Signals, systems and processes.

Lecture Notes.

Michel Besserve
Max Planck Institute for Intelligent Systems, Tübingen, Germany
mbesserve@gmail.com
michelbesserve.com
June 2022

Contents

Foreword

These notes constitute elements regarding the mathematical representation of time series. They have been used for teaching second year Master students at the Neural Information Processing program of the Graduate Training Center of Neuroscience of Tübingen, Germany. This program is oriented towards students that want to approach Neuroscience from a computational perspective. Given the variety of backgrounds, and the diversity of aspects of time series computational neuroscientists may encounter, lectures where aimed at:

- Introduce and learn to manipulate the mathematical representations of signals (convolution, difference equations, z-transform, point-processes);
- Introduce key principles behind classical signal processing tools: filtering, spectral analysis, wavelets;
- Understand the capabilities and limitations of time series analysis approaches and how to use them safely.

To achieve this purpose, I have attempted to gather in a coherent way a broad range of topics that are typically taught in different fields. Notably, I tried to put in coherence the discrete and continuous time perspectives on signals and systems, but also link deterministic and stochastic aspects, such that the later can be explained without resorting to heavy formalism. Mathematical refreshers are introduced in order to allow students with a B.Sc. from various scientific fields (Physics, Computer Science, Biochemistry, Mathematics) to follow the technical developments without referring to other textbooks.

I am very grateful to Kaidi Shao for helping me giving these lectures. Constructive feedback on these notes is of course welcome.

Part 1

Mathematical tools for signal processing

Chapter 1

Calculus

This chapter is a first mathematical refresher of the basic calculus rules that we need to manipulate sequences. Sequences are the simplest way to represent discrete time signals and will be used extensively in the next chapters. Along the way, we will also provide refreshers for complex numbers and vector spaces.

1 Complex numbers

While the theory behind complex numbers is very elegant, we will stick to a few calculus rules. We can postulate the existence of an imaginary unit **i** such that $\mathbf{i}^2 = -1$, and that any complex number can be considered as a linear combination of 1 and **i** with sum and product rules generalizing the ones in $\mathbb{R}$.

For any complex number z, we can write its algebraic representation:

$$z = a.1 + b.\mathbf{i},$$

with a and b **unique** real numbers.

a is called the *real part* $\mathrm{Re}(z)$, and b is called the *imaginary part* $\mathrm{Im}(z)$

The *conjugate* of a complex number $z = a.1 + b.\mathbf{i}$ is the number

$$\bar{z} = a.1 - b.\mathbf{i}.$$

As a consequence we get

$$\mathrm{Re}(z) = \frac{z+\bar{z}}{2} \text{ and } \mathrm{Im}(z) = \frac{z-\bar{z}}{2\mathbf{i}}$$

Instead of using the algebraic coordinates (a,b), we can use the polar representation (ρ,φ)

$$z = \rho(\cos\varphi + \mathbf{i}\sin\varphi)$$

$\rho \geq 0$ is unique and is called the *modulus* $\|z\|$, and φ is defined up to 2π, and is called the *argument* $\arg(z)$

$$\rho = |z| = \sqrt{a^2 + b^2}$$

Euler's formula is also particularly useful:

$$z = \rho \exp(\mathbf{i}\varphi)$$

where exp is the exponential function. It has the property $\exp(0) = 1$, and for any real or complex numbers (a,b) the exponential function satisfies

$$\exp(a+b) = \exp(a)\exp(b)$$

and form Euler's formula we can retrieve that for a real θ we have

$$\exp(\mathbf{i}\theta) = \cos\theta + \mathbf{i}\sin\theta$$

Then product operations are made easy with the exponential expression:

$$|z_1.z_2| = |z_1|.|z_2| \text{ and } \arg(z_1.z_2) = \arg(z_1) + \arg(z_2)$$

and trigonometric expression can be replaced by exponentials using:

$$\cos(\varphi) = \frac{e^{\mathbf{i}\varphi} + e^{-\mathbf{i}\varphi}}{2} \text{ and } \sin(\varphi) = \frac{e^{\mathbf{i}\varphi} - e^{-\mathbf{i}\varphi}}{2\mathbf{i}}$$

This polar representation has an intuitive interpretation in the complex plane as can be seen in Fig. 1.1, where we draw the complex number z in the complex plane and show the defined quantities, as well as the opposite and conjugate of z. In addition to this geometric representation, it is also useful to remember the values of trigonometric quantities for classical angles, as represented in Fig. 1.2.

EXERCISE 1. *Retrieve the following identities:*

$$\cos^2\theta + \sin^2\theta = 1 \tag{1.1}$$

$$\cos^2\theta = \frac{1}{2}(1 + \cos(2\theta)) \tag{1.2}$$

$$\sin(2\theta) = 2\sin(\theta)\cos(\theta) \tag{1.3}$$

EXERCISE 2. *Rewrite the following trigonometric formulas as sums and products of elementary functions of a and functions of b*

$$\cos(a+b), \sin(a+b).$$

PROOF. First deduce an expression of cos using exponential representations and manipulate them. □

EXERCISE 3. *Write* $\cos(x) + \sqrt{3}sin(x)$ *in the form* $A\cos(x+\phi)$

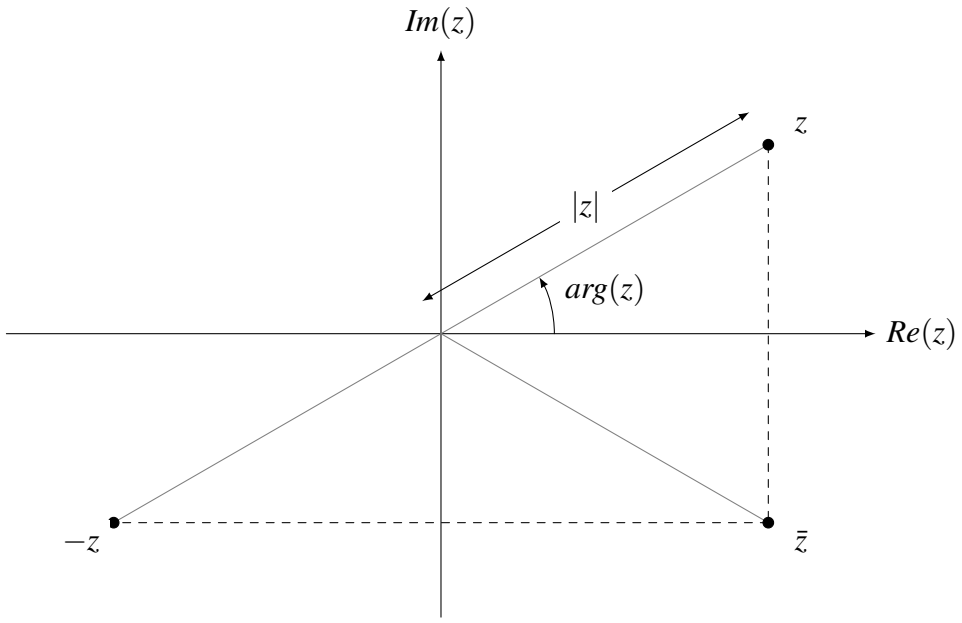

FIGURE 1.1. Representation of complex numbers on the complex plane, together with conjugate and opposite.

2 Finite sums

Manipulating sums easily is fundamental in discrete time signal processing. A compact way to write a sum of terms is using the "Sigma" symbol:

$$\sum_{i=i_0}^{n} x_i = x_{i_0} + x_{i_0+1} + \cdots + x_n$$

This notation is extensively used and can be generalized to multiple indices, and also index sets that are more complex than intervals. This notation as well as many formulas for rewriting sums using different indexing are justified by the *associativity* and *commutativity* of the summation operation. This implies that it does not matter in which order we sum the terms in such expressions. The distributivity with the product operation ($a \cdot (b+c) = ab + ac$)) can also used to rewrite such sums.

The following rules are very useful and can be retrieved easily:

(1) $\sum_{i=i_0}^{n} (ax_i + by_i) = a\sum_{i=i_0}^{n} x_i + b\sum_{i=i_0}^{n} y_i$
(2) $\sum_{i=i_0}^{n}\sum_{j=j_0}^{m} a_{i,j} = \sum_{j=j_0}^{m}\sum_{i=i_0}^{n} a_{i,j}$ (The Rectangle)
(3) $(\sum_{i=i_0}^{n} x_i)(\sum_{j=j_0}^{m} y_j) = \sum_{i=i_0}^{} \sum_{j=j_0}^{} x_i y_j = \sum_{j=j_0}^{} \sum_{i=i_0}^{} x_i y_j$

Check for example https://www.math.ucdavis.edu/ for interesting results and Wikipedia.

EXERCISE 4. *Rewrite* $\sum_{1 \le i,j \le n} a_{i,j}$ *with two sums in two different ways.*

EXERCISE 5. *Use multiple sums to find the expression of* $\sum_{k=1}^{N} k$.

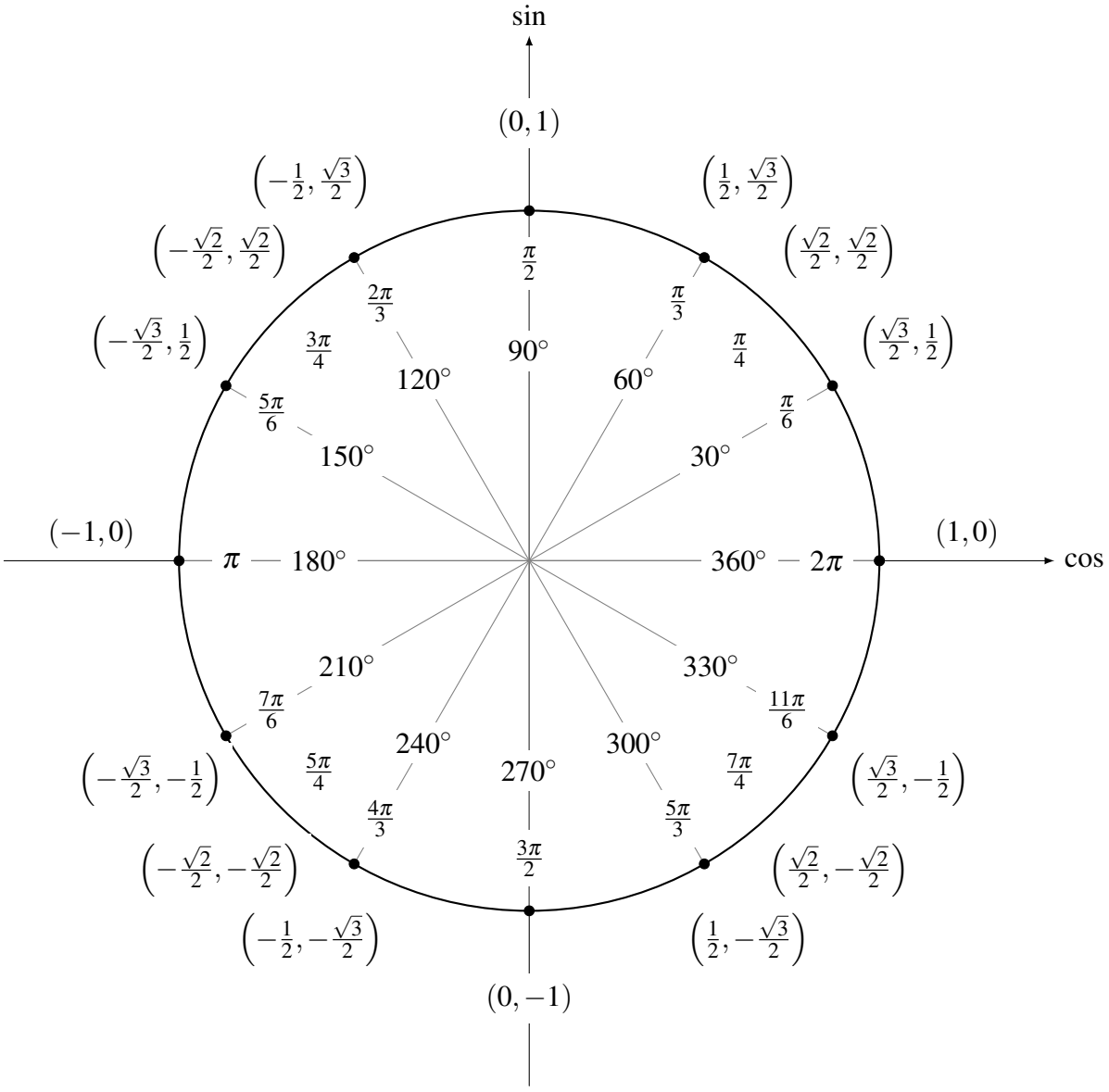

FIGURE 1.2. Representation of angles and their sine and cosine on the unit circle modified from http://www.texample.net.

PROOF. Tip: represent this sum as a double sum over a triangle. □

3 Sequences

3.1 Definition

Sequences are an "infinite list" of objects $\{\cdots, a_0, a_1, \cdots\}$ indexed by integers in a index set I, that we will denote $(a_n)_{n\in I}$. The index set can be any countable set and in the remaining of the course, most sequences will be indexed by $\mathbb{Z} = \{\cdots, -1, 0, 1, \cdots\}$, the set of negative and positive integers. This first section on sequences deals with sequences of real or complex numbers. Sequences can be generalized to much more complex objects as we will see along the chapters.

3.2 Convergence

Convergence is the key to give a meaning to the values of a sequences at the infinities. (a_n) converges to a finite *limit* l informally means: *"For any arbitrary level of accuracy, we can choose N large enough, such that the values for $n \geq N$ are close to the limit within the previously fixed level of accuracy."* Formally this is defined as follows.

DEFINITION 6 (Convergence). For all ***real*** numbers $\varepsilon > 0$, it exists a positive integer N such that for any positive integer $n \geq N$, $|a_n - l| < \varepsilon$.

We denote this property as

$$\lim_{n\to+\infty} a_n = l \quad \text{or} \quad a_n \underset{n\to+\infty}{\longrightarrow} l$$

3.3 Divergence

Instead of guarantying to go arbitrary close, divergence provides guaranties to go arbitrary far from 0. The definition can essentially be written by replacing "$|a_n - l| < \varepsilon$" by "$|a_n| > \varepsilon$" in the previous section.

4 Series

A series $\sum_{n\geq n_0} a_n$ denotes the formal object to study the sum of the terms of a sequence (a_n) indexed by non-zero positive integers. It is called convergent whenever the sequence of partial sums

$$S_N = \sum_{n=n_0}^{N} a_n$$

converges. Then the value of the series is

$$\sum_{n=n_0}^{+\infty} a_n = \lim_{N\to+\infty} S_N$$

EXERCISE 7. *Study the geometric series $\sum_{n=0} \rho^n$, for ρ complex number.*

SOLUTION. We observe that

$$(1-\rho)\sum_{n=0}^{N}\rho^n = \sum_{n=0}\rho^n - \sum_{n=0}\rho^{n+1}$$

Thus

$$(1-\rho)\sum_{n=0}^{N}\rho^n = \sum_{n=0}^{N}\rho^n - \sum_{n=1}^{N+1}\rho^n = 1-\rho^{N+1}$$

Thus for $|\rho| \neq 1$ (possibly complex), we get

$$\sum_{n=0}^{N}\rho^n = \frac{1-\rho^{N+1}}{(1-\rho)}.$$

This is a classical result for geometric progressions. Now assume $|\rho| < 1$, we get convergence of the partial sums to

$$\sum_{n=0}^{+\infty}\rho^n = \frac{1}{1-\rho}.$$

□

5 Sequence spaces

5.1 Vector spaces

A F-vector space V is a set equiped with two operations:

- addition "+" (commutative, associative): for u, v in V, $u+v \in V$
- scalar multiplication ".": for $u \in V$ and λ in a field F (typically $\mathbb{R}$ or $\mathbb{C}$), $\lambda.u \in V$
- check Wikipedia for all conditions on these operations (associativities, distributivities,...)

Elements of V are called vectors, elements of the field F are called scalars.

The following additional conditions are needed:

- there is a unique zero element $\mathbf{0}$ in V such that $u+\mathbf{0} = u$ for all u in V
- $1.u = u$, for all u in V
- $\alpha.(\lambda.u) = (\alpha\lambda).u$, for all u in V
- for all u in V there is an inverse element $-u$ such that $u+(-u) = 0$

A *subspace* of V is a subset of V stable by the addition and scalar multiplication operations.

A subspace of a vector space is a vector space!!!

5.2 Vector spaces of sequences

Sequences can form vector spaces when equipped with the two following operations:

- element-wise addition
- element-wise scalar multiplication

In particular, the sets of all sequences of real or complex numbers with indices in $\mathbb{N}$ or in $\mathbb{Z}$ are real or complex vector spaces. They are denoted $\mathbb{R}^{\mathbb{N}}, \mathbb{C}^{\mathbb{Z}}, \ldots$. Those spaces are too large to be "usable". The following sequence spaces are subsets of those large spaces.

5.3 ℓ^p spaces

For $p \geq 1$ or 2, $\ell^p(\mathbb{Z})$ is the subspace of all elements of $\mathbb{C}^{\mathbb{Z}}$ such that

$$\|x\|_p = \left(\sum_k |x_k|^p \right)^{1/p} < +\infty.$$

Three important examples of subspaces are

- $\ell^1(\mathbb{Z})$ the subspace of sequences such that $\sum_k |x_k|^< + \infty$
- $\ell^2(\mathbb{Z})$ the subspace of sequences such that $\sum_k |x_k|^2 < +\infty$
- $\ell^\infty(\mathbb{Z})$ the subspace of bounded sequences i.e. such that $\|x\|_\infty = \sup_n |x_n| < +\infty$.

Chapter 2

Functions of a real variable

In the same way discrete time signals are usually represented by sequences, continuous time signals are usually represented by functions. We recall a variety of definitions related to functions of a real variable. At the end of the chapter, we introduce distributions, which can be seen as a useful generalization of functions.

1 Definition and basic properties

A function f with *domain* X subset of $\mathbb{R}$ and *codomain* Y subset of $\mathbb{R}$ or $\mathbb{C}$ is denoted

$$f : X \to Y$$

and associates to each value in $x \in X$ a unique value in Y, $f(x)$. If the domain is symmetric around 0, f is *even* if $f(-t) = f(t)$, and *odd* if $f(-t) = -f(t)$, for all $t \in X$. If the domain is $\mathbb{R}$, f is T-periodic if $f(t+T) = f(t)$, for all $t \in \mathbb{R}$. The *support* of a function is the closure of the domain for which $f(t) \neq 0$

1.1 Function manipulation

Mirroring, scaling (stretch/compress), translation, periodization.

EXERCISE 8. *Using a "rectangle" function* $rect(t) = \begin{cases} 0 & |t| > 1/2 \\ 1/2 & |t| = 1/2 \\ 1 & |t| < 1/2 \end{cases}$*, plot the following functions: rectangle, rectangle rescaled, rectangle translated and rescaled.*

1.2 Convergence

f converges to a finite *limit* l means as t tends to t_0: *"For an arbitrary level of accuracy, we can choose a distance to* t_0 *small enough, such that for any* t *closer than this distance,* $f(t)$ *is close to the limit within the previously fixed accuracy."*

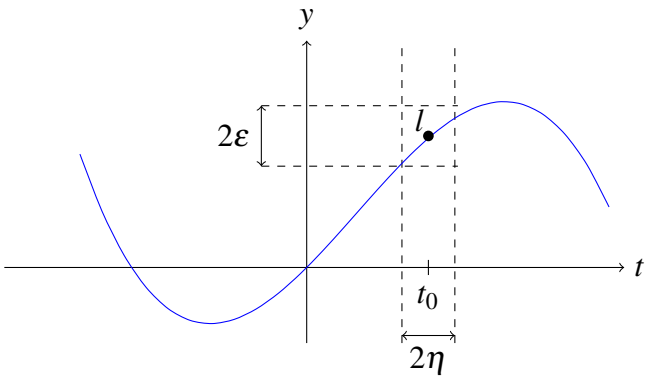

FIGURE 1.1. Illustration of the convergence definition.

DEFINITION 9 (Convergence). For all ***real*** numbers $\varepsilon > 0$, it exists a real positive number η such that for any t in the domain such that $|t - t_0| < \eta$, $|f(t) - l| < \varepsilon$.

We denote this property

$$\lim_{t \to t_0} f(t) = l \text{ or } f(t) \underset{t \to t_0}{\longrightarrow} l$$

An illustration of this definition is provided on Fig. 1.1. Instead of the formal definition, sequence characterization if very useful.

THEOREM 10. *For f defined in the neighborhood of t_0, and $l \in \mathbb{C}$, $\lim_{t \to t_0} f(t) = l$ if and only if for all sequences $(t_n) \to t_0$, $(f(t_n)) \to l$.*

As a consequence, convergence of a function of the real variable is reduced to the study of convergence of particular sequences.

1.3 Continuity

Continuity essentially states that the function converges to its own value.

$f(t)$ continuous in t_0 whenever $f(t) \underset{t \to t_0}{\longrightarrow} f(t_0)$

Many classical functions are continuous on their domain. In particular, all polynomials (such as integer powers), log, exp, cos, sin, rational fractions, are continuous on their domain. Simple but important exceptions are Heaviside's, rectangular and sign function which have a few points of discontinuity!

TABLE 1. Derivative and primitive of classical functions

primitive	function	derivative
$-\cos$	$\sin$	$\cos$
$\sin$	$\cos$	$-\sin$
$\exp$	$\exp$	$\exp$
$\lvert t\rvert\log(\lvert t\rvert)-\lvert t\rvert$	$\log(\lvert t\rvert)$	$\frac{1}{\lvert t\rvert}$
$\frac{t^{\alpha+1}}{\alpha+1}$, $\alpha\neq -1$	t^{α}	$\alpha t^{(\alpha-1)}$, $\alpha\neq 0$
	$\lvert t\rvert$?	

2 Derivative

2.1 Definition of derivative

For a function $f(t)$, the derivative in t_0 is the following limit whenever it exists

$$f'(t_0)=\frac{df(t)}{dt}_{|t_0}=\lim_{h\to 0}\frac{f(t_0+h)-f(t_0)}{h}$$

2.2 Combination rules

The following rules are used all the time:

- $(\lambda f(t)+g(t))'=(\lambda f'(t)+g'(t))$
- $\frac{df(g(t))}{dt}=g'(t)f'(g(t))$
- $\frac{d(f(t).g(t))}{dt}=f'.g+f.g'$

2.3 Primitive (antiderivate)

F is a primitive function or antiderivative of a function f on a domain if $F'=f$. For any primitive of f

$$\int_a^b f(t)dt=F(b)-F(a)$$

2.4 Classical derivatives

EXERCISE 11. *Compute the derivatives of the following functions* $\frac{1}{1+t^2}$, $\exp(-t^2)$, $\tan(t)$

EXERCISE 12. *Show that* $sinc(t)=\begin{cases}\frac{\sin t}{t}, t\neq 0\\ 1, t=0\end{cases}$ *, is continuous. Draw this function.*

3 Integration

Informally, integrals measure the area under the graph of the function. It can be rigorously defined using the Riemann or Lebesgue integral. Both definitions coincide for the functions we will use (piecewise continuous). Many function spaces based on the Lebesgue integral require the functions to be *measurable,* unless mentioned otherwise this will always be the case for us.

We will denote the integral on E

$$\int_E f(t)dt$$

3.1 Elementary properties

(1) f with complex values is said *integrable* on E if and only if $\int_E |f(t)|dt$ is finite. Then $\int_E f(t)dt$ is also finite.
(2) For non-overlapping subsets E and F, $\int_{E\cup F} f = \int_E f + \int_F f$.
(3) Linearity: $\int_E (\alpha f(t) + g(t))dt = \alpha \int_E f(t)dt + \int_E g(t)dt$
(4) If $f \leq g$ on E, $\int_E f dt \leq \int_E g dt$
(5) If f integrable, $|\int_E f dt| \leq \int_E |f| dt$
(6) If it exists g integrable such that $|f| \leq g$, then f is integrable and $\int_E |f| dt \leq \int_E g dt$
(7) If f and g differ only in a countable number of points, their integrability properties are identical (they are equal *almost everywhere*)

3.2 Integrating with primitives

If g is a primitive of f integrable on an interval $[a,b]$ (more generally, g need only be *absolutely continuous*, not defined here)

$$\int_{[a,b]} f dt = \int_a^b f dt = -\int_b^a f dt = [g]_a^b = g(b) - g(a)$$

One can derive from that the integration by parts formula: for f, g continuously differentiable on E (more generally absolutely continous)

$$\int_a^b f g' dt = [fg]_a^b - \int_a^b f' g dt$$

3.3 Change of variables

Given φ continuously differentiable achieving a one to one mapping between intervals X and Y,

$$\int_Y f(y)dy = \int_X f(\varphi(x)).|\varphi'(x)|dx$$

EXERCISE 13. *Apply this formula to basic transformation: mirroring, time rescaling, shifting.*

3.4 Indefinite integrals

Difficulties regarding integration arise when $t \to +/-\infty$ or at points of divergence of the function.To study integrability, with usually resort to either bounding the function with and integrable one, and/or use a primitive of f on an interval that approaches the integration interval, and study the limit. The following result is helpful

THEOREM 14. *If the following integral exists for all* $\varepsilon > 0$

$$I_\varepsilon = \int_{a+\varepsilon}^{b} |f(t)|dt$$

and $\lim_{\varepsilon\to 0} I_\varepsilon = I < +\infty$*, then* f *is integrable and* $I = \int_a^b |f(t)|dt$.

EXERCISE 15. *Study the integrability fo* $1/t^2, \cos(t)/t^2$ *on* $[1, +\infty)$.

3.5 Advanced properties

Advance properties such as continuity and derivative under the integral sign and Fubini's theorem will be introduced when needed in the next chapters.

4 Function spaces

Given an interval I we can define the following vector spaces of functions:

- $C^p(I)$ is the space of p times continuously differentiable functions on I.
- $C^\infty(I)$ is the space of infinitely continuously differentiable functions on I.
- $L^1(I)$ is the space of measurable integrable functions.
- $L^2(I)$ is the space of measurable square integrable functions.

Obviously $C^p(I) \subset C^q(I)$ for $p \geq q$. Whenever I is a closed bounded interval (of the form $[a,b]$), we have $C^p(I) \in L^1(I) \cap L^2(I)$ as functions are continuous on a closed interval and thus bounded. Otherwise, there is in general no inclusion rule between the above spaces. However, we can mention in addition $L^1_{loc}(\mathbb{R})$, the space of measurable functions integrable on any bounded interval, which is an important subspace of $L^1(\mathbb{R})$.

5 Norms, convergence and uniform continuity

We will have to consider issue of convergence for sequences of functions to investigate questions of the form

$$f_n \to_{n\to+\infty} f$$

There are different notions of convergence for functions associated to various types of results.

5.1 Point-wise convergence

Maybe the simplest convergence we can think about is to consider a fixed time t_0 and check whether the series defined at this points converges to $f(t_0)$ for all t_0.

5.2 Mean square convergence

Most types of convergence rely on defining a distance and go back to the real number case. Take the L^2 space of square integrable functions and the associated norm $\|f\|_2 = \int |f(t)|^2 dt$ which matches requirements for being a distance. We can consider using the *mean squared error* $\|f_n - f\|_2^2$ as a distance between f and its approximation and check it converges to zero.

$$\|f_n - f\|_2^2 \to_{n\to+\infty} 0$$

5.3 Uniform convergence

Instead of the L^2 norm we can use the so-called infinite norm $\|f\|_\infty = \sup_{t\in\mathbb{R}} |f(t)|$. Then the following

$$\|f_n - f\|_\infty \to_{n\to+\infty} 0$$

is called uniform convergence. It guaranties the same "speed" of convergence everywhere.

5.4 Uniform continuity

Another related concept (but still of different use) is "uniform continuity", which guaranties continuity at the same "speed" when approaching any point on an interval.

Part 2

Fourier analysis of signals and systems

Chapter 3

Fourier Series

1 Introduction

The goal of the following chapters is to describe signals that evolve in time by their frequency content. Frequencies describe changes that occur in the same way, "a cycle", at regular interval across time and are thus homogenous to the inverse of a time unit. The SI unit for frequency is the Hz (hertz) corresponding to one cycle per second. Most classical ranges of frequencies in the course will be at the scale of Hz or kHz (1000Hz), classical notation for the frequency variable are greek letters ν or ξ. Instead of counting cycles, it is possible to measure frequencies in terms of the angular displacement (in radians) per units of time. The classical angular frequency unit is rd/s (radian per second) with the conversion $1\text{Hz} = 2\pi\text{rad}/s$, a classical notation for the angular frequency is the greek letter ω. In this course, we choose to use the frequency variable (i.e. counting cycles not angles), be aware that 50% of the literature favors a angular frequency description, which implies a few easy changes in the associated equations.

2 Definition and properties

2.1 Definition

A Fourier series is a function of the time variable that takes the form

$$\sum_{n=-\infty}^{+\infty} c_n \exp(i2\pi nt/T),\ c_n \in \mathbb{C}$$

If it exists, this description is unique:$\sum_{n=-\infty}^{+\infty} c_n \exp(i2\pi nt/T) = 0$ if and only if $c_n = 0$, for all n.

An immediate consequence of the definition is that the function is T-periodic. To avoid to deal with a useless number of zero coefficients, T is most of the time chosen to be the minimal period. For example, if $c_n = 0$ for all n odd, then

$$\sum_{n=-\infty}^{+\infty} c_n \exp(i2\pi nt/T) = \sum_{n=-\infty}^{+\infty} c_{2n+1} \exp(i2\pi(2n+1)t/T) + \sum_{n=-\infty}^{+\infty} c_{2n} \exp(i2\pi 2nt/T)$$
$$= \sum_{n=-\infty}^{+\infty} c_{2n} \exp(i2\pi nt/(T/2))$$

has period $T/2$.

2.2 The case of real signals

Assume the Fourier series is real, we have

$$\sum_{n=-\infty}^{+\infty} c_n \exp(i2\pi nt/T) = \overline{\sum_{n=-\infty}^{+\infty} c_n \exp(i2\pi nt/T)}, \tag{2.1}$$

and the right-hand-side can be rewritten

$$\sum_{n=-\infty}^{+\infty} \bar{c_n} \exp(-i2\pi nt/T) = \sum_{n'=-\infty}^{+\infty} \bar{c}_{-n'} \exp(i2\pi n't/T)$$

By identifying the coefficients on both sides of equation 2.1 we get

$$c_{-n} = \bar{c}_n$$

Euler's formula can be used to derive an equivalence between the complex exponential and a cosine-sine representation which is most frequently used for real signals:

$$a_0/2 + \sum_{n=1}^{+\infty} \left[a_n \cos(2\pi nt/T) + b_n \sin(2\pi nt/T)\right].$$

Now only positive integer indices are used and a_n and b_n are real numbers (provides the Fourier series is real)!

We can write

$$\sum_{n=-\infty}^{+\infty} c_n \exp(i2\pi nt/T) = c_0 + \sum_{n=1}^{+\infty} \left[c_n \exp(i2\pi nt/T) + c_{-n} \exp(-i2\pi nt/T)\right]$$
$$= c_0 + \sum_{n=1}^{+\infty} \left[(c_n + c_{-n}) \cos(2\pi nt/T) + \mathbf{i}(c_n - c_{-n}) \sin(2\pi nt/T)\right]$$

such that for real signals

$$\sum_{n=-\infty}^{+\infty} c_n \exp(i2\pi nt/T) = c_0 + \sum_{n=1}^{+\infty} \left[2\mathrm{Re}(c_n) \cos(2\pi nt/T) - 2\mathrm{Im}(c_n) \sin(2\pi nt/T)\right]$$

and we get the following relationship between the complex and sine=cosine representation

$$a_n = 2\mathrm{Re}(c_n),\ b_n = -2\mathrm{Im}(c_n),\ c_n = \frac{1}{2}(a_n - \mathbf{i}b_n).$$

2.3 Computing Fourier coefficients

Given a T-periodic function f, we try to approximate it by a Fourier series such that

$$f(t) = \sum_{n=-\infty}^{+\infty} c_n \exp(i2\pi nt/T) = a_0/2 + \sum_{n=1}^{+\infty} [a_n \cos(2\pi nt/T) + b_n \sin(2\pi nt/T)]$$

in some sense (see next section). We will assume this is possible, that f is integrable on a period and derive the expression of the Fourier coefficients ($\int_T$ denotes the integral on a one period interval, which is independent on the considered interval).

$$c_n = \frac{1}{T}\int_T f(t)\exp(-2\pi\mathbf{i}nt/T)dt$$

$$a_n = \frac{2}{T}\int_T f(t)\cos(2\pi nt/T)dt$$

$$b_n = \frac{2}{T}\int_T f(t)\sin(2\pi nt/T)dt$$

Disregarding the issues pertaining to the validity of the approximation, these will define the *Fourier coefficients* of the function f. Note that if f is integrable on a period interval, the above coefficients are finite and well defined!

2.4 Spectral representation

Given a periodic function integrable on a period interval, the Fourier coefficient provide the first representation of a signal in the frequency domain. We will represent the amplitude and phase spectra of these signals as a superposition of spectral lines. For $|n| = 1$, it corresponds to the fundamental frequency (if the chosen period was minimal), for $|n| > 1$, they are called the harmonics.

EXERCISE 16. *Compute the Fourier coefficients and plot the spectral representation of the rectangular function. Using Matlab, plot the partial sum of Fourier series for a variable number of coefficients.*

3 Rules

We assume

$$\begin{aligned} f &\leftrightarrow \{(c_k, k/T)\} \\ g &\leftrightarrow \{(c'_k, k/T)\} \end{aligned}$$

3.1 Linearity

$$f + \lambda g \leftrightarrow \{\left(c_k + \lambda c'_k, k/T\right)\}$$

3.2 Shifting

$$\begin{array}{rcl} f \quad (t-a) & \leftrightarrow & \{(c_k \exp(-\mathbf{i}2\pi ak/T), k/T)\} \\ f.\exp(\mathbf{i}2\pi k_0 t/T) & \leftrightarrow & \{\left(c_{k-k_0}, k/T\right)\} \end{array}$$

3.3 Scaling

$$f(at) \leftrightarrow \{(c_k, a.k/T)\}$$

3.4 Using symmetry

- If f is even, $b_n = 0$ for all n
- If f is odd, $a_n = 0$ for all n

EXERCISE 17. *What can be concluded if $f(t + T/2) = -f(t)$ for all t (T is being the period).*

4 Theorems

When and how can we say $f_N(t) = \sum_{n=-N}^{+N} c_n \exp(i2\pi nt) \rightarrow_{N\rightarrow+\infty} f(t)$? This question was historically not an easy one to answer. In this section we provide several answers depending on the type of convergence we are considering. This answers will be given in the following vector spaces of functions.

DEFINITION 18. We can define the following vector spaces:

- $L^2_p(-T/2, T/2)$ is the space of T-periodic functions square integrable on $[-T/2, T/2]$ (f s.t. $\int_{-T/2}^{T/2} |f(t)|^2 dt < +\infty$).
- $L^1_p(-T/2, T/2)$ is the space of T-periodic functions integrable on $[-T/2, T/2]$ (f s.t. $\int_{-T/2}^{T/2} |f(t)| dt < +\infty$).

4.1 Point-wise convergence

Maybe the simplest convergence we can think about is to consider a fixed time t_0 and check whether the series defined at this points converges to $f(t_0)$.

THEOREM 19. *(Dirichlet's theorem) Let f be in $L^1_p(-T/2,T/2)$, if $\lim_{t\to t_0,t<t_0} = f(t_0^-)$ and $\lim_{t\to t_0,t>t_0} = f(t_0^+)$ and the left and right-hand derivatives exist at t_0, then*

$$f_N(t_0) \to \frac{1}{2}[f(t_0^+) + f(t_0^-)]$$

This implies that pointwise convergence is achieved at any point where the function is continuous and derivable.

4.2 Best approximation and Parseval's equality

The second type of convergence we can consider is using the *mean squared error* as a distance (averaged on a period) between f and its approximation.

THEOREM 20. *If $f \in L^2_p(-T/2,T/2)$,*

(1) For all N, f_N is the best approximation of f as a linear combination of the functions $\{\exp(i2\pi nt/T)\}_{|n|\leq N}$ in the mean square sense:

$$\frac{1}{T}\int_{-T/2}^{T/2} |f(t) - f_N(t)|^2 dt = \min_{\alpha} \frac{1}{T}\int_{-T/2}^{T/2} |f(t) - \sum_{|n|\leq N} \alpha_n \exp(i2\pi nt/T)|^2 dt.$$

(2) This approximation converges to f in the mean square sense:

$$\frac{1}{T}\int_{-T/2}^{T/2} |f(t) - f_N(t)|^2 dt \to_{N\to+\infty} 0,$$

and

$$\sum_{n=-\infty}^{+\infty} |c_n|^2 = \frac{1}{T}\int_{-T/2}^{T/2} |f(t)|^2 dt.$$

For real signals we have

$$|a_0|^2/4 + \sum_{n=1}^{+\infty} \left(|a_n|^2 + |b_n|^2\right)/2 = \frac{1}{T}\int_{-T/2}^{T/2} |f(t)|^2 dt.$$

The intuition behind Parseval is that the power of a periodic signal is the sum of the powers of each sinusoidal components.

4.3 Uniform convergence

A stronger notion of convergence provides guaranties on the approximation that are uniform on a whole interval. The speed of convergence of $f_N(t)$ to $f(t)$ does not depend on t.

DEFINITION 21. (f_n) converges to f uniformly on S if for all ***real*** numbers $\varepsilon > 0$, it exists a positive integer N such that for all $t \in S$ and all $n \geq N$, $|f_n(t) - f(t)| < \varepsilon$.

THEOREM 22. *Assume that f has period T and is continuous on $\mathbb{R}$, is differentiable on $[-T/2, T/2]$ except possibly at finite number of points; and that f' is piecewise continuous.*

(1) *The Fourier coefficients of f satisfy*

$$\sum_{n=-\infty}^{+\infty} |c_n| < +\infty.$$

(2) *The Fourier series of f converges uniformly to f on $\mathbb{R}$.*

5 Gibbs phenomenon

5.1 Example

Gibbs phenomenon is a good illustration of how the above notions of convergence play a role in practice. Consider the 1-periodic waveform

$$f(t) = \begin{cases} 1 & 0 \leq t < 1/2 \\ -1 & -1/2 \leq t < 0 \end{cases}.$$

We can show its Fourier series is

$$\sum_{k=0}^{+\infty} \frac{4}{\pi(2k+1)} \sin(2\pi(2k+1)t)$$

As a consequence the partial sum writes:

$$\sum_{k=0}^{N} \frac{4}{\pi(2k+1)} \sin(2\pi(2k+1)t)$$

Using the MatLab script “squarefunFS.m”, we can observe that, close to the discontinuities, there is always and over- and undershoot of the approximation which does not decrease in amplitude with N. Indeed, although point-wise and mean squared convergence is guarantied by the above theorems, uniform convergence is not guarantied due to the discontinuities.

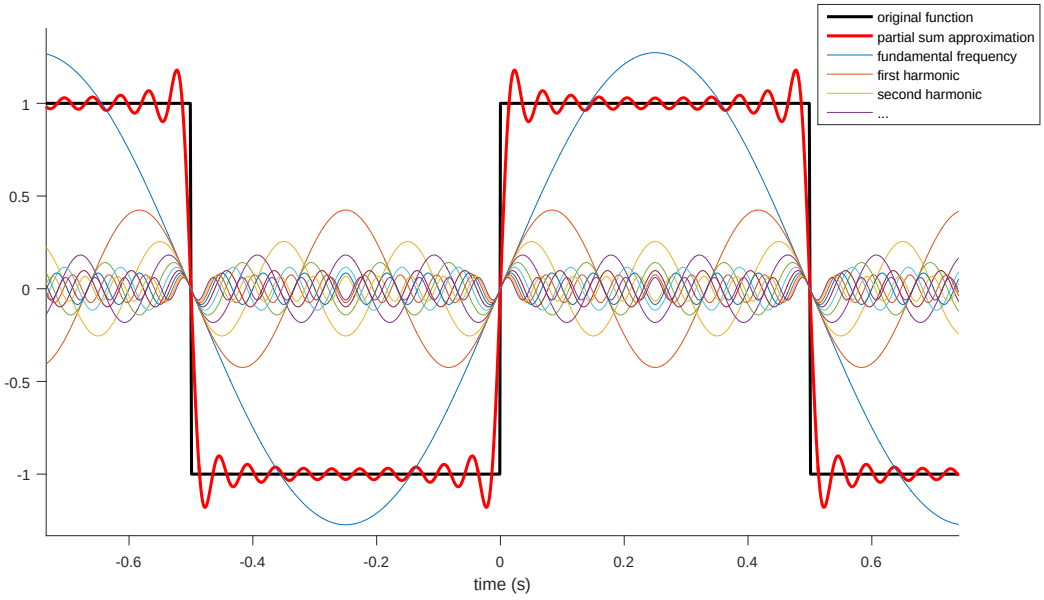

FIGURE 5.1. Illustration of the Gibbs phenomenon for a square wave form.

5.2 Overshoot estimation

The overshoot can be understood using the following reasoning: assume N is large and we evaluate the partial sum at the point $\frac{1}{4N}$.

$$f_N(\frac{1}{4N}) = \sum_{k=0}^{N} \frac{4}{\pi(2k+1)} \sin(\frac{2\pi(2k+1)}{4N}) = \sum_{k=0}^{N} \frac{2}{N} \frac{\sin(\frac{\pi(2k+1)}{2N})}{\pi(2k+1)/(2N)}$$

The right-hand side is a Riemann sum which tends to the following integral for N large

$$\sum_{k=0}^{N} \frac{2}{N} \frac{\sin(\frac{\pi(2k+1)}{2N})}{\pi(2k+1)/(2N)} \to 2\int_0^1 \frac{\sin(\pi x)}{\pi x} dx = 2/\pi(\int_0^\pi \frac{\sin(y)}{y} dy) = 1 + 2.(0.089489872...)$$

Due to an estimation of the right hand side (Antoni Zygmund, Trigonometrical series, 1955). We thus have asymptotically a systematic overshoot at $t = \frac{1}{4N}$ of approximately 9% of the amplitude of the discontinuity at 0. This shows there is no uniform convergence of the series on an interval including $t = 0$.

Chapter 4

Continuous Time Fourier Transform

Most of the signals that we can observed are at least not exactly periodic. As a consequence, they cannot be described as a sum of trigonometric functions with frequencies that are multiples of a fundamental frequencies. The continuous Fourier transform operates a fundamental switch from a description of a signal with a countable number of frequencies to a description where frequency values can cover the whole real axis (and are thus not-countable). While this thus leads to a much more abstract frequency representation, this is the price to pay to be able to describe signals that are "localized" in time.

1 Introduction

We seek a way to extend the above frequency analysis to non-periodic functions. We get an intuition by looking at the Fourier coefficients when the period of the function tends to ∞:

$$c_k(f) = \frac{1}{T}\int_{-T/2}^{T/2} f(t)\exp(-2\pi \mathbf{i} kt/T)dt.$$

Then the integral tends to spread on the whole real axis, and the frequencies k/T are getting denser over the frequency axis, such that in the limit we can speculate that a continuous frequency parameter would be appropriate.

2 Definition and properties

2.1 Definition

We thus propose the following expression for the (forward) *Fourier transform*

$$(\mathscr{F}f)(\xi) = \hat{f} = \int_{\mathbb{R}} f(t)\exp(-\mathbf{i}2\pi\xi t)dt.$$

To reconstruct the orgininal signal from its Fourier transform, again inspired by Fourier series, we just turn the summation of the series into an integral with respect to the continuous frequency variable, to get the *inverse Fourier transform*

$$\left(\overline{\mathscr{F}}\hat{f}\right)(t) = \int_{\mathbb{R}} \hat{f}(\xi)\exp(\mathbf{i}2\pi\xi t)d\xi.$$

2.2 First Fourier transform

Let us compute the Fourier transform of

$$\mathrm{Rect}_T(t) = \begin{cases} 1 & |t| \leq T/2 \\ 0 & otherwise \end{cases}.$$

First we observe that the function is in L^1, which implies

$$|\mathrm{Rect}_T(t)\exp(\mathbf{i}2\pi\xi t)| = |\mathrm{Rect}_T(t)|$$

is integrable, such that the Fourier transform can be computed and

$$\widehat{\mathrm{Rect}_T}(\xi) = \int_{-T/2}^{T/2} \exp(-2\pi\mathbf{i}\xi t)dt$$

Thus for $\xi = 0$, we get

$$\widehat{\mathrm{Rect}_T}(0) = T$$

and for $\xi \neq 0$, we get by integration of the exponential

$$\widehat{\mathrm{Rect}_T}(\xi) = \left[\frac{\exp(-2\pi\mathbf{i}\xi t)}{-2\pi\mathbf{i}\xi}\right]_{-T/2}^{T/2} = \frac{\sin(\pi T\xi)}{\pi\xi},$$

one can easily check, as predicted by the above theorem, that this function is continuous even in 0, and we can write it as a cardinal sine

$$\widehat{\mathrm{Rect}_T}(\xi) = T\mathrm{sinc}(\pi\mathrm{T}\xi).$$

2.3 First properties

The above example illustrated that functions from $L^1(\mathbb{R})$ can be Fourier transformed, and come up with interesting properties summarized in the following theorem.

THEOREM 23. *Let f be in $L^1(\mathbb{R})$, then $\hat{f}$ satisfies the following conditions:*

(1) *$\hat{f}$ is continuous and bounded on $\mathbb{R}$.*
(2) $\sup_{\xi\in\mathbb{R}} |\hat{f}(\xi)| \leq \int_{\mathbb{R}} |f(t)|dt$

Note

Continuity is a direct consequence of the following classical continuity under integral result (see [**Gasquet and Witomski(1999)**, Proposition 14.2.1]).
PROPOSITION (Continuity under the integral sign). *Given an open interval (a,b) (limits can be infinities), if*

(1) *For all $\xi \in (a,b)$, $t \mapsto f(t,\xi)$ is integrable on $\mathbb{R}$ (i.e. in $L^1(\mathbb{R})$).*
(2) *For almost all $t \in \mathbb{R}$, the function $\xi \mapsto f(t,\xi)$ is continuous.*
(3) *There exists $g \in L^1(\mathbb{R})$ such that*

$$|f(t,\xi)| \leq g(t) \text{ almost everywhere (w.r.t. variable } t), \text{for all } \xi \in (a,b).$$

Then the mapping $h(\xi) \mapsto \int_{\mathbb{R}} f(t,\xi)dt$ is continuous on (a,b).

In addition, if the Fourier transform is also integrable, we can invert the Fourier transform at continuity points.

THEOREM 24. *(Inversion theorem) If f and $\hat{f}$ are both in $L^1(\mathbb{R})$, then $\left(\bar{\mathscr{F}}\hat{f}\right)(t) = f(t)$ at all points where f is continuous.*

2.4 First calculation rules

Let $f,g \in L^1(\mathbb{R})$

(1) Linearity
$$\widehat{\lambda f + g} = \lambda \hat{f} + \hat{g}$$
(2) Time shift
$$\widehat{f(t-\tau)} = \hat{f}(\xi).\exp(-2\pi \mathbf{i}\xi\tau)$$
(3) Frequency shift
$$\widehat{f(t)e^{2\mathbf{i}\pi f_0 t}} = \hat{f}(\xi - f_0)$$
(4) Scaling
$$\widehat{f(at)} = \frac{1}{|a|}\hat{f}(\xi/a), a \neq 0$$

2.5 Conjugation and parity rules

For $f \in L^1(\mathbb{R})$

(1) $\overline{\mathscr{F}(f)} = \overline{\mathscr{F}}(\overline{f})$.
(2) $\mathscr{F}(f)(-\xi) = \overline{\mathscr{F}}(f) = \overline{\mathscr{F}}(f(-t)) = \overline{\mathscr{F}}(f_\sigma)$ with $(f_\sigma : t \mapsto f(-t))$.
(3) f even $\Rightarrow \hat{f}$ even.
(4) f odd $\Rightarrow \hat{f}$ odd.
(5) f real and even $\Rightarrow \hat{f}$ real and even.

2.6 More direct Fourier transform computations

EXERCISE 25. *Compute the Fourier transform of*

$$f(t) = e^{-at}H(t), Re(a) > 0.$$

Deduce from this result the Fourier transform of

$$f(t) = e^{at}H(-t), Re(a) > 0,$$

and

$$f(t) = e^{-a|t|}, a > 0.$$

2.7 Derivation rules

PROPOSITION 26. *The following rules link derivation and Fourier transform (see for example* [**Gasquet and Witomski(1999)**, Proposition 17.2.1]*):*

(1) *If $t^k f(t) \in L^1(\mathbb{R})$, $k = 0, 1, \ldots, n$, then $\hat{f}$ is n times differentiable, and*

$$\hat{f}^{(k)}(\xi) = \mathscr{F}\left[(-2\mathbf{i}\pi t)^k f(t)\right](\xi), k = 1, 2, \ldots, n$$

(2) *If $f \in C^n(\mathbb{R}) \cap L^1(\mathbb{R})$ and if all the derivatives $f^{(k)}$, $k = 0, 1, \ldots, n$ are in $L^1(\mathbb{R})$, then*

$$\widehat{f^{(k)}}(\xi) = (2\mathbf{i}\pi\xi)^k \hat{f}(\xi), k = 0, 1, \ldots, n$$

(3) *If f has bounded support, $\hat{f} \in C^\infty(\mathbb{R})$.*

Note

This result is based on the the following classical result (see [**Gasquet and Witomski(1999)**, Proposition 14.2.2]).
PROPOSITION (Derivative under the integral sign). *Given an open interval (a,b) (limits can be infinities), if*

(1) *For all $\xi \in (a,b)$, $t \mapsto f(t,\xi)$ is mesurable on $\mathbb{R}$.*
(2) *For almost all $t \in \mathbb{R}$, the function $\xi \mapsto f(t,\xi)$ is continuously differentiable on (a,b).*
(3) *There exists $g \in L^1(\mathbb{R})$ such that*

$$|\frac{df}{d\xi}(t,\xi)| \leq g(t) \text{ almost everywhere (w.r.t. variable } t), \text{for all } \xi \in (a,b).$$

Then the mapping $h(\xi) \mapsto \int_{\mathbb{R}} f(t,\xi)dt$ is differentiable on (a,b) and

$$\frac{dh}{d\xi}(\xi) = \int_{\mathbb{R}} \frac{df}{d\xi}(t,\xi)dt$$

2.8 Application: Fourier transform of a Gaussian

We will compute the Fourier transform of a Gaussian by using a differential equation.

Let $f(t) = \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)$ be the Gaussian probability density of variance one. By computing the time derivative we get the differential equation

$$\frac{df}{dt}(t) = -tf(t).$$

Computing the Fourier transform on the right side we get, using the first derivation result

$$\mathscr{F}\left[-t.f(t)\right](\xi) = \frac{1}{2\pi\mathbf{i}}\mathscr{F}\left[(-2\pi\mathbf{i}t).f(t)\right] = \frac{1}{2\pi\mathbf{i}}\frac{d\hat{f}}{d\xi}.$$

Doing the same for the right hand side.

$$\mathscr{F}\left[\frac{df}{dt}(t)\right](\xi) = (2\pi\mathbf{i}\xi).\hat{f}(\xi).$$

As a consequence, matching both sides

$$(2\pi\mathbf{i}\xi).\hat{f}(\xi) = \frac{1}{2\pi\mathbf{i}}\frac{d\hat{f}}{d\xi}(\xi),$$

leading to the differential equation for the Fourier transform

$$\frac{d\hat{f}}{d\xi}(\xi) = -(2\pi)^2\xi.\hat{f}(\xi). \tag{2.1}$$

Using the time domain differential equation above, we already know one solution of this equation is

$$g_1(\xi) = \frac{1}{\sqrt{2\pi}} \exp(-(2\pi\xi)^2/2).$$

Since the differential equation is linear,

$$g_2(\xi) = \exp(-(2\pi\xi)^2/2)$$

is also a solution. To make sure we find the solution that is the actual Fourier transform $\hat{f}$ of our original function f. We can first describe the family of solutions of this differential equation. Assume g is a solution of the above differential equation, we can define $h = g/g_2$ without problem ($1/g_2$ is positive infinitely differentiable), then we can check that

$$\frac{d\left[g/g_2\right]}{d\xi}(t) = \frac{d\left[g.\exp((2\pi\xi)^2/2)\right]}{d\xi}(t) = \left[\frac{dg}{d\xi} + (2\pi)^2\xi g(\xi)\right]\exp(-(2\pi\xi)^2/2) = 0,$$

because g is a solution of the differential equation 2.1. Thus by integrating this result we get that all continuously differentiable solution of equation 2.1 are of the form

$$g(\xi) = Cg_2(\xi),$$

with C a constant value. As a consequence $\hat{f}(\xi) = C\exp(-(2\pi\xi)^2/2)$, and to find the constant we use the fact that by definition of the Fourier transform

$$\hat{f}(0) = \int_{\mathbb{R}} f(t)dt.$$

For our particular choice, f is a probability density, thus $\hat{f}(0) = 1$, so that we finally get

$$\hat{f}(\xi) = \exp(-(2\pi\xi)^2/2).$$

3 Convolution and Fourier transform

3.1 Convolution of continuous time functions

$$(f*g)(t) = \int_{\mathbb{R}} f(\tau)g(t-\tau)d\tau = \int_{\mathbb{R}} f(t-\tau)g(\tau)d\tau$$

An easy result regarding existence is for function in $L^1(\mathbb{R})$

PROPOSITION 27. *There are three classical cases for which the convolution is well defined (*[**Gasquet and Witomski(1999)**, Proposition 20.2.1 and 20.3.2]*)*

(1) If f and g are in $L^1(\mathbb{R})$, then $f*g$ is defined almost everywhere and belongs to $L^1(\mathbb{R})$.
(2) If $f \in L^1(\mathbb{R})$ and $g \in L^1_{loc}(\mathbb{R})$, if g is bounded, then $f*g$ is defined everywhere and is bounded.
(3) If $f \in L^1(\mathbb{R})$ and $g \in L^2(\mathbb{R})$, then $f*g$ is defined almost everywhere and belongs to $L^1(\mathbb{R})$.

EXERCISE 28. *Compute the convolution product $Rect_T * Rect_T$*

3.2 Properties of convolution

As far as the convolutions are well defined, we have

(1) $f*g = g*f$.
(2) $f*(g*h) = (f*g)*h = f*g*h$.
(3) $(f+\lambda g)*h = f*h+\lambda g*h$.

3.3 Fourier transform of convolutions and elementwise products

Again in $L^1(\mathbb{R})$ things are easy ([**Gasquet and Witomski(1999)**, Proposition 23.1.1])

PROPOSITION 29. *If f and g are in $L^1(\mathbb{R})$,*

(1) $\widehat{f*g}(\xi) = \hat{f}(\xi).\hat{g}(\xi)$, for all $\xi \in \mathbb{R}$.

(2) If in addition $\hat{f}$ and $\hat{g}$ are in $L^1(\mathbb{R})$, then

$$\widehat{f.g}(\xi) = \hat{f} * \hat{g}(\xi), \text{ for all } \xi \in \mathbb{R}.$$

4 Time-frequency trade-offs

4.1 Fourier transform in $L^2(\mathbb{R})$

We can extend the definition of the Fourier transform to $L^2(\mathbb{R})$ (functions such that $\int |f(t)|^2 dt < +\infty$), even if the function is not integrable. The main results are gathered in the following proposition.

PROPOSITION 30. *The extension of the Fourier transfrom for $f \in L^2(\mathbb{R})$ can be computed as*

$$\mathscr{F} f(\xi) = \lim_{n \to +\infty} \int_{-n}^{n} e^{-2\mathbf{i}\pi\xi t} f(t) dt.$$

Similarily, the inverse FT is

$$\overline{\mathscr{F}} f(\xi) = \lim_{n \to +\infty} \int_{-n}^{n} e^{2\mathbf{i}\pi\xi t} f(t) dt.$$

In addition,

(1) $\mathscr{F} f \in L^2(\mathbb{R})$
(2) $\overline{\mathscr{F}} f \in L^2(\mathbb{R})$
(3) $\mathscr{F}\overline{\mathscr{F}} f = \overline{\mathscr{F}}\mathscr{F} f = f$, *almost everywhere.*
(4) $\mathscr{F}\mathscr{F} f = \mathscr{F}\mathscr{F} f = f_\sigma = f(-t)$, *almost everywhere.*
(5) *Definitions conincide on* $L^1 \cap L^2$.

In addition, we have the isometry properties

(1) If $f, g \in L^2(\mathbb{R})$, then $\int_{\mathbb{R}} f(t)\bar{g}(t)dt = \int_{\mathbb{R}} (\mathscr{F} f)(\xi)\overline{(\mathscr{F} g)(\xi)} d\xi$.
(2) (Parseval-)Plancherel Theorem: $\int_{\mathbb{R}} |f(t)|^2 dt = \int_{\mathbb{R}} |\mathscr{F} f(\xi)|^2 d\xi$.

EXERCISE 31. *Compute the Fourier transform of $\frac{1}{a+2\mathbf{i}\pi\xi}$, for $a \in \mathbb{C}, Re(a) > 0$.*

4.2 Uncertainty principle

There are incompatibilities between the properties of time and frequency description that have deep implications, for example in physics, but also for signal processing. The most intuitive result is called the uncertainty principle (see [**Gasquet and Witomski(1999)**, Proposition 23.1.2] for example).

Let $f(t)$ such that $f, tf, \xi\hat{f}$ are in $L^2(\mathbb{R})$, and $\int |f(t)|^2 dt > 0$, for any t_0, ξ_0 (typically the "centers of mass" of the energy density are used), we can define the equivalent of a variance,

representing the spread of the energy distribution across time and frequency around t_0 and ξ_0 respectively:

$$\begin{aligned} \text{Effective duration } \sigma_f \text{ such that } \sigma_f^2 &= \frac{\int (t-t_0)^2 |f(t)|^2 dt}{\int |f(t)|^2 dt}, \\ \text{Effective bandwidth } \sigma_{\hat{f}} \text{ such that } \sigma_{\hat{f}}^2 &= \frac{\int (\xi-\xi_0)^2 |\hat{f}(\xi)|^2 dt}{\int |f(t)|^2 dt}. \end{aligned}$$

PROPOSITION 32. *Let f such that $f, tf, \xi\hat{f}$ are in $L^2(\mathbb{R})$, and $\int |f(t)|^2 dt > 0$, then*

$$\sigma_f . \sigma_{\hat{f}} \geq \frac{1}{4\pi}.$$

EXERCISE 33. *Compute the duration bandwidth product for Gaussian signals around 0. Compare to the inequality above. Illustrate the uncertainty principle by generating time-frequency atoms at various frequencies and time points based on Gaussian signals.*

Chapter 5

Discrete time Fourier transforms

1 Discrete Fourier Transform (DFT)

The Discrete Time Fourier (DFT) transform is the last step in our progressive transition from the abstract analysis of continuous time signals and transforms to transformations that can be handled numerically on our computer. While the previous chapter, by discretizing the time axis, let to a Fourier transform defined on a compact frequency interval instead of the whole real axis, the DFT is making two last important simplifications:

(1) the time horizon for which we can record and store a signal is finite, so we while consider signal that contain a finite number of points (which are assumed to result from equispaced sampling of an original continuous time signal),
(2) the number of frequencies at which we can evaluate a DTFT (defined in the previous chapter) is also finite, so we will only look at the Fourier transform on an equispaced grid of frequency points.

This constraint lead to define again a new linear transformation, but this time between finite dimensional vector spaces, such that this transformation can be easily studied as a matrix acting on the signal through the matrix-vector product. We summarize the elementary properties of this transform in this chapter, but the readers who would like to have a more detailed study and more concrete examples can refer to [**Butz(2015)**].

1.1 Definition

Assume the signal is now finite length N

$$x = [x[0], x[1], ..., x[N-1]]^T.$$

We define the discrete Fourier transform as

$$x \mapsto (F_N x)[n] = \sum_{k=0}^{N-1} x[k] \exp(-2\pi \mathbf{i} nk/N),\ n = 0..N-1.$$

Note that F_N is a linear transformation turn a length N column vector into another length N vector. As a consequence, F_N can be described as a square $N \times N$ matrix and $(F_N)_{k,n} = \exp(-2\pi \mathbf{i} nk/N)$. If we write down the coefficients the matrix takes the following shape:

$$F_N = \begin{bmatrix} 1 & 1 & \cdots & 1 & \cdots & 1 \\ 1 & e^{-\mathbf{i}2\pi/N} & & e^{-\mathbf{i}2\pi k/N} & & e^{\mathbf{i}2\pi/N} \\ \vdots & & & \vdots & & \vdots \\ 1 & e^{-\mathbf{i}2\pi n/N} & \cdots & e^{-\mathbf{i}2\pi kn/N} & \cdots & e^{\mathbf{i}2\pi n/N} \\ \vdots & & & \vdots & & \vdots \\ 1 & e^{\mathbf{i}2\pi/N} & \cdots & e^{\mathbf{i}2\pi k/N} & \cdots & e^{\mathbf{i}2\pi(N-1)/N} \end{bmatrix}_{\substack{n \\ k}}$$

PROPOSITION 34. *F_N is the DTFT of $x_{|0..N-1}$, the signal x truncated to interval $[0..N-1]$ sampled at frequencies $k/N, k = 0,..,N-1$:*

$$F_N x_{|0..N-1}[k] = \widetilde{x_{|0..N-1}}(k/N).$$

PROOF. Immediate by writing down the right hand side of the equation using the DTFT definition. □

Note that due to the 1-periodicity of the DTFT, it is then known on an infinite frequency grid.

1.2 Properties

PROPOSITION 35. *F_N is an invertible matrix and*

(1) $F_N^{-1} = \frac{1}{N}\overline{F_N}$.
(2) $\|F_N x\|^2 = N\|x\|^2$ (Parseval's equality).

PROOF. (1) is immediate by writing down the j,k term of the matrix product $\frac{1}{N}\overline{F_N}F_N$

$$\begin{aligned} (\frac{1}{N}\overline{F_N}F_N)_{j,k} = & \quad \frac{1}{N}\sum_{n=0}^{N-1}(\overline{F_N})_{j,n}(F_N)_{n,k} \\ = & \quad \frac{1}{N}\sum_n \exp(2\pi\mathbf{i}jn/N)\exp(-2\mathbf{i}\pi nk/N) \\ = & \quad \frac{1}{N}\sum_{n=0}^{N-1}\exp(2\pi\mathbf{i}n(j-k)/N) \end{aligned}$$

From the last expression we get that if $j = k$, all coefficient of the sum are 1, leading to the results that all diagonal coefficients of the product $\frac{1}{N}\overline{F_N}F_N$ are 1. If $j \neq k$, the last expression shows a geometric progression that we can rewrite

$$(\frac{1}{N}\overline{F_N}F_N)_{j,k} = \frac{1}{N}\frac{1-\exp(2\pi\mathbf{i}N(j-k)/N)}{1-\exp(2\pi\mathbf{i}(j-k)/N)} = 0$$

(2) results from the expression of the vector norm

$$\|F_N x\|^2 = \left(\overline{F_N x}\right)^T .(F_N x) = \left(\bar{x}^T \overline{F_N}^T\right).(F_N x)$$

Using point (1) of the proposition and $\overline{F_N}^T = \overline{F_N}$, we get

$$\|F_N x\|^2 = \left(\overline{F_N x}\right)^T .(F_N x) = N\left(\bar{x}^T\right).(x) = \|x\|^2$$

□

1.3 Circular convolution

For two discrete sequences x, y of period N, their circular convolution is the N-periodic sequence $x \circ y$ such that

$$(x \circ y)[n] = \sum_{k=0}^{N-1} x[k] y[n-k].$$

We can "periodize" finite length vectors to generalize this convolution to vectors of same length N. We then have the following relations for the DFT

PROPOSITION 36. *Given x, y two vectors of length N*

$$F_N(x \circ y) = (F_N x) \boldsymbol{.} (F_N y) \quad \textit{and} \quad F_N(x \boldsymbol{.} y) = (F_N x) \circ (F_N y),$$

where the "$\boldsymbol{.}$" indicates the element-wise product of the two vectors.

PROOF. We prove the first equality:

$$F_N(x \circ y)[j] = \sum_{n=0}^{N-1} \sum_{k=0}^{N-1} x[k] y[n-k] e^{-\mathbf{i}2\pi nj/N}$$

by change of variable $n \mapsto m + k$

$$F_N(x \circ y)[j] = \sum_{m=-k}^{N-1-k} \sum_{k=0}^{N-1} x[k] y[m] e^{-\mathbf{i}2\pi (m+k) j/N}$$

by factorizing the terms depending on k only

$$F_N(x \circ y)[j] = \sum_{k=0}^{N-1} x[k] e^{-\mathbf{i}2\pi kj/N} \sum_{m=-k}^{N-1-k} y[m] e^{-\mathbf{i}2\pi mj/N}$$

by using the fact that the y vector is periodized in the circular convolution equation we get

$$F_N(x \circ y)[j] = (F_N x)[j] . \left(\sum_{m=-k}^{-1} y[m+N] e^{-\mathbf{i}2\pi mj/N} + \sum_{m=0}^{N-1-k} y[m] e^{-\mathbf{i}2\pi mj/N} \right)$$

which leads by change of variable to

$$F_N(x \circ y)[j] = (F_N x)[j] . \left(\sum_{n=-k+N}^{-1+N} y[n] e^{-\mathbf{i}2\pi (n-N) j/N} + \sum_{m=0}^{N-1-k} y[m] e^{-\mathbf{i}2\pi mj/N} \right)$$

noticing that $e^{-\mathbf{i}2\pi(n-N)j/N} = e^{-\mathbf{i}2\pi(n)j/N}e^{\mathbf{i}2\pi Nj/N} = e^{-\mathbf{i}2\pi(n)j/N}$, we can bring back together both terms of the sum, leading to the entry-wise product

$$F_N(x \circ y)[j] = (F_N x)[j].\left(\sum_{m=0}^{N-1} y[m]e^{-\mathbf{i}2\pi mj/N}\right) = (F_N x)[j].(F_N y)[j].$$

The second equation of the proposition is then straightforward to deduce using the inverse DFT matrix. □

1.4 Calculation rules

The calculation rules for the DFT follow some rules already defined for the DTFT, except for the scaling and shifting rules! Scaling becomes a complex matter, while for shifting rules we have to bear in mind that, as a consequence of circular convolution rules, the shift that has to be considered is a circular shift, which corresponds to a shift of the periodized signal, in the time domain as well as in the Fourier domain.

Let $x, y \in \mathbb{R}^N$, let x_P, x_P indicate their periodized version (with period N)

(1) Linearity
$$F_N(\lambda x + y) = \lambda F_N x + F_N y$$
(2) Time shift
$$\left[F_N\left(x_P[k-k_0]\right)\right][n] = (F_N x)[n] \cdot \exp(-2\pi\mathbf{i}nk_0/N),\, k_0 \in \mathbb{Z}$$
(3) Frequency shift (note the left-hand-side dot indicates an entrywise product)
$$F_N\left[x[n] \cdot e^{2\mathbf{i}\pi k_0 n/N}\right] = (F_N x)_P[k-k_0]$$

1.5 Zero padding

We saw that the DTFT of a signal with support restricted to $[0..N-1]$ can be sampled at points k/N with the DFT, but what if we want a more precise grid? The solution is simple: we just add M zeros at the end of our signal vector and then apply the DFT on these $N+M$ samples. Just writing the definition of the DFT and the definition of the DTFT, we can easily identify that such zero padded DFT computes the values of the target DTFT on a grid with a finer step

$$\left(F_{N+M}\mathrm{pad}(x_{0..N-1}, M)\right)[k] = \tilde{x}(\frac{k}{N+M}).$$

1.6 Complexity and Fast Fourier Transform (FFT)

DFT can be done with matrix vector multiplication with complexity N^2, however a much faster algorithm is available for the case $N = 2^k$, k integer with complexity $N \log N$.

Part 3

Linear time invariant systems (LTI)

Chapter 6

Analog and digital filters

The purpose of this lecture is to describe families of classical LTI systems that are used to design filters, both in continuous and discrete time. The classical family for designing analog filters (continuous time LTI systems) is described using linear differential equations while digital filters (in discrete time) are described using difference equations. We show how these two families of systems can be studied in the Fourier domain and introduce two new transforms to simplify this analysis: the Laplace and z-Transforms.

1 Introduction

We saw in the previous lecture the linear time invariant (LTI) filters are characterized by the input-output relationship of the form

$$y = h * x,$$

where h is the impulse response of the system. This relationship holds for both continuous and discrete time systems, and the convolution-Fourier Transform relationship for both cases allows to fully describe the system by it *frequency response*. Indeed, the output for a unit complex exponential sinusoid $x[n]$ of normalized frequency ν is

$$y[n] = \hat{h}(\nu)x[n]$$

in discrete time and

$$\hat{h}(f)x(t)$$

in continuous time for a complex exponential sinusoid of frequency f (in Hz), such that the input-ouptut amplitude ratios and phase differences allow to fully characterize the Fourier transform of the impulse response, and thus the impulse response itself using the Fourier inversion formula. In this lecture we will see how this framework is used to analyze digital (discrete time) and analog (continuous time) filters. Due to the above results, in the remaining of the lecture, we will often name *frequency response* the Fourier transform of the impulse response of the system.

2 Analog filters

2.1 Differential equations

The most frequently encontered continuous time LTI systems are determined by a linear differential equation. A differential equation relates linearly the derivatives of the input to the derivative of the output such that it takes the general form

$$\sum_{k=0}^{M} a_k \frac{d^k y}{dt^k} = \sum_{k=0}^{N} b_k \frac{d^k x}{dt^k}.$$

2.2 Example of the first order low pass filter

We consider the system governed by the first order differential equation

$$y(t) + \tau \frac{dy}{dt} = x(t).$$

Assume the input is in L^1 as well as the output and its derivative, then we can apply the Fourier transform to both sides of the equation to get

$$\hat{y} + \tau 2\pi \mathbf{i} \xi \hat{y}(\xi) = \hat{x}(\xi).$$

This leads to the solution

$$\hat{y}(\xi) = \frac{1}{1 + \tau 2\pi \mathbf{i} \xi} \hat{x}(\xi).$$

Interestingly, we can easily deduce from the chapter on continuous time Fourier transform that $\frac{1}{1+\tau 2\pi \mathbf{i}\xi} = \frac{1/\tau}{1/\tau + 2\pi \mathbf{i}\xi}$ is the Fourier transform of $\frac{1}{\tau}.e^{-t/\tau}u(t)$, where $u(t)$ is the Heavyside function. As a consequence, since $\frac{1}{\tau}.e^{-t/\tau}u(t)$ is in L^1 we can apply the Fourier convolution rule such that

$$y = \frac{1}{\tau}.\left(e^{-t/\tau}u(t)\right) * x.$$

where $h(t) = \frac{1}{\tau}.e^{-t/\tau}u(t)$ thus appears to be the impulse response function of the system, which turns out to be causal. As a consequence, the input-output relationship of the system can take the explicit integral form

$$y(t) = \int_0^{+\infty} \frac{1}{\tau} e^{-s/\tau} x(t-s) ds.$$

2.3 General description using the Laplace Transform

In the above example, if we define the Laplace Transform of x as the function X for imaginary numbers such that

$$X(2\pi \mathbf{i}\xi) = \hat{x}(\xi), \xi \in \mathbb{R}$$

and in the same way, the Laplace Transform of y is

$$Y(2\pi\mathbf{i}\xi) = \hat{y}(\xi),\ \xi \in \mathbb{R}.$$

Following again the Fourier transform properties, we express the Laplace Transform of the derivative of the output

$$2\pi\mathbf{i}\xi.Y(2\pi\mathbf{i}\xi) = \widehat{\frac{dy}{dt}}(\xi),\ \xi \in \mathbb{R}.$$

so that we obtain that the input output relationship can be writtern for imaginary number $p = 2\pi\mathbf{i}\xi$ as

$$Y(p) = \frac{1}{1+\tau p}X(p).$$

The rational fraction

$$H(p) = \frac{1}{1+\tau p}$$

is called the transfer function of the system, and we easily get from this definition that H is the Laplace Transform of h, since

$$H(2\pi\mathbf{i}\xi) = \frac{1}{1+\tau 2\pi\mathbf{i}\xi} = \hat{h}(\xi).$$

This study of continuous time systems using the Laplace Transform can be generalized to arbitrary linear differential equation. Although we assumed the derivative of the output was in L^1 to do the above reasoning, we will see in the next chapter that the Fourier-derivation rule can be generalized to a space of distribution (extending the concept of function), much more general than L^1, such that the derivative in the sense of distribution can be apply to the inputs and outputs without having to worry about these details. Without entering into the details here, we can just formalize the derivative by defining the Laplace Transform of a signal $x(t)$ as a function of a complex variable p (not necessarily imaginary)

$$x(t) \mapsto \mathscr{L}[x(t)] = X(p).$$

and stating that the time derivative of x has then the Laplace transform.

$$\mathscr{L}\left[\frac{dx}{dt}\right] = pX(p).$$

The precise definition of the Laplace Transfrom will not be investigated here, but we just need to know that it is a linear transformation, and that

$$\mathscr{L}[h*x](p) = H(p)X(p),$$

such that in the same way as it is done for the Fourier transform, the input-output relationship of an LTI system can be expressed as a multiplication. Using these properties, the general expression of the input-output relationship of the differentiall equation

$$\sum_{k=0}^{M} a_k \frac{d^k y}{dt^k} = \sum_{k=0}^{N} b_k \frac{d^k x}{dt^k}.$$

in the Laplace domain is

$$\sum_{k=0}^{M} a_k p^k Y(p) = \sum_{k=0}^{N} b_k p^k X(p)$$

leading to the transfer function of the system take the form of the rational fraction

$$\frac{Y(p)}{X(p)} = \frac{\sum_{k=0}^{N} b_k p^k}{\sum_{k=0}^{M} a_k p^k}.$$

This formalism is very used in the analysis of continuous time systems as it provides mathematical tools to compute the output of a system given its input, however we will not elaborate on that during this course. As many books and paper describe systems in the Laplace domain, we have to keep in mind that we convert Laplace transforms to Fourier transforms using the change of variables $p = 2\pi\mathbf{i}\xi$.

3 Digital filters

3.1 Difference equations

In the realm of discrete time signals, the time derivatives used in differential equations have to be replaced by something relevant to sequences to describe LTI systems. Although there is absolutely no direct connection between them, somewhat natural replacement it to exchange k-th order derivative by a time lag if k samples in the discription of the input output relationship, so that we get the general form of a difference equation

$$\sum_{k=0}^{M} a_k y[n-k] = \sum_{k=0}^{N} b_k x[n-k].$$

From this equation, we can easily deduce that systems described by such equation are LTI (note the coefficients a_k and b_k are constant). This equation can also be rewritten as a formula to evaluate the output at time n from present and past input and output values

$$y[n] = \frac{1}{a_0}\left(\sum_{k=0}^{N} b_k x[n-k] - \sum_{\mathbf{k=1}}^{M} a_k y[n-k]\right).$$

such that causality of the system becomes clear. Note that in the previous equation, a_0 can be set to one without loss of generality.

3.2 Example of the numerical derivative

Consider now as a first example the difference equation

$$y[n] = x[n] - x[n-1]$$

This is clearly a difference equation as defined above where the only non-zero coefficients are

$$b_0 = 1, b_1 = -1, a_0 = 1.$$

Intutively, this filter can be seen (up to a multiplicative coefficient) as an approximation of the time derivative slowly varying continuous signal. Indeed, assume a signal $x(t)$ is sampled at a rate of $\Delta t = 1$ second, if the derivative is approximately constant between two samples, we have

$$x(t) \approx x(t - \Delta t) + \Delta t . \frac{dx}{dt}(t)$$

such that by sampling at $t = n.\Delta t = n$ we get

$$\frac{dx}{dt}(n) \approx x(n) - x((n-1)).$$

To study the filter, we can compute the impulse response of the filter. In this straightforward example, the impulse response appears directly in the difference equation, which appears as a convolution product for which only two terms are non-zero. We thus easily conclude that

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

Alternatively, this formula can be found by using a discrete Dirac as input and compute the output by applying repetitively the difference equation.

Once we know the impulse response, we can infer the frequency response of the filter by simply computing the Fourier transform of h.

$$\tilde{h}(\nu) = 1 - \exp(-\mathbf{i} 2\pi\nu).$$

Such that the square modulus is

$$|\tilde{h}|^2(\nu) = \tilde{h}(\nu)\overline{\tilde{h}}(\nu) = 2(1 - \cos(2\pi\nu)).$$

From this expression, we see that this filter tends to enhance high frequencies ($|\tilde{h}|^2(1/2) = 4$) while low frequencies are attenuated ($|\tilde{h}|^2(0) = 0$). This corresponds to the expected behavior of a derivative, which vanished for constant signals but gets large we the input signal exhibits rapid changes.

3.3 Example of the first order recurrence equation

Consider now as a first example the difference equation

$$y[n] = \alpha y[n-1] + x[n].$$

This is clearly a difference equation as defined above where the only non-zero coefficients are (beware of the minus sign!)

$$b_0 = 1, a_1 = -\alpha, a_0 = 1.$$

To study this system, we can evaluate the impulse response by using a discrete Dirac as input. Repetitive application of the difference equation results in the output impulse response

$$h[n] = \begin{cases} \alpha^n, & n \geq 0 \\ 0 & \text{otherwise} \end{cases}.$$

which is a causal impulse response (vanishing for negative times). We observe that the system in not BIBO stable for $|\alpha| \geq 1$, so we will restrict ourselves to the case $-1 < \alpha < 1$. We observe that the largest is $|\alpha|$, the longest in duration is the impulse response, in the sense that it will take much more time to get close to zero. For α positive, the shape of the impulse response is simply monotonously exponentially decaying, while for negative α, the impulse response alternates between positive and negative values. The frequency response of the filter is then given by

$$\tilde{h}(\nu) = \sum_{n=0}^{+\infty} \alpha^n \exp(-\mathbf{i}2\pi n\nu).$$

Noticing the terms of the sum form a geometric progression of common ration $\alpha\exp(-\mathbf{i}2\pi\nu)$, we get (using $|\alpha| < 1$)

$$\tilde{h}(\nu) = \frac{1}{1-\alpha\exp(-\mathbf{i}2\pi\nu)}.$$

The modulus square of the frequency response is then

$$|\tilde{h}|^2(\nu) = \tilde{h}(\nu)\overline{\tilde{h}}(\nu) = \frac{1}{1-\alpha\exp(-\mathbf{i}2\pi\nu)}\frac{1}{1-\alpha\exp(+\mathbf{i}2\pi\nu)} = \frac{1}{1+\alpha^2-2\alpha\cos(2\pi\nu)}$$

We see that when α is positive, the filter tends to be low pass, while it is high-pass for negative α.

3.4 General description using the z-Transform

In the same we the Laplace transform was defined, we can define the z-transform by extending the Discrete Time Fourier Transform to the complex plane using a change of variables. Formally, we can define the Fourier transform of the discrete signal x as the (Laurent) series

$$\mathscr{Z}[x] = X(z) = \sum_{k=-\infty}^{+\infty} x[k]z^{-k}, z \in \mathbb{C}$$

We thus see that whenever the DTFT $\tilde{x}$ is well defined we a connection to the z-Transform using the change of variable $z = e^{2\pi\mathbf{i}\nu}$

$$\tilde{x}(\nu) = \sum_{k=-\infty}^{+\infty} x[k]\exp(2\pi\mathbf{i}\nu)^{-k} = X(e^{2\pi\mathbf{i}\nu}).$$

As a consequence of the definition, the z-Transform of the a k_0-laged signal is

$$\mathscr{Z}[x(n-k_0)] = z^{-k_0}X(z).$$

which extend the shifting rule of the DTFT. In addition, the z-Transform also extend the discrete convolution rule of the DTFT, such that

$$\mathscr{Z}[h*x] = H(z).X(z), \text{ with } H(z) = \mathscr{Z}[h]$$

such that an LTI system with impulse response h can be described as multiplication of the input z-Transform by a *transfer function* $H(z)$, the z-transform of the impulse response.

The general form of the transfer function for the above difference equation can be easily deduced from these elementary properties by applying the transform to both sides of the quality

$$\sum_{k=0}^{M} a_k z^{-k} Y(z) = \sum_{k=0}^{N} b_k z^{-k} X(z).$$

leading to

$$\frac{Y(z)}{X(z)} = H(z) = \frac{\sum_{k=0}^{N} b_k z^{-k}}{\sum_{k=0}^{M} a_k z^{-k}}.$$

The z-Transform formalism can be used to analyze the frequency response and impulse response of the filter.

EXAMPLE 37. If we take our previous example of difference equation

$$y[n] = \alpha y[n-1] + x[n].$$

The z-Transform allows us to quickly retrieve the frequency response of the system. Indeed the z-Transfrom of the difference equation leads to

$$\frac{Y(z)}{X(z)} = \frac{1}{1-\alpha z^{-1}}$$

and by applying the change of variable $z = e^{2\pi \mathbf{i} \nu}$, we get the expected frequency response.

4 Introduction to filter design

4.1 Filter stability

When designing a filter, an important constraint is BIBO stability (otherwise the output of the filter would simply not be usable). The filters based on differential equations (resp. difference equations) have a transfer function that take the form of a rational fraction in p (resp. in z). And we can characterize the stabilities of these filters based on the poles of the rational fraction (the zeros of the denominator).

- a continuous time filter based on a differential equation with transfer function $H(p)$ is stable if and only if all its poles p_k verify $\mathrm{Re}(p_k) < 0$
- a discrete time filter based on a difference equation with transfer function $H(z)$ is stable if and only if all its poles z_k verify $|z_k| < 1$.

Implication: stability of FIR filters.

4.2 Analog filter design

Traditionally, analog filters are designed based on a template that specifies the constraints put on the filter. Then a filter is chosen from a particular family of filters, in order to fit in the template. In analog filter design, the classical unit is the pulsation (typically in rad/s) instead of the frequency (typically in Hz). The pulsation variable ω is simply defined from the frequency variable ξ as

$$\omega = 2\pi\xi \quad ,$$

which simplifies the writing of transfer functions.

4.2.1 The canonical case of low-pass filters

Templates are usually defined for low pass filters, and then a transformation is applied to generate other types of filters. Figure 4.1 shows such a template, where ω_c indicates the cut-off pulsation (maximum pulsation that the filter should let pass) and ω_s indicates the stopband pulsation (minimum pulsation that the filter should suppress).

To fit in such a template, one can choose the filter among several families: Butterworth, Tchebytchev, Bessel... Each of these families favors different expected properties of the filter. The family of Butterworth filters is such that,

$$|H(\mathbf{i}\omega)|^2 = \frac{1}{1+(\omega/\omega_c)^{2n}}$$

where n indicates the order of the filter. One beneficial property of these filters is their "maximal flattness": the gain is monotonically decreasing and the all derivatives of the Gain up to $n-1$ order vanish at zero frequency.

4.2.2 From low-pass to other classical filters

There is an automatic way to design other filters based on low-pass templates. First a normalized low pass template can be defined by normalizing the frequency axis by reparameterizing by $\Omega = \omega/\omega_c$. This leads to a normalized transfer function H(p') (that we can compute by reparametrizing by $p' = p/\omega_c$). Then the transfer function H(p) of other types of filters satisfying their own template can be set by applying the following transformations:

- normalized lowpass to highpass: $p' = \omega_c/p$
- normalized lowpass to bandpass: $p' = \frac{p^2+\omega_L.\omega_H}{p(\omega_H-\omega_L)}$
- normalized lowpass to bandstop (notch): $p' = \frac{p(\omega_H-\omega_L)}{p^2+\omega_L.\omega_H}$

where ω_c is the cutoff frequency of the highpass filter, ω_L and ω_H are the lower and upper cutoff frequencies of the bandpass and notch filters.

4.3 From analog to digital IIR filters

As the theory and methodology of analog filter design is well established, it is common to design digital filters based by designing and analog filter and then using a procedure to convert

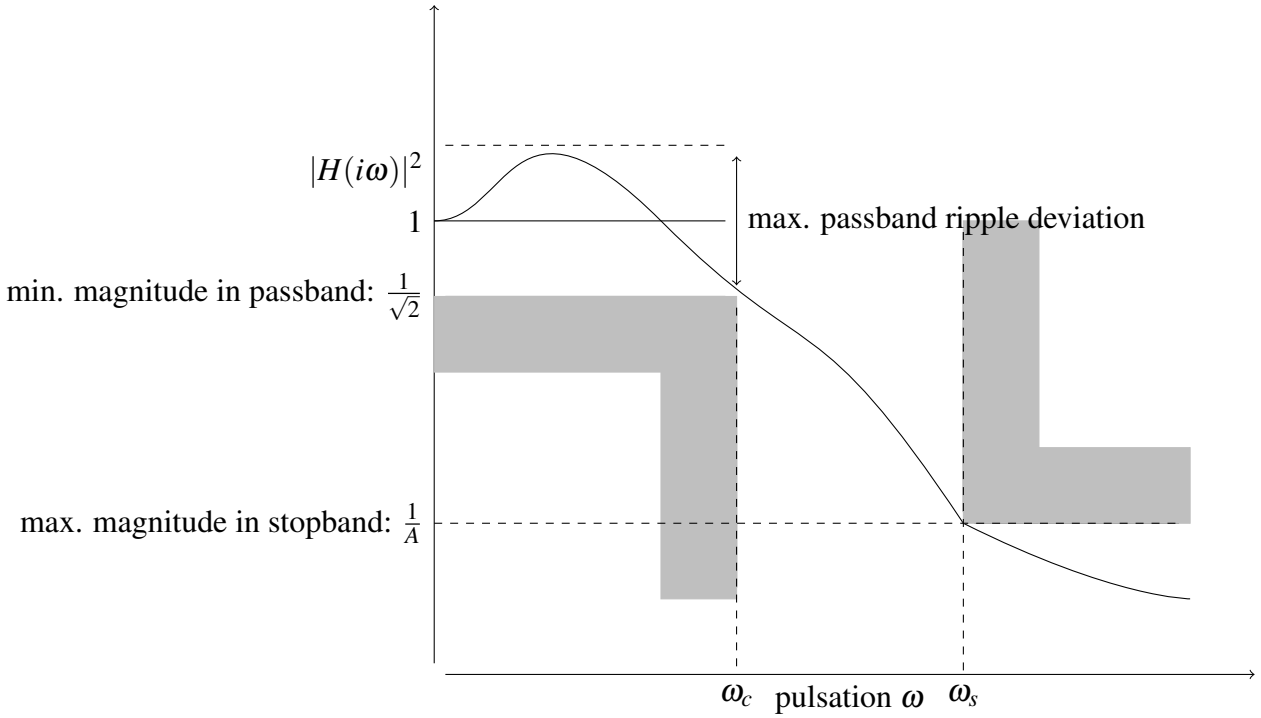

FIGURE 4.1. Design template of a low pass filter.

it to a digital filter. A classical way of doing this conversion is the bilinear transform: given the transfer function of the analog filter $H(p)$, the transfer function $G(z)$ of the corresponding digital filter, approximating the result of an analog filter sampled at period T, is computing by applying the change of variable

$$p = \frac{2}{T}\frac{1-z^{-1}}{1+z^{-1}} \quad .$$

One benefit of the bilinear transformation is that it preserves the stability properties of the filter, as in maps the left half-plane of the Laplace domain to the unit disk of the Z-domain.

Chapter 7

Advanced concepts in filtering

We have seen earlier that distributions extend the concept of function in an elegant and useful way. This enabled us for example to define the derivative of a discontinuous function in a simple and intuitive way. It would be also useful to be able to use distributions for the Fourier analysis of signals and systems. For this purpose, we need to generalize the concepts of Fourier transform and convolution to distributions. Such generalization is possible but requires additional definitions, as the previously introduced space of distributions $\mathscr{D}'$ needs to be restricted to define the Fourier transform properly.

1 Refresher and new distributions

We have seen a distribution $T \in \mathscr{D}'$ is defined as a continuous linear functional acting on test functions belonging to $\mathscr{D}$, the space of C^∞ functions with bounded support such that

$$T: \begin{array}{ccc} \mathscr{D} & \to & \mathbb{C} \\ \varphi & \mapsto & \langle T, \varphi \rangle \end{array}.$$

1.1 Operations on distributions

We have seen that we can compute the derivative of distributions by extending properties of regular distributions. In the same way, we can define other operations on distribution by "transferring" the operation to the test function such that the definition matches what is obtained when applying the operation to regular distributions. In this way, we get the following definitions for

- time shifting/translation: if we denote $\tau_a f = f(t-a)$ for functions, we define

$$\langle \tau_a T, \varphi \rangle = \langle T, \tau_{-a} \varphi \rangle.$$

- mirroring: if we denote $f_\sigma : t \mapsto f(-t)$ for functions, we define
$$\langle T_\sigma, \varphi \rangle = \langle T, \varphi_\sigma \rangle.$$
- multiplication by a C^∞ function: the multiplication is just applied to the test function
$$\langle f.T, \varphi \rangle = \langle T, f.\varphi \rangle.$$

EXAMPLE 38. The shifted Dirac that we will abusively denote $\delta(t-a)$, is defined as
$$\delta_a = \tau_a \delta.$$
Using the definition of distribution translation when applying a test function, we get
$$\langle \tau_a \delta, \varphi \rangle = \langle \delta, \tau_{-a}\varphi \rangle = \langle \delta, \varphi(t+a) \rangle = \varphi(a).$$
Thus instead of returning the value of the test function in 0, the shifted Dirac returns the value in a as expected (i.e. where the Dirac is located).

1.2 A distribution based on the Cauchy principal value

Since the inverse function $1/t$ is not locally integrable due to the divergence of the integral in 0, it is not possible to consider it as a regular distribution. However, we will see that it is still possible to define a distribution that in some sense is the inverse of t. While $1/t$ is the derivative of $\log(|t|)$ is the classical sense of functions, we define a new distribution: $p.v.(\frac{1}{t})$ as the derivative of $\log(|t|)$ in the sense of distributions. Note that this is a proper definition as $\log(|t|)$ is a regular distribution, since log is locally integrable because the integrals in 0 both on the left an right-hand side converge (use its primitive $t\log t - t$ to check it).

We can actually show that
$$\langle p.v.(\frac{1}{t}), \varphi \rangle = \lim_{\varepsilon \to 0^+} \int_{|t|>\varepsilon} \varphi(t)/t \, dt.$$
which justifies the name principal value, as the result is the Cauchy principal value of the improper integral $\int \varphi(t)/t\,dt$, i.e. the limit of the integral when the integration set approaches 0 from above and below at the same speed.

PROOF. We just apply the definition of this distribution using a test function φ
$$\langle p.v.(\frac{1}{t}), \varphi \rangle = \langle \log(|t|)', \varphi \rangle = -\langle \log(|t|), \varphi' \rangle = -\int \log(|t|)\varphi'(t)dt$$
using integration by parts we get (note that φ has finite support and thus vanishes at infinity)
$$\begin{aligned}
\langle p.v.(\frac{1}{t}), \varphi \rangle &= -\lim_{\varepsilon \to 0^+} \int_{|t|>\varepsilon} \log(|t|)\varphi'(t)dt \\
&= \lim_{\varepsilon \to 0^+} \left\{ -[\log(|t|)\varphi(t)]_\varepsilon^{+\infty} - [\log(|t|)\varphi(t)]_{-\infty}^{-\varepsilon} + \int_{|t|>\varepsilon} \frac{\varphi(t)}{t} dt \right\} \\
&= \lim_{\varepsilon \to 0^+} \left\{ \log(\varepsilon)\varphi(\varepsilon) - \log(\varepsilon)\varphi(-\varepsilon) + \int_{|t|>\varepsilon} \frac{\varphi(t)}{t} dt \right\}
\end{aligned}$$

Since $\varphi$is smooth it has a derivative in 0 such that when $\varepsilon \to 0$, $\log(\varepsilon)\left[\varphi(\varepsilon)-\varphi(-\varepsilon)\right] \to \log(\varepsilon)\left[2\varepsilon\varphi'(\varepsilon)\right] \to 0$, which leads to the expected result. □

We can see that this distribution corresponds to the inverse of t in some sense due to the following result.

PROPOSITION 39. *The product distribution* $t.\left[p.v.(\frac{1}{t})\right]$ verifies

$$t.\left[p.v.(\frac{1}{t})\right] = \mathbf{1}$$

where $\mathbf{1}$ stands for the constant function of value 1.

PROOF. Evaluating its value on a test function φ we get

$$\langle t\,.\,p.v.(\frac{1}{t}),\varphi\rangle = \langle p.v.(\frac{1}{t}),t.\varphi\rangle = \lim_{\varepsilon\to 0^+}\int_{|t|>\varepsilon} t\frac{\varphi(t)}{t}\,dt = \int 1.\varphi(t)\,dt = \langle T_{\mathbf{1}},\varphi\rangle.$$

where $T_{\mathbf{1}}$ is simply the regular distribution corresponding to the constant function $\mathbf{1}$. Using the bijection between regular distributions and locally integrable function we thus get the equality

$$t.\left[p.v.(\frac{1}{t})\right] = \mathbf{1}$$

in the sense of distributions. □

2 Tempered distributions

Let us guess what could be a formula for the Fourier transform of a distribution using the regular distribution T_f associated to a locally integrable function f

$$\left\langle T_f,\varphi\right\rangle = \int_{\mathbb{R}} f(x)\varphi(x)dx.$$

PROPOSITION 40. *For* $f,g\in L^1(\mathbb{R})$*, then* $f.\hat{g}$ *and* $\hat{f}.g$ *are in* $L^1(\mathbb{R})$ *and*

$$\int_{\mathbb{R}} f(t)\hat{g}(t)dt = \int_{\mathbb{R}} \hat{f}(\xi)g(\xi)d\xi$$

This result is based on Fubini's Theorem stating that we can permute the order of integration of double integrals.

THEOREM 41 (Fubini). *If $f : \mathbb{R}\times\mathbb{R} \to \mathbb{C}$ is measurable then (with all terms possibly being infinite)*

$$\int_{R\times R} |f(x,y)|dxdy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |f(x,y)|dy \right) dx = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |f(x,y)|dx \right) dy$$

If $\int_{R\times R} |f(x,y)|dxdy < +\infty$, i.e. f is integrable, we have

$$\int_{R\times R} f(x,y)dxdy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y)dy \right) dx = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y)dx \right) dy$$

PROOF. $g \in L^1(\mathbb{R})$ implies that $\hat{g}$ is bounded, thus $f.\hat{g} \in L^1(\mathbb{R})$, and the same reasoning holds for $\hat{f}.g$. The remaining of the proof stems from Fubini's theorem using the function $h(t,\xi) = f(t)g(\xi)\exp(-2\mathbf{i}\pi t\xi)$. □

The consequence of the above proposition is that a natural way to extend the Fourier transform of distributions should be that the Fourier transform $\hat{T}$ of T should satisfy

$$\langle \hat{T}, \varphi \rangle = \langle T, \hat{\varphi} \rangle$$

for appropriate test functions φ. However, the problem is that $\hat{\varphi}$does not necessarily belong to $\mathscr{D}$. We will replace $\mathscr{D}$ by a larger space of functions $\mathscr{S}$ called the *Schwartz class* which is stable by Fourier transformation.

2.1 The Schwartz class $\mathscr{S}$

We say that a function $f(t)$ *decays rapidly* if

$$\lim_{|t|\to\infty} t^p f(t) = 0, \forall p \in \mathbb{N}.$$

Gaussians are examples of rapidly decaying functions.

DEFINITION 42. The Schwartz class $\mathscr{S}$ is the space of functions such that

$$f \in \mathscr{S} \Leftrightarrow \begin{cases} f \in C^\infty(\mathbb{R}) \\ f^{(k)} \text{ decays rapidly} \end{cases}, \text{ for all } k \in \mathbb{N}$$

Gaussians belong to the Schwartz class. This space is adapted for the Fourier transform as functions are L^1 as well as several of their transformations.

PROPOSITION 43. [**Gasquet and Witomski(1999)**, p. 173]

$$f \in \mathscr{S} \Rightarrow \begin{cases} f^{(k)} \in L^1(\mathbb{R}), & \textit{for all } k \in \mathbb{N} \\ P(t).f(t) \in L^1(\mathbb{R}) & \textit{for any polynomial } P \end{cases}.$$

The interest of $\mathscr{S}$ for the Fourier transform can be summarized by this table.

time domain	$\rightarrow$	frequency domain
Rapidly decaying	$\Leftrightarrow$	C^∞
C^∞	$\Leftrightarrow$	Rapidly decaying

As a consequence, imposing both smoothness and rapidly decaying properties guaranties stability of the space when applying the Fourier transform and we get the following result.

THEOREM 44. [**Gasquet and Witomski(1999)**, p. 173] *$\mathscr{S}$ is Fourier transform invariant:*

$$f \in \mathscr{S} \Rightarrow \hat{f} \in \mathscr{S}$$

2.2 Tempered distributions

The tempered distributions are defined using test functions from the Schwartz space. The notion of continuity on $\mathscr{S}$ is needed and amounts to define convergence in $\mathscr{S}$ as

$$f_n \rightarrow_{\mathscr{S}} 0 \text{ whenever } \lim_{n\rightarrow+\infty} \sup_{\mathbb{R}} |x^p f_n^{(q)}(t)| = 0.$$

This notion of convergence behaves well with respect to several important operations.

PROPOSITION 45. [**Gasquet and Witomski(1999)**, p. 288] $f_n \rightarrow_{\mathscr{S}} 0$ *implies*

- $f_n' \rightarrow_{\mathscr{S}} 0$
- $P(t)f_n(t) \rightarrow_{\mathscr{S}} 0$, *for any ploynomial* P
- $f_n \rightarrow 0$ *in* $L^1(\mathbb{R})$
- $\hat{f}_n \rightarrow_{\mathscr{S}} 0$

DEFINITION 46. The space of tempered distributions $\mathscr{S}'$ is the set of continuous linear functionals on $\mathscr{S}$, i.e. linear functions of the form

$$T : \begin{array}{ccc} \mathscr{S} & \rightarrow & \mathbb{C} \\ \varphi & \mapsto & \langle T, \varphi \rangle \end{array},$$

such that for any sequence $\varphi_n \rightarrow_{\mathscr{S}} 0$ implies $\langle T, \varphi \rangle \rightarrow 0$.

Due to its more restrictive definition, the space of tempered distributions $\mathscr{S}'$ is a subspace of $\mathscr{D}'$. From the stability properties of $\mathscr{S}$ can be derived similar properties for tempered distributions.

PROPOSITION 47. *If T is a tempered distribution*

- $t^p.T$ *is tempered*
- $T^{(k)}$ *is tempered*

One very interesting property of tempered distribution is its ability to describe functions that are not integrable, as long as they do not increase too fast.

DEFINITION 48. f measurable is slowly increasing if it exists an integer N and a positive constant C such that

$$|f(t)| \leq C|1+t^2|^N, \text{ for all } t.$$

PROPOSITION 49. *If f is measurable slowly increasing, f is a tempered distribution.*

Also, sum of Diracs, as long as their amplitude does not increase too fast, are also tempered.

PROPOSITION 50. *Let $T = \sum_{n=-\infty}^{+\infty} y[n]\delta(t-nT)$, with y slowly increasing, then T is tempered.*

An important case is $\Delta_T = \sum_{n\in\mathbb{Z}} \delta(t-nT)$, the Dirac comb of period T.

3 Fourier transform of distributions

We just saw that the family of tempered distribution is actually very large and includes many functions. One big benefit of considering a function being a tempered distribution is that the Fourier transform of tempered distributions can always be defined. This means that we get rid of the requirement for a function to be in L^1 or L^2 we computing the Fourier transform. Of course, in addition to that we have the benefit of being able to compute Fourier transform of distributions that are even not functions.

3.1 Definition and properties

The stability of test functions in $\mathscr{S}$ with respect to the Fourier transform allows us to naturally extend the Fourier transform to tempered distributions as follows.

DEFINITION 51. Let T be a tempered distribution, $\hat{T}$ is the tempered distribution such that for any $\varphi \in \mathscr{S}$,

$$\langle \hat{T}, \varphi \rangle = \langle T, \hat{\varphi} \rangle.$$

Importantly, this definition is compatible with regular distributions for which the Fourier transform in the sense of distributions and in the sense of functions are matching (using Prop. 40).

PROPOSITION 52. *If $f \in L^1$ or L^2*

$$\widehat{T_f} = T_{\hat{f}}.$$

It is also straightforward to check that the linearity of the Fourier transform extends to distributions.

PROPOSITION 53. *The Fourier Transform is linear on $\mathscr{S}'$*

Moreover, Fourier inversion becomes straightforward since one can easily check that

$$\mathscr{F}\left[\mathscr{F}\left[T\right]\right] = T_\sigma$$

This definition of the Fourier transform allows us to easily calculate the Fourier transform of some elementary distributions.

PROPOSITION 54. *We have the following Fourier transforms*

- $\hat{\delta} = \mathbf{1}$
- $\hat{\mathbf{1}} = \delta$

Where $\mathbf{1}$ *denotes the constant function with value 1.*

PROOF. For the first case, we apply the definition of the Fourier transform on a test function φ.

$$\langle \hat{\delta}, \varphi \rangle = \langle \delta, \hat{\varphi} \rangle = \hat{\varphi}(0) = \int 1.\varphi(t)dt.$$

We thus obtain that the Fourier transform is the regular distribution $T_{\mathbf{1}}$, corresponding to a constant function of value 1. For the second case, we can just use

$$\mathscr{F}\left[\mathscr{F}\left[\delta\right]\right] = \delta_\sigma = \delta$$

to get the result. □

3.2 Example: Fourier transform of the Dirac comb

The Dirac comb introduced above is a tempered distribution, so its Fourier transform can be defined. We can actually show that

$$\widehat{\Delta}_a = \frac{1}{a}\Delta_{1/a}.$$

such that the Fourier transform of a Dirac comb is also a Dirac comb. The rigorous proof of this result is beyond the scope of the course, but we can provide an intuitive sketch of it.

PROOF. We can first apply the Fourier transform shifting rules to each Dirac of the sum

$$\widehat{\Delta}_a = \sum_{n\in\mathbb{Z}} \widehat{\delta(t-na)} = \sum_{n\in\mathbb{Z}} \exp(-2\pi \mathbf{i} na\xi).$$

In addition, we notice that the Dirac comb is a peridic distribution of period a. Although we did not study the theory of periodic distribution, let us assume that such distribution can be developed using its Fourier series, such that

$$\Delta_a = \sum_{k\in\mathbb{Z}} c_k \exp(2\pi \mathbf{i} kt/a)$$

with the Fourier coefficients

$$c_k = \frac{1}{a}\int_{-\frac{1}{2a}}^{\frac{1}{2a}} \Delta_a(t).\exp(-2\pi \mathbf{i} kt/a)dt.$$

We notice, that on the interval of integration, only the Dirac in zero is not vanishing, such that

$$c_k = \frac{1}{a}\int_{-\frac{1}{2a}}^{\frac{1}{2a}} \delta(t).\exp(-2\pi \mathbf{i} kt/a)dt = \frac{1}{a}.$$

We conclude that

$$\Delta_a = \frac{1}{a}\sum_{k\in\mathbb{Z}} \exp(2\pi \mathbf{i} kt/a).$$

Taking another look at the Fourier transform we finally obtain

$$\widehat{\Delta}_a = \sum_{n\in\mathbb{Z}} \exp(-2\pi \mathbf{i} na\xi) = \sum_{n\in\mathbb{Z}} \exp(2\pi \mathbf{i} na\xi) = \frac{1}{a}\Delta_{1/a}.$$

□

3.3 Calculation rules

There is a direct generalization of classical calculation rules for the Fourier transform. Given a tempered distribution T

- $\hat{T}^{(k)} = \widehat{\left[(-2\mathbf{i}\pi t)^k.T\right]}$
- $\widehat{T^{(k)}}(\xi) = (2\mathbf{i}\pi\xi)^k.\hat{T}$
- $\tau_a\hat{T} = \widehat{\left[e^{2\mathbf{i}\pi at}.T\right]}$
- $\widehat{\tau_a T} = e^{-2\mathbf{i}\pi a\xi}.\hat{T}$

Note that the third result in this list applied to a constant implies rigorously that the Fourier transform of a sinusoid at frequency f_0 is a sum of two Diracs at frequencies $f_0$and $-f_0$!

3.4 Convolution and Fourier Transform

The convolution can also be generalized to distributions based on regular distribution. Assume $f, g \in L^1$, the regular distribution T_{f*g} verifies

$$\begin{aligned}\langle T_{f*g}, \varphi\rangle &= \int f*g(t)\varphi(t)dt = \int\left(\int f(t-\tau).g(\tau)\,d\tau\right)\varphi(t)dt \\ &= \int\left(\int f_\sigma(\tau-t)\varphi(t)dt\right)g(\tau)\,d\tau = \langle T_g, \varphi * f_\sigma\rangle.\end{aligned}$$

As a consequence, on can define the convolution of a distribution T with a function f as

$$\langle f*T, \varphi\rangle = \langle T, \varphi * f_\sigma\rangle.$$

EXAMPLE 55. Let us consider the shifted Dirac

$$\delta_a = \tau_a\delta$$

For $f \in \mathscr{S}$, $f*\delta_a = f(t-a)$.

PROOF. Using the definition of convolution for distributions

$$\begin{aligned}\langle f * \delta_a, \varphi \rangle &= \langle \delta_a, \varphi * f_\sigma \rangle = \varphi * f_\sigma(a) = \int \varphi(\tau) f_\sigma(a-\tau) d\tau \\ &= \int \varphi(\tau) f(\tau - a) d\tau = T_{f(t-a)}\end{aligned}$$

□

In the case of tempered distribution, we then have again a Fourier-convolution result.

PROPOSITION 56. *If $\psi \in \mathscr{S}$ and $T \in \mathscr{S}'$ then*

$$\widehat{\psi * T} = \hat{\psi}.\hat{T}$$

and

$$\widehat{\psi.T} = \hat{\psi} * \hat{T}.$$

4 Application 1: Sampling

4.1 Idealized sampling of a continuous time signal

(Regular) Sampling is the operation of converting a continuous time signal into a discrete time signal by keeping only the values on a periodic grid with period T leading to $x_d[k] = x(kT)$.

The idealized sampled signal is defined as

$$\uparrow_T x = \sum_{k\in\mathbb{Z}} x(kT)\delta(t-kT) = \sum_{k\in\mathbb{Z}} x(kT)\delta_{kT}(t)$$

Note that whenever it is well defined, using Fourier-shifting rules

$$\widehat{\uparrow_T x}(\xi) = \sum_{k\in\mathbb{Z}} x(kT)\widehat{\delta_{kT}(t)} = \sum_{k\in\mathbb{Z}} x(kT)e^{-\mathbf{i}2\pi kT\xi} = \widetilde{x_d}(\nu = \xi T).$$

which justifies a posteriori the definition of the idealized sampling, since its continuous time Fourier transform corresponds (up to as scaling of the frequency axis) to the Discrete Time Fourier Transform of the sequence resulting from sampling the signal.

4.2 Nyquist criterion

In addition, we can see that idealized sampling just amounts to multiplying by a Dirac comb in the time domain

$$\uparrow_T x = x(t).\Delta_T.$$

Thus if x is regular enough, we can apply the Fourier-convolution rules

$$\widehat{\uparrow_T x}(\xi) = \frac{1}{T}.\hat{x} * \Delta_{1/T}.$$

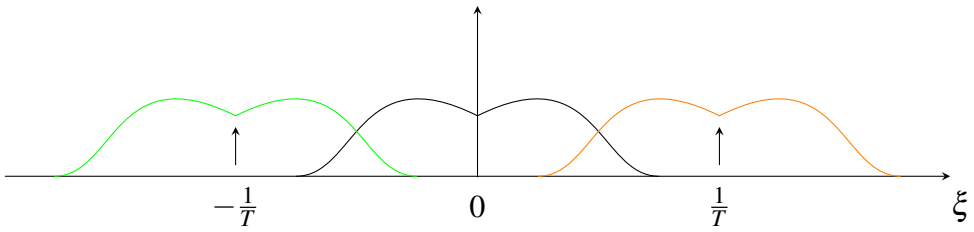

FIGURE 4.1. Illustration of the effect of sampling. The original Fourier transform (in black) curve is periodized by adding copies (in color) of itself shifted at multiples of the sampling frequency $\frac{1}{T}$, such that these shifted spectra can overlap.

By using the definition of the Dirac comb and the shifted Dirac convolution studied above, we get

$$\widehat{\uparrow_T x}(\xi) = \frac{1}{T}.\sum_{k\in\mathbb{Z}} \hat{x} * \delta_{k/T} = \frac{1}{T}.\sum \hat{x}(\xi - k/T)$$

which shows that shifted "images" of the original Fourier transform appear centered at the multiples of the sampling frequency $\frac{1}{T}$. In the general case, these images can overlap with the original Fourier transform (see illustration Fig. 4.1), such that information about the original continuous time signal is lost.

An natural consequence of the above equality is that we can establish a condition called the Nyquist criterion, such that no information is lost during sampling: When we sample a continuous signal at period T, all the information of the original Fourier transform is preserved (in the interval $[-\frac{1}{2T}, \frac{1}{2T}]$) whenever the original signal was band-limited such that $\hat{x}(\xi)$ vanishes for $|\xi| \leq \frac{1}{2T}$. $f_n = \frac{1}{2T}$ is called the Nyquist frequency. An illustration of a satisfied Nyquist criterion is shown on Fig. 4.2. A practical consequence of this result is that hardware systems measuring signals usually include an analog low-pass filter with a cut-off frequency below the Nyquist frequency before converting the analog signal to a numeric signal.

5 Application 2: Hilbert transform

Another important application where distributions are useful is the Hilbert transform, which is used in practice to extract the instanteneous phase and amplitude of a signal.

5.1 Definition

Let x be a continuous signal

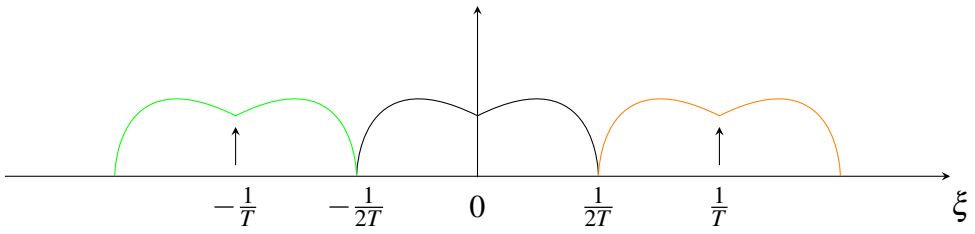

FIGURE 4.2. Illustration of the Nyquist criterion. If the original Fourier transform (in black) is band limited such that its maximum frequency is $< \frac{1}{2T}$, the Nyquist frequency, then the shifted images of the original Fourier transform do not overlap and no frequency information is lost.

$$\mathscr{H}(x)(t) = \frac{1}{\pi} p.v. \int \frac{x(\tau)}{t-\tau} dt.$$

It can thus be seen as a filter implementing the convolution by a distribution

$$\mathscr{H}(x) = \frac{1}{\pi} p.v.(\frac{1}{t}) * x.$$

5.2 Fourier transform of $p.v.(\frac{1}{t})$

It can be proven that $p.v.(\frac{1}{t})$ is a tempered distribution, such that its Fourier transform can be defined. Assuming this result, it is possible to show that this Fourier transform is

$$\widehat{p.v.(\tfrac{1}{t})}(\xi) = -\mathbf{i}\pi.\,\mathrm{sign}(\xi).$$

PROOF. We provide here just a sketch of the proof. First, we note that the distribution $t.\left[p.v.(\frac{1}{t})\right]$, is tempered as product of a tempered distribution with a polynomial and we proved before the equality

$$t.\left[p.v.(\frac{1}{t})\right] = \mathbf{1}$$

in the sense of distributions. Now if we compute the Fourier transform of the derivative of $p.v.(\frac{1}{t})$ using Fourier-derivation rules, we get

$$\frac{d}{d\xi}\left[\widehat{p.v.(\tfrac{1}{t})}(\xi)\right] = \widehat{(-2\pi\mathbf{i}t).p.v.(\tfrac{1}{t})} = (-2\pi\mathbf{i}).\hat{\mathbf{1}} = (-2\pi\mathbf{i}).\delta(\xi).$$

If we integrate this derivative (with respect to ξ), although we did not study the integration of distribution, let us assume that they behave similarly to functions, such that

$$\widehat{p.v.(\tfrac{1}{t})}(\xi) = (-2\pi\mathbf{i}).\int \mathbf{1} d\xi + \lambda = (-2\pi\mathbf{i}).u(\xi) + \lambda,$$

where λ is a constant to be determined and u is the Heaviside function. The constant is determined by the fact that $p.v.(\frac{1}{t})$ is odd, implying that its Fourier transform is odd, leading to

$$\widehat{p.v.(\tfrac{1}{t})}(\xi) = (-\pi\mathbf{i}).\text{sign}(\xi).$$

□

Whenever the convolution rule holds (for example for $x \in \mathscr{S}$), we get the effect of the Hilbert transform in the Fourier domain (note the $1/\pi$ rescaling of the distribution)

$$\widehat{\mathscr{H}[x]}(\xi) = -\mathbf{i}\,\text{sign}(\xi).\hat{x}(\xi).$$

From this Fourier domain expression, we can also deduce that by applying twice the Hilbert transform we get

$$\mathscr{F}\{\mathscr{H}[\mathscr{H}[x]]\}(\xi) = \mathbf{i}^2\,\text{sign}^2(\xi).\hat{x}(\xi) = -\hat{x}(\xi),$$

such that

$$\mathscr{H}[\mathscr{H}[x]] = -x.$$

5.3 Application to sinusoids

Assuming the we can apply the Fourier domain expression to sinusoids, we can evaluate the Hilbert transform if a sine wave a positive frequency f_0 as follows

$$\mathscr{H}[\cos(2\pi f_0 t + \phi)] = \frac{1}{2}\left(\mathscr{H}\left[e^{\mathbf{i}2\pi f_0 t + \mathbf{i}\phi}\right] + \mathscr{H}\left[e^{-\mathbf{i}2\pi f_0 t - \mathbf{i}\phi}\right]\right)$$

since both complex exponentials have a Dirac Fourier transform, one in the positive frequencies and the other in the negative frequencies, we obtain

$$\begin{aligned}\mathscr{H}[\cos(2\pi f_0 t + \phi)] &= \frac{1}{2}\left(-\mathbf{i}.e^{\mathbf{i}2\pi f_0 t + \mathbf{i}\phi} + \mathbf{i}.e^{-\mathbf{i}2\pi f_0 t - \mathbf{i}\phi}\right) = \frac{1}{2\mathbf{i}}\left(e^{\mathbf{i}2\pi f_0 t + \mathbf{i}\phi} - e^{-\mathbf{i}2\pi f_0 t - \mathbf{i}\phi}\right) \\ &= \sin(2\pi f_0 t + \phi).\end{aligned}$$

For this general result we deduce that

$$\mathscr{H}[\cos(2\pi f_0 t)] = \sin(2\pi f_0 t)$$

and

$$\mathscr{H}[\sin(2\pi f_0 t)] = -\cos(2\pi f_0 t)$$

5.4 Bedrosian theorem

On important application of the Hilbert transform is to extract the time varying properties of a phase and/or amplitude modulated sinusoidal signal. Assume we observe a modulated signal

$$x(t) = a(t)\cos(2\pi f_0 t + \phi(t))$$

where $a(t)$ is a time varying amplitude and $\phi(t)$ is a time varying phase that modulates the sinusoid properties. We would like to retrieve a, f_0 and ϕ from the observed signal x. In

principle, if we could generalize the above results that hold for constant amplitude and phase, we could build the complex signal

$$\begin{aligned} x(t)+\mathbf{i}\mathscr{H}(x)(t) &= a(t)\cos(2\pi f_0 t+\phi(t))+\mathbf{i}a(t)\sin(2\pi f_0 t+\phi(t)) \\ &= a(t)\exp(\mathbf{i}2\pi f_0 t).\exp(\mathbf{i}\phi(t)), \end{aligned}$$

such that the amplitude and phase properties of the signal are easily extracted using the modulus and argument of this complex signal. This is however a rather non-trivial question in when amplitude and phase are time varying since the above representation can be ambiguous. For example, in an extreme case we could imagine that $x(t)=a(t)$ and $f_0=0$ and $\phi(t)=0$. Fortunately, the Hilbert transform can help retrieving such representation under some mild assumptions thanks to the following Bedrosian theorem [**Xu and Yan(2006)**].

THEOREM 57. *For $f,g\in L^2$, if $\hat{f}$ vanishes for $\xi>a$ and $\hat{g}$ vanishes for $\xi<a$, then*

$$\mathscr{H}(f(t)g(t))=f(t)\mathscr{H}(g(t)).$$

5.5 Analytic signal

Indeed, using the above theorem, the Hilbert transform can be used to build the so-called *analytic signal* of the modulated signal $x(t)$, defined by the equation

$$x_a(t)=x(t)+\mathbf{i}\mathscr{H}(x)(t).$$

If we assume that our modulated signal is such that the modulating signal $a(t)\exp(\mathbf{i}\phi(t))=a(t)[\cos(\phi(t))+\mathbf{i}\sin(\phi(t))]$ is band limited such that both real and imaginary part of its Fourier Transform vanishes for $\xi\geq f_0$. Let us compute $\mathscr{H}(x)$.

$$\mathscr{H}[a(t)\cos(2\pi f_0 t+\phi(t))]=\mathscr{H}[a(t)\cos(2\pi f_0 t)\cos(\phi(t))-a(t)\sin(2\pi f_0 t)\sin(\phi(t))]$$

using a classical trigonometric identity. Under the band-limited assumption on the modulating signal, we can then apply the Bedrosian theorem, such that

$$\mathscr{H}[a(t)\cos(2\pi f_0 t+\phi(t))]=a(t)\cos(\phi(t))\mathscr{H}[\cos(2\pi f_0 t)]-a(t)\sin(\phi(t))\mathscr{H}[\sin(2\pi f_0 t)].$$

Finally, applying Hilbert transform on the sinusoids and another trigonometric identity, we get

$$\begin{aligned} \mathscr{H}[a(t)\cos(2\pi f_0 t+\phi(t))] &= a(t)\cos(\phi(t))\sin(2\pi f_0 t)+a(t)\sin(\phi(t))\cos(2\pi f_0 t) \\ &= a(t)\sin(2\pi f_0 t+\phi(t)). \end{aligned}$$

As a consequence, the analytic signal has the expected form

$$x_a(t)=a(t)\exp(\mathbf{i}2\pi f_0 t).\exp(\mathbf{i}\phi(t)).$$

Importantly, this analysis emphasizes the fact that we need to satisfy the assumptions of the Bedrosian theorem to be able to extract proper amplitude and phase information from a given signal. Using classical shifting rules of the Fourier transform, we can see that this means that we need to make sure that the Fourier transform of the signal has its support included in a relatively narrow frequency band, which can be thought as centered around the carrier frequency f_0. However, note that the frequency f_0 plays only an auxiliary role in this

analysis, since we can as well modify f_0 by altering $\phi(t)$ with a commensurate linear phase change. In practice, applying the Hilbert transform properly on real data that does not have a narrow band spectrum thus requires a preliminary band pass filtering to avoid ambiguity in the definition of the phase and amplitude.

Chapter 8

Time-varying Fourier analysis

1 Short-Term Fourier Transform

Time frequency analysis methods are motivated by the fact that many empirical signals are non-stationary and exhibit transient activity. In order to make Fourier analysis focus on the local properties of the signal came the idea to select a portion of the signal with a sliding window, and then perform Fourier analysis.

$$STFT\{x(t)\}(\tau,\xi) = \mathbf{X}(\tau,\xi) = \int_{\mathbb{R}} x(t)w(t-\tau)e^{-\mathbf{i}2\pi\xi t}dt$$

where $w(t)$ is a window function.

One shortcoming of this approach is that it imposes a fixed resolution through the window function. This can be seen by rewriting the STFT as a convolution:

$$\mathbf{X}(\tau,\xi) = e^{-\mathbf{i}2\pi\xi t}\int_{\mathbb{R}} x(t)w(t-\tau)e^{-\mathbf{i}2\pi\xi(t-\tau)}dt = e^{-\mathbf{i}2\pi\xi t}x * f_{\tau,\xi}(\tau)$$

with $f_{\tau,\xi}(t) = w(-t)e^{\mathbf{i}2\pi\xi t}$. So, up to a phase shift, the STFT appears as using a family of band-pass filters with fix bandwidth to analyze the input signal.

2 Continuous time wavelet transform

To provide and adaptive resolution, wavelet define a family of function based on a mother wavelet ψ such that

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right), b \in \mathbb{R}, a > 0\,.$$

The wavelet coefficients of a signal x are then

$$C_x(a,b) = \int_{\mathbb{R}} x(t)\,\bar{\psi}_{a,b}(t)dt$$

An example of wavelets originally used was the Morlet wavelet:

$$\psi(t) = e^{-\frac{t^2}{2}} \cos 5t$$

One can see that the family of wavelet does not have a rigid envelope and allows more flexibility with respect to the time resolution of the analysis.

In addition, the wavelet transform appears as a generalization of the Fourier transform in the sense that it comes with a reconstruction formula.

THEOREM 58. *Let ψ be in $L^1(\mathbb{R}) \cap L^1(\mathbb{R})$, satisfying the following conditions:*

(1) $\int |\psi(\hat{\xi})|^2/|\xi| = K < +\infty$.
(2) $\|\psi\|_2 = 1$

Then

$$x(t) = \frac{1}{K} \int \int C_x(a,b)\psi_{a,b}(t) da/a^2 db$$

Moreover we have the following conservation of energy formula.

$$\frac{1}{K} \int \int |C_x(a,b)|^2 da/a^2 db = \int |x(t)|^2 .dt$$

We notice that the condition implies that the mother wavelet is zero mean. An example satisfying the condition is the analytic Morlet wavelet.

$$\psi_\sigma(t) = c + \sigma \pi^{-1/4} e^{-t^2/2} (e^{\mathbf{i}\sigma t} - \kappa_\sigma)$$

Part 4

Stochastic processes

Chapter 9

Introducing randomness in time series

This chapter will be devoted to random signals. After a small refresher on random variables, the basic tools to extended Fourier analysis and Linear Time Invariant systems to these random signals will be introduced.

1 Refresher on probabilities

1.1 Random variables

A random variable (RV) is a measurable function of the form

$$\begin{aligned} X\colon \Omega &\to \mathbb{C}^n \\ \omega &\mapsto X(\omega) \end{aligned}$$

where Ω is a probability space called the set of possible outcomes. The probability is the measure of the probability space Ω, i.e. a function which for a given a measurable subset S of Ω associates the probability $P\{S\} \in [0,1]$, and such that $P(\Omega) = 1$. For a given outcome $\omega \in \Omega$, $X(\omega)$ is called a realization of the random variable.

1.2 Probability density

A random variable X with values in $\mathbb{C}^n$ admits a density p_X whenever for any measurable set $S \subset \mathbb{C}^n$,

$$P\{X \in S\} = \int_S p_X(x)dx.$$

Note that for a real univariate random variable (i.e. with values in $\mathbb{R}$), this implies

$$P\{a \leq X \leq b\} = \int_a^b p_X(x)dx.$$

which also implies

$$p_X(x) = \lim_{\Delta x} \frac{P\{x \leq X \leq x + \Delta x\}}{\Delta x}.$$

In the case of multivariate random variables, we can consider them as a concatenation of univariate variables. The multivariate density is then called the *joint density* of the variables, while the univariate distributions are called the *marginal densities*. For example in the case of two random variables X and Y, $p_{X,Y}$ is the joint density and p_X is the marginal density of X. Note that the definitions of densities implies

$$p_X(x) = \int p_{X,Y}(x,y)dy.$$

such that computing the marginal density amounts to integrate with respect to the variables on their domain of definition (we also say that Y has been marginalized out).

As we will only consider variable admitting a probability density in the remaining of the chapter, many probabilistic quantities will be defined based on this density, while the may have a more general definition. In particular, we can define the *expectation* or *expected value* for any measurable function f of a possibly multivariate random variable.

$$\mathbb{E}[f(X)] = \int f(x)p_X(x)dx.$$

Note that in this definition of expectation, we implicitly assume that we integrate with respect to the joint distribution of all RV's present inside the expectation (this will always be the case for us).

1.3 Independence

Two RV's are said to be independent whenever, for any measurable function f and g (defined on their respective domain)

$$\mathbb{E}[f(X)g(Y)] = \int\int f(x)g(Y)p_{X,Y}(x,y)dxdy = \mathbb{E}[f(X)]\mathbb{E}[g(Y)].$$

We can show that this implies for all values x, y of the RV's that the joint probability density factorizes

$$p_{X,Y}(x,y) = p_X(x)p_Y(y).$$

More generally, we will say that a collection of random variables $\{X_k\}_{k\in[1..K]}$ are jointly independent whenever for all measurable f_k

$$\mathbb{E}[\prod_{k\in[1..K]} f_k(X_k)] = \prod_{k\in[1..K]} \mathbb{E}[f_k(X_k)].$$

which implies again that the joint factorizes in the product of the marginal densities

$$p_{X_1,..,X_K}(x_1,..,x_k) = p_{X_1}(x_1)\cdot p_{X_2}(x_2)\cdots p_{X_K}(x_K).$$

2 (Discrete-time) Stochastic processes

A stochastic process is a collection of random variables indexed by time. We will focus on discrete time stochastic processes indexed by integer time indices.

$$\{X[k]\colon k \in \mathbb{Z}\}.$$

This means that for a given time k, $X[k]$ is simply a random variable. From another point of view, given one outcome $\omega \in \Omega$, the realization of X , $\{X[k](\omega)\colon k \in \mathbb{Z}\}$ is simply a deterministic sequence of values. We can imagine one realization is the time course of the signal when we do one single experimental recording. If we do another experiment, even in the exact same conditions, the recorded signal will be different at least due to the measurement noise. The stochastic process is a way to model these fluctuations across experiments, by assuming that they are generated by drawing from a proability distribution (defined by a density in our case). In particular, we will assume that for any collection of time indices of any length $\{t_1,...,t_K\}$, the corresponding collection of RV's observed at these time points admits a joint probability density

$$p_{X[t_1],..,X[t_K]}(x_1,...,x_K).$$

2.1 Stationary processes

A *strictly stationary* stochastic process [**Koopmans(1995)**, page 38] has all its properties that are time invariant. This can be stated by saying that for any collection of time indices of any length $\{t_1,...,t_K\}$, the joint probability density satisfies, for any integer time lag τ

$$p_{X[t_1+\tau],..,X[t_K+\tau]}(x_1,...,x_K) = p_{X[t_1],..,X[t_K]}(x_1,...,x_K).$$

i.e. when translating all time indices with the same time lag, the probability density stays the same. This assumption is quite strong, and in many cases we can use a weaker definition of stationarity.

A real valued stochastic process is *weakly stationary* or *second order stationary* whenever for any time point t and lag τ

$$\mathbb{E}[X[t]] = \mathbb{E}[X[t+\tau]] = \mu_X < +\infty$$

and

$$\mathbb{E}[X[t]X[t+\tau]] = \mathbb{E}[X[0]X[\tau]] < +\infty.$$

Hence weak stationarity restricts the time invariance condition to the first and second order statistics of the process (mean and variance/covariance). We can extend this definition to real-valued multivariate stochastic processes by imposing stationarity of the individual components and imposing time invariance of the covariance between each components. For example, a weakly stationary bivariate process (X,Y) satisfies X weakly stationary, Y weakly stationary and for all t,τ integers

$$\mathbb{E}[X[t]Y[t+\tau]] = \mathbb{E}[X[0]Y[\tau]].$$

Finally, a weakly stationary complex-values (and possibly multivariate) process can simply be defined by building a real-valued process that concatenates the real and imaginary parts of the original process.

2.2 Spectral properties of stationary processes

We now consider weakly stationary processes that are possibly complex valued. For a given weakly stationary process X, we can define the auto-correlation function as

$$\gamma_{XX}(\tau) = \mathbb{E}\left[\overline{X}[0] \cdot X[\tau]\right].$$

Note that while this is the conventional name for it in signal processing, this is NOT a correlation that is computed, since the mean of both variables is not removed, and there is no normalization by the standard deviation of the variables. Note also that this terminology is not the only one: some textbooks [**Koopmans(1995)**, page 38] name the same quantity *auto-covariance* while others have a different definition for auto-covariance [**Brockwell and Davis(1991)**, Definition 1.3.1] (where the removal of the mean of both signals is implemented, as in a classical covariance). In the same way, we can define the cross-correlation function between to two jointly weakly stationary processes X, Y as

$$\gamma_{XY}(\tau) = \mathbb{E}\left[\overline{X}[0] \cdot Y[\tau]\right].$$

Note that due to the stationarity of the processes, we have the following symmetry properties for the auto- and cross-correlation

$$\gamma_{XX}(-\tau) = \mathbb{E}\left[\overline{X}[0] \cdot X[-\tau]\right] = \mathbb{E}\left[\overline{X}[\tau] \cdot X[0]\right] = \overline{\gamma_{XX}}(\tau)$$

and

$$\gamma_{XY}(-\tau) = \mathbb{E}\left[\overline{X}[0] \cdot Y[-\tau]\right] = \mathbb{E}\left[\overline{X}[\tau] \cdot Y[0]\right] = \overline{\gamma_{YX}}(\tau).$$

Note in the above expression that the order of the variables is permuted in the last equation.

We then define the Power Spectral Density (PSD) as the DTFT of the auto-correlation

$$S_X(\nu) = \widetilde{\gamma_{XX}}(\nu) = \sum_{k\in\mathbb{Z}} \gamma_{XX}[k] \exp(-2\mathbf{i}\pi k\nu).$$

The PSD is a generalization of the squared modulus of the Fourier transform for deterministic signals. We note that using the symmetry properties of the auto-correlation function we get for the conjugate of the PSD

$$\overline{S_X(\nu)} = \sum_{k\in\mathbb{Z}} \overline{\gamma_{XX}}[k] \exp(2\mathbf{i}\pi k\nu) = \sum_{k\in\mathbb{Z}} \gamma_{XX}[-k] \exp(2\mathbf{i}\pi k\nu) = S_X(\nu),$$

such that S_X is real valued (even if X is complex). Moreover, it is possible to show that the PSD is positive.

An interesting case of stationary process is when the RV's at all time points are jointly independent. Let us assume the real-valued weakly stationary process ε consists in independent identically distributed (i.i.d.) variables such that for all time points k

$$\mathbb{E}[\varepsilon[k]] = 0$$

and

$$\mathbb{E}\left[|\varepsilon[k]|^2\right] = \sigma_\varepsilon^2.$$

Since the RV's at different time points are independent, the auto-correlation is

$$\gamma_{\varepsilon\varepsilon}(k) = \begin{cases} \sigma_\varepsilon^2 & \text{if } k = 0 \\ 0 & \text{otherwise.} \end{cases}$$

As a consequence, the PSD of ε is constant

$$S_\varepsilon(\nu) = \sigma_\varepsilon^2.$$

A signal having such property is called a stationary white noise because its PSD is constant, as a reference to the spectral properties of the white light which contains a wide collection of frequencies (note that for a process to be white noise the i.i.d. property is not necessary: it is enough that the autocorrelation vanished for non-zero lags). The Power Spectral Density inherits its name from the fact that it describes the distribution of the power of the signal over the frequency domain. In particular, if we integrate the PSD over a frequency interval of length 1, we get

$$\int_{-1/2}^{1/2} S_X(\nu)d\nu = \gamma_{XX}(0) = \mathbb{E}\left[|X[0]|^2\right].$$

which is the total power of the signal, i.e. the energy per time sample..

2.3 Filtering of stationary processes

One important question is what happens to a process when it is filtered by an LTI system. As expected by combining the time invariance properties of the filter and the stationary process, the output of the system will also be stationary.

PROPOSITION 59. [**Brockwell and Davis(1991)**, Theorem 4.4.1.] *The output of a stable LTI system (with impulse response $h \in \ell^1(\mathbb{Z})$) for a weakly stationary input is weakly stationary.*

In addition, we can deduce the output PSD from the input PSD using the following result.

PROPOSITION 60. [**Brockwell and Davis(1991)**, Theorem 4.4.2.] *The output of an LTI system of impulse response h with an input of PSD S_X has a PSD such that*

$$S_Y(\nu) = |\widetilde{h}(\nu)|^2 \cdot S_X(\nu)$$

PROOF. We just provide a sketch of the proof. First let's compute the cross-correlation between input and output

$$\begin{aligned} \gamma_{XY}(\tau) &= \mathbb{E}\left[\overline{X}[0] \cdot Y[\tau]\right] = \mathbb{E}\left[\overline{X}[0] \cdot \sum_k h[k]X[\tau-k]\right] \\ &= \sum_k h[k]\mathbb{E}\left[\overline{X}[0] \cdot X[\tau-k]\right] = \sum_k h[k]\gamma_{XX}(\tau-k) \\ &= h * \gamma_{XX} \end{aligned}$$

by linearity of the expectation. Next we compute to output autocorrelation

$$\begin{aligned}\gamma_{YY}(\tau) &= \mathbb{E}\left[\overline{Y}[0]\cdot Y[\tau]\right] = \mathbb{E}\left[\overline{Y}[0]\cdot\sum_k h[k]X[\tau-k]\right] \\ &= \sum_k h[k]\mathbb{E}\left[\overline{Y}[0]\cdot X[\tau-k]\right] = h*\gamma_{YX}.\end{aligned}$$

and as a consequence we get (using the above symmetry properties of the cross-correlation, σ indicating mirroring of the function with respect to 0)

$$\gamma_{YY} = h*\overline{\gamma_{XY\sigma}}.$$

Now if we come back to the definition of Fourier transform, we have

$$\widetilde{\overline{x}_\sigma} = \sum_k \overline{x}[-k]\exp(-2\mathbf{i}\pi\nu k) = \sum_k \overline{x}[k]\exp(2\mathbf{i}\pi\nu k) = \overline{\widetilde{x}}$$

and as a consequence of this result and of the Fourier-convolution rules we

$$S_Y(\nu) = \widetilde{h}(\nu)\cdot\widetilde{\overline{\gamma_{XY\sigma}}}(\nu) = \widetilde{h}(\nu)\cdot\overline{\widetilde{\gamma_{XY}}(\nu)} = \widetilde{h}(\nu)\cdot\overline{\widetilde{h}(\nu)\cdot\widetilde{\gamma_{XX}}(\nu)} = |\widetilde{h}(\nu)|^2 S_X(\nu).$$

□

This result provides a simple way to generate of random signal with a given PSD (and auto-correlation), by simply applying a filter with the desired characteristics to an input white noise.

2.4 Introduction to spectral estimation

The PSD is an important property of stationary processes, such that we are led to estimate it from empirical data. While it is possible to estimate the auto-correlation and then compute the DTFT to get the PSD estimate, a popular family of approach rely on the use of the Fourier transform of the observed data. Since empirically we have only access to N consecutive samples $X_{0,..,N-1}$ of the signal, we can evaluate the modulus square of the DFT F_N of these data points. If we look at the average behavior of the DFT by evaluating the expectation, we get

$$\begin{aligned}\mathbb{E}\left|F_N X_{0,..,N-1}\right|^2(n) &= \mathbb{E}\left|\sum_k X[k]e^{-2\mathbf{i}\pi kn/N}\right|^2 \\ &= \mathbb{E}\sum_l\sum_k X[k]\overline{X}[l]e^{-2\mathbf{i}\pi(k-l)n/N} \\ &= \sum_l\sum_k \mathbb{E}\left[X[k]\overline{X}[l]\right]e^{-2\mathbf{i}\pi(k-l)n/N} \\ &= \sum_l\sum_k \gamma_{XX}(k-l)e^{-2\mathbf{i}\pi(k-l)n/N}\end{aligned}$$

such that using a change of variables, we can sum over the sub-diagonals of the (k,l) table, for which the term of the sum is constant. In this way we get

$$\mathbb{E}|F_N X_{0,..,N-1}|^2(n) = \sum_{k=-N+1}^{N-1} (N-|k|)\gamma_{XX}(k)e^{-2\mathbf{i}\pi kn/N}.$$

If we thus define the periodogram as the squared modulus DFT of the N-samples of X normalized by N

$$P_N(X) = \frac{1}{N}|F_N X_{0,..,N-1}|^2.$$

We see that

$$\mathbb{E}P_N(X) = \sum_{k=-N+1}^{N-1} (1-\frac{|k|}{N})\gamma_{XX}(k)e^{-2\mathbf{i}\pi kn/N}$$

such that when N gets large, the expectation of the periodogram becomes a reasonable approximation of the PSD. Indeed, assume for simplification that the auto-correlation vanishes for large lags $> N_0$, the term $(1-\frac{|k|}{N})$ tends to one for all lags $\leq N_0$ and

$$\mathbb{E}P_N(X)[n] \rightarrow_{N\rightarrow+\infty} S_X(n/N)$$

such that we can estimate the PSD at frequency grid points n/N using the DFT. A simple example where we have a strict equality even for small N is the i.i.d stochastic process ε defined above. For this case we have

$$\mathbb{E}P_N(\varepsilon) = \gamma_{XX}(0) = \sigma_\varepsilon^2.$$

Unfortunately, while the expectation of the periodogram is well behaved, the variance of the estimate does not tend to zero for large N (it actually stays constant!). To reduce the variance of periodogram estimates, several approaches have been developed. The idea behind the Welch periodogram is to average the estimate of consecutive windows, such that fluctuations due to the variance of the estimate average out using the law of large number, such that the overall variance tends to zero when the number of windows gets large.

Chapter 10

Models of stationary processes

1 Introduction

Incorporating randomness in signals allows, to some extent, to capture properties of real-world data:

- "exogenous" (unobserved) influences on the observed signals (e.g. synaptic noise when looking at a neuron's membrane potential...),
- measurement errors.

LTI systems offer a way to design parametric models of the observations that incorporate these factors, such that we can fit observed data and exploit models for e.g. doing forecasts or get insights about the underlying generating mechanisms.

We will go through several selected examples of such models addressing different purposes.

2 Modeling one (stationary) signal

2.1 Innovation representation

We will focus on a simple case (see [**Proakis and Manolakis(2007)**, chapter 12] for a general definition). Assume a weakly stationary (zero mean) process X possesses a PSD

$$S_X(\nu) = \frac{1}{|1-\alpha\exp(-2\pi\mathbf{i}\nu)|^2}\,,\ |\alpha|<1$$

Then we recognize that this can be interpreted as the spectrum of the filtering of a white PSD. We can go further and explicitly compute the white noise as

$$\varepsilon[n] = X[n] - \alpha X[n-1]\,.$$

Then indeed we can regenerate X from ε with the DE,

$$X[n] = \alpha X[n-1] + \varepsilon[n] \tag{2.1}$$

From which we can deduce the impulse response convolution:

$$X[n] = \sum_{k=0}^{+\infty} \alpha^k \varepsilon[n-k]$$

Moreover $S_\varepsilon = 1$ implies ε is white with unit variance. Interestingly, this implies that

$$\text{For } n > k, \mathbb{E}[\varepsilon[n]X[k]] = 0\,.$$

Hence the name *innovation representation* for eq. (2.1).

2.2 Autoregressive (AR) process

Consider the (discrete) stationary stochastic process s.t.

$$X[n] = \sum_{k=1}^{p} a_k X[n-k] + \varepsilon[k]\,,$$

with ε a Gaussian process such that for any (n,k) such that $k > n$, we have

$$\mathbb{E}[\varepsilon[k]X[n]] = 0 \quad \text{and} \quad \mathbb{E}[\varepsilon[k]\varepsilon[n]] = 0$$

and such that $\varepsilon[k] \sim \mathcal{N}(0, \sigma^2)$.

The signal defined in this way is a stationary Gaussian process as a result of filtering a stationary Gaussian (white noise) process (provided the filter is stable).

The process ε does NOT represent measurement errors, but exogenous influences on the system dynamics that cannot be inferred from the past time points. It is called **innovation**.

2.3 Linking AR coefficients to signal statistics

We can exploit the innovation properties: for $k = 1..p$

$$\mathbb{E}[\bar{X}[n-m]X[n]] = \sum_{k=1}^{p} a_k \mathbb{E}[\bar{X}[n-m]X[n-k]]$$

leading to the system

$$Ca = c$$

with

$$C = \begin{bmatrix} \gamma_{XX}(0) & \gamma_{XX}(-1) & \cdots & \gamma_{XX}(-p+1) \\ \gamma_{XX}(1) & \gamma_{XX}(0) & \cdots & \gamma_{XX}(-p+2) \\ & & \ddots & \\ \gamma_{XX}(p-1) & \gamma_{XX}(p-2) & \cdots & \gamma_{XX}(0) \end{bmatrix}, c = \begin{bmatrix} \gamma_{XX}(1) \\ \gamma_{XX}(2) \\ \vdots \\ \gamma_{XX}(p) \end{bmatrix}$$

which allows to estimate a.

In addition

$$\mathbb{E}\left[\bar{X}[n]X[n]\right] = \sum_{k=1}^{p} a_k \mathbb{E}\left[\bar{X}[n]X[n-k]\right] + \mathbb{E}\left[\bar{X}[n]\varepsilon[n]\right]$$

leading to

$$\gamma_{XX}(0) = \sum_{k=1}^{p} \gamma_{XX}(-k)a_k + \sigma^2$$

which completes the set of $p+1$ Yule-Walker equations sufficient to estimate all parameters of the model from the data.

2.4 Model estimation

One can simply estimate autocorrelations from observations and solve the Yule-Walker equation with the empirical estimates of C and c.

Alternatively, one can perform linear regression of X from its p past samples.

The estimates based on time averages are well-behaved under mild assumptions, despite samples not being i.i.d.

2.5 Applications

Such simple, easy to estimate, model can be exploited for various purposes:

- forecasting,
- parametric power spectral density estimation,
-

3 System identification

System identification is the very general problem of identifying properties of a system from observations of its outputs and (possibly) inputs.

Highly complex real-world system may be high-dimensional, non-linear, noisy and time varying, such that the problem can be challenging or even impossible to solve.

In this section, we address a simple *linear time-invariant univariate* case.

3.1 Finite Impulse Response (FIR) models

We introduce Finite Impulse Response models (FIR, see e.g. [**Ljung(1999)**]) can be introduced as a white noise filter.

Consider the (discrete) linear system s.t.

$$Y[n] = \sum_{k=0}^{p} b_k X[n-k] + \eta[n],$$

with white noise η independent from X (can represent a measurement noise), and weakly stationary input X.

Then we ask whether b can be estimated from the joint observation of (X, Y).

3.2 Parameter estimation

We can write for $m = 0..p$,

$$\mathbb{E}\left[\bar{X}[n-m]Y[n]\right] = \sum_{k=0}^{p} b_k \mathbb{E}\left[\bar{X}[n-m]X[n-k]\right],$$

leading to the system

$$Ab = d$$

$$A = \begin{bmatrix} \gamma_{XX}(0) & \gamma_{XX}(-1) & \cdots & \gamma_{XX}(-p) \\ \gamma_{XX}(1) & \gamma_{XX}(0) & \cdots & \gamma_{XX}(-p+1) \\ & & \ddots & \\ \gamma_{XX}(p) & \gamma_{XX}(p-1) & \cdots & \gamma_{XX}(0) \end{bmatrix}, d = \begin{bmatrix} \gamma_{XY}(0) \\ \gamma_{XY}(1) \\ \vdots \\ \gamma_{XY}(p) \end{bmatrix}$$

Again, estimation can be made by estimation of cross- and auto-correlations, as well as linear regression.

4 Estimating a signal contaminated by colored noise

Then final question we will address is: given an noisy output of a known system, can I estimate its input.

This is a particular case of **inverse problem**, which is typically ill-posed. However, we can typically impose constraints and additional assumptions to find an optimal solution, in some sense.

4.1 Problem formulation

Assume weakly stationary processes (zero mean) x and η. We observe

$$y = h * x + \eta$$

with h impulse response of a stable filter (not necessarily causal). x is the information we want to recover, η is a noise disturbance.

A Wiener filter with impulse response g aims at recovering x using the estimate $g * y$. The criterion used in the Minimum Mean Squared Error (MMSE).

$$\min_g \mathbb{E}|x[0] - (g * y)[0]|^2$$

(note time point zero is arbitrary due to stationarity).

4.2 Cross spectral density

Using previously defined cross-correlation

$$\gamma_{XY}(\tau) = \mathbb{E}\left[\overline{X}[0] \cdot Y[\tau]\right].$$

We define the Cross Spectral Density (CSD) as

$$S_{XY}(\nu) = \widetilde{\gamma_{XY}}(\nu) = \sum_{k\in\mathbb{Z}} \gamma_{XY}[k] \exp(-2\mathbf{i}\pi k\nu).$$

and get easily $S_{XY}(\nu) = \overline{S_{YX}(\nu)}$.

PROPOSITION 61. *Given Z stationary. Output Y of an LTI system of impulse response h with input X verifies*

$$S_{ZY}(\nu) = \widetilde{h}(\nu) \cdot S_{ZX}(\nu) \quad \textit{and} \quad S_{YZ}(\nu) = \overline{\widetilde{h}(\nu)} \cdot S_{XZ}(\nu).$$

4.3 Solution to Wiener filtering

Expressing MMSE with cross-spectral densities, we get

$$\widetilde{g}(\nu) = \frac{1}{\widetilde{h}(\nu)} \left(\frac{|\widetilde{h}(\nu)|^2 S_x(\nu)}{|\widetilde{h}(\nu)|^2 S_x(\nu) + S_\eta(\nu)} \right)$$

This expresses a trade between inverting the filtering process performed by h and tempering the amplification of the noise.

Chapter 11

Point processes

Point processes are continuous time stochastic processes describing the properties of sequences of events. They have a wide range of applications:

- Neuroscience: analysis of spike trains.
- Medical research: survival analysis, assessment of treatment effects, epidemiology,...
- Physics: earthquake models,...
- Online networks: propagation of information, influence analysis,...

1 Counting processes

1.1 Definition

$(N(t))$ is a counting process if $N(t)$ counts the total number of events that have occurred up to time t, if $t \geq 0$.

For negative times (when needed), $N(t)$ is the opposite of the number of points occurring between t and 0.

Hence, it must satisfy:

- $N(t) \geq 0$ for all $t \geq 0$,
- $N(t) \leq 0$ for all $t \leq 0$,
- $N(t)$ is integer valued.

We define the increment $N((s,t]) = N(t) - N(s)$ the number of events that occur in the interval $(s,t]$.

Process with independent increments $N(t)$ has independent increments if for any set of disjoint intervals I_1,...,I_n, $(N(I_k))$ are jointly independent random variables.

In particular, for $s < t$, $N(s)$ and $N((s,t])$ are independent.

Process with stationary increments The distribution of event in a bounded interval, $N((s,t])$, depends only on the length of this interval.

This implies that for all τ

$$\mathbb{E}N((s+\tau,t+\tau]) = \mathbb{E}N((s,t])\,,$$

and

$$\mathrm{var}N((s+\tau,t+\tau]) = \mathrm{var}N((s,t])\,.$$

2 Homogeneous Poisson process (HPP)

2.1 Definition

An HPP $N(t)$ with intensity (rate) $\lambda > 0$ is the counting process satisfying

- $N(0) = 0$,
- the independent increments property,
- for $s < t$, $N((s,t])$ is Poisson distributed with mean $\lambda|t-s|$

$$\mathbb{P}(N((s,t]) = n) = e^{-\lambda|t-s|}\frac{(\lambda|t-s|)^n}{n!}, \quad n \geq 0\,.$$

For $\lambda = 1$, it is called the standard Poisson process.

2.2 Inter-event interval

The HPP definition implies that the distribution of $t_{k+1} - t_k$ for any k is exponential with mean $1/\lambda$, characterized by the CDF

$$\mathbb{P}(t_{k+1} - t_k > s) = e^{-\lambda s} \quad s \geq 0.$$

which corresponds to the density

$$p(s) = \lambda e^{-\lambda s} \quad s \geq 0.$$

This exponential distribution implies the memoryless property: for $t > s$

$$\mathbb{P}(t_{k+1} - t_k > t | t_{k+1} - t_k > s) = \mathbb{P}(t_{k+1} - t_k > t - s)\,,$$

i.e. the time elapsed since the last event does not provide information about when the next event will happen.

2.3 Variance of an HPP

This can be deduced from the fact that $N((s,t])$ is Poisson distributed with mean $\lambda|t-s|$, we get the variance

$$\mathrm{var}N((s,t]) = \lambda|t-s|$$

Increases linearly with the rate and duration! The Fano factor, i.e. the ratio of variance to expectation, is 1.

2.4 Simulating an HPP

Given inter event intervals are Poisson distributed, we can generate the N first events (for $t \geq 0$) by

- drawing N i.i.d. exponential variables (z_k) with mean $1/\lambda$,
- define $t_n = \sum_{k=1}^{n} z_k$.

3 Inhomogeneous Poisson process (HPP)

3.1 Definition

An IPP $N(t)$ with (deterministic) intensity (rate) $\lambda(t) > 0$ is the counting process satisfying

- $N(0) = 0$,
- the independent increment property,
- for $s < t$, $N((s,t])$ is Poisson distributed with mean $\Lambda(s,t) = \int_s^t \lambda(u)du$

$$\mathbb{P}(N((s,t]) = n) = e^{-\Lambda(s,t)} \frac{(\Lambda(s,t))^n}{n!}, \quad n \geq 0.$$

3.2 Variance of an IPP

This can be deduced from the fact that $N((s,t])$ is Poisson distributed with mean $\Lambda(s,t)$, we get the variance

$$\mathrm{var}N((s,t]) = \Lambda(s,t)$$

The Fano factor is still one.

3.3 Beyond Poisson

There exist a large variety of other point processes. Among them:

- Cox processes are a generalization of IPP where $\lambda(t)$ is stochastic,
- Hawkes processes have a $\lambda(t)$ that depends explicitly on past events.

Both lead to Fano factors that are typically different from one.

3.4 Mathematics of general counting processes

Let $\mathscr{H}_t$ be an filtration such that process $N(t)$ is adapted to it (include all events that can happen to $N(t)$ up to time t). The conditional intensity process can be heuristically defined as

$$\lambda(t)dt = P(N(t,t+dt] = 1|\mathscr{H}_{t^-}) = \mathbb{E}[N(t,t+dt] = 1|\mathscr{H}_{t^-}]$$

For the Hawkes process with kernel k and predictable intensity μ, we have

$$\lambda(t) = \mu(t) + \int_{-\infty}^{t} k(t-s) dN_s .$$

4 Empirical properties of point processes

4.1 Multi-trial framework

We will formally use s, the derivative of $N(t)$ in the sense of distributions, recorded on $k = 1, ..., K$ trials.

$$s^{(k)}(t) = \sum_{k=1}^{K} \delta(t - t_j^{(k)})$$

4.2 Mean intensity

$$\bar{\lambda}(t) = \langle s(t) \rangle = \lim_{K \to +\infty} \sum_k s^{(k)}(t)$$

where $\langle . \rangle$ indicates the empirical average.

In order to smooth this empirical measure, we can use binning, or a smoothing kernel.

4.3 Autocorrelation function

$$R_{ss}(\tau) = \left\langle \frac{1}{T} \int_0^T s(t) s(t-\tau) dt \right\rangle$$

We can decompose this quantity as

$$R_{ss}(\tau) = \bar{\Lambda} + \left\langle \frac{1}{T} \int_0^T (\bar{\lambda}(t) - \bar{\Lambda})(\bar{\lambda}(t-\tau) - \bar{\Lambda}) dt \right\rangle + \left\langle \frac{1}{T} \int_0^T (s(t) - \bar{\lambda}(t))(s(t-\tau) - \bar{\lambda}(t-\tau)) dt \right\rangle$$

Bibliography

[Brockwell and Davis(1991)] PJ Brockwell and RA Davis. *Time series: theory and methods*. Springer, 1991.

[Butz(2015)] T Butz. *Fourier transformation for Pedestrians*. Springer, 2015.

[Gasquet and Witomski(1999)] C Gasquet and P Witomski. *Fourier Analysis and Applications: Filtering, Numerical Computation, Wavelets,*. Springer, 1999.

[Koopmans(1995)] Lambert H Koopmans. *The spectral analysis of time series*. Academic press, 1995.

[Ljung(1999)] L Ljung. *System Identification, Theory for the User*. 1999.

[Proakis and Manolakis(2007)] JG Proakis and DG Manolakis. *Digital Signal Processing*. 2007.

[Xu and Yan(2006)] Yuesheng Xu and Dunyan Yan. The bedrosian identity for the hilbert transform of product functions. *Proceedings of the American Mathematical Society*, 134(9):2719–2728, 2006.