growing, see the statistics reminders in Section 1.5.1.

Definition 6.4 (Biased autocorrelation estimator). *The biased time estimator of the autocorrelation function is defined by:*

$$\begin{aligned}\hat{\gamma}_X^N[k] &= \frac{1}{N} \sum_{\ell=0}^{N-k-1} X_{\ell+k} X_{\ell}^* \quad \text{for } 0 \leq k < N, \\ \hat{\gamma}_X^N[k] &= \left(\hat{\gamma}_X^N[-k] \right)^* \quad \text{for } -N < k < 0.\end{aligned}$$

The difference with the unbiased estimator is that the normalization factor $\frac{1}{N-|k|}$ is replaced by $\frac{1}{N}$. This estimator is biased since:

$$\mathbb{E}(\hat{\gamma}_X^N[k]) = \mathbb{E} \left(\frac{N-|k|}{N} \hat{\gamma}_X^{N,unb}[k] \right) = \frac{N-|k|}{N} \gamma_X[k]. \quad (146)$$

The bias can be written as:

$$\begin{aligned}\text{bias}(\hat{\gamma}_X^N[k]) &= \mathbb{E}(\hat{\gamma}_X^N[k]) - \gamma_X[k] \\ &= \frac{N-|k|}{N} \gamma_X[k] - \gamma_X[k] \\ &= -\frac{|k|}{N} \gamma_X[k].\end{aligned}$$

The bias equals 0 for $k = 0$ and tends to increase for $|k|$ close to N . The biased estimator has a smaller variance as compared with the unbiased estimator. Indeed,

$$\text{Var}(\hat{\gamma}_X^N[k]) = \left(\frac{N-|k|}{N} \right)^2 \text{Var}(\gamma_X^{unb}[k]) \quad (147)$$

is close to 0 when k is nonzero and N is sufficiently large.

In statistics, the mean square error (MSE) is commonly used to compare estimators. It reads as the sum of the squared bias and the variance. The MSE of the unbiased and biased estimators of the autocorrelation function can be written as:

$$\text{MSE}(\hat{\gamma}_X^{N,unb}[k]) = 0 + \text{Var}(\hat{\gamma}_X^{N,unb}[k]) \quad (148)$$

and

$$\text{MSE}(\hat{\gamma}_X^N[k]) = \left(\text{bias}(\hat{\gamma}_X^N[k]) \right)^2 + \text{Var}(\hat{\gamma}_X^N[k]) \quad (149)$$

$$= \frac{k^2}{N^2} \left(\gamma_X[k] \right)^2 + \text{Var}(\hat{\gamma}_X^N[k]). \quad (150)$$

The biased estimator is often chosen because its MSE is lower than that of the unbiased estimator. Indeed, for most random processes modeling physical phenomena, $|\gamma_X[k]|$ is rapidly decreasing when $|k|$ is growing (intuitively, this corresponds to the fact that two samples taken at close times are more correlated than two samples taken far apart). Thus, the squared bias within (150) is of small magnitude:

- when $|k|$ is low, $\frac{k^2}{N^2}$ is close to 0;
- when $|k|$ is moderate to large, $(\gamma_X[k])^2$ is small.

Since the variance of the biased estimator is much smaller than that of the unbiased estimator, the MSE of the former is the smallest.

Going further...

Let $X^N = (X_0, X_1, \dots, X_{N-1})$ be the random process restricted to N times. It is a random vector including the N random variables used to define the above estimators of the mean and autocorrelation function.

Property. *The biased autocorrelation estimator can be written as a convolution product:*

$$\hat{\gamma}_X^N = \frac{1}{N} X^N * \bar{X}^N$$

where $\forall k \in \{0, \dots, N\}$, $\bar{X}_{-k}^N := (X_k^N)^*$.

Proof. For $0 \leq k < N$,

$$\begin{aligned} \hat{\gamma}_X^N[k] &= \frac{1}{N} \sum_{\ell=0}^{N-k-1} X_{\ell+k} X_{\ell}^* \\ &= \frac{1}{N} \sum_{m=k}^{N-1} X_m X_{m-k}^* \quad \text{with } m = \ell + k \\ &= \frac{1}{N} \sum_{m=k}^{N-1} X_m^N \bar{X}_{k-m}^N \\ &= \frac{1}{N} (X^N * \bar{X}^N)[k]. \end{aligned}$$

Moreover, the function $\hat{\gamma}_X^N[k]$ is Hermitian symmetric. Exploiting the definition of $\bar{X}^N[k]$ for $k < 0$, one can easily check that $X^N * \bar{X}^N$ is also Hermitian symmetric. Therefore, if $-N < k < 0$, we have

$$\begin{aligned} \hat{\gamma}_X^N[k] &= \left(\hat{\gamma}_X^N[-k] \right)^* \\ &= \frac{1}{N} \left(X^N * \bar{X}^N[-k] \right)^* \\ &= \frac{1}{N} (X^N * \bar{X}^N)[k]. \end{aligned}$$

□

Definition 6.5 (Second-order ergodicity). *The stationary process X_n is second-order ergodic if the unbiased estimator of its autocorrelation function converges in L^2 towards the autocorrelation function:*

$$\forall k \in \mathbb{N}, \quad \hat{\gamma}_X^{N,nb}[k] \xrightarrow[N \rightarrow \infty]{L^2} \gamma_X[k]$$

or equivalently, if

$$\forall k \in \mathbb{N}, \quad \text{Var}(\gamma_X^{N,nb}[k]) \xrightarrow[N \rightarrow \infty]{} 0.$$

Second-order ergodicity is an essential property for estimating the auto-correlation of a stationary random process using time estimators.

6.3 ESTIMATION OF THE PSD: PERIODOGRAM

Estimation of the power spectral density of a random process from a single realization is a classical problem. The goal is to provide a characterization of the process in the frequency domain.

Example 6.2. *In climatology, scientists aim to find the periodicity of climate variation due to the earth oscillations. A more specific objective is to distinguish the climate variations related to human activity compared to other factors.*

One of the most widely used measurements is the concentration of deuterium in ice cores dating back several thousand years. The signal corresponding to the deuterium concentration as a function of time is modeled by a random process. Obviously, only one realization of this process is available. By estimating the PSD of the process, one can find the main periodic features which correspond to the peaks of the PSD signal.

6.3.1 Definition

The classical estimator of the power spectral density is called periodogram.

Definition 6.6 (Periodogram). *The periodogram is an estimator of the PSD defined as the Fourier transform of the biased autocorrelation estimator for N time samples:*

$$\hat{\Gamma}_X^N(\nu) = \sum_{k=-N+1}^{N-1} \hat{\gamma}_X^N[k] e^{-i2\pi k\nu}$$

or equivalently:

$$\hat{\Gamma}_X^N(\nu) = \frac{1}{N} \left| \sum_{k=0}^{N-1} X_k e^{-i2\pi k\nu} \right|^2.$$

Proof. Let us denote by $X^N = (X_0, X_1, \dots, X_{N-1})$ the random process X_n restricted to N time samples. The equivalence between both definitions is derived by using $\hat{\gamma}_X^N = \frac{1}{N} X^N * \bar{X}^N$:

$$\begin{aligned} \hat{\Gamma}_X^N(\nu) &= \mathcal{F} \gamma_X^N(\nu) \\ &= \frac{1}{N} \mathcal{F} \{X^N * \bar{X}^N\}(\nu) \\ &= \frac{1}{N} \mathcal{F} X^N(\nu) \mathcal{F} \bar{X}^N(\nu) \\ &= \frac{1}{N} \left(\mathcal{F} X^N(\nu) \right) \left(\mathcal{F} X^N(\nu) \right)^* \\ &= \frac{1}{N} \left| \mathcal{F} X^N(\nu) \right|^2. \end{aligned}$$

□

Recall the definition of power spectral density given in section 5.2.3:

$$\forall \nu \in \mathbb{R}, \quad \Gamma_X^N(\nu) = \lim_{N \rightarrow \infty} \mathbb{E} \left(\frac{1}{2N+1} \left| \sum_{k=-N}^N X_k e^{-i2\pi k\nu} \right|^2 \right).$$

We can notice that the expectation and time limit operators have been removed in the definition of the periodogram estimator. These two simplifications have important consequences:

- **Reduced spectral resolution (bias):** Removing the limit operator is equivalent to restricting the random process to N time samples. The periodogram is therefore a biased estimator. The bias tends to 0 when $N \rightarrow \infty$.
- **High variance:** Removing the expectation operator is equivalent to computing an empirical average for a single realization. Therefore, the variance of the periodogram is very large. It can further be shown that the periodogram is not a consistent estimator, since its variance does not tend to 0 for $N \rightarrow \infty$.

From a practical viewpoint, given a realization $x[n]$ of the process observed for N time samples, one can compute the periodogram by first evaluating the DFT of $x[n]$ (using zero padding to refine the plot as seen in section 3.6) and then computing the squared modulus of the DFT.

Example 6.3 (Variance of the periodogram). *We consider a random process obtained by filtering a white noise. The theoretical PSD is plotted in bold in Fig. 32. Three realizations of the periodogram are also represented, obtained from three different realizations of the process. The respective signal lengths are $N = 40$ samples (left figure) and $N = 400$ samples (right figure). We notice that:*

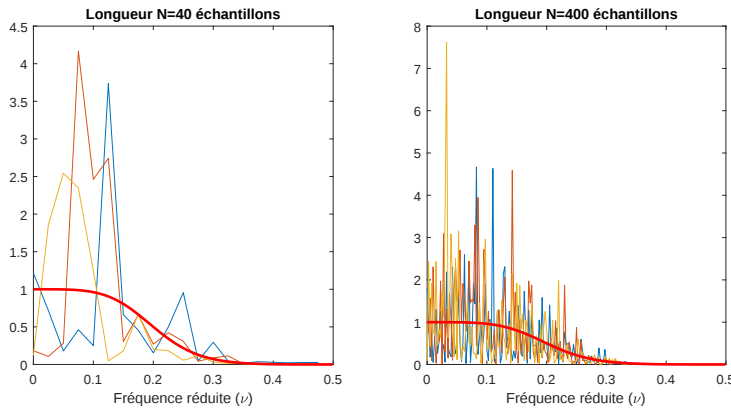


Figure 32: Illustration of the periodogram variance: comparison of the theoretical PSD (bold line) with realizations of the periodogram for $N = 40$ and $N = 400$ samples.

- The PSD estimate using the periodogram is very different from the theoretical PSD. This illustrates the large variance of the periodogram estimator.
- The periodogram plot strongly depends on the realization of the random process. Moreover, the variance does not decrease when the number of time samples increases, see the right figure. This illustrates that the estimator is not consistent.

6.3.2 Statistical properties of the periodogram

Bias and spectral resolution

Property. The periodogram is **biased**:

$$\mathbb{E}(\hat{\Gamma}_X^N(\nu)) = (\Gamma_X \circledast T_N)(\nu)$$

where $T_N(\nu) = \frac{1}{N} \frac{\sin^2(\pi N \nu)}{\sin^2(\pi \nu)}$ is the Fourier transform of the triangle window $t_N[k] = \frac{N-|k|}{N} \mathbb{1}_{|k| \leq N}$.

Proof. By definition, the periodogram can be rewritten as:

$$\hat{\Gamma}_X^N(\nu) = \sum_{k=-N+1}^{N-1} \hat{\gamma}_X^N[k] e^{-i2\pi k \nu}.$$

According to (146), we get:

$$\begin{aligned} \mathbb{E}(\hat{\Gamma}_X^N(\nu)) &= \sum_{k=-N+1}^{N-1} \mathbb{E}(\hat{\gamma}_X^N[k]) e^{-i2\pi k \nu} \\ &= \sum_{k=-N+1}^{N-1} \frac{N-|k|}{N} \gamma_X[k] e^{-i2\pi k \nu} \\ &= \sum_{k \in \mathbb{Z}} t_N[k] \gamma_X[k] e^{-i2\pi k \nu}. \end{aligned}$$

$\mathbb{E}(\hat{\Gamma}_X^N(\nu))$ reads as the DTFT of the entrywise product of functions t_N and γ_X . The latter DTFT is equal to the circular convolution of the DTFTs of t_N and γ_X , see section 3.3.2. $\square$

Summarizing,

- The bias is related to the fact that N time-samples are considered,
- Spectral resolution is related to the bias. Recall that the spectral resolution is defined as the width of the main lobe of T_N , see section 3.6.2. Here, it is equal to $\frac{2}{N}$ since T_N vanishes for $\nu = \pm \frac{1}{N}$,
- The bias is asymptotically zero (when $N \rightarrow \infty$). To improve spectral resolution, one needs to increase the number of time-samples N .

Going further...

Asymptotic law of the estimator

Property. Let X_n be a Gaussian process whose PSD is nonzero: $\forall \nu \in [-\frac{1}{2}, \frac{1}{2}]$, $\Gamma_X(\nu) \neq 0$. In the asymptotic regime ($N \rightarrow \infty$), the periodogram follows an exponential law:

$$p_{\hat{\Gamma}_X^\infty(\nu)}(x) = \frac{1}{\Gamma_X(\nu)} e^{-\frac{x}{\Gamma_X(\nu)}} \mathbf{1}_{x \geq 0}.$$

The asymptotic mean and variance of the periodogram read:

$$\mu_{\hat{\Gamma}_X^\infty(\nu)} = \Gamma_X(\nu), \quad \text{Var}(\hat{\Gamma}_X^\infty(\nu)) = \Gamma_X(\nu)^2.$$

Because the standard deviation is equal to the value we are trying to estimate, we conclude that the periodogram estimator has poor accuracy.

6.3.3 Variants of the periodogram method

As previously mentioned, the variance of the periodogram is very high. Several extensions were proposed to reduce the variance by averaging techniques, among which Bartlett's periodogram. As shown in Fig. 33, Bartlett's method consists of dividing the signal into K consecutive windows, and averaging the periodograms related to each of these windows.

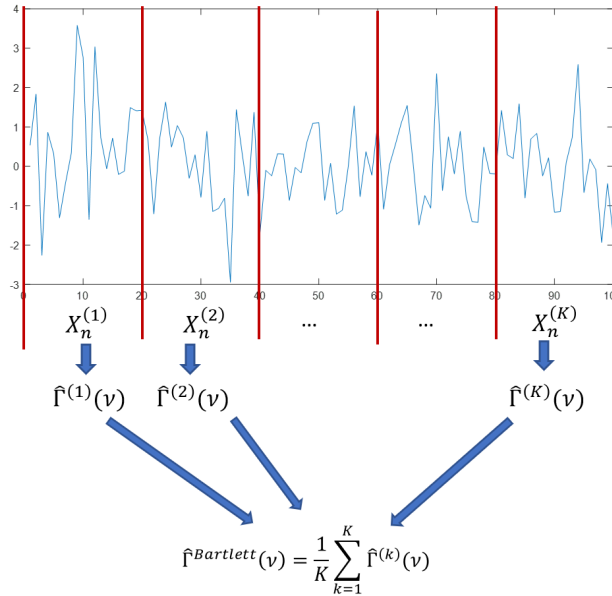


Figure 33: Principle of Bartlett's periodogram.

Definition 6.7. Let $X_0, X_1, \dots, X_{N-1}$ be N consecutive time-samples of process X_n , with $N = KL$ ^(a). The Bartlett periodogram on K windows of length L is defined by:

$$\hat{\Gamma}_X^{Bart}(\nu) = \frac{1}{K} \sum_{k=0}^{K-1} \Gamma^{(k)}(\nu) \quad \text{where} \quad \Gamma^{(k)}(\nu) = \frac{1}{L} \left| \sum_{\ell=0}^{L-1} X_{\ell+kL} e^{-i2\pi\nu\ell} \right|^2.$$

^a without loss of generality, the initial time corresponds to $n = 0$.

Let us express the bias of Bartlett's periodogram. The random process X_n being stationary, $\mathbb{E}(\hat{\Gamma}^{(k)}(\nu))$ does not depend on k . So $\forall k, \forall \nu, \mathbb{E}(\hat{\Gamma}^{(k)}(\nu)) = \mathbb{E}(\hat{\Gamma}^{(0)}(\nu))$. The bias reads:

$$\begin{aligned} \text{bias}(\hat{\Gamma}_X^{Bart}(\nu)) &= \mathbb{E}(\hat{\Gamma}_X^{Bart}(\nu)) - \Gamma_X(\nu) \\ &= \mathbb{E}(\hat{\Gamma}^{(0)}(\nu)) - \Gamma_X(\nu) \\ &= \text{bias}(\hat{\Gamma}_X^L(\nu)) \end{aligned}$$

and identifies with the bias of the standard periodogram calculated for L samples. According to section 6.3.2, **the spectral resolution is equal to $\frac{2}{L}$. It is less accurate than that of the standard periodogram calculated for N time samples** (equal to $\frac{2}{N}$). The latter is K times smaller: $\frac{2}{N} = \frac{1}{K} \frac{2}{L}$. This loss of accuracy is related to the fact that Bartlett's method is based on the calculation of periodograms for windows of reduced size, that is, $L = \frac{N}{K}$.

On the other hand, **the periodogram's variance can be strongly reduced using Bartlett's method.** To set this idea on a simple example, consider a white noise in the strict sense. For ν and $j \neq k$, the random variables $\hat{\Gamma}^{(j)}(\nu)$ and $\hat{\Gamma}^{(k)}(\nu)$ are independent and have equal variances. Thus,

$$\begin{aligned} \text{Var}(\hat{\Gamma}_X^{Bart}(\nu)) &= \frac{1}{K^2} \sum_{k=0}^{K-1} \text{Var}(\hat{\Gamma}^{(k)}(\nu)) \\ &= \frac{1}{K^2} \sum_{k=0}^{K-1} \text{Var}(\hat{\Gamma}^{(0)}(\nu)) \\ &= \frac{1}{K} \text{Var}(\hat{\Gamma}^{(0)}(\nu)) \\ &= \frac{\text{Var}(\hat{\Gamma}_X^L(\nu))}{K}. \end{aligned}$$

According to section 6.3.2, $\text{Var}(\hat{\Gamma}_X^L(\nu)) \rightarrow \Gamma_X(\nu)^2$ when $L \rightarrow \infty$, thus $\text{Var}(\hat{\Gamma}_X^{Bart}(\nu)) \rightarrow \frac{\Gamma_X(\nu)^2}{K}$. In the asymptotic regime (K is fixed and $L \rightarrow \infty$), the variance of Bartlett's periodogram is K times smaller than that of the standard periodogram, calculated for N samples.

In practice, choosing the number of windows K implies a trade-off between the bias and variance, as illustrated hereafter.

Example 6.4. Consider three realizations of a random process whose PSD is displayed in bold in Fig. 34, for 10 000 time samples.

- The top figure shows three realizations of the standard periodogram. The variance is too large to correctly retrieve the PSD.
- The center figure displays realizations of Bartlett's periodogram. Averaging is carried out over 50 windows of length 200. The variance is moderate while the bias remains very small: the curve of $\hat{\Gamma}_X^{Bart}(\nu)$, shown in dotted line, almost identifies with that of $\Gamma_X(\nu)$. Bartlett's estimator appears to be a good trade-off.
- The bottom figure displays realizations of Bartlett's periodogram. Averaging is carried out over 500 windows of length 20. The variance is very small whereas the bias is large.

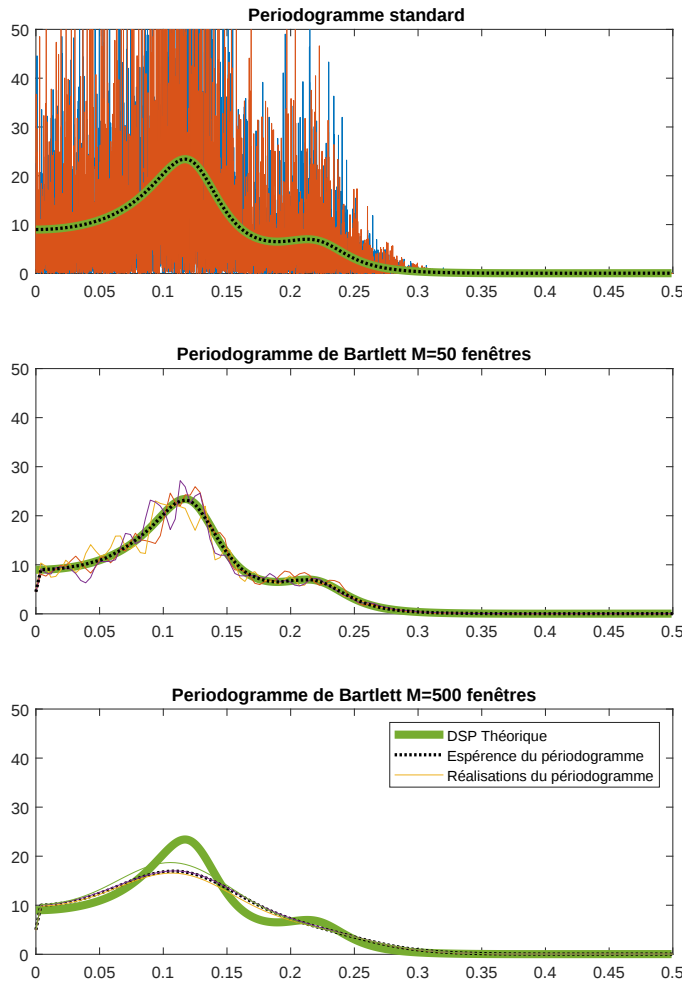


Figure 34: Bartlett's periodogram: increasing the number of windows yields a reduction of the variance and an increase of bias.

6.4 PARAMETRIC ESTIMATION OF THE PSD

6.4.1 Principle

The periodogram, studied in section 6.3, is a non-parametric estimator having a very large variance. In the parametric approach, the stationary process X_n is described by a model with few parameters (e.g., the autoregressive model). By estimating these parameters from a realization $x[n]$ of the random process, one can then retrieve an estimate of the PSD. This approach requires to make an assumption on the parametric model describing the random process. On the other hand, it yields estimators having a lower variance.

Parametric estimation can be applied, e.g., to speech processing using AR models or to vibration analysis problems, which lead to estimate the parameters of noisy sinusoids.

6.4.2 Case of a signal modeled by an AR process

According to Theorem 5.3, the AR model of order p can be written using the recurrence equation:

$$X_n = \sum_{k=1}^p a_k X_{n-k} + W_n \quad (151)$$

where W_n is a generative white noise of variance σ^2 . The model involves $p + 1$ parameters: $a_1, \dots, a_p, \sigma^2$.

Estimation of AR model parameters from experimental data

The most classical approach aims to estimate the autocorrelation function of X_n . The Yule-Walker equations (145) provide a linear relation between the model parameters and the autocorrelation coefficients:

$$\forall k \geq 0, \quad \gamma_X[k] = \sum_{i=1}^p a_i \gamma_X[k-i] + \sigma^2 \delta[k].$$

Let us gather these equations for $k = 0, \dots, p$. Since $\gamma_X[k]$ is an even function, we get the following linear system:

$$\begin{pmatrix} \gamma_X[1] & \gamma_X[2] & \cdots & \gamma_X[p] & 1 \\ \gamma_X[0] & \gamma_X[1] & \cdots & \gamma_X[p-1] & 0 \\ \gamma_X[1] & \gamma_X[0] & \cdots & \gamma_X[p-2] & 0 \\ \vdots & \vdots & \ddots & \vdots & \\ \gamma_X[p-1] & \gamma_X[p-2] & \cdots & \gamma_X[0] & 0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \\ \sigma^2 \end{pmatrix} = \begin{pmatrix} \gamma_X[0] \\ \gamma_X[1] \\ \gamma_X[2] \\ \vdots \\ \gamma_X[p] \end{pmatrix}$$

In practice, $\gamma_X[k]$ is replaced with the biased estimator $\hat{\gamma}_X^N[k]$, see definition 6.4. Solving the linear system yields estimates $\hat{a}_k$ and $\hat{\sigma}^2$ of the model parameters.

Estimation of the PSD

The PSD of an AR process has a closed form expression, see (144). The parametric estimator of the PSD thus reads:

$$\hat{\Gamma}_X^{\text{AR}}(\nu) = \frac{\hat{\sigma}^2}{\left| 1 - \sum_{k=1}^p \hat{a}_k e^{-2\pi i k \nu} \right|^2}.$$

Figure 35 illustrates the reduction of variance obtained using parametric estimation. The PSD of the process is shown in thick line. The top figure displays realizations of the parametric estimator while the bottom one displays realizations of the standard periodogram.

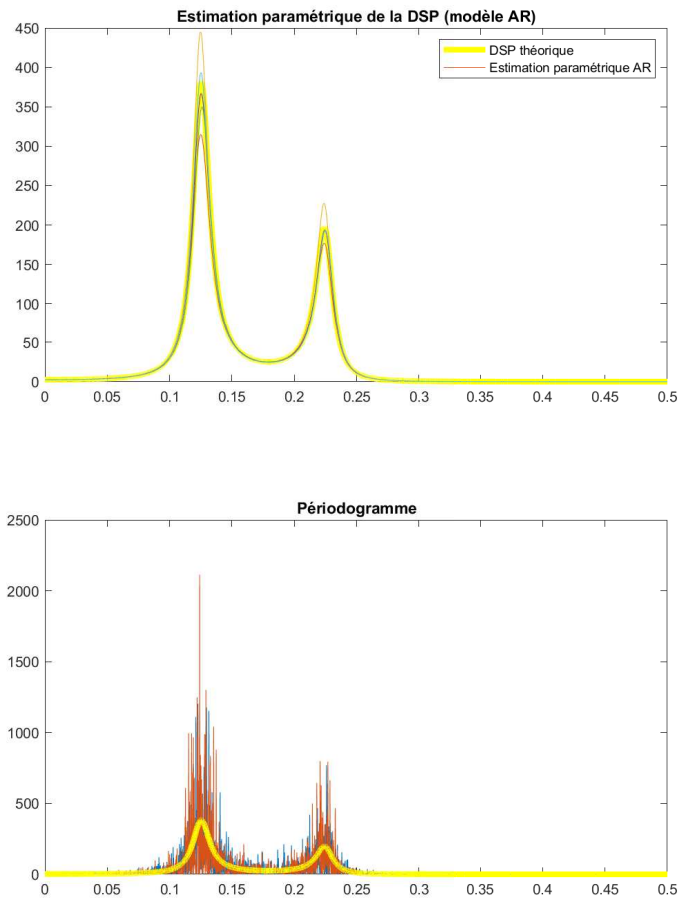


Figure 35: Comparison of the parametric PSD estimator with the standard periodogram.

Example 6.5. Application to speech processing

Some speech sounds can be described using an AR model. These are unvoiced phonemes, for which the vocal cords do not vibrate. The main unvoiced sounds include fricative consonants: /f/, /s/, etc. and whispered vowels. The AR model can be interpreted as follows:

- The turbulent airflow at the entrance of the vocal tract is seen as a white noise,
- The vocal tract resonances are related to the poles of the AR model.

Estimation of the AR model parameters related to specific phonemes allows one to identify the phoneme (speech recognition) or to carry out speech synthesis. Fig. 36 illustrates the speech processing tasks.

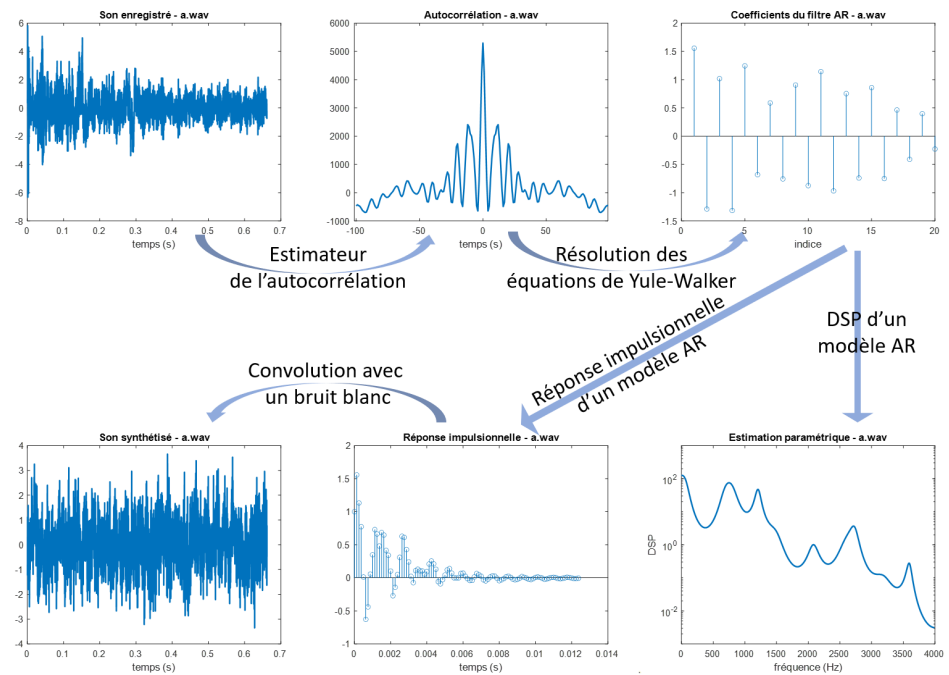


Figure 36: Speech processing and speech synthesis.

6.5 OPTIMAL LINEAR ESTIMATOR

In many applications, one intends to estimate a random process X at time n from a random process Y observed for P time samples. Obviously, estimating X_n is possible only when both processes X and Y are correlated. The autocorrelation function of Y and the cross-correlation between X and Y are assumed known. Moreover, **the random variables are assumed centered and real valued.**

Classical frameworks include:

- prediction: predict the value of X_n at time n from observations at P previous times: $X_{n-1}, \dots, X_{n-P}$,

- denoising: retrieve X_n at time n from noisy observations $Y_k = X_k + W_k$ for different time moments k ,
- deconvolution: retrieve X_n at time n from blurred measurements $Y_k = (h * X)_k$, where h is the impulse response of a low-pass filter modeling the blur.

Hereafter, we use the simplified notations $U = X_n \in \mathbb{R}$ for the random variable to be estimated and $\mathbf{Z} \in \mathbb{R}^P$ for the random vector gathering the P observations Y_k .

6.5.1 Definition and properties

Definition 6.8 (Optimal linear estimator). *Let $\mathbf{Z}$ be an $\mathbb{R}^P$ -valued centered random vector and U be a real valued centered random variable. A **linear estimator** of U from $\mathbf{Z}$ is a weighted sum*

$$\hat{U}(\mathbf{h}) = \mathbf{h}^t \mathbf{Z} = \sum_{p=1}^P h_p Z_p \quad (152)$$

where vector $\mathbf{h}$ gathers the weights h_p .

An **optimal linear estimator** reads $\hat{U}(\mathbf{h}_{\text{opt}}) = \mathbf{h}_{\text{opt}}^t \mathbf{Z}$ where $\mathbf{h}_{\text{opt}}$ minimizes the variance of the estimation error:

$$\mathbf{h}_{\text{opt}} = \underset{\mathbf{h}}{\operatorname{argmin}} \mathbb{E} \left[\left(U - \hat{U}(\mathbf{h}) \right)^2 \right]. \quad (153)$$

Theorem 6.1. *Let $\Sigma_{\mathbf{Z}} = \mathbb{E}(\mathbf{Z}\mathbf{Z}^t)$ denote the covariance matrix of $\mathbf{Z}$. Assuming $\Sigma_{\mathbf{Z}}$ is invertible, the optimal weight vector can be written as:*

$$\mathbf{h}_{\text{opt}} = \Sigma_{\mathbf{Z}}^{-1} \mathbb{E}(\mathbf{Z}U). \quad (154)$$

Proof. The variance of the estimation error reads:

$$\begin{aligned} \mathbb{E} \left[\left(U - \hat{U}(\mathbf{h}) \right)^2 \right] &= \mathbb{E} \left((U - \mathbf{h}^t \mathbf{Z})^2 \right) \\ &= \mathbb{E} (\mathbf{h}^t \mathbf{Z} \mathbf{Z}^t \mathbf{h}) - 2\mathbb{E} (\mathbf{h}^t \mathbf{Z} U) + \mathbb{E} (U^2) \\ &= \mathbf{h}^t \Sigma_{\mathbf{Z}} \mathbf{h} - 2\mathbf{h}^t \mathbb{E}(\mathbf{Z}U) + \mathbb{E}(U^2). \end{aligned}$$

Minimizing this variance leads to solve a quadratic optimization problem:

$$\mathbf{h}_{\text{opt}} = \underset{\mathbf{h}}{\operatorname{argmin}} \left\{ J(\mathbf{h}) := \mathbf{h}^t \Sigma_{\mathbf{Z}} \mathbf{h} - 2\mathbf{h}^t \mathbb{E}(\mathbf{Z}U) + \mathbb{E}(U^2) \right\}.$$

By assumption, matrix $\Sigma_{\mathbf{Z}}$ is invertible. Being a covariance matrix, it is symmetric positive definite. Therefore, function J is strictly convex, and the optimization problem admits a unique solution, for which the gradient of J vanishes:

$$\nabla_{\mathbf{h}} J(\mathbf{h} = \mathbf{h}_{\text{opt}}) = \mathbf{0} \iff 2\Sigma_{\mathbf{Z}} \mathbf{h}_{\text{opt}} - 2\mathbb{E}(\mathbf{Z}U) = \mathbf{0},$$

thus $\mathbf{h}_{\text{opt}} = \Sigma_{\mathbf{Z}}^{-1} \mathbb{E}(\mathbf{Z}U)$. $\square$

The solution $\hat{U}$ related to the weight vector $\mathbf{h}_{\text{opt}}$ can be interpreted as the orthogonal projection of $U \in L^2(\Omega, \mathcal{F}, \mathcal{P})$ on the subspace of $L^2(\Omega, \mathcal{F}, \mathcal{P})$ spanned by $Z_1, \dots, Z_P$. Corollary 6.1 states that the residual of the orthogonal projection, $U - \hat{U}$, is orthogonal to this subspace (according to the inner product of $L^2(\Omega, \mathcal{F}, \mathcal{P})$).

Corollary 6.1. *The estimation error is uncorrelated with the observations:*

$$\forall p \in \{1, \dots, P\}, \quad \mathbb{E}[(U - \hat{U}(\mathbf{h}_{\text{opt}})) Z_p] = 0.$$

Proof. By noticing that $\mathbf{h}_{\text{opt}}$ is constant (deterministic), we have:

$$\begin{aligned} \mathbb{E}(\mathbf{Z}(U - \hat{U}(\mathbf{h}_{\text{opt}}))) &= \mathbb{E}(\mathbf{Z}U - \mathbf{Z}\mathbf{h}_{\text{opt}}^t\mathbf{Z}) \\ &= \mathbb{E}(\mathbf{Z}U - \mathbf{Z}\mathbf{Z}^t\mathbf{h}_{\text{opt}}) \\ &= \mathbb{E}(\mathbf{Z}U) - \mathbb{E}(\mathbf{Z}\mathbf{Z}^t)\mathbf{h}_{\text{opt}} \\ &= \mathbb{E}(\mathbf{Z}U) - \mathbf{\Sigma}_\mathbf{Z}\mathbf{h}_{\text{opt}} \\ &= 0 \end{aligned}$$

according to (154). □

6.5.2 Prediction of a random process

The optimal linear estimation technique is now applied to the prediction of a random process. In various fields (e.g., prediction of the evolution of financial stocks), one intends to predict the value of a process X_n at time n ($U = X_n$) from the past P values, denoted $\mathbf{Z} = (X_{n-1}, X_{n-2}, \dots, X_{n-P})$. The autocorrelation function γ_X is assumed known.

Theorem 6.2. *The optimal linear predictor of X_n from $\mathbf{Z} = (X_{n-1}, X_{n-2}, \dots, X_{n-P})^t$ reads:*

$$\hat{X}_n = \mathbf{h}_{\text{opt}}^t \mathbf{Z} = \sum_{p=1}^P [h_{\text{opt}}]_p X_{n-p} \quad (155)$$

where

$$\mathbf{h}_{\text{opt}} = \begin{pmatrix} \gamma_X[0] & \gamma_X[1] & \cdots & \gamma_X[P-1] \\ \gamma_X[1] & \gamma_X[0] & \cdots & \gamma_X[P-2] \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_X[P-1] & \gamma_X[P-2] & \cdots & \gamma_X[0] \end{pmatrix}^{-1} \begin{pmatrix} \gamma_X[1] \\ \gamma_X[2] \\ \vdots \\ \gamma_X[P] \end{pmatrix}.$$

Proof. The expression of $\mathbf{h}_{\text{opt}}$ follows from Theorem 6.1 and from $\gamma_X[k] = \mathbb{E}(X_{n+k}X_n) = \gamma_X[-k]$, the process X being real valued. Note that since X_n is stationary, $\mathbb{E}(\mathbf{Z}X_n)$ and $\mathbf{\Sigma}_\mathbf{Z} = \mathbb{E}[\mathbf{Z}\mathbf{Z}^t]$ do not depend on n . □

Remark. (155) is a convolution relation. The predictor can thus be seen as a linear filter of impulse response $\mathbf{h}_{\text{opt}}$.

Example 6.6. On Fig. 37, the upper plot shows $\gamma_X[k]$. We can notice that the samples spaced about 40 points apart are strongly correlated. The bottom plot displays the coefficients $\mathbf{h}_{\text{opt}}$ for $P = 50$.

Fig. 38 shows a realization $x[n]$ of the process X_n (bold curve). The predictions made with $P = 5$ samples (top) and $P = 50$ samples (bottom) are drawn in thin lines. The prediction is improved using 50 samples, since the high correlation between distant samples is taken into account.

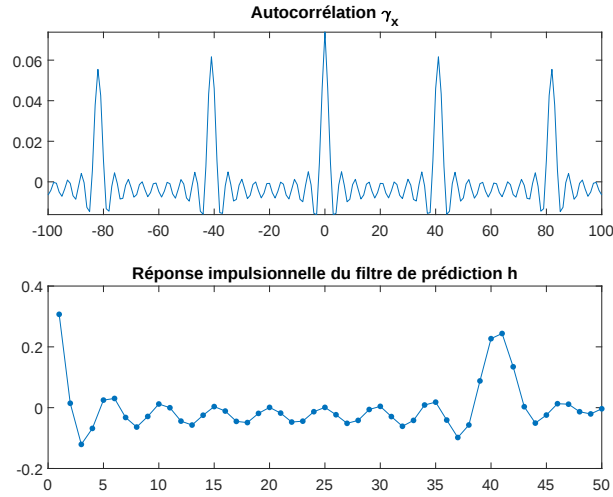


Figure 37: Top: autocorrelation of the process (γ_X). Bottom: impulse response of the prediction filter ($\mathbf{h}$).

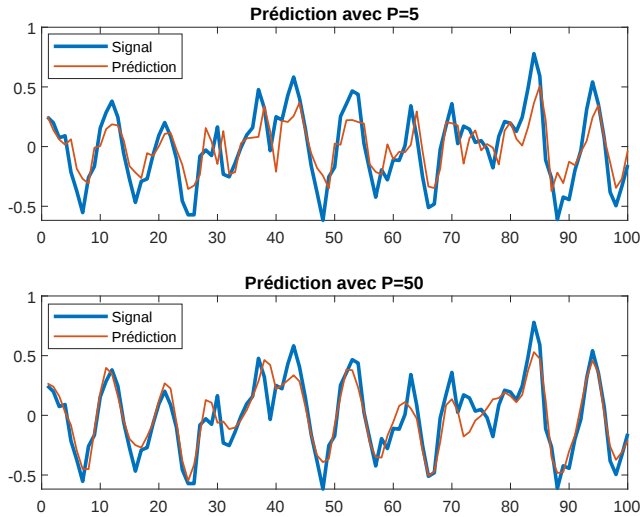


Figure 38: Prediction of the process using $P = 5$ (top) and $P = 50$ (bottom) samples.

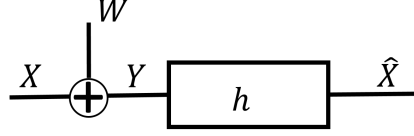


Figure 39: Denoising model: search for the optimal filter h to apply to noisy data Y_n .

6.5.3 Wiener filter for denoising

In this section, optimal linear estimation is applied to signal denoising. The noisy signal is modeled by the random process Y_n , equal to the sum of an (unknown) noiseless process X_n and an unknown noise process W_n :

$$Y_n = X_n + W_n$$

The objective is to reconstruct X_n by filtering the signal Y_n . The model and filtering operation are represented in Fig. 39. We make the following assumptions:

- W and X are uncorrelated,
- the autocorrelation functions γ_X and γ_W are known.

Case of a finite impulse response filter

We first address the reconstruction of X_n from P past observations. We therefore seek to apply to process Y a filter h of duration P :

$$\hat{X}_n = (h * Y)_n = \sum_{p=0}^{P-1} h_p Y_{n-p}.$$

This equation can be rewritten $\hat{X}_n = \mathbf{h}^t \mathbf{Z}$ with $\mathbf{Z} = (Y_n, Y_{n-1}, \dots, Y_{n-P+1})^t$. According to Theorem 6.1, the optimal filter is given by $\mathbf{h}_{\text{opt}} = \mathbf{\Sigma}_Z^{-1} \mathbb{E}(\mathbf{Z}X)$, with

$$\mathbf{\Sigma}_Z = \begin{pmatrix} \gamma_Y[0] & \gamma_Y[1] & \cdots & \cdots & \gamma_Y[P-1] \\ \gamma_Y[1] & \gamma_Y[0] & \cdots & \cdots & \gamma_Y[P-2] \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \gamma_Y[P-1] & \gamma_Y[P-2] & \cdots & \cdots & \gamma_Y[0] \end{pmatrix}$$

and

$$\boldsymbol{\gamma}_{\mathbf{Z}X} = \begin{pmatrix} \gamma_{YX}[0] \\ \gamma_{YX}[-1] \\ \vdots \\ \gamma_{YX}[1-P] \end{pmatrix}.$$

Since W and X are uncorrelated, the entries of the above matrices can be written using functions γ_X and γ_W : for $k = 0, \dots, P-1$, $\gamma_Y[k] = \gamma_X[k] + \gamma_W[k]$ and $\gamma_{YX}[-k] = \gamma_X[-k] = (\gamma_X[k])^*$. Since X_n is real valued, we then have $\gamma_{YX}[-k] = \gamma_X[k]$.

Case of an infinite impulse response filter

We now generalize the above result to infinite impulse response filters. In other words, we aim to predict X_n from an infinite number of time observations, Y_{n-p} for $p \in \mathbb{Z}$. Theorem 6.3 states that the frequency response of the optimal filter can be directly deduced from the power spectral densities of X and W .

Theorem 6.3 (Wiener filter). *The optimal linear filter for denoising, called **Wiener filter**, reads as the convolution*

$$\hat{X}_n = \sum_{p \in \mathbb{Z}} [\mathbf{h}_{\text{opt}}]_p Y_{n-p}$$

where the transfer function is:

$$H_{\text{opt}}(\nu) = \frac{\Gamma_X(\nu)}{\Gamma_X(\nu) + \Gamma_W(\nu)} = \frac{1}{1 + \Gamma_W(\nu)/\Gamma_X(\nu)}.$$

Sketch of proof. The case of an infinite impulse response does not correspond to the framework of theorem 6.1, where the number of observations Y_{n-p} is finite. This said, some arguments remain valid in infinite dimension, such as the fact that $\hat{X}_n$ is the orthogonal projection of X_n on the (closure of the) subspace of $L^2(\Omega, \mathcal{F}, \mathcal{P})$ spanned by variables Y_{n-p} . Hence, the estimation error and the observations are uncorrelated (corollary 6.1). So, we have:

$$\begin{aligned} \mathbb{E}((X_n - \hat{X}_n)Y_{n-p}) &= 0 \quad \forall p, n \in \mathbb{Z} \\ \mathbb{E}(X_n Y_{n-p}) - \mathbb{E}(\hat{X}_n Y_{n-p}) &= 0 \\ \mathbb{E}(X_n(X_{n-p} + W_{n-p})) - \mathbb{E}(\hat{X}_n Y_{n-p}) &= 0 \\ \mathbb{E}(X_n X_{n-p}) + \mathbb{E}(X_n W_{n-p}) - \mathbb{E}(\hat{X}_n Y_{n-p}) &= 0 \\ \gamma_X[p] + 0 - \gamma_{\hat{X}Y}[p] &= 0, \end{aligned}$$

thus $\gamma_X = \gamma_{\hat{X}Y}$. According to the interference formula, $\gamma_{\hat{X}Y} = h_{\text{opt}} * \gamma_Y$. Moreover, $\gamma_Y = \gamma_X + \gamma_W$ since $Y_n = X_n + W_n$ and W and X are uncorrelated. Finally,

$$\gamma_X = h_{\text{opt}} * \gamma_Y = h_{\text{opt}} * (\gamma_X + \gamma_W).$$

Applying the DTFT to the previous expression, we get:

$$\Gamma_X(\nu) = H_{\text{opt}}(\nu)(\Gamma_X(\nu) + \Gamma_W(\nu)),$$

hence the expression of $H_{\text{opt}}(\nu)$. $\square$

The expression of the frequency response can be interpreted as follows: the filter attenuates the signal within the frequency bands where the PSD of the noise is greater than that of the random process X . One of the main limitations of this approach is that the knowledge of power spectral densities of the target processes is required.

Example 6.7. Consider a noisy process $Y_n = X_n + W_n$ and assume that the PSD $\Gamma_X(\nu)$ and $\Gamma_W(\nu)$ are known, see Fig. 40(a). The frequency response $H_{\text{opt}}(\nu)$ of the Wiener filter is displayed in dotted line. It adapts to the signal-to-noise ratio in each frequency band: when the signal power is important compared to the noise, the gain of the filter is high, thus the frequency content of Y_n is preserved. In the opposite case, the gain of the filter is low and the noise is attenuated. Fig. 40(b) displays a realization of the three processes in the time domain. In the top figure, the noisy process Y_n is shown together with the reference signal (that is, the noiseless process X_n used to generate simulated data Y_n); in the bottom figure, the denoised process $\hat{X}_n$ is compared with X_n .

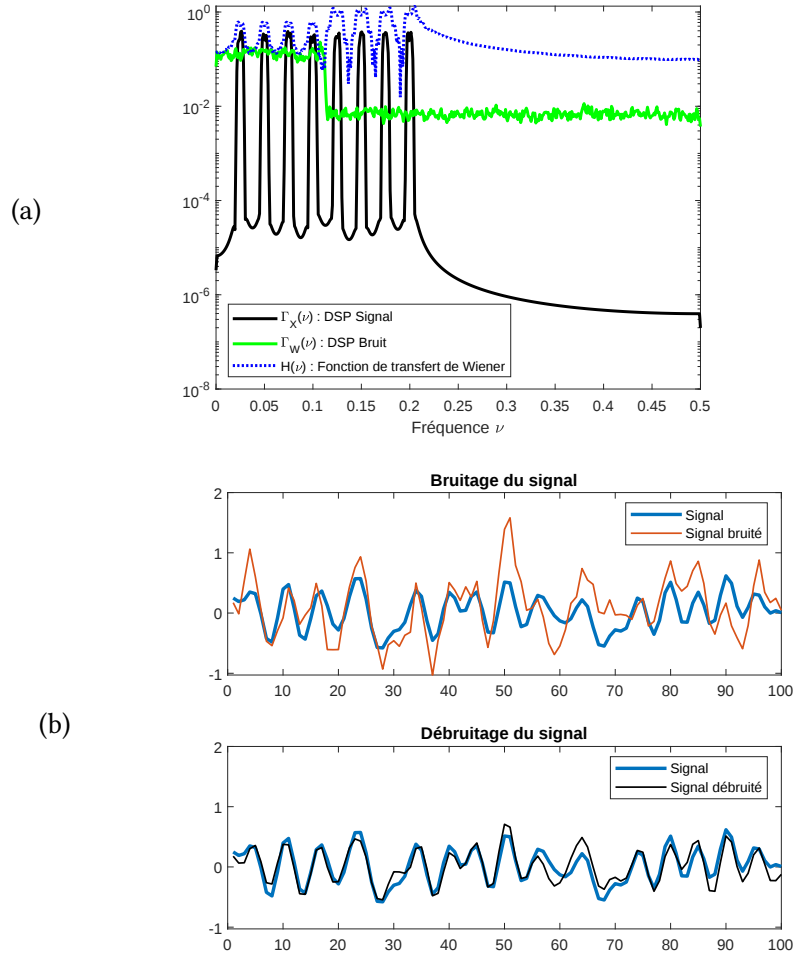


Figure 40: Denoising using the Wiener filter. (a) PSD of the signal and noise processes (Γ_X and Γ_W) and frequency response of the Wiener filter (H_{opt}). (b) Time representation of Y_n (noisy process), X_n (unknown noiseless process) and $\hat{X}_n$ (denoised result).

6.6 CONCLUSION

In this chapter, we first derived estimators of the mean of a process, the biased and unbiased estimators of its autocorrelation function, and the estimators of its DSP, namely the standard periodogram and the Bartlett periodogram. The standard periodogram method has significant limitations in terms of bias (related to spectral resolution) and variance, which make the estimator imprecise. Thanks to its windowing effect, Bartlett's method makes it possible to significantly reduce the variance of the estimator. DSP parametric estimation techniques are very effective when the process of interest can be modeled using a few parameters (*e.g.*, for autoregressive processes). Finally, optimal linear filtering techniques are useful to address various prediction, denoising and deconvolution problems. Their principle is to seek for the best filter in the sense of the mean square error. This leads to solving an optimization problem whose unknowns are the coefficients of the filter. This optimization problem being quadratic, its resolution boils down inverting a linear system.

The various techniques presented are illustrated by simulations provided on the edunao page of the course, within the framework of sound synthesis and denoising of an audio signal. The codes corresponding to these examples are available on edunao.

INDEX

- L^p space, 14
- ℓ^p space, 14
- anti-aliasing filter, 75
- autocorrelation, 80, 95
- autocorrelation function
 - estimator, 95
- autocovariance, 81
- autoregressive model (AR), 90
- Bartlett periodogram, 101
- BIBO stability, 32
- causality, 32
- continuous time signal, 25
- convolution, 26
- cross-correlation, 26, 81
- cross-covariance, 81
- denoising, 110
- finite difference equation, 90
- digital signal, 25
- Dirac comb, 17, 46
- discrete Fourier transform, 53, 99
- discrete time Fourier transform,
 - 49
- discrete time signal, 25
- distributions, 16
- energy, 26
- ergodicity, 94, 97
- estimator of the mean, 94
- filtering of a random process, 88,
 - 111
- Finite-dimensional distributions,
 - 79
- FIR filter, 34
- Fourier series representation, 48
- Fourier transform, 40
- frequency response, 55
- Gaussian process, 80
- Hilbert basis, 15
- IIR filter, 36
- impulse response, 33, 89
- interference formula, 88
- Kronecker function, 14
- mean, 80
- observation, 78
- optimal linear estimator, 106
- parametric estimation, 104
- periodogram, 98
- Poisson summation formula, 45
- power, 26, 83
- power spectral density, 83, 98
- prediction, 108
- quantization, 25
- random process, 21
- random vector, 21
- realization, 78
- finite difference equation, 36
- sampling, 67
- Schwartz space, 19
- Shannon's theorem, 70
- signal-to-noise ratio, 26, 83
- spectral resolution, 61
- spectrum, 40
- speech, 106
- stationarity, 81
- stochastic signal, 26, 78
- support, 19
- trajectory, 78
- transfer function, 55
- translation invariance, 31
- variance, 83

white noise, 86

Wiener filter, 110

windowing, 59

Yule-Walker equation, 92

zero padding, 58, 62, 99

zero-order hold, 75

BIBLIOGRAPHY

- [1] S. Mallat. *A wavelet tour of signal processing*. Elsevier, Amsterdam, troisième édition, déc. 2008.
- [2] Jutten C. *et al.* *L'héritage de Fourier, 250 ans après*. Interstices, revue scientifique en ligne, <https://interstices.info/1-heritage-de-fourier>, Le Chesnay, nov. 2018.
- [3] L. Schwartz. *Théorie des distributions*. Hermann, 1966.
- [4] F. Golse. *Distributions, analyse de Fourier, équations aux dérivées partielles*. Éditions de l'École Polytechnique, Palaiseau, oct. 2012.
- [5] M. Vetterli, J. Kovacevic, and V. K. Goyal. *Foundations of Signal Processing*. Cambridge Univ. Press, Cambridge, UK, oct. 2014.