Analysis

Lecture notes for the Master degree in Stochastics and Data Sciences

Academic Year 2023-2024

Bertrand Lods

Comments: These lecture notes were originally written in for the Analysis, Course B, lectures of the academic year 2015-2016 of the Master Degree in Stochastics and Data Sciences of Turin University. Some editing and corrections have been provided for the current Academic Year 2023-2024.

The notes cover the course in detail, and are therefore sufficient for the preparation of the exam. They also include material not covered during the course (Complements sections).

The proofs that students should prepare and "know" for the exams are indicated with the symbol

✍

in the left margin.

These lectures notes have been written with the support from the project: "Start up della Laurea Magistrale in Stochastics and Data Science: pubblicizzazione, internazionalizzazione ed eccellenza interdisciplinare" funded by fondazione CRT.

First printing, December 2017

Contents

1. Normed spaces

In all the notes, the vector spaces we are considering are $\mathbb{R}$-spaces but everything extends in some straightforward way to vector spaces built over the complex field $\mathbb{C}$. We recall more generally that, given a field $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$, a $\mathbb{K}$-vector space X is defined as a set which is closed under finite vector addition and scalar multiplication by elements of $\mathbb{K}$. In a more formal way, X is a $\mathbb{K}$-vector space if there exist two operations, addition $+$ and scalar multiplication, for which the following conditions hold for any elements $x, y, z \in X$ and any scalars $\alpha, \beta \in \mathbb{K}$:

1. *(Commutativity)*
$$x + y = y + x$$
2. *(Associativity of vector addition)*
$$(x + y) + z = x + (y + z).$$
3. There is an element $0_X \in X$ such that
$$0_X + x = x + 0_X = x \qquad \forall x \in X$$
4. For any $x \in X$, there is $-x \in X$ such that $x + (-x) = 0_X$.
5. *(Associativity of scalar multiplication)*
$$\alpha(\beta x) = (\alpha\,\beta)x.$$
6. *(Distributivity of scalar sums)*
$$(\alpha + \beta)x = \alpha\,x + \beta x.$$
7. *(Distributivity of vector sums)*
$$\alpha(x + y) = \alpha\,x + \alpha y.$$

8. *(Scalar multiplication identity)*

$$1\,x = x.$$

It is important to understand that, in the above point 6, the symbol $+$ in the left-hand side stands for the "classical" addition in the field $\mathbb{K}$ whereas the same symbol $+$ on the right-hand-side stands for the (generally more involved) addition on the space X. In the same way, the mulplication in $\mathbb{K}$ is denoted by $\alpha\,\beta$ and there is no possible confusion with the scalar multiplication $\alpha\,x$ since, in the first case, both α, β lie in $\mathbb{K}$ whereas, in the second case, we multiply a "vector" $x \in X$ by a scalar $\alpha \in \mathbb{K}$ and the result is a vector $\alpha\,x \in X$.

1.1 The norm and its properties

1.1.1 General properties

We start with the definition of seminorms and norms

Definition 1.1.1 A seminorm on X is a function $\mathcal{N} : X \to \mathbb{R}$ such that, for any vectors $x, y \in X$ and any $\alpha \in \mathbb{R}$, the following properties hold:

1. ***Nonnegativity***: $\mathcal{N}(x) \geqslant 0$,
2. ***Homogeneity***: $\mathcal{N}(\alpha x) = |\alpha|\mathcal{N}(x)$,
3. ***Triangle Inequality***: $\mathcal{N}(x+y) \leqslant \mathcal{N}(x) + \mathcal{N}(y)$.

A seminorm is then a ***norm*** if it also holds:

1. ***Uniqueness***: $\mathcal{N}(x) = 0$ if and only if $x = 0$.

If $\mathcal{N}$ is a norm, we rather use the notation $\|x\| = \mathcal{N}(x)$. A vector space X endowed with a norm $\|\cdot\|$ is called a normed space and denoted by $(X, \|\cdot\|)$.

■ **Example 1.1** The absolute value function $|x|$ $(x \in \mathbb{R})$ is a norm on the real line $\mathbb{R}$. In the same way, the Euclidean norm

$$\|\boldsymbol{x}\| = \sqrt{\sum_{i=1}^{N} x_i^2}, \qquad \boldsymbol{x} = (x_1, \ldots, x_N) \in \mathbb{R}^N$$

is a norm on $\mathbb{R}^N$ $(N \geqslant 1)$. ■

■ **Example 1.2** Given a sequence of real $\boldsymbol{x} = (x_n)_{n\in\mathbb{N}} = (x_1, x_2, \ldots)$, we introduce the space of all absolutely summable sequences,

$$X = \ell^1(\mathbb{N}) = \left\{ \boldsymbol{x} = (x_n)_{n\in\mathbb{N}} \subset \mathbb{R} \text{ such that } \|\boldsymbol{x}\|_1 := \sum_{n=1}^{\infty} |x_n| < \infty \right\}.$$

This set is a vector space and the mapping $\|\cdot\|_1 : X \to \mathbb{R}_+$ defined as

$$\|\,\boldsymbol{x}\,\|_1 = \sum_{n\geqslant 1} |x_n| \qquad \forall \boldsymbol{x} = (x_n)_n \subset \ell^1(\mathbb{N}) \tag{1.1}$$

is a norm on X. ■

Proposition 1.1.1 Let $I = [a, b] \subset \mathbb{R}$ be a given interval. We introduce the space $\mathscr{C}(I)$

as the set of all continuous functions $f : I \to \mathbb{R}$. Let

$$\|f\|_1 = \int_a^b |f(t)|\mathrm{d}t \qquad \forall f \in \mathscr{C}(I)$$

and

$$\|f\|_\infty = \sup_{t\in I} |f(t)|, \qquad f \in \mathscr{C}(I).$$

Then, $(\mathscr{C}(I), \|\cdot\|_1)$ and $(\mathscr{C}(I), \|\cdot\|_\infty)$ are two normed spaces.

Proof. The proof that $\|\cdot\|_\infty$ is a norm on $\mathscr{C}(I)$ is an easy task. To prove that $\|\cdot\|_1$ is also a norm on $\mathscr{C}(I)$, one first has to notice that $\|f\|_1 < \infty$ for any $f \in \mathscr{C}(I)$ which is a classical result of integral calculus (the Riemann integral of a continuous function on some finite interval is well-defined and finite). Then, it is very easy to prove that $\|\cdot\|_1$ is a seminorm on $\mathscr{C}(I)$. To prove it is actually a norm, one has to show that, given $f \in \mathscr{C}(I)$,

$$\int_a^b |f(t)|\mathrm{d}t = 0 \implies f(t) = 0 \qquad \forall t \in [a,b].$$

This comes from the fact that the function $t \in [a,b] \mapsto |f(t)|$ is a continuous and nonnegative function and that the integral of a continuous and nonnegative result is zero if and only if the function is vanishing everywhere ***(Check this arguing by contradiction!)***. ■

We can also introduce derivatives to build function spaces:

Proposition 1.1.2 Let $I = [a,b] \subset \mathbb{R}$ be a given interval. We introduce the space $\mathscr{C}^1(I)$ as the space of all continuously differentiable functions $f : I \to \mathbb{R}$ (i.e. f is differentiable over I and its derivative f' is continuous over I). Let

$$\|f\|_{1,1} = \int_a^b |f(t)|\mathrm{d}t + \int_a^b |f'(t)|\mathrm{d}t \qquad \forall f \in \mathscr{C}(I)$$

and

$$\|f\|_{1,\infty} = \sup_{t\in I} (|f(t)| + |f'(t)|) \qquad f \in \mathscr{C}^1(I).$$

Then, $(\mathscr{C}^1(I), \|\cdot\|_{1,1})$ and $(\mathscr{C}^1(I), \|\cdot\|_{1,\infty})$ are normed spaces

Proof. The proof is left as an **Exercise**. ■

Remark 1.1.1 Notice that, if one simply defines

$$\mathcal{N}(f) = \sup_{t\in I} |f'(t)| \qquad f \in \mathscr{C}^1(I)$$

one sees that $\mathcal{N}$ is a *semi-norm* on $\mathscr{C}^1(I)$ but not a norm ! Indeed, any constant function $f \equiv c$ is such that $\mathcal{N}(f) = 0$ so the uniqueness property is not satisfied.

1.1.2 Topology of normed spaces – convergence of sequences

Let recast here the standard definitions of topological spaces in the special case of normed spaces:

Definition 1.1.2 Let $(X, \|\cdot\|)$ be a normed space. For any $x_0 \in X$ and $r > 0$, the open ball centered at x_0 with radius $r > 0$ is defined as

$$B(x_0, r) = \{x \in X \; ; \; \|x - x_0\| < r\}$$

while the closed ball centered at x_0 with radius $r > 0$ is

$$B_c(x_0, r) = \{x \in X \; ; \; \|x - x_0\| \leqslant r\}.$$

A subset $U \subset X$ is an *open set* if, for any $x \in U$ there exists $r > 0$ such that $B(x, r) \subset U$. A subset $U \subset X$ is closed if its complementary $U^c = X \setminus U$ is an open subset of X.

One checks that $\varnothing$ and X are both open and closed subsets of $(X, \|\cdot\|)$.

■ **Example 1.3** Given $x \in X$ and $\varrho > 0$, the ball $B(x, \varrho)$ is an open set of $(X, \|\cdot\|)$ whereas $B_c(x, \varrho)$ is a closed set. ***(Check this!)*** ■

Notice that there exist subsets of X which are neither open nor closed.

Open and closed sets enjoy important properties whose proof is left as an **Exercise**:

Proposition 1.1.3 A finite intersection of open sets of $(X, \|\cdot\|)$ is an open set of $(X, \|\cdot\|)$. Any union of open sets of $(X, \|\cdot\|)$ is an open set of $(X, \|\cdot\|)$.

A finite union of closed subsets of $(X, \|\cdot\|)$ is a closed set of $(X, \|\cdot\|)$. Any intersection of closed subsets of $(X, \|\cdot\|)$ is a closed subset of $(X, \|\cdot\|)$.

We recall several examples in $\mathbb{R}$

■ **Example 1.4** Let $X = \mathbb{R}$ endowed with the natural norm $|\cdot|$. Let $a < b$ be given. One checks easily that

1. The sets (a, b), $(b, +\infty)$ and $(-\infty, a)$ are open subsets of $(\mathbb{R}, |\cdot|)$;
2. the sets $[a, b]$, $[b, +\infty)$ and $(-\infty, a]$ and $\{a\}$ are closed sets of $(\mathbb{R}, |\cdot|)$;
3. the sets $[a, b)$ and $(a, b]$ are neither open nor closed subsets of $(\mathbb{R}, |\cdot|)$.

Moreover, one checks that

$$[0, 1] = \bigcap_{n \geqslant 1} \left(-\frac{1}{n} \; ; \; 1 + \frac{1}{n}\right),$$

which illustrates the fact that an arbitrary (infinite) intersections of open subsets is not necessarily an open subset. In the same way, since

$$(0, 1) = \bigcup_{n \geqslant 1} \left[\frac{1}{n} \; ; \; 1 - \frac{1}{n}\right]$$

one sees that an arbitrary union of closed subsets is not necessarily a closed subset. ■

We will assume the concept of sequence to be familiar. We just recall that a sequence of elements of X is a function $f : \mathbb{N} \to X$, i.e. $f(n) \in X$ for any $n \in \mathbb{N}$ but a more useful way to see a sequence is to regard it as an ordered list of elements of X which can be written as $(x_n)_n \subset X$ (here of course, $x_n = f(n)$, $n \in \mathbb{N}$). A subsequence of $(x_n)_n$ is then a sequence of the form $(x_{n_k})_k$ where $(n_k)_k \subset \mathbb{N}$ is a strictly increasing sequence of $\mathbb{N}$ [1].

[1] Equivalently, a subsequence of $(x_n)_n$ is a sequence of the form $(x_{\varphi(n)})_n$ where $\varphi : \mathbb{N} \to \mathbb{N}$ is a strictly increasing function.

We introduce now the notion of convergence with respect to the norm

Definition 1.1.3 — Convergent sequence. Let $(X, \|\cdot\|)$ be a normed space. A sequence $(x_n)_n \subset X$ of elements in X is said to converge to $x \in X$ if

$$\lim_{n\to\infty} \|x_n - x\| = 0$$

i.e., if for any $\varepsilon > 0$, there exists $N = N(\varepsilon) \in \mathbb{N}$ such that $\|x_n - x\| < \varepsilon$ for any $n \geqslant N$. We write then

$$\lim_n x_n = x \qquad \text{or} \qquad x_n \to x.$$

The point x is then called the limit of $(x_n)_n$ in $(X, \|\cdot\|)$.

One has the following whose proof is left as an **Exercise**:

Theorem 1.1.4 Let $(X, \|\cdot\|)$ be a normed space and let $(x_n)_n$ be a convergent sequence in X. Then:

1. the limit $x = \lim_n x_n$ is unique;
2. any subsequence of $(x_n)_n$ also converges to x.
3. if $(x_n)_n \subset X$ is a convergent sequence with $\lim_n x_n = x$ then
$$\lim_n \|x_n\| = \|x\|,$$
in particular
$$\sup_n \|x_n\| < \infty.$$
4. Continuity of vector addition: if $(x_n)_n, (y_n)_n \subset X$ are converging sequence of X and if $(\alpha_n)_n \subset \mathbb{R}$ is a convergent sequence of real numbers with
$$\lim_n x_n = x \in X, \quad \lim_n y_n = y \in X \quad \lim_n \alpha_n = \alpha \in \mathbb{R}$$
then $(x_n + \alpha_n y_n)_n \subset X$ is a convergent sequence with
$$\lim_n (x_n + \alpha_n y_n) = x + \alpha y.$$

Proof. Notice that the proof of point (iii) comes from the *Reverse Triangle Inequality*:

$$|\|x\| - \|y\|| \leqslant \|x - y\| \qquad \forall x, y \in X.$$

The proof is left as an **Exercise**. ■

The use of sequences allows to have a more tractable characterization of closed sets

Proposition 1.1.5 A subset $C \subset X$ is a closed subset of $(X, \|\cdot\|)$ whenever the limit of any convergent sequence $(x_n)_n \subset C$ belongs to C.

In other words, C is closed if and only if once a sequence of elements of C is converging, the limit cannot escape C. This explains the terminology...

✍ *Proof.* Assume first that C is closed and let $(x_n)_n \subset C$ be a convergent sequence with $x = \lim_n x_n$. Let us show that $x \in C$. We argue by contradiction assuming that $x \notin C$. Since $X \setminus C$ is open, there exists $r > 0$ such that $B(x,r) \subset X \setminus C$. By definition of convergence, there exists $N \in \mathbb{N}$ such that $\|x_n - x\| < r$ for any $n \geqslant N$. In other words, $x_n \in B(x,r)$ for any $n \geqslant N$. This contradicts the fact that $(x_n)_n \subset C$ since $B(x,r) \subset X \setminus C$. This proves that $x \in C$.

Conversely, assume that the limit of any convergent sequence of elements in C belongs to C. We show that C is closed arguing by contradiction. Let us assume that C is not closed, i.e. $X \setminus C$ is not open and, by definition, there exists $x \in X \setminus C$ such that, for any $r > 0$, $B(x,r) \not\subset X \setminus C$. This means that, for any $r > 0$, $B(x,r) \cap C \neq \varnothing$. For any $n \geqslant 1$, we pick then $x_n \in C \cap B(x, \frac{1}{n})$. This sequence is clearly a sequence of elements of C and, for any $n \geqslant 1$, $\|x_n - x\| < \frac{1}{n}$ so that $x = \lim_n x_n$. Since $x \notin C$, we just constructed a sequence of elements of C which converges to a limit not belonging to C. This is a contradiction. ■

We can then characterize the closure of a given set:

Definition 1.1.4 Let $(X, \|\cdot\|)$ be a normed space and let $Y \subset X$. The closure $\overline{Y}$ is given by

$$\overline{Y} = \left\{ x \in X \,,\, \exists (x_n)_n \subset Y \quad \text{with } x = \lim_n x_n \right\}.$$

Remark 1.1.2 Observe that $Y \subset \overline{Y}$. It can also be proved that $\overline{Y}$ is a closed set of $(X, \|\cdot\|)$ and that $\overline{Y}$ is the *smallest* closed subset of $(X, \|\cdot\|)$ which contains Y, i.e.

$$\overline{Y} = \bigcap_{\substack{C \text{ closed} \\ Y \subset C}} C$$

Check this as an Exercise.

As an example, we show here that closed balls coincide with the closure of open balls:

Lemma 1.1.6 Let $(X, \|\cdot\|)$ be a normed space and let $x_0 \in X$, $r > 0$. Then,

$$B_c(x_0, r) = \overline{B(x_0, r)}$$

where the closure is meant with respect tos $\|\cdot\|$.

Proof. Since the closed ball $B_c(x_0,r)$ is closed and contains $B(x_0,r)$, it contains $\overline{B(x_0,r)}$, i.e. $\overline{B(x_0,r)} \subset B_c(x_0,r)$ holds. Prove then the other inclusion and let $x \in B_c(x_0,r)$. We have to prove that $x \in \overline{B(x_0,r)}$. Of course, if $\|x - x_0\| < r$, there is nothing to prove. Let us then assume that $\|x - x_0\| = r$. Since $r > 0$, we can define $u = \frac{1}{r}(x - x_0)$. Clearly, $\|u\| = 1$ i.e. u is the unit vector in the direction of $x - x_0$. One has $x = x_0 + ru$. For any $n \in \mathbb{N}$ set

$$x_n = x_0 + r\left(1 - \frac{1}{n}\right)u.$$

Then, $\|x_n - x_0\| = |r - \frac{r}{n}|\,\|u\| = r(1 - \frac{1}{n}) < r$. Thus, $x_n \in B(x_0,r)$ for any $r > 0$ and

$$\|x_n - x\| = \left\|\frac{r}{n}u\right\| = \frac{r}{n}$$

so that $\lim_n x_n = x$. We thus proved that x is the limit of some sequence $(x_n)_n$ contained in $B(x_0,r)$, i.e. $x \in \overline{B(x_0,r)}$. ■

We introduce the following

Definition 1.1.5 Let $(X, \|\cdot\|)$ be a normed space and let $Y \subset X$ be given. We say that Y is *dense* in X if

$$\overline{Y} = X$$

We have the following useful characterization:

Lemma 1.1.7 Let $(X, \|\cdot\|)$ be a normed space and $Y \subset X$. Then, Y is dense in X if and only if, for any non-empty open set $\mathcal{O} \subset X$, it holds $Y \cap \mathcal{O} \neq \varnothing$.

Proof. The proof is left as an ***Exercise.*** ■

Remark 1.1.3 The meaning of the above is just that, if Y is dense in X, then it *meets* any open subset of X or, more precisely, it meets any ball of X.

Remark 1.1.4 From the definition of closure of Y, we see that Y is dense in X if and only if any element $x \in X$ can be approximated by a sequence in Y, i.e. for any $x \in X$ there is $\{x_n\}_n \subset Y$ such that $\lim_n \|x_n - x\| = 0$.

Remark 1.1.5 As well known, in $(\mathbb{R}, |\cdot|)$, we have that $\mathbb{Q}$ is dense in $\mathbb{R}$, i.e. $\overline{\mathbb{Q}} = \mathbb{R}$. One also have that

$$\overline{\mathbb{R}\setminus\mathbb{Q}} = \mathbb{R}.$$

As explained by Lemma 1.1.7, this exactly means that, between two real numbers $a < b$, there exist $q \in \mathbb{Q}$ and $x \in \mathbb{R}\setminus\mathbb{Q}$ such that

$$a < q < b, \qquad a < x < b$$

since the open interval (a,b) should meet both $\mathbb{Q}$ and $\mathbb{R}\setminus\mathbb{Q}$ ***(Get convinced that Lemma 1.1.7 exactly says this)***.

As we defined the notion of closure, we can define the *interior* of a set $Y \subset X$ as follows

Definition 1.1.6 Let $(X, \|\cdot\|)$ be a normed space and let $Y \subset X$ be given. A point $y \in Y$ is called an interior point of Y if there is $\varepsilon > 0$ such that $B(y,\varepsilon) \subset Y$. We define then the interior of Y as

$$\mathrm{Int}(Y) = \{y \in Y\ ;\ y \text{ is interior to } Y\} = \{y \in Y\ ;\ \exists \varepsilon > 0 \text{ such that } B(y,\varepsilon) \subset Y\}.$$

Remark 1.1.6 Observe that $\mathrm{Int}(Y) \subset Y \subset \overline{Y}$ and $\mathrm{Int}(Y)$ is an open subset of X. It can also be proved that $\mathrm{Int}(Y)$ is the *biggest* open subset of $(X, \|\cdot\|)$ contained in Y, i.e.

$$\mathrm{Int}(Y) = \bigcup_{\substack{U \text{ open} \\ U \subset Y}} U$$

Check this as an Exercise.

The link between closure and interior is made in the following

Lemma 1.1.8 Let $(X, \|\cdot\|)$ be a normed space and let $Y \subset X$ be given. Then

$$\overline{X \setminus Y} = X \setminus \mathrm{Int}(Y)$$

i.e. the closure of the complement is the complement of the interior.

Proof. The proof is left as an ***Exercise.*** ■

We introduce now the notion of limit point

Definition 1.1.7 Let $(X, \|\cdot\|)$ be a given normed space and let $(x_n)_n \subset X$ be a given sequence in X. We say that $x \in X$ is a *limit point* of the sequence $(x_n)_n$ if it is the limit of a subsequence of $(x_n)_n$, i.e. there exists a subsequence $(x_{\varphi(n)})_n$ such that

$$\lim_n \|x_{\varphi(n)} - x\| = 0.$$

■ **Example 1.5** If $X = \mathbb{R}$ endowed with the absolute value, the sequence $((-1)^n)_n$ has exactly two limit points -1 and 1. The sequence $(n)_n$ has no limit point in $\mathbb{R}$. ■

The link between the closure and the notion of limit points is given by the following whose proof is left as an **Exercise**:

Proposition 1.1.9 Given $Y \subset X$. The closure $\overline{Y}$ is the set of all limit points of sequences of Y.

Exercise 1.1 Given a sequence $(x_n)_n \subset X$, prove that $x \in X$ is a limit point of $(x_n)_n$ if and only if

$$\mathrm{card}\left(\{k \in \mathbb{N};:\ x_k \in B(x, \varepsilon)\}\right) = \infty \qquad \forall \varepsilon > 0,$$

where card(A) denote the cardinal of the set of A. ■

1.1.3 Equivalent norms – Product spaces

One can define the notion of equivalent norms

Definition 1.1.8 Let X be a vector space over $\mathbb{R}$. We say that two norms $\mathcal{N}_1$ and $\mathcal{N}_2$ on X are equivalent if there exist two positive constants $C_1, C_2 > 0$ such that

$$\mathcal{N}_1(x) \leqslant C_1 \mathcal{N}_2(x) \qquad \text{and} \qquad \mathcal{N}_2(x) \leqslant C_2 \mathcal{N}_1(x) \qquad \forall x \in X.$$

Remark 1.1.7 Geometrically, two norms $\mathcal{N}_1, \mathcal{N}_2$ are equivalent if the unit ball centered in 0 for $\mathcal{N}_1$ contains a (non-empty) open ball centered in 0 for $\mathcal{N}_2$ and viceversa.

As we shall see later on, on some finite dimensional spaces, *all the norms are equivalent*. This is no more the case for infinite dimension spaces as illustrated by the following

■ **Example 1.6** Let $X = \mathscr{C}(I)$ with $I = [0, 1]$ and let $\|\cdot\|_1, \|\cdot\|_\infty$ be the norms defined in Proposition 1.1.1. Consider, for any $k \in \mathbb{N}$, the function

$$f_k(t) = \max\{(1 - kt), 0\}, \qquad t \in I,$$

so that $(f_k)_k \subset X$. Then, one has

$$\|f_k\|_1 = \frac{1}{2k}, \qquad \|f_k\|_\infty = 1 \qquad \forall k \geqslant 1.$$

In particular, there cannot be a positive constant $C > 0$ such that $\|f_k\|_\infty \leqslant C\|f_k\|_1$ for any k. This shows that $\|\cdot\|_1$ and $\|\cdot\|_\infty$ are not equivalent. ∎

Proposition 1.1.10 Let $(X_1, \|\cdot\|_1)$ and $(X_2, \|\cdot\|_2)$ be two normed spaces. Define

$$X = X_1 \times X_2 = \{\boldsymbol{x} = (x_1, x_2), \quad x_i \in X_i,\ i = 1, 2\}$$

and let us define

$$\|\boldsymbol{x}\|_+ = \|x_1\|_1 + \|x_2\|_2, \qquad \|\boldsymbol{x}\|_e = \sqrt{\|x_1\|_1^2 + \|x_2\|_2^2}$$
$$\text{and} \qquad \|\boldsymbol{x}\|_{\max} = \max\left(\|x_1\|_1, \|x_2\|_2\right) \qquad \forall \boldsymbol{x} = (x_1, x_2), \in X.$$

Then, $(X, \|\cdot\|_+)$, $(X, \|\cdot\|_{\mathrm{e}})$ and $(X, \|\cdot\|_{\max})$ are equivalent normed spaces.

Proof. It is easy to check that $(X, \|\cdot\|_+)$, $(X, \|\cdot\|_e)$ and $(X, \|\cdot\|_{\max})$ are normed spaces. Moreover, since

$$\|\boldsymbol{x}\|_{\max} \leqslant \|\boldsymbol{x}\|_e \leqslant \|\boldsymbol{x}\|_+ \leqslant 2\|\boldsymbol{x}\|_{\max} \qquad \forall \boldsymbol{x} \in X$$

the three normeds are equivalent. ∎

1.1.4 Continuity of functions

We can define the notion of continuity of mappings between two normed spaces:

Definition 1.1.9 Let $(X, \|\cdot\|_1)$ and $(Y, \|\cdot\|_2)$ be two normed spaces. A mapping $f : X \to Y$ is said to be continuous at $x_0 \in X$ if, for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\|x_0 - x\|_1 < \delta \implies \|f(x_0) - f(x)\|_2 < \varepsilon \qquad (x \in X).$$

We say that f is continuous over X if f is continous at *any* $x_0 \in X$.

We have the strongest definition:

Definition 1.1.10 Let $(X, \|\cdot\|_1)$ and $(Y, \|\cdot\|_2)$ be two normed spaces. A mapping $f : X \to Y$ is said to be uniformly continuous over X if, for any $\varepsilon > 0$ there exists $\delta > 0$ such that, for any $x_0, x \in X$ it holds

$$\|x_0 - x\|_1 < \delta \implies \|f(x_0) - f(x)\|_2 < \varepsilon\,.$$

It is important to distinguish the two above notions. Clearly, an uniformly continuous function $f : X \to Y$ is continuous over X but the converse is not true. The difference between the two notions lies in the fact that, in the definition of uniform continuity, the choice of δ depends *only* on ε and not on the point x_0 (the same δ will "work" for any x_0 whereas, in the definition of continuity, different x_0's would yield different choices of δ). For instance, the mapping $x \mapsto x^2$ is continuous over $\mathbb{R}$ but is not uniformly continuous over $\mathbb{R}$ (***Check this***). On the contrary, the mapping $x \in [0, \infty) \mapsto \sqrt{x}$ is uniformly continuous over $[0, \infty)$ thanks to the inequality

$$|\sqrt{x} - \sqrt{y}| \leqslant \sqrt{|x - y|} \qquad \forall x \geqslant 0,\ y \geqslant 0$$

(***Check the inequality and its consequence on uniform continuity***). Continuity can also be characterized in terms of converging sequences:

Proposition 1.1.11 Let $(X, \|\cdot\|_1)$ and $(Y, \|\cdot\|_2)$ be two normed spaces and let $f : X \to Y$ be a given mapping. Then, f is continuous at $x \in X$ if and only if, for any sequence $(x_n)_n \subset X$ with $\lim_n x_n = x$ it holds $\lim_n f(x_n) = f(x)$ (in Y), i.e. f is continuous at $x \in X$ if and only if

$$\lim_n \|x_n - x\|_1 = 0 \implies \lim_n \|f(x_n) - f(x)\|_2 = 0.$$

Proof. The proof is left as an **Exercise**. ■

With this characterization, it is very easy to prove the following result.

Proposition 1.1.12 Let $(X_i, \|\cdot\|_i)$, $i = 1, 2, 3$ be three normed spaces and let $f : X_1 \to X_2$ and $g : X_2 \to X_3$ be continuous. Then,

$$g \circ f : X_1 \to X_3$$

is continuous where $g \circ f(x) = g(f(x))$ for any $x \in X_1$.

1.2 Fundamental example: Lebesgue spaces

1.2.1 The definition of L^1 as a set of equivalence classes

Let (S, Σ, μ) be a given measure space. We recall that $\mathcal{L}^1(S, \Sigma, \mu)$ denotes the set of all measurable functions $f : S \to \mathbb{R}$ such that

$$\int_S |f| \mathrm{d}\mu < \infty.$$

It is clear that $\mathcal{L}^1(E, \Sigma, \mu)$ is a vector space and the null element $\mathbf{0}$ of $\mathcal{L}^1(S, \Sigma, \mu)$ is the function $f \equiv 0$, i.e. $f(s) = 0$ *for all* $s \in S$.

It is tempting to define a norm on $\mathcal{L}^1(S)$ like

$$\|f\|_1 = \int_S |f| \mathrm{d}\mu$$

and, one checks easily that this defines a *semi-norm* on $\mathcal{L}^1(S, \Sigma, \mu)$ but that

$$\|f\|_1 = 0 \iff \int_S |f| \mathrm{d}\mu = 0 \iff f = 0 \qquad \mu - \text{a. e. on } S$$

In particular, f is not necessarily the null element of $\mathcal{L}^1(S, \Sigma, \mu)$. Therefore, $\|\cdot\|_1$ is not a norm on $\mathcal{L}^1(S, \Sigma, \mu)$. To avoid this flaw, we actually change the set of functions we wish to investigate. This is done in the following way, introducing the following equivalence relation:

Definition 1.2.1 Let $f, g \in \mathcal{L}^1(S, \Sigma, \mu)$. We will say that f is equivalent to g and write

$$f \sim g$$

if $f - g$ is zero μ-a.e. on S, i.e. $\mu(\{s \in S\,;\, f(s) \neq g(s)\}) = 0$. It is readily seen that $\sim$ is an equivalence relation i.e. for any $f, g, h \in \mathcal{L}^1(S, \Sigma, \mu)$:

- $f \sim f$;
- $f \sim g$ if and only if $g \sim f$;
- $f \sim g$ and $g \sim h$ implies $f \sim h$.

In particular, given $f \in \mathcal{L}^1(S, \Sigma, \mu)$, we introduce the equivalence class of f as

$$[f]_\mu = \{g \in \mathcal{L}^1(S, \Sigma, \mu)\,;\, f \sim g\}.$$

We see from the above that the collection of all equivalence classes $[f]_\mu$ of $\mathbb{R}$-valued measurable functions $f \in \mathcal{L}^1(S, \Sigma, \mu)$ is a partition of $\mathcal{L}^1(S, \Sigma, \mu)$, i.e.

$$\mathcal{L}^1(S, \Sigma, \mu) = \bigcup_{f \in \mathcal{L}^1(S,\Sigma,\mu)} [f]_\mu \qquad \text{with disjoint union.}$$

We introduce then the set

$$L^1(S, \mu) = \left\{[f]_\mu, f \in \mathcal{L}^1(S, \Sigma, \mu)\right\}.$$

Notice that, strictly speaking, any element in $L^1(S, \mu)$ is a ***subset*** of $\mathcal{L}^1(S, \Sigma, \mu)$, i.e. elements of $L^1(S, \mu)$ are sets. We endow $L^1(S, \mu)$ with a linear structure as follows: for any $f, g \in \mathcal{L}^1(S, \Sigma, \mu)$ and any $\alpha \in \mathbb{R}$, one can define the sum:

$$[f]_\mu + [g]_\mu = [f + g]_\mu$$

and the multiplication

$$\alpha[f]_\mu = [\alpha f]_\mu.$$

This makes $L^1(S, \mu)$ a vector space and the null element of $L^1(S, \mu)$ is the class

$$[0]_\mu = \{g \in \mathcal{L}^1(S, \Sigma, \mu)\,,\, g \sim \mathbf{0}\} = \{g \in \mathcal{L}^1(S, \Sigma, \mu)\,, g = 0 \;\; \mu - \text{a.e.}\}.$$

Now, the space $L^1(S, \mu)$ can be endowed with a norm:

$$\|[f]\|_1 = \int_S |f| \mathrm{d}\mu \qquad \forall f \in \mathcal{L}^1(S, \Sigma, \mu).$$

Notice that the norm is well-defined in the sense that it does not depend on the representative f chosen for the class $[f]_\mu$: indeed, if $g \in [f]_\mu$, then $[f]_\mu = [g]_\mu$ and $\|[f]_\mu\|_1 = \|[g]_\mu\|_1$ ***(Important: try to understand that !)***.

Of course, from now, we will abuse notations and **identify** equivalence classes $[f]_\mu$ with functions f and will simply write

$$f \in L^1(S, \mu) \Longleftrightarrow \int_S |f| \mathrm{d}\mu$$

but we should always keep in mind that, doing so, we actually identify two functions which are equal μ-a.e. on S as a unique function. With this identification, we denote

$$\|f\|_1 = \int_S |f| \mathrm{d}\mu$$

and has the following

Proposition 1.2.1 $(L^1(S,\mu), \|\cdot\|_1)$ is a normed vector space.

1.2.2 L^p-spaces, $1 \leqslant p \leqslant \infty$.

In all the sequel, (S,Σ,μ) is a given, *fixed*, measure space. Having always in mind the above identification between measurable functions f and its equivalent class $[f]_\mu$, we introduce the following

Definition 1.2.2 Let $p \in \mathbb{R}$ be given with $1 < p < \infty$. We set

$$L^p(S,\mu) = \big\{f \;:\; S \to \mathbb{R}\,,\ f \text{ measurable such that } |f|^p \in L^1(S,\mu)\big\}$$

and

$$\|f\|_p = \left(\int_S |f|^p \mathrm{d}\mu\right)^{1/p} \qquad \forall f \in L^p(S,\mu).$$

We shall check later on that, with such a definition, $L^p(S,\mu)$ is vector space and $\|\cdot\|_p$ is a norm.

Definition 1.2.3 We set

$$L^\infty(S,\mu) = \Big\{f \;:\; S \to \mathbb{R}\,,\ f \text{ measurable such that there is } C > 0$$
$$\text{such that } |f(x)| \leqslant C \;\mu - \text{a.e. on } S\Big\}.$$

and

$$\|f\|_\infty = \inf\{C\,,\ |f(x)| \leqslant C \;\mu - \text{a.e. on } S\}.$$

■ **Example 1.7** If $S = \mathbb{N}$ and μ is the counting measure, i.e. $\mu(A) = \mathrm{card}(A)$ for any subset $A \subset \mathbb{N}$, then one sees that

$$L^p(S,\mu) = \ell^p(\mathbb{N}) \qquad 1 \leqslant p \leqslant \infty$$

is the spaces of sequences :

$$\ell^p(\mathbb{N}) = \{\boldsymbol{x} = (x_n)_n\ ;\ \|\boldsymbol{x}\|_p^p := \sum_n |x_n|^p < \infty\}.$$

■

Remark 1.2.1 Notice that, if $f \in L^\infty(S,\mu)$ then we have

$$|f(x)| \leqslant \|f\|_\infty \qquad \mu\text{-a.e. on } S. \tag{1.2}$$

Indeed, there exists a sequence $(C_n)_n$ such that $C_n \to \|f\|_\infty$ and for each n,

$$|f(x)| \leqslant C_n \qquad \mu\text{-a.e. on } S.$$

Therefore, there is $E_n \subset \Sigma$ with $\mu(E_n) = 0$ and

$$|f(x)| \leqslant C_n \qquad \forall x \in S \setminus E_n.$$

Setting then $E = \bigcup_n E_n$, one has $\mu(E) = 0$ and

$$|f(x)| \leqslant C_n \qquad \forall n \in \mathbb{N}\,,\, \forall x \in S \setminus E.$$

In particular,

$$|f(x)| \leqslant \|f\|_\infty \qquad \forall x \in S \setminus E.$$

From now on, given $1 < p < \infty$, we indicate by q the conjugate exponent of p defined by the identity

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

We begin with the following

Lemma 1.2.2 — Young's inequality. Let $p > 1$ and let $q > 1$ be its conjugate exponent. Then, for any nonnegative $a, b \in \mathbb{R}$, it holds

$$ab \leqslant \frac{1}{p}a^p + \frac{1}{q}b^q.$$

Proof. Though very simple, the result is very important for applications. The proof uses the convexity of the exponential function. Remember that a given function $f\ : \mathbb{R} \to \mathbb{R}$ is said to be convex if, for any $x, y \in \mathbb{R}$ and any $\theta \in [0,1]$,

$$f(\theta x + (1-\theta)y) \leqslant \theta f(x) + (1-\theta) f(y).$$

Moreover, if f is twice differentiable and $f''(x) \geqslant 0$ for any $x \in \mathbb{R}$, then f is convex. This in particular shows that the exponential function is convex over $\mathbb{R}$. The result is obvious if $a = 0$ or $b = 0$, so assume that $a > 0$ and $b > 0$. One has

$$ab = \exp\left(\log\left(ab\right)\right) = \exp\left(\log a + \log b\right).$$

Then, using the well-known multiplicative property of logarithm:

$$ab = \exp\left(\frac{1}{p}\log\left(a^p\right) + \frac{1}{q}\log\left(b^q\right)\right).$$

Since the exponential function is strictly increasing and convex, using $\frac{1}{p} + \frac{1}{q} = 1$ one gets

$$ab \leqslant \frac{1}{p}\exp\left(\log\left(a^p\right)\right) + \frac{1}{q}\exp\left(\log\left(b^q\right)\right) = \frac{a^p}{p} + \frac{b^q}{q}$$

and this ends the proof. ∎

Remark 1.2.2 For $p = 2$, one sees that $q = 2$ and Young's inequality simply states that

$$2ab \leqslant a^2 + b^2$$

which is an obvious consequence of the fact that $(a-b)^2 \geqslant 0$. The above Young's inequality can be seen as a generalization, for any $p > 1$, of this obvious estimate when $p = 2$. Moreover, given $\varepsilon > 0$, one sees that, replacing a with $a\varepsilon$ and b with b/ε, Young's inequality yields

$$ab \leqslant \frac{\varepsilon^p}{p} a^p + \frac{1}{q\,\varepsilon^q} b^q \qquad \forall a, b, \varepsilon > 0.$$

An important consequence of the above inequality is the following

Theorem 1.2.3 — Hölder inequality. Let $1 \leqslant p \leqslant \infty$ and $1/p + 1/q = 1$. Assume that $f \in L^p(S, \mu)$ and $g \in L^q(S, \mu)$. Then, $fg \in L^1(S, \mu)$ and

$$\|fg\|_1 \leqslant \|f\|_p \|g\|_q. \tag{1.3}$$

✍ *Proof.* The conclusion is obvious if $p = 1$ or $p = \infty$. We assume that $1 < p < \infty$. We recall Young's inequality:

$$ab \leqslant \frac{1}{p} a^p + \frac{1}{q} b^q \qquad \forall a, b > 0.$$

Applying this to $a = |f(s)|$ and $b = |g(s)|$, we get

$$|f(s)g(s)| \leqslant \frac{1}{p}|f(s)|^p + \frac{1}{q}|g(s)|^q \qquad \text{for } \mu\text{-a.e. on } S$$

Integrating this over S we get

$$\|fg\|_1 \leqslant \frac{1}{p}\|f\|_p^p + \frac{1}{q}\|g\|_q^q$$

which shows that $fg \in L^1(S, \mu)$. To prove (1.3), we replace f by λf (with $\lambda > 0$) in the above

$$\lambda \|fg\|_1 \leqslant \frac{\lambda^p}{p}\|f\|_p^p + \frac{1}{q}\|g\|_q^q \qquad \forall \lambda > 0,$$

ie.

$$\|fg\|_1 \leqslant \frac{\lambda^{p-1}}{p}\|f\|_p^p + \frac{1}{\lambda\,q}\|g\|_q^q \qquad \forall \lambda > 0.$$

Choosing $\lambda = \|f\|_p^{-1}\,\|g\|_q^{q/p}$ we get (1.3). ■

Actually, by induction, one can prove the following

Corollary 1.2.4 — Generalized Holder inequality. Given $n \in \mathbb{N}$, let $(p_1, \ldots, p_n) \in (\mathbb{R}_+)^n$ be such that

$$\sum_{i=1}^{n} \frac{1}{p_i} = 1.$$

For any $i = 1, \ldots, n$, let $f_i \in L^{p_i}(S, \mu)$. Then,

$$f = \prod_{i=1}^{n} f_i \in L^1(S, \mu) \qquad \text{and} \qquad \|f\|_1 \leqslant \prod_{i=1}^{n} \|f_i\|_{p_i}.$$

Proof. The proof is made by induction over n and is left as an **Exercise**. ■

Theorem 1.2.5 For any $1 \leqslant p \leqslant \infty$, $L^p(S,\mu)$ is a vector space and $\|\cdot\|_p$ is a norm.

Remark 1.2.3 For $1 < p < \infty$, the triangular inequality

$$\|f+g\|_p \leqslant \|f\|_p + \|g\|_p \qquad \forall f,g \in L^p(S,\mu)$$

is known as ***Minkowski's inequality***.

✍ *Proof.* The fact that, given $f \in L^p(S,\mu)$ one has $\lambda f \in L^p(S,\mu)$ is obvious as well as the homogeneity and uniqueness property for $\|\cdot\|_p$. So, to prove that $L^p(S,\mu)$ is a vector space, we just need to prove that, given $f,g \in L^p(S,\mu)$, the sum $f+g \in L^p(S,\mu)$ while proving that $\|\cdot\|_p$ is a norm amounts in proving the above Minkowski's inequality. Let us then fix $f,g \in L^p(S,\mu)$.

Since the function $r \mapsto r^p$ is convex over $\mathbb{R}^+$ ***(Check this !)*** one sees that, for any $x,y \in \mathbb{R}$ it holds that

$$|\tfrac{1}{2}x + \tfrac{1}{2}y|^p \leqslant \frac{1}{2}|x|^p + \frac{1}{2}|y|^p$$

or, equivalently, that

$$|x+y|^p \leqslant 2^{p-1}\left(|x|^p + |y|^p\right) \qquad \forall x,y \in \mathbb{R}. \tag{1.4}$$

This implies in particular that, for μ-a. e. $s \in S$

$$|f(s)+g(s)|^p \leqslant 2^{p-1}\left(|f(s)|^p + |g(s)|^p\right)$$

and, integrating over S this yields

$$\int_S |f+g|^p \mathrm{d}\mu \leqslant 2^{p-1}\left(\int_S |f|^p \mathrm{d}\mu + \int_S |g|^p \mathrm{d}\mu\right),$$

i.e. $f+g \in L^p(S,\mu)$ with

$$\|f+g\|_p^p \leqslant 2^{p-1}\left(\|f\|_p^p + \|g\|_p^p\right).$$

Let us now prove Minkowski's inequality. If $\|f+g\|_p = 0$ there is nothing to prove so let us assume $\|f+g\|_p \neq 0$. We write

$$\|f+g\|_p^p = \int_S |f+g|^p \mathrm{d}\mu = \int_S |f+g|\,|f+g|^{p-1}\mathrm{d}\mu$$

which results, since $|f+g| \leqslant |f| + |g|$, in

$$\|f+g\|_p^p \leqslant \int_S |f|\,|f+g|^{p-1}\mathrm{d}\mu + \int_S |g|\,|f+g|^{p-1}\mathrm{d}\mu. \tag{1.5}$$

Now, the function $\psi = |f+g|^{p-1}$ belongs to $L^q(S,\mu)$ where q is the conjugate exponent of p. Indeed,

$$|\psi|^q = \left(|f+g|^{p-1}\right)^q = |f+g|^p \qquad \text{since} \quad q(p-1) = p$$

and therefore

$$\|\psi\|_q = \left(\int_S |\psi|^q \mathrm{d}\mu\right)^{1/q} = \left(\int_S |f+g|^p \mathrm{d}\mu\right)^{1/q} = \|f+g\|_p^{\frac{p}{q}} < \infty. \tag{1.6}$$

This shows that $|\psi|^q \in L^1(S,\mu)$ i.e. $\psi \in L^q(S,\mu)$. Since $|f| \in L^p(S,\mu)$ one can apply Holder's inequality with $|f|$ and ψ to get

$$\int_S |f|\,|f+g|^{p-1}\mathrm{d}\mu = \|\,f\psi\,\|_1 \leqslant \|f\|_p\,\|\psi\|_q = \|f\|_p\,\|f+g\|_p^{\frac{p}{q}}$$

where we used (1.6). In the same way,

$$\int_S |g|\,|f+g|^{p-1}\mathrm{d}\mu \leqslant \|g\|_p\,\|f+g\|_p^{\frac{p}{q}}.$$

Thus, turning back to (1.5),

$$\|f+g\|_p^p \leqslant \|f\|_p\,\|f+g\|_p^{\frac{p}{q}} + \|g\|_p\,\|f+g\|_p^{\frac{p}{q}}$$

and, dividing by $\|f+g\|_p^{\frac{p}{q}} \neq 0$, we get

$$\|f+g\|_p^{p-\frac{p}{q}} \leqslant \|f\|_p + \|g\|_p$$

which is Minkowski's inequality since $p - \frac{p}{q} = 1$. ■

We illustrate the above results for the special case of general probability space $(\Omega,\mathcal{F},\mathbb{P})$. We have the following

Proposition 1.2.6 Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability triple. Then, the following holds

$$L^\infty(\Omega,\mathcal{F},\mathbb{P}) \subset L^p(\Omega,\mathcal{F},\mathbb{P}) \subset L^q(\Omega,\mathcal{F},\mathbb{P}) \subset L^1(\Omega,\mathcal{F},\mathbb{P}) \qquad \forall 1 \leqslant q \leqslant p \leqslant \infty.$$

Proof. We first prove that $L^\infty(\Omega,\mathcal{F},\mathbb{P}) \subset L^p(\Omega,\mathcal{F},\mathbb{P})$ for any $1 \leqslant p \leqslant \infty$. Let $X \in L^\infty(\Omega,\mathcal{F},\mathbb{P})$ be given. One has

$$\int_\Omega |X|^p\mathrm{d}\mathbb{P} \leqslant \|X\|_\infty^p \int_\Omega \mathrm{d}\mathbb{P} = \|X\|_\infty \mathbb{P}(\Omega) = \|X\|_\infty^p,$$

i.e $X \in L^p(\Omega,\mathcal{F},\mathbb{P})$ with $\|X\|_p \leqslant \|X\|_\infty$.

Let now assume that $X \in L^p(\Omega,\mathcal{F},\mathbb{P})$ and let $1 \leqslant q < p$. One has to compute $\int_\Omega |X|^q\mathrm{d}\mathbb{P}$. Set $Z = |X|^q$ and $r = p/q > 1$. Clearly, $Z \in L^r(\Omega,\mathcal{F},\mathbb{P})$. Moreover

$$\int_\Omega |X|^q\mathrm{d}\mathbb{P} = \int_\Omega Z\mathrm{d}\mathbb{P} = \int_\Omega ZY\mathrm{d}\mathbb{P}$$

where $Y = \mathbf{1}_\Omega$. Let $r' \geqslant 1$ such that $\frac{1}{r} + \frac{1}{r'} = 1$. Notice that $\int_\Omega Y^{r'}\mathrm{d}\mathbb{P} = \int_\Omega Y\mathrm{d}\mathbb{P} = \mathbb{P}(\Omega) = 1$ so that $Y \in L^{r'}(\Omega,\mathcal{F},\mathbb{P})$. According to Holder inequality $ZY \in L^1(\Omega,\mathcal{F},\mathbb{P})$ with

$$\int_\Omega |X|^q\mathrm{d}\mathbb{P} = \|ZY\|_1 \leqslant \|Z\|_r\|Y\|_{r'} = \|Z\|_r.$$

This proves that $X \in L^q(\Omega,\mathcal{F},\mathbb{P})$ with $\|X\|_q \leqslant \|Z\|_r^{1/q} = \|X\|_p$. ■

Remark 1.2.4 Notice that the above result remains valid if (S, Σ, μ) is a given measure space with $\mu(S) < \infty$.

In the special case in which

$$S = \mathbb{N}, \qquad \mathcal{F} = \mathcal{P}(\mathbb{N})$$

and μ is the counting measure defined by

$$\mu(A) = \sum_{k \in A} \delta_k(A), \quad A \subset \mathbb{N},$$

using the classical identification between functions f over $\mathbb{N}$ and sequences $\boldsymbol{x} = (x_n)_n$ given by

$$\begin{cases} f \ : & \mathbb{N} \to \mathbb{R} \\ & n \mapsto f(n) =: x_n \end{cases}$$

we see that

$$L^1(S,\mu) = \mathcal{L}^1(S,\mu) = \ell^1(\mathbb{N}) = \left\{ \boldsymbol{x} = (x_n)_n;\ \|\boldsymbol{x}\|_1 := \sum_{n=1}^{\infty} |x_n| < \infty \right\}$$

i.e. $\ell^1(\mathbb{N})$ is nothing that some $L^1(S,\mu)$ for some special choice of S and μ. Notice that the counting measure μ is indeed such that

$$\int_{\mathbb{N}} |f(n)| \mathrm{d}\mu(n) = \sum_{n=1}^{\infty} |f(n)| = \sum_{n=1}^{\infty} |x_n|$$

with the identification above. This allows in particular to define and prove the following

Proposition 1.2.7 Let $p \geqslant 1$ be given. We define the set

$$\ell^p(\mathbb{N}) = \left\{ \boldsymbol{x} = (x_n)_{n \in \mathbb{N}} \subset \mathbb{R} \text{ such that } \sum_{n=1}^{\infty} |x_n|^p < \infty \right\}.$$

Then, $\ell^p(\mathbb{N})$ is a vector space. Moreover, if

$$\|\boldsymbol{x}\|_p := \left(\sum_{n=1}^{\infty} |x_n|^p \right)^{\frac{1}{p}} \qquad \forall \boldsymbol{x} = (x_n)_n \in \ell^p(\mathbb{N})$$

then, $(\ell^p(\mathbb{N}), \|\cdot\|_p)$ is a normed space.

Exercise 1.2 Let $X = \mathscr{C}([0,1])$ be the space of all continuous functions $f : [0,1] \to \mathbb{R}$. For any $p > 1$, define

$$\|f\|_p = \left(\int_0^1 |f(t)|^p \mathrm{d}t \right)^{\frac{1}{p}}, \qquad f \in X.$$

Use Young's inequality to prove that $(X, \|\cdot\|_p)$ is a normed space. Show also that, if $f, g \in X$, then

$$\|fg\|_1 \leqslant \|f\|_p \|g\|_q \qquad \frac{1}{p} + \frac{1}{q} = 1.$$

1.3 Another example: the space of linear applications

In normed space, there is a very useful characterization of continuous linear applications: recall that, given X, Y two vector spaces, a mapping $L : X \to Y$ is said to be linear if

$$L(\lambda x + y) = \lambda L(x) + L(y) \qquad \forall x, y \in X,\ \lambda \in \mathbb{R}.$$

One has the following

Proposition 1.3.1 Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be two normed spaces and let $L : X \to Y$ be a linear application. The following are equivalent:

1. L is continuous on X;
2. L is continuous at $x = 0$;
3. there is a positive constant $C > 0$ such that

$$\|L(x)\|_Y \leqslant C\|x\|_X \qquad \forall x \in X.$$

Proof. It is clear that $i) \implies ii)$. Assume now that L is continuous at 0 and let us apply the definition of continuity with $\varepsilon = 1$. There exists a positive constant $\delta > 0$ such that $\|x - 0_X\|_X \leqslant \delta \implies \|L(x) - L(0_Y)\|_Y < \varepsilon = 1$. Since L is linear, $L(0_X) = 0_Y$ so that it holds

$$\|x\|_X \leqslant \delta \implies \|L(x)\|_Y < 1.$$

Let now $x \in X \setminus \{0_X\}$ be given. Set $y = \frac{\delta}{\|x\|_X} x$. By the homogeneity of the norm, $\|y\|_X = \delta$ so that $\|L(y)\|_Y < 1$. Using the linearity of L and again the homogeneity of the norm, this reads

$$\|L(x)\|_Y < \frac{1}{\delta}\|x\|$$

and proves $iii)$ with $C = \frac{1}{\delta}$. Let us prove now that $iii) \implies i)$. By linearity, for any $x, y \in X$, it holds

$$\|L(x) - L(y)\|_X = \|L(x - y)\|_X \leqslant C\|x - y\|_Y$$

and this clearly implies the continuity of L at any $x \in X$ (actually, L is uniformly continuous on X). ■

Exercise 1.3 Assume that X, Y are two finite dimensional spaces (not necessarily with $\dim X = \dim Y$). Prove that any linear application $L : X \to Y$ is continuous.

Definition 1.3.1 If $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ are two normed spaces, we denote by $\mathscr{L}(X, Y)$ the space of continuous linear applications from X to Y. If $Y = X$, we simply denote $\mathscr{L}(X) = \mathscr{L}(X, X)$.

■ **Example 1.8** Given two normed spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$, let us define

$$\|L\|_{\mathscr{L}(X,Y)} = \|L\|_{\text{op}} = \sup_{\|x\|_X \leqslant 1} \|L(x)\|_Y \qquad \forall L \in \mathscr{L}(X,Y).$$

Then, according to Proposition 1.3.1, $\|L\|_{\mathscr{L}(X,Y)}$ is finite for any $L \in \mathscr{L}(X,Y)$. With this, one check easily as an **Exercise** that $\|\cdot\|_{\mathscr{L}(X,Y)}$ is a norm on $\mathscr{L}(X,Y)$. Moreover, one can check ***(Do it!)*** that, by definition,

$$\|L(x)\|_Y \leqslant \|L\|_{\mathscr{L}(X,Y)} \|x\|_X \qquad \forall x \in X.$$

■

Proposition 1.3.2 If $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ are two vector spaces and X is of *finite dimension*, prove that any linear application $L : X \to Y$ is continuous.

Proof. The proof is a relatively simple **Exercise**. ■

Remark 1.3.1 If $\dim(X) = n$ and $\dim(Y) = p$, the space $\mathcal{L}(X,Y)$ can be identified with the space $\mathcal{M}_{n\times p}(\mathbb{R})$ of matrices with n lines and p rows. ***(Explain why)***.

■ **Example 1.9** Let $X = \mathscr{C}([0,1]) = \mathscr{C}([0,1],\mathbb{R})$ be endowed with the sup-norm

$$\|f\|_\infty = \sup_{t\in[0,1]} |f(t)| \qquad \forall f \in X.$$

Introduce the linear mapping $L : X \to X$ defined by

$$L(f)(x) = \int_0^x f(t)\mathrm{d}t \qquad \forall x \in [0,1],\ f \in X.$$

One checks easily that L is well-defined (i.e. $L(f) \in X$ for any $f \in X$) and linear. Moreover, for any $x \in [0,1]$, it holds

$$|L(f)(x)| \leqslant \int_0^x |f(t)|\mathrm{d}t \leqslant \|f\|_\infty \int_0^x \mathrm{d}t = x\|f\|_\infty \leqslant \|f\|_\infty.$$

This means that $\|L(f)\|_\infty \leqslant \|f\|_\infty$ for any $f \in X$. Therefore, L is continuous (Prop. 1.3.1) and $\|L\|_{\mathscr{L}(X)} \leqslant 1$. ■

Exercise 1.4 Prove that, with the notations of the previous example, one exactly has $\|L\|_{\mathscr{L}(X)} = 1$. (*Hint:* Exhibit $f \in X$ with $\|f\|_\infty = 1 = \|L(f)\|_\infty$). ■

Exercise 1.5 Let us again consider $X = \mathscr{C}([0,1])$ and endowed it now with the norm $\|\cdot\|_1$ defined in Proposition 1.1.1. Prove that the mapping $L : X \to \mathbb{R}$ defined by $L(f) = f(0)$ is linear but not continuous (where $\mathbb{R}$ is endowed with its usual norm), i.e. $L \notin \mathscr{L}(X,\mathbb{R})$. What happens if we endow X with $\|\cdot\|_\infty$? ■

Exercise 1.6 Let $X = \mathscr{C}^1([0,1])$ be endowed with the sup-norm

$$\|f\|_\infty = \sup_{t\in[0,1]} |f(t)|, \qquad f \in X.$$

Let $T : f \in X \mapsto T(f) = f'(0) \in \mathbb{R}$. Considering the sequence of functions

$$f_n(x) = \frac{\sin(n^2 x)}{n}, \qquad x \in [0,1], \quad n \in \mathbb{N},$$

prove that $T \notin \mathscr{L}(X, \mathbb{R})$. ■

1.4 Compactness

We introduce the concept of compact subset of a normed space:

Definition 1.4.1 Let $(X, \|\cdot\|)$ be a normed space and let $K \subset X$. We say that K is compact if every sequence $(x_n)_n \subset K$ contains a subsequence which converges to some $x \in K$.

Remark 1.4.1 Of course, if K is compact then K is closed.

One has the following first property

Lemma 1.4.1 If K is a compact subset of a normed space $(X, \|\cdot\|)$, then K is closed and there exists $M > 0$ such that $\sup_{x\in K} \|x\| \leqslant M$, i.e. K is bounded.

Proof. The fact that a compact set is closed is clear from the definition. Let us now prove that K is bounded. Assume the contrary, i.e.

$$\sup_{x\in K} \|x\| = \infty.$$

This means that there is a sequence $(x_n)_n \subset K$ such that $\lim_n \|x_n\| = \infty$. Being K compact, there is a subsequence $(x_{\varphi(n)})_n$ of $(x_n)_n$ which converges to some $x \in K$. We know that then $\sup_n \|x_{\varphi(n)}\| < \infty$ which contradicts the previous convergence to infinity. ■

Proposition 1.4.2 — Product of compact spaces. Let $(X_1, \|\cdot\|_1)$ and $(X_2, \|\cdot\|_2)$ be two compact normed spaces and let $X = X_1 \times X_2$. Then, $(X, \|\cdot\|_+)$ and $(X, \|\cdot\|_{\max})$ are compact normed spaces where $\|\cdot\|_+$ and $\|\cdot\|_{\max}$ have been defined in Proposition 1.1.10.

Proof. Let $(\xi_n)_n \subset X$ be a given sequence. It means that there are two sequences $(x_n)_n \subset X_1$ and $(y_n)_n \subset X_2$ such that $\xi_n = (x_n, y_n)$ for any $n \in \mathbb{N}$. Since $(X_1, \|\cdot\|_1)$ is compact, there is a subsequence $(x_{\varphi(n)})_n \subset X_1$ which converges to $x \in X_1$. Consider then the sequence $(y_{\varphi(n)})_n \subset X_2$. Since $(X_2, \|\cdot\|_2)$ is compact, there is a subsequence $(y_{\phi(\varphi(n))})_n$ of $(y_{\varphi(n)})_n$ which converges to $y \in X_2$. Then, the subsequence $(x_{\phi(\varphi(n))})_n$ of $(x_{\varphi(n)})_n$ is still convergent in X_1 to x and one sees easily that the subsequence

$$(\xi_{\phi(\varphi(n))})_n = (x_{\phi(\varphi(n))}, y_{\phi(\varphi(n))})_n \subset X_1 \times X_2$$

is converging in X to $\boldsymbol{x} = (x, y)$. This proves that X is compact. ■

Remark 1.4.2 Of course, the above result readily extends to any finite product of compact normed spaces.

On $\mathbb{R}$, it is easy to describe a large class of compact sets:

Lemma 1.4.3 Let $\mathbb{R}$ be endowed with the absolute value, $|\cdot|$. Any interval $[a,b] \subset \mathbb{R}$ is compact.

Proof. Without loss of generality, we only prove the compactness of $I = [0,1]$. Let $(x_n)_n$ be a given sequence in I. Let us define two sequences $(a_n)_n$ and $(b_n)_n$ inductively as follows: set $a_0 = 0$ and $b_0 = 1$ and assume $a_1 \leqslant \ldots \leqslant a_n \leqslant b_n \ldots \leqslant b_1$ to be constructed. Then, one sets

$$a_{n+1} = a_n \qquad \text{and} \qquad b_{n+1} = \frac{a_n + b_n}{2} \quad \text{if } \left\{ k \in \mathbb{N}\,;\, x_k \in \left[a_n, \frac{a_n+b_n}{2}\right]\right\} \text{ is infinite}$$

otherwise, one sets $a_{n+1} = \frac{a_n+b_n}{2}$ and $b_{n+1} = b_n$. With such a construction, the sequence $(a_n)_n$ is nondecreasing and the sequence $(b_n)_n$ is nonincreasing. Moreover, for any $n \in \mathbb{N}$,

$$|a_n - b_n| = 2^{-n} \qquad \text{and} \qquad \{k \in \mathbb{N}\,;\, x_k \in [a_n, b_n]\} \text{ is infinite.}$$

One sees that then that the two sequence are convergent to some common limit x ***(Check all these details)***. For any $\varepsilon > 0$ there exists $n \in \mathbb{N}$ such that

$$x - \varepsilon \leqslant a_n \leqslant b_n \leqslant x + \varepsilon.$$

In particular, $\{k \in \mathbb{N}\,;\, x_k \in (x-\varepsilon, x+\varepsilon)\}$ is infinite so that x is a limit point of $(x_n)_n$ (see Exercise 1.1). This shows that I is compact . ■

Proposition 1.4.4 Let $(X, \|\cdot\|)$ be a normed space and let K be a compact subset of X. If $A \subset K$ is a closed subset then A is compact.

Proof. The proof is left as an easy **Exercise**. ■

Corollary 1.4.5 — Heine-Borel Theorem. A subset K of $\mathbb{R}^N$ (where $\mathbb{R}^N$ is endowed with, say, the usual Euclidean normed) is compact if and only if it is closed and bounded.

✍ *Proof.* The fact that a compact subset of $\mathbb{R}^N$ is closed and bounded is a well-know property true in any normed space (see Lemma 1.4.1). Conversely, let $K \subset \mathbb{R}^N$ be closed and bounded. Being bounded, there exists $R > 0$ such that

$$K \subset [-R, R]^N.$$

Since K is closed, from the previous Proposition, it suffices to prove that $[-R,R]^N$ is a compact subset of $\mathbb{R}^N$. This is a simple consequence of the fact that $[-R,R]$ is a compact subset of $\mathbb{R}$ (Lemma 1.4.3) together with Corollary 1.4.2. ■

Remark 1.4.3 The previous Corollary can be reformulated as follows: every bounded sequence of $\mathbb{R}^N$ has a convergent subsequence.

1.4.1 Compactness and continuous functions

Continuous functions over compact normed spaces enjoy nice and useful properties. We first state the following

Proposition 1.4.6 Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ and let $f : X \to Y$ be continuous. If $K \subset X$ is a compact subset of X then $f(K)$ is a compact subset of Y.

Proof. Let $(y_n)_n$ be a sequence of $f(K)$. It means that there is a sequence $(x_n)_n \subset K$ such that $y_n = f(x_n)$ for any n. Being K compact, there is a subsequence $(x_{\varphi(n)})_n$ of $(x_n)_n$ which converges to $x \in K$, i.e. $\lim_n \|x_{\varphi(n)} - x\|_X = 0$. Being f continuous, one has $\lim_n \|f(x_{\varphi(n)}) - f(x)\|_Y = 0$, i.e. $\lim_n \|y_{\varphi(n)} - f(x)\|_Y = 0$. Since $y = f(x) \in f(K)$, we get that any sequence of $f(K)$ has a subsequence which converges to a limit in $f(K)$, i.e. $f(K)$ is compact. ■

As a consequence

Theorem 1.4.7 Let $(X, \|\cdot\|)$ be a normed space and let $K \subset X$ be compact. Let $f : K \to \mathbb{R}$ be continous. Then, f assumes its maximum and minimum on K.

Proof. According to Proposition 1.4.6, $f(K)$ is a compact subset of $\mathbb{R}$ and, from Heine-Borel Theorem, $f(K)$ is bounded and closed. This shows that f is bounded. The fact that it reaches is minimum and maximum value is a simple consequence of the fact that $f(K)$ is closed. Indeed, let

$$M = \sup\{f(x)\,,\, x \in K\}.$$

By definition, there is a sequence $(x_n)_n \subset K$ such that $\lim_n f(x_n) = M$. The sequence $(f(x_n))_n$ lies in $f(K)$ which is compact and os is closed. Thus, its limit M also lies in $f(K)$, i.e. there is $x \in K$ such that $f(x) = M$. This shows that M is a maximum value of f. One proceeds in the same way with the minimum value. ■

One also has

Theorem 1.4.8 — Heine Theorem. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be two normed spaces and let $K \subset X$ be compact. Assume that $f : K \to Y$ is continuous. Then, f is uniformly continuous on K.

Proof. Suppose f is not uniformly continuous. This means that there exists $\varepsilon_0 > 0$ such that

$$\forall \delta > 0, \exists x, y \in K, \quad \text{with } \|x - y\|_X < \delta \quad \text{and} \quad \|f(x) - f(y)\|_Y \geqslant \varepsilon_0.$$

Now choosing $\delta = \frac{1}{n}, n \in \mathbb{N}$ this allows to build two sequences $(x_n)_n$ and $(y_n)_n$ such that

$$\|x_n - y_n\|_X < \frac{1}{n} \qquad \text{and} \qquad \|f(x_n) - f(y_n)\|_Y \geqslant \varepsilon_0 \qquad \forall n \in \mathbb{N}.$$

Since K is compact, then we can extract a subsequence of x_n, that we call $(x_{\varphi(n)})_n$, which converges to some $x_0 \in K$. It follows that $(y_{\varphi(n)})_n$ converges also to x_0 ***(Explain this)***.

Because f is continuous we get that $\lim_{n\to\infty} f(x_{\varphi(n)}) = f(x_0) = \lim_{n\to\infty} f(y_{\varphi(n)})$ in Y, i.e.

$$\lim_{n\to\infty} \|f(x_{\varphi(n)}) - f(y_{\varphi(n)})\|_Y = 0$$

but this contradicts the fact that $\|f(x_{\varphi(n)}) - f(y_{\varphi(n)})\|_Y \geqslant \epsilon_0$ for each $n \in \mathbb{N}^+$. ■

1.5 Finite Dimensional Spaces

We aim here to characterize the compact subsets of finite dimensional normed spaces. We begin with the following

Proposition 1.5.1 Let $(X, \|\cdot\|)$ be an finite dimensional normed vector space, $\dim X = d$ and let $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\}$ be a basis for X. Then, there are positive constants $C_0, C_1 > 0$ such that

$$C_0 \sum_{i=1}^{d} |x_i| \leqslant \left\| \sum_{i=1}^{d} x_i \boldsymbol{e}_i \right\| \leqslant C_1 \sum_{i=1}^{d} |x_i| \qquad \forall\, (x_1, \ldots, x_d) \in \mathbb{R}^d.$$

Proof. We aim to prove that there are positive constants $C_0, C_1 > 0$ such that

$$C_0 \|\underline{x}\|_1 \leqslant \|\boldsymbol{x}\| \leqslant C_1 \|\underline{x}\|_1, \qquad \forall \boldsymbol{x} \in X$$

where $\boldsymbol{x}$ is written on the basis $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\}$ as $\boldsymbol{x} = \sum_{i=1}^{d} x_i \boldsymbol{e}_i$ and where we denote by $\|\underline{x}\|_1 = \sum_{i=1}^{d} |x_i|$ the norm of the vector $\underline{x} = (x_1, \ldots, x_d)$ of $\mathbb{R}^d$. First, using the triangle inequality, one has

$$\|\boldsymbol{x}\| = \| \sum_{i=1}^{d} x_i \boldsymbol{e}_i \| \leqslant \sum_{i=1}^{d} \|x_i \boldsymbol{e}_i\| = \sum_{i=1}^{d} |x_i| \|\boldsymbol{e}_i\|.$$

Setting $C_1 = \max_{1 \leqslant j \leqslant d} \|\boldsymbol{e}_j\|$, one has $C_1 < \infty$ since we have a finite number of vectors $\boldsymbol{e}_j$ and clearly

$$\|\boldsymbol{x}\| \leqslant C_1 \sum_{i=1}^{d} |x_i| = C_1 \|\underline{x}\|_1.$$

Let us prove the converse inequality. We denote by $\mathbf{S}_1$ the closed unit sphere of $(\mathbb{R}^d, \| \cdot \|_1)$, i.e.

$$\mathbf{S}_1 = \left\{ \underline{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d \, ; \, \|\underline{x}\|_1 = 1 \right\}.$$

Since $\mathbf{S}_1$ is closed and bounded in $(\mathbb{R}^d, \| \cdot \|_1)$, it is compact. Consider the mapping

$$f \,:\, \mathbb{R}^d \to \mathbb{R}^+$$

given by

$$f(\underline{x}) = \left\| \sum_{i=1}^{d} x_i \boldsymbol{e}_1 \right\|, \qquad \forall \underline{x} = (x_1, \ldots, x_d).$$

Using the continuity of the norm $\| \cdot \|$ (as an application from X to $\mathbb{R}$) we see easily that f is continuous over $\mathbb{R}^d$ (i.e. if $\{\underline{x}_n\}_n$ converges in $\mathbb{R}^d$ to $\underline{x}$ then $\lim_n f(\underline{x}_n) = f(\underline{x})$ in $\mathbb{R}$). Being $\mathbf{S}_1$ compact, Heine-Borel Theorem (Corollary 1.4.5) asserts that

$$\inf_{\underline{x} \in \mathbf{S}_1} f(\underline{x}) = \min_{\underline{x} \in \mathbf{S}_1} f(\underline{x}) = C_0 \geqslant 0$$

i.e. there exists $\underline{\boldsymbol{\alpha}} = (\alpha_1, \ldots, \alpha_d) \in \mathbf{S}_1$ such that

$$C_0 = f(\underline{\boldsymbol{\alpha}}).$$

Let us prove that $C_0 > 0$. Otherwise, $\|\sum_{i=1}^d \alpha_i \boldsymbol{e}_i\| = 0$ i.e. $\sum_{i=1}^d \alpha_i \boldsymbol{e}_i = 0$ (the null element of X). Being the family $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\}$ a basis of X, the vectors $\boldsymbol{e}_i$ are linearly independent and this means that $\alpha_i = 0$ for any $i = 1, \ldots, d$. This is impossible since $\|\underline{\boldsymbol{\alpha}}\|_1 = \sum_{i=1}^d |\alpha_i| = 1$. Therefore $C_0 > 0$ and, by definition, for any vector $\underline{x} = (x_1, \ldots, x_d) \in \mathbf{S}_1$, one has

$$\left\|\sum_{i=1}^d x_i \boldsymbol{e}_i\right\| \geqslant C_0.$$

This proves the desired inequality on the sphere $\mathbf{S}_1$. To deduce the inequality over $\mathbb{R}^d$, we just consider, for any $\underline{x} \in \mathbb{R}^d$ non zero the normalized vector

$$\underline{y} = (y_1, \ldots, y_d) = \frac{1}{\|\underline{x}\|_1}(x_1, \ldots, x_d) \in \mathbf{S}_1$$

so that $\left\|\sum_{i=1}^d y_i \boldsymbol{e}_i\right\| \geqslant C_0$ which reads

$$\frac{1}{\|\underline{x}\|_1}\left\|\sum_{i=1}^d x_i \boldsymbol{e}_i\right\| \geqslant C_0$$

and the proof is achieved. ■

The above Proposition actually asserts that, if $\dim(X) = d$, *any* norm of $\|\cdot\|$ is "related" to the $\|\cdot\|_1$ norm of $\mathbb{R}^d$. This easily translates in the following

Proposition 1.5.2 If X is a finite dimensional vector space, all norms over X are equivalent.

Proof. As before, assume $\dim(X) = d$ and let $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\}$ be a basis of $\mathbb{R}^d$. Let $\boldsymbol{x} \in X$ be written on the basis $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\}$ as $\boldsymbol{x} = \sum_{i=1}^d x_i \boldsymbol{e}_i$ and denote by $\|\underline{x}\|_1 = \sum_{i=1}^d |x_i|$ the norm of the vector $\underline{x} = (x_1, \ldots, x_d)$ of $\mathbb{R}^d$. Let $\mathcal{N}_1$ and $\mathcal{N}_2$ be two norms on X. From the previous proposition (applied to $(X, \mathcal{N}_1)$) there are positive constants $C_0, C_1 > 0$ such that

$$C_0\|\underline{x}\|_1 \leqslant \mathcal{N}_1(\boldsymbol{x}) \leqslant C_1\|\underline{x}\|_1 \qquad \forall \boldsymbol{x} \in X$$

whereas there are two positive constants $C_0', C_1' > 0$ such that

$$C_0'\|\underline{x}\|_1 \leqslant \mathcal{N}_2(\boldsymbol{x}) \leqslant C_1'\|\underline{x}\|_1 \qquad \forall \boldsymbol{x} \in X.$$

It is easy to see then that

$$\mathcal{N}_1(\boldsymbol{x}) \leqslant \frac{C_1}{C_0'}\mathcal{N}_2(\boldsymbol{x}) \qquad \text{and} \qquad \mathcal{N}_2(\boldsymbol{x}) \leqslant \frac{C_1'}{C_0}\mathcal{N}_2(\boldsymbol{x})$$

which proves the result. ■

The above Proposition also allows to identify – in a continuous way – a finite dimension space $(X, \|\cdot\|)$ and the space $\mathbb{R}^d$ (d being the dimension of X). Indeed, introducing a basis $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\}$ a basis of X, the mapping

$$\Phi \, : \, X \to \mathbb{R}^d$$

which, to some $\boldsymbol{x} = \sum_{i=1}^d x_i \boldsymbol{e}_i \in X$ associates $\Phi(\boldsymbol{x}) = \underline{x} = (x_1, \ldots, x_d)$, we see that Φ is a bijection from X to $\mathbb{R}^d$ which is continuous whose inverse is also continuous. This results in the following whose proof is a simple ***Exercise***.

Corollary 1.5.3 If $(X, \|\cdot\|)$ is a finite dimensional vector space and $K \subset X$ is closed and bounded then K is compact.

Again, this is very specific to *finite dimensional spaces* and, as we shall see, this actually characterizes finite dimensional spaces. Indeed, in infinite dimensional normed spaces, the closed unit ball cannot be compact. This shows that, in infinite dimensional spaces, the compact subsets do not coincide with closed and bounded subsets !!

We first state the following technical lemma:

Lemma 1.5.4 — Riesz Lemma. Let $(X, \|\cdot\|)$ be a normed vector space and let Y be a closed subspace of X (i.e. Y is closed in X and Y is a linear subspace of X). If $Y \neq X$ then, for any $\varepsilon \in (0,1)$, there exists $x \in X$ with $\|x\| = 1$ such that

$$\inf_{y \in Y} \|x - y\| \geqslant 1 - \varepsilon.$$

Remark 1.5.1 The Lemma asserts that, if $Y \neq X$ is a closed subspace, then, for any $\varepsilon \in (0,1)$, there is some unit vector $x \in X$ such that $\mathrm{dist}(x, Y) \geqslant 1 - \varepsilon$.

✍ *Proof.* Let $z \in X \setminus Y$. Since Y is closed and $z \notin Y$, one has

$$\alpha = \mathrm{dist}(z, Y) = \inf_{y \in Y} \|z - y\| > 0$$

(Explain why!). Pick $\varepsilon \in (0,1)$. There exists $\overline{y} \in Y$ such that $\|\overline{y} - z\| \leqslant \frac{\alpha}{1-\varepsilon}$ (otherwise, we would get $\mathrm{dist}(z, Y) \geqslant \frac{\alpha}{1-\varepsilon} > \alpha$!). Notice that, $\overline{y} \neq z$ so that $r := \|\overline{y} - z\| > 0$. Set

$$x := \frac{1}{r}(z - \overline{y}).$$

Clearly, $\|x\| = 1$. Let $y \in Y$ be given. One can write

$$\|x - y\| = \frac{1}{r}\|z - \overline{y} - r\,y\|$$

and, since Y is a linear subspace, $\overline{y} + ry \in Y$ so that $\|z - \overline{y} - ry\| \geqslant \alpha$. Therefore $\|x - y\| \geqslant \frac{\alpha}{r} \geqslant 1 - \varepsilon$ by assumption on $r = \|z - \overline{y}\|$. Since this is true for any $y \in Y$, this proves the result. ∎

We also need the following

Lemma 1.5.5 If $(X, \|\cdot\|)$ is a normed space, any linear subspace of finite dimension is closed.

Proof. Let $V \subset X$ be of finite dimension, say $\dim(V) = d$ and let $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\}$ be a basis of V. Let $\{\boldsymbol{x}_n\}$ be a sequence in V which converges to some $\boldsymbol{x} \in X$. In particular, $\sup_n \|\boldsymbol{x}_n\| < \infty$. From Proposition 1.5.1, if we introduce the components vectors $\underline{x}_n \in \mathbb{R}^d$ such that $\underline{x}_n = (x_1^{(n)}, \ldots, x_d^{(n)})$ we see that the sequence $\{\underline{x}_n\}_n$ is bounded in $(\mathbb{R}^d, \|\cdot\|_1)$. In particular, it admits a subsequence $\{\underline{x}_{\varphi(n)}\}_n$ converges to some $\underline{x} = (x_1, \ldots, x_d)$ in $(\mathbb{R}^d, \|\cdot\|_1)$. Using again Proposition 1.5.1, we deduce that $\{\underline{x}_{\varphi(n)}\}_n$ converges to the element $\boldsymbol{y} = \sum_{i=1}^d x_i \boldsymbol{e}_i$. Since it also converges to $\boldsymbol{x}$ we get

$$\boldsymbol{x} = \sum_{i=1}^d x_i \boldsymbol{e}_i$$

and in particular $\boldsymbol{x} \in V$. This proves that V is closed. ■

We can prove the following fundamental result

Theorem 1.5.6 — Riesz Theorem. A normed space $(X, \|\cdot\|)$ is finite dimensional if and only if the closed unit ball $B_c(0,1) = \{x \in X \; ; \; \|x\| \leqslant 1\}$ of X is compact.

✍ *Proof.* We already saw that in finite dimensional spaces the closed and bounded subsets are compact. Let us assume that $B_c(0,1)$ is compact. Argue by contradiction assuming that X is infinite dimensional. Let us then pick $\boldsymbol{e}_0 \in X$ with $\|\boldsymbol{e}_0\| = 1$ and let $F_0 = \mathrm{Span}(\boldsymbol{e}_0) = \{t\boldsymbol{e}_0\,,\, t \in \mathbb{R}\}$. According to the previous Lemma, F_0 is a closed linear subspace of X (since it is of dimension 1). Being X infinite dimensional, $F_0 \neq X$. According to Riesz Lemma, there exists $\boldsymbol{e}_1 \notin F_0$ such that $\|\boldsymbol{e}_1\| = 1$ and $\min_{x \in F_0} \|x - \boldsymbol{e}_1\| \geqslant 1/2$. Set then $F_1 := \mathrm{Span}(\boldsymbol{e}_0, \boldsymbol{e}_1)$. One constructs inductively a sequence [2] $(\boldsymbol{e}_n)_n \subset X$ such that $\|\boldsymbol{e}_n\| = 1$ and such that, the finite dimensional space $F_n = \mathrm{Span}(\boldsymbol{e}_0, \ldots, \boldsymbol{e}_n)$ satisfies

$$\boldsymbol{e}_n \notin F_{n-1}, \qquad \text{and} \inf_{x \in F_{n-1}} \|x - \boldsymbol{e}_n\| \geqslant \frac{1}{2}.$$

In particular, $(\boldsymbol{e}_n)_n \subset B_c(0,1)$ which is compact so that a subsequence of $(\boldsymbol{e}_n)_n$ should converge. But, for any $n > m \in \mathbb{N}$, $\|\boldsymbol{e}_n - \boldsymbol{e}_m\| \geqslant \frac{1}{2}$ (since $\boldsymbol{e}_n \notin F_m$). This is a contradiction. ■

Remark 1.5.2 Of course, in the above statement, the choice of the *unit* ball is arbitrary. You can check easily ***(Do it !)*** that the closed unit ball is compact if and only if every closed ball is compact which, again, is equivalent to the fact that any closed and bounded subset of X is compact.

1.6 Problems

Exercise 1.7 Let $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ two normed spaces and $L : E \to F$ a linear mapping such that $(L(x_n))_n$ *is bounded in* F *for any sequence* $(x_n)_n \subset E$ *which converges to* $0 \in E$. Prove that L is continuous. ■

Exercise 1.8 Let $(X, \|\cdot\|)$ be a normed space. Let $Y \subset X$ be a linear subspace with $Y \neq X$. Prove that $\mathrm{Int}(Y) = \varnothing$. ■

Exercise 1.9 Compute the norm of the following linear applications in $\mathscr{L}(X)$:
- For $X = \ell^\infty(\mathbb{N})$ with the usual norm, consider the shift-mapping defined by $S : \boldsymbol{x} = (x_n)_n \mapsto S(\boldsymbol{x})$ where $S_{n+1}(\boldsymbol{x}) = x_n$ and $S_0(\boldsymbol{x}) = 0$.
- For $X = \mathscr{C}([0,1])$ endowed with $\|\cdot\|_\infty$, consider $Tf(x) = f(x)g(x)$ where $g \in X$ is given.

■

[2] Notice that this is an infinite sequence since the space X is of infinite dimension so that, for any $n \in \mathbb{N}$, $F_n \neq X$

Exercise 1.10 Compute the nom of the following linear applications in $\mathscr{L}(X,\mathbb{R})$:

- For $X = \mathscr{C}([0,1])$ endowed with $\|\cdot\|_\infty$, consider $T(f) = \int_0^1 f(x)g(x)\,dx$ where $g \in X$ is a given mapping with $g(1/2) = 0$ and $g(x) \neq 0$ for any $x \neq 1/2$.
- For $X = \ell^2(\mathbb{N})$ with the usual norm, consider the mapping
$$u(\boldsymbol{x}) = \sum_{n=1}^{\infty} a_n x_n$$
for some $(a_n)_n \in X$.
- For $X = \ell^1(\mathbb{N})$ with the usual norm, consider the mapping
$$u(\boldsymbol{x}) = \sum_{n=1}^{\infty} a_n x_n$$
for some $(a_n)_n \in \ell^\infty(\mathbb{N})$.
- For X the set of all real convergent sequences $\boldsymbol{x} = (x_n)_n$ endowed with the norm $\|\boldsymbol{x}\|_\infty = \sup_n |x_n|$, consider the mapping $u(\boldsymbol{x}) = \lim_k x_k \in \mathbb{R}$.

Preamble We admit in the first two exercises the change of variables formula (polar coordinate) in $\mathbb{R}^N$: if $F(\boldsymbol{x}) = \Phi(\|\boldsymbol{x}\|)$ depends only on the norm $\|\boldsymbol{x}\|$ (we say F is a radial function), then [3]
$$\int_{\mathbb{R}^N} F(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = |\mathbb{S}^{N-1}| \int_0^\infty \Phi(\varrho)\varrho^{N-1}\mathrm{d}\varrho$$
where $|\mathbb{S}^{N-1}|$ is the surface area the unit sphere $\mathbb{S}^{N-1} = \{\boldsymbol{y} \in \mathbb{R}^N \,;\, \|\boldsymbol{y}\| = 1\}$ given by
$$|\mathbb{S}^{N-1}| = \frac{2\pi^{N/2}}{\Gamma(N/2)}$$
where Γ is the Gamma function.

Exercise 1.11 Let $1 \leqslant p < \infty$ be given.

1. For which values of α does the mapping
$$x \in \mathbb{R}^N \mapsto \frac{1}{(1+\|x\|^2)^\alpha}$$
belong to $L^p(\mathbb{R}^n)$?
2. For which values of β does the mapping
$$x \in \mathbb{R}^N \mapsto \frac{1}{\|x\|^\beta} e^{-\frac{\|x\|^2}{2}}$$
belong to $L^p(\mathbb{R}^n)$?

[3] More generally, if $F \in L^1(\mathbb{R}^N)$, denoting any $\boldsymbol{x} \in \mathbb{R}^N$, $\boldsymbol{x} \neq 0$ as $\boldsymbol{x} = \varrho\,\sigma$, with $\varrho = \|\boldsymbol{x}\|$ and $\|\sigma\| = 1$, we have
$$\int_{\mathbb{R}^N} F(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \int_0^\infty \varrho^{N-1}\mathrm{d}\varrho \int_{\mathbb{S}^{N-1}} F(\varrho\,\sigma)\mathrm{d}\sigma$$
where $\mathrm{d}\sigma$ denotes the surface Lebesgue integral over $\mathbb{S}^{N-1}$.

3. Let now $1 \leqslant q < p \leqslant +\infty$. Using the previous two points, find a mapping f belonging to $L^q(\mathbb{R}^n)$ but not belonging to $L^p(\mathbb{R}^n)$ and a mapping $g \in L^p(\mathbb{R}^n)$ with $g \notin L^q(\mathbb{R}^n)$.

■

Exercise 1.12 Let $\alpha > 0$ and $\beta > 0$. Set

$$f(x) = (1 + \|x\|^\alpha)^{-1} \left(1 + |\log \|x\||^\beta\right)^{-1}, \qquad \forall x \in \mathbb{R}^N.$$

Under what conditions on α, β does f belong to $L^p(\mathbb{R}^N)$? ■

Exercise 1.13 (**Generalized Holder's inequality**) Let $f_1, \ldots, f_n$ be measurable functions over (S, Σ, μ) such that

$$f_i \in L^{p_i}(S, \mu) \qquad \forall i = 1, \ldots, n$$

where $1 \leqslant p_i \leqslant \infty$ and $\sum_{i=1}^n \frac{1}{p_i} = 1$. Set

$$f = \prod_{i=1}^n f_i.$$

1. Show by induction over n that $f \in L^1(S, \mu)$ with

$$\|f\|_1 \leqslant \prod_{i=1}^n \|f_i\|_{p_i}.$$

2. Deduce that if $f \in L^p(S, \mu) \cap L^q(S, \mu)$ with $1 \leqslant p \leqslant q \leqslant \infty$, then for any $f \in L^r(S, \mu)$ for any $p \leqslant r \leqslant q$ with

$$\|f\|_r \leqslant \|f\|_p^\alpha \, \|f\|_q^{1-\alpha} \qquad \text{where } \alpha \in [0, 1] \text{ is such that } \frac{1}{r} = \frac{\alpha}{p} + \frac{1-\alpha}{q}.$$

■

2. Banach spaces

2.1 Cauchy sequences-Banach spaces

An important related concept is the one of Cauchy sequence:

Definition 2.1.1 Let $(X, \|\cdot\|)$ be a normed space. A sequence $(x_n)_n \subset X$ of elements in X is a Cauchy sequence if, for any $\varepsilon > 0$, there exists $N = N(\varepsilon) \in \mathbb{N}$ such that

$$\|x_n - x_m\| < \varepsilon \qquad \forall n, m \geqslant N.$$

Remark 2.1.1 One checks easily that every Cauchy sequence $(x_n)_n \subset X$ is bounded. Indeed, by definition, for any $\varepsilon > 0$, there is $N \in \mathbb{N}$ such that

$$\|x_n - x_m\| < \varepsilon \qquad \forall n, m \geqslant N.$$

For, say, $\varepsilon = 1$ and $m = N$ we see that, for any $n \geqslant N$,

$$\|x_n\| = \|x_n - x_N + x_N\| \leqslant \|x_n - x_N\| + \|x_N\| \leqslant 1 + \|x_N\|.$$

Thus, with $C_1 = 1 + \|x_N\|$,

$$\sup_{n \geqslant N} \|x_n\| \leqslant C_1.$$

Setting now $C_2 := \max\left(\|x_1\|, \ldots, \|x_{N-1}\|\right)$, one sees that C_2 is finite since it is the maximum of only a finite number of real numbers. By definition,

$$\|x_n\| \leqslant C_2 \qquad \forall n < N.$$

Therefore,

$$\sup_n \|x_n\| \leqslant C = \max(C_1, C_2) < \infty$$

i.e. $(x_n)_n$ is bounded.

The triangle inequality readily implies the following

Lemma 2.1.1 Let $(X, \|\cdot\|)$ be a normed space. Any convergent sequence is a Cauchy sequence.

✍ *Proof.* Let $(x_n)_n \subset X$ be a convergent sequence in X and let x be its limit. By definition, given $\varepsilon > 0$, there is $N \in \mathbb{N}$ such that $\|x_n - x\| < \varepsilon$ for any $n \geqslant N$. Pick then $n, m \geqslant N$, one has from the triangle inequality

$$\|x_n - x_m\| \leqslant \|x_n - x\| + \|x_m - x\| < 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we proved that the sequence is a Cauchy sequence (since we can rename 2ε as ε !). ■

It is very important to understand that the converse result is not true in general: there exists Cauchy sequence which cannot converge.

■ **Example 2.1** Consider the set of rational numbers $\mathbb{Q}$ endowed with the norm $|\cdot|$. There exist Cauchy sequences of rational numbers that do not converge to a rational number. For example, let $(x_n)_n$ be the following rational numbers based on the decimal expansion of π:

$$x_1 = 3, \quad x_2 = 3.1, \quad x_3 = 3.14, \ldots.$$

Then, $(x_n)_n$ is a Cauchy sequence in $\mathbb{Q}$ however, $(x_n)_n$ is not a convergent sequence in $\mathbb{Q}$ since there is no rational $q \in \mathbb{Q}$ with $x_n \to q$ (the limit is $\pi \notin \mathbb{Q}$). ■

✍ ■ **Example 2.2** We introduce here the set $c_{00}(\mathbb{N})$ of all sequences of real numbers that have only finitely many nonzero components, i.e.

$$c_{00} = \{x = (x_n)_n \subset X \,,\, \exists N \in \mathbb{N} \,,\, x_k = 0 \qquad \forall k \geqslant N\}\,.$$

It is clear that $c_{00}(\mathbb{N})$ is a subset of the space $\ell^1(\mathbb{N})$ introduced in Example 1.2. In particular, it is a norm space with respect to the norm induced by the one in $\ell^1(\mathbb{N})$ and defined in (1.1). Now, for any $n \in \mathbb{N}$, let $\boldsymbol{x}^{(n)}$ be the sequence given by

$$\boldsymbol{x}^{(n)} = (2^{-1}, 2^{-2}, \ldots, 2^{-n}, 0, 0, \ldots)$$

i.e. $x^{(n)} = (x_k^{(n)})_k$ with $x_k^{(n)} = 2^{-k}$ if $k \leqslant n$ and $x_k^{(n)} = 0$ if $k > n$. Clearly $\boldsymbol{x}^{(n)} \in c_{00}(\mathbb{N})$ for any $n \in \mathbb{N}$. Therefore, the family $\left(\boldsymbol{x}^{(n)}\right)_n$ is a sequence of elements of $c_{00}(\mathbb{N})$ (pay attention that this is a sequence whose elements are again sequences !!). Moreover, if $m < n$, then

$$\|\boldsymbol{x}^{(n)} - \boldsymbol{x}^{(m)}\|_1 = \sum_{k=1}^{\infty} |x_k^{(n)} - x_k^{(m)}| = \sum_{k=m+1}^{n} 2^{-k}.$$

One sees then that (***Explain this***) $(\boldsymbol{x}^{(n)})_n$ is a Cauchy sequence in $(c_{00}(\mathbb{N}), \|\cdot\|_1)$. However, it does not converge in $(c_{00}(\mathbb{N}), \|\cdot\|_1)$. Indeed, assume the contrary, i.e. there exists $\boldsymbol{x} = (x_k)_k \in c_{00}(\mathbb{N})$ such that $\lim \boldsymbol{x}^{(n)} = \boldsymbol{x}$. It is not difficult to check that, necessarily, for any $k \in \mathbb{N}$, it must hold

$$\lim_{n\to\infty} x_k^{(n)} = \boldsymbol{x}_k$$

and therefore that $\boldsymbol{x}_k = 2^{-k}$ for any $k \in \mathbb{N}$. However, one sees that the sequence $\boldsymbol{x} = (2^{-k})_k$ does not belong to $c_{00}(\mathbb{N})$. This proves the claim. ■

Exercise 2.1 Let $(X, \|\cdot\|)$ be a normed space and let $(x_n)_n$ be a sequence such that the series $\sum_{n=1}^{\infty} \|x_{n+1} - x_n\|$ converges. Prove that $(x_n)_n$ is a Cauchy sequence (one can look at Theorem 2.2.2 hereafter where this result is proven in some peculiar case). ■

One has the following

Lemma 2.1.2 Let $(X, \|\cdot\|)$ be a given normed space and let $(x_n)_n$ be a Cauchy sequence in X. If $(x_n)_n$ admits a limit point then it is convergent.

Proof. Let x be a limit point of $(x_n)_n$ and let $\varepsilon > 0$. Let us pick $N_0 \in \mathbb{N}$ such that $\|x_n - x_m\| < \varepsilon/2$ for any $n, m \geqslant N_0$. Let now $N_1 > N_0$ be given such that $\|x_{N_1}, x\| < \varepsilon/2$ (the existence of such N_1 is clear since x is a limit point). Then, for any $n \geqslant N_1$, one has

$$\|x_n - x\| \leqslant \|x_n - x_{N_1}\| + \|x_{N_1} - x\| < \varepsilon$$

which proves that the sequence converges to x. ■

We introduce now the notion of complete normed space

Definition 2.1.2 A normed space $(X, \|\cdot\|)$ is said to be ***complete*** if *any* Cauchy sequence is convergent in X. A complete normed space $(X, \|\cdot\|)$ is called a ***Banach space***.

■ **Example 2.3** The examples we already deal with showed that $\mathbb{Q}$ endowed with the norm $|\cdot|$ is not a complete space and $c_{00}(\mathbb{N})$ endowed with the $\ell^1(\mathbb{N})$-norm is not a complete space. ■

Exercise 2.2 Prove that, introducing now the supremum norm

$$\|\boldsymbol{x}\|_\infty := \sup_n |x_n|, \qquad \boldsymbol{x} = (x_n)_n \in c_{00},$$

then again, $(c_{00}, \|\cdot\|_\infty)$ is not a Banach space. ■

The most fundamental example of complete normed space is the set of real numbers

Theorem 2.1.3 If $X = \mathbb{R}$ is endowed with the usual norm $|\cdot|$ for any $x, y \in \mathbb{R}$, then $(\mathbb{R}, |\cdot|)$ is a complete normed space. In other words, any Cauchy sequence in $\mathbb{R}$ is convergent.

Proof. Let $(x_n)_n \subset \mathbb{R}$ be a Cauchy sequence in $\mathbb{R}$. One checks easily that $(x_n)_n$ is a bounded sequence in $\mathbb{R}$. Therefore, a well-known consequence of this is that it admits a subsequence which converge. In other words, the Cauchy sequence $(x_n)_n$ has a limit point and we conclude with Lemma 2.1.2. ■

A fundamental example of infinite dimensional complete space is the following:

Proposition 2.1.4 The space $(\ell^1(\mathbb{N}), \|\cdot\|_1)$ introduced in Example 1.2 is a Banach space.

Proof. Let $(\boldsymbol{x}^{(n)})_n$ be a Cauchy sequence in $\ell^1(\mathbb{N})$. Recall that, for any $n \in \mathbb{N}$, $\boldsymbol{x}^{(n)}$ is sequence given by

$$\boldsymbol{x}^{(n)} = (\boldsymbol{x}_k^{(n)})_k.$$

Since it is a Cauchy sequence, for any $\varepsilon > 0$ there is $N > 0$ such that

$$\|\boldsymbol{x}^{(n)} - \boldsymbol{x}^{(m)}\|_1 = \sum_{k=1}^{\infty} |\boldsymbol{x}_k^{(n)} - \boldsymbol{x}_k^{(m)}| < \varepsilon \qquad \forall n, m > N.$$

It is easy to see then that, for any $k \in \mathbb{N}$, one has

$$|\boldsymbol{x}_k^{(n)} - \boldsymbol{x}_k^{(m)}| \leqslant \|\boldsymbol{x}^{(n)} - \boldsymbol{x}^{(m)}\|_1 < \varepsilon$$

i.e., for any $k \in \mathbb{N}$ fixed, the sequence $(\boldsymbol{x}_k^{(n)})_n \subset \mathbb{R}$ is a Cauchy sequence in $\mathbb{R}$ (endowed with its natural distance). Since $\mathbb{R}$ is complete by the previous Theorem, it must converge i.e. for any $k \in \mathbb{N}$,

$$x_k = \lim_{n \to \infty} \boldsymbol{x}_k^{(n)}$$

exists in $\mathbb{R}$. We introduce then the sequence $\boldsymbol{x} = (x_k)_k$ and let us show that $\boldsymbol{x} \in \ell^1(\mathbb{N})$ and that $\boldsymbol{x} = \lim_n \boldsymbol{x}^{(n)}$ for the $\ell^1(\mathbb{N})$ -distance. To do so, we fix $\varepsilon > 0$. Because $(\boldsymbol{x}^{(n)})_n$ is a Cauchy sequence in $\ell^1(\mathbb{N})$, there exists $N > 0$ such that

$$\|\boldsymbol{x}^{(n)} - \boldsymbol{x}^{(m)}\|_1 < \varepsilon \qquad \forall n, m > N.$$

Let us fix $n > N$ and let us fix an integer $k_0 > 0$. Since $\lim_{m \to \infty} \boldsymbol{x}_k^{(m)} = \boldsymbol{x}_k$ for any $k \geqslant 1$ and since the k_0 is finite one has

$$\sum_{k=1}^{k_0} |\boldsymbol{x}^{(n)} - \boldsymbol{x}_k| = \sum_{k=1}^{k_0} \lim_{m \to \infty} |\boldsymbol{x}_k^{(n)} - \boldsymbol{x}_k^{(m)}| \leqslant \lim_{m \to \infty} \|\boldsymbol{x}^{(n)} - \boldsymbol{x}^{(m)}\|_1 < \varepsilon.$$

Since this is true for *any* $k_0 > 0$ and

$$\|\boldsymbol{x}^{(n)} - \boldsymbol{x}\|_1 = \lim_{k_0 \to \infty} \sum_{k=1}^{k_0} |\boldsymbol{x}_k^{(n)} - \boldsymbol{x}_k|$$

one deduces that $\|\boldsymbol{x}^{(n)} - \boldsymbol{x}\|_1 < \varepsilon$ for any $n \geqslant N$ which proves that the sequence is converging to $\boldsymbol{x}$ and in particular that $\boldsymbol{x} \in \ell^1(\mathbb{N})$. ∎

More generally, one has the following

> **Proposition 2.1.5** For any $p \in [1, \infty)$, the normed space $(\ell^p(\mathbb{N}), \|\cdot\|_p)$ is a Banach space. Moreover, $(\ell^\infty(\mathbb{N}), \|\cdot\|_\infty)$ is also a Banach space.

Proof. For $p = 1$, we already saw in Proposition 2.1.4 that $(\ell^1(\mathbb{N}), \|\cdot\|_1)$ is complete. Let us now consider the case $1 < p < \infty$. Assume $(\boldsymbol{x}^{(n)})_n$ is a Cauchy sequence in $\ell^p(\mathbb{N})$, i.e. for any $n \in \mathbb{N}$, $\boldsymbol{x}^{(n)} = (\boldsymbol{x}_k^{(n)})_k \in \ell^p(\mathbb{N})$ and, for any $\varepsilon > 0$, there is $N \in \mathbb{N}$ such that

$$\|\boldsymbol{x}^{(n)} - \boldsymbol{x}^{(m)}\|_p^p = \sum_{k=1}^{\infty} \left|\boldsymbol{x}_k^{(n)} - \boldsymbol{x}_k^{(m)}\right|^p < \varepsilon \qquad \forall n, m \geqslant N.$$

As in the proof of Proposition 2.1.4, for any $k \in \mathbb{N}$, the sequence $(\boldsymbol{x}_k^{(n)})_n \subset \mathbb{R}$ is a Cauchy sequence in $\mathbb{R}$ and it converges to $x_k \in \mathbb{R}$. Set then $\boldsymbol{x} = (x_k)_k$. We aim to show that $\boldsymbol{x} \in \ell^p(\mathbb{N})$ and $\|\boldsymbol{x}_n - \boldsymbol{x}\|_p \to 0$ as $n \to \infty$. For any $n > N$, let us fix $k_0 > 0$. One has

$$\sum_{k=1}^{k_0} \left|\boldsymbol{x}_k^{(n)} - x_k\right|^p = \sum_{k=1}^{k_0} \lim_{m \to \infty} |\boldsymbol{x}_k^{(n)} - \boldsymbol{x}_k^{(m)}|^p \leqslant \lim_{m \to \infty} \|\boldsymbol{x}_k^{(n)} - \boldsymbol{x}_k^{(m)}\|_p^p < \varepsilon.$$

Since k_0 is arbitrary, letting $k_0 \to \infty$ yields $\|\boldsymbol{x}_k^{(n)} - \boldsymbol{x}\|_p^p < \varepsilon$ for any $n \geqslant N$ and this completes the proof for $1 < p < \infty$. For $p = \infty$, the proof is left as an **Exercise**. ■

2.1.1 Examples revisited

We show in this section that the examples of normed spaces introduced previously are actually Banach spaces. We use the notations of Section 1.2:

Proposition 2.1.6 For any compact interval $I \subset \mathbb{R}$, the normed space $(\mathscr{C}(I), \|\cdot\|_\infty)$ is a Banach space.

Proof. Call $X = \mathscr{C}(I)$ and take a sequence $(f_n)_n \subset X$ which is a Cauchy sequence for $(X, \|\cdot\|_\infty)$. This means that, for any $\varepsilon > 0$, there is $N \in \mathbb{N}$ such that

$$\|f_n - f_m\|_\infty < \varepsilon \qquad \forall n, m > N.$$

By definition, we have

$$|f_n(x) - f_m(x)| < \varepsilon \qquad \forall n, m > N, \qquad \forall x \in I.$$

The important thing, as usual for the $\|\cdot\|_\infty$ is that the integer N is the *same* for all $x \in I$. For a given $x \in I$, the sequence $(f_n(x))_n$ is a Cauchy sequence in $\mathbb{R}$ and, therefore, converges to some $f(x)$. Since the interval I is compact and $\|\cdot\|_\infty$ induces the uniform convergence topology one checks without difficulty ***(Do it!)*** that $f \in \mathscr{C}(I)$ and $\lim_n \|f_n - f\|_\infty = 0$ which proves the completeness of $(X, \|\cdot\|_\infty)$. ■

Most interesting maybe is the following

Proposition 2.1.7 For any compact interval $I \subset \mathbb{R}$, the normed space $(\mathscr{C}(I), \|\cdot\|_1)$ is ***not*** a Banach space.

Proof. The proof consists in showing that there is a Cauchy sequence $(f_n)_n$ of $(\mathscr{C}(I), \|\cdot\|_1)$ which is not convergent. Set for simplicity $I = [0,1]$. Consider the piecewise linear function described by

$$f_n(x) = \begin{cases} 1, & \text{if } 0 \leqslant x \leqslant \frac{1}{2} \\ -nx - \frac{n}{2} + 1, & \text{if } \frac{1}{2} \leqslant x \leqslant \frac{1}{2} + \frac{1}{n} \\ 0, & \text{if } \frac{1}{2} + \frac{1}{n} \leqslant x \leqslant 1\ . \end{cases}$$

Then, for any $m > n$, it holds

$$\|f_n - f_m\|_1 \leqslant \int_{1/2}^{1/2+1/n} |f_n(x) - f_m(x)| \mathrm{d}x \leqslant \frac{1}{n} \qquad \text{as } n \to \infty$$

(check this carefully). Therefore $(f_n)_n$ is a Cauchy sequence for the $\|\cdot\|_1$-norm. Suppose now it converges to $f \in \mathscr{C}([0,1])$. Then

$$\int_0^{1/2} |f(x) - f_n(x)| \mathrm{d}x \leqslant \|f - f_n\|_1 \to 0$$

and, by continuity, we would get $f(x) = 1$ for any $x \in [0, 1/2]$. Similarly we see $f(x) = 0$ for any $x \in [1/2, 1]$. The function f is not continuous which is a contradiction. ■

For general L^p-spaces, one has the following

Theorem 2.1.8 — Fischer–Riesz. Let (S, Σ, μ) be a given measure space. For any $1 \leqslant p \leqslant \infty$, $L^p(S, \Sigma, \mu)$ is a Banach space.

Before proving the result, we establish two preliminary results of independent interest. The first one is a result which is actually a general property of Banach spaces but, for L^p-spaces, we will first prove the property *and after that* prove it is a Banach

Lemma 2.1.9 Let (S, Σ, μ) be a given measure space and let $1 \leqslant p < \infty$. Assume that $\{f_n\}_n \subset L^p(S, \Sigma, \mu)$ is a sequence such that

$$\sum_{n=1}^{\infty} \|f_n\|_p < \infty.$$

Then, the series

$$\sum_{n=1}^{\infty} f_n \text{ converges almost every where and in} L^p(S, \Sigma, \mu),$$

which means that there exists $f \in L^p(S, \Sigma, \mu)$ such that

$$\lim_{N \to \infty} \left\| f - \sum_{n=1}^{N} f_n \right\|_p = 0$$

and $\lim_{N\to\infty} \sum_{n=1}^{N} f_n(s) = f(s)$ for μ-almost every $s \in S$.

Proof. Let $\{f_n\}_n \subset L^p(S, \Sigma, \mu)$ be given and such that

$$\sum_{n=1}^{\infty} \|f_n\|_p < \infty.$$

We introduce the following functions

$$F_N(s) = \sum_{n=1}^{N} f_n(s), \qquad G_N(s) = \sum_{n=1}^{N} |f_n(s)| \quad s \in S, \quad N \in \mathbb{N}.$$

Clearly, for μ-almost every $s \in S$, the sequence $\{G_N(s)\}_N \subset [0, \infty]$ is nondecreasing and one can define

$$G(s) = \sum_{n=1}^{\infty} |f_n(s)| = \sup_N G_N(s) \qquad \mu - \text{ a.e. } s \in S.$$

Notice that it could be $G(s) = \infty$. However, thanks to Minkowski inequality

$$\|G_N\|_p \leqslant \sum_{n=1}^{N} \|f_n\|_p \leqslant \sum_{n=1}^{N} \|f_n\|_p < \infty$$

i.e

$$\sup_N \int_S G_N^p \mathrm{d}\mu < \infty, \qquad \text{and} \qquad \lim_N G_N^p = G^p \quad \text{a. e.}$$

According to the monotone convergence theorem, one deduces that

$$\int_S G^p \mathrm{d}\mu < \infty, \text{ i. e. } G \in L^p(S, \Sigma, \mu).$$

This implies in particular that

$$G(s) < \infty \qquad \text{for } \mu\text{-a. e. } s \in S.$$

This means that the numerical series

$$\sum_{n=1}^{\infty} |f_n(s)| < \infty \quad \text{for } \mu\text{-a. e. } s \in S$$

and therefore the numerical series

$$\sum_{n=1}^{\infty} f_n(s) \quad \text{converges in } \mathbb{R} \text{ for } \mu\text{-a. e. } s \in S$$

(absolutely converging series are converging). We denote by $f(s)$ the limit of the above series, one has and it holds, by definition

$$\lim_{N \to \infty} F_N(s) = f(s) \qquad \text{for } \mu\text{-a. e. } s \in S.$$

Let us now prove that the convergence also holds in $L^p(S, \Sigma, \mu)$. Notice that it holds $|F_N(s)| \leqslant G_N(s) \leqslant G(s)$ for μ-almost every $s \in S$ which results in

$$|f(s)| \leqslant G(s) \text{ for } \mu\text{-a. e. } s \in S$$

and $|F_N(s) - f(s)| \leqslant 2G(s)$ for μ-a. e. $s \in S$ or equivalently

$$|F_N(s) - f(s)|^p \leqslant (2G(s))^p \quad \text{for } \mu\text{-a. e. } s \in S.$$

The function $2G$ belonging to $L^1(S, \Sigma, \mu)$, one deduces now from the Dominated Convergence Theorem that

$$\lim_N \int_S |F_N - f|^p \mathrm{d}\mu = 0$$

i.e. $\lim_N \|F_N - f\|_p = 0$. This proves the result. ■

✍ *Proof of Fischer–Riesz Theorem.* From now on, we simply write $L^p(S, \mu)$ instead of $L^p(S, \Sigma, \mu)$. For the proof, we distinguish between the cases $1 \leqslant p < \infty$ and $p = \infty$. We first prove the result for $p = \infty$.

The case $p = \infty$. Assume that $(f_n)_n$ is a Cauchy sequence in $L^\infty(S, \mu)$. Given an integer $k \geqslant 1$ there is an integer N_k such that

$$\|f_m - f_n\|_\infty \leqslant \frac{1}{k} \qquad \forall n, m \geqslant N_k \tag{2.1}$$

so that, there is a null set $E_k \in \Sigma$ such that

$$|f_m(x) - f_n(x)| \leqslant \frac{1}{k} \qquad \forall x \in S \setminus E_k, \ \forall m, n \geqslant N_k.$$

Indeed, given $m, n \geqslant N_k$, by definition, there is a null set $E_k(m,n)$ for which $|f_m(x) - f_n(x)| \leqslant \frac{1}{k}$ for $x \in S \setminus E_k(m,n)$. Taking then $E_k = \bigcup_{n,m \geqslant N_k} E_k(m,n)$, one has E_k null set and the above property holds on $S \setminus E_k$ for *all* $n, m \geqslant N_k$.

Then we let $E = \bigcup_{k=1}^{\infty} E_k$ so that $\mu(E) = 0$ and get that, for any $x \in S \setminus E$ the sequence $(f_n(x))_n$ is a Cauchy sequence (in $\mathbb{R}$). Thus, for any $x \in S \setminus E$, $\lim_n f_n(x)$ exists and we denote it by $f(x)$. Passing to the limit in (2.1) as $m \to \infty$ we obtain then that

$$|f(x) - f_n(x)| \leqslant \frac{1}{k} \qquad \forall x \in S \setminus E\,, \forall n \geqslant N_k$$

from which we see that $f \in L^\infty(S,\mu)$ with

$$\|f - f_n\|_\infty \leqslant \frac{1}{k} \qquad \forall n \geqslant N_k.$$

Letting then $n \to \infty$ and then $k \to \infty$, we get that $\lim_n \|f - f_n\|_\infty = 0$. This proves that $(L^\infty(S,\mu), \|\cdot\|_\infty)$ is complete.

The case $1 \leqslant p < \infty$. Let us assume now that $1 \leqslant p < \infty$ and let $(f_n)_n$ be a Cauchy sequence in $L^p(S,\mu)$. By induction, it is easy to construct a sequence $(n_k)_k$ such that $n_{k+1} \geqslant n_k$ and

$$\|f_{n_{k+1}} - f_{n_k}\|_p \leqslant \frac{1}{2^k} \qquad \forall k \geqslant 1.$$

Notice that, setting $\varphi(k) = n_k$, the mapping $\varphi : \mathbb{N} \to \mathbb{N}$ is strictly increasing (since $n_{k+1} > n_k$) and therefore $(f_{n_k})_k$ is a subsequence $(f_{\varphi(k)})_k$ of the original sequence $(f_n)_n$. Set $g_1 = f_{n_1}$ and

$$g_{k+1} := f_{n_{k+1}} - f_{n_k}, \qquad k \geqslant 1.$$

It holds that $g_k \in L^p(S,\mu)$ with

$$\sum_{k=1}^{\infty} \|g_k\|_p \leqslant \sum_{k=1}^{\infty} \|f_{n_{k+1}} - f_{n_k}\|_p \leqslant \sum_{k=1}^{\infty} 2^{-k} < \infty.$$

From the previous Lemma, there is some $f \in L^p(S,\mu)$ such that

$$\lim_{k\to\infty} \|f - \sum_{k=1}^{N} g_k\|_p = 0 \quad \text{and} \quad \lim_{N\to\infty} \sum_{k=1}^{N} g_k(s) = f(s) \quad \text{for } \mu\text{-a. e. } s \in S.$$

Since

$$f_{n_N} = \sum_{k=1}^{N} g_k$$

one found a subsequence $\{f_{N_k}\}_k$ of $\{f_n\}$ which converges almost everywhere and in $L^p(S,\mu)$. Therefore, the Cauchy sequence $\{f_n\}_n$ has a limit point and, as known, the whole sequence converge. This proves the result. ■

Notice that, in the above proof, we showd the following which has is own interest:

Proposition 2.1.10 Let $1 \leqslant p \leqslant \infty$ and let $(f_n)_n$ be a Cauchy sequence in $L^p(S,\mu)$ which converges to f. Then, there is a subsequence (f_{n_k}) which converges μ-almost everywhere to f.

Another class of Banach spaces is the following

Proposition 2.1.11 Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be two normed spaces and assume that $(Y, \|\cdot\|_Y)$ is a Banach space. Then, the space $(\mathscr{L}(X,Y), \|\cdot\|_{\mathscr{L}(X,Y)})$ is a Banach space.

✍ *Proof.* Let $(L_n)_n$ be a sequence of linear and continuous applications from X to Y. Assume that $(L_n)_n$ is a Cauchy sequence for $\|\cdot\|_{\mathscr{L}(X,Y)}$, i.e. for any $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$\|L_n - L_m\|_{\mathscr{L}(X,Y)} < \varepsilon \qquad \forall n, m \geqslant N.$$

By definition of the norm, it holds

$$\|L_n(x) - L_m(x)\|_Y < \varepsilon \|x\|_X \qquad \forall n, m \geqslant N \ \forall x \in X.$$

In particular, for any $x \in X$, the sequence $(L_n(x))_n$ is a Cauchy sequence in Y and since Y is complete, the sequence converges to $L(x)$. This defines a mapping $L : X \to Y$ and it is not difficult to prove that L is linear ***(check this)***. Let us check then that L is continuous and that $\|L_n - L\|_{\mathscr{L}(X,Y)} \to 0$ as $n \to \infty$. Taking the limit as $m \to \infty$ in the previous inequality, one gets that

$$\|L_n(x) - L(x)\|_Y \leqslant \varepsilon \, \|x\|_X \qquad \forall n \geqslant N \ \forall x \in X. \tag{2.2}$$

In particular, for any fixed $n \geqslant N$ and any given $x \in X$

$$\|L(x)\|_Y \leqslant \|L_n(x)\|_Y + \varepsilon \, \|x\|_X$$

and, since L_n is continuous, there exists $C_n > 0$ such that

$$\|L(x)\|_Y \leqslant C_n \|x\|_X + \varepsilon \|x\|_X.$$

This shows that $L \in \mathscr{L}(X,Y)$ ***(Explain why)***. Now, (2.2) exactly means that $\|L_n - L\|_{\mathscr{L}(X,Y)} < \varepsilon$ for any $n \geqslant N$ which proves that $(L_n)_n$ converges to L in $(\mathscr{L}(X,Y), \|\cdot\|_{\mathscr{L}(X,Y)})$. ■

A first useful consequence of the previous Propositon is the following:

Corollary 2.1.12 Let $(X, \|\cdot\|_X)$ be a Banach space and let $L \in \mathscr{L}(X)$ be such that

$$\|L\|_{\mathscr{L}(X)} < 1.$$

Then, the linear mapping $I - L$ is invertible (i.e. $I - L$ is bijective and its inverse $(I-L)^{-1} \in \mathscr{L}(X)$) and

$$(I - L)^{-1} = \sum_{n=0}^{\infty} L^n$$

where I is the identity mapping of X and L^n is defined inductively by $L^0 = I$ and $L^{n+1} = L \circ L^n = L^n \circ L$ for any $n \geqslant 0$.

✍ *Proof.* For any $n \in \mathbb{N}$, set

$$A_n := \sum_{k=0}^{n} L^k.$$

Clearly, A_n is a linear mapping as a composition and sum of linear mappings. For the same reason, it is continuous, i.e. $A_n \in \mathscr{L}(X)$. For any $m > n$, it holds

$$\|A_m - A_n\|_{\mathscr{L}(X)} = \left\| \sum_{k=n+1}^{m} L^k \right\|_{\mathscr{L}(X)}$$

and the property of the norm yields

$$\|A_m - A_n\|_{\mathscr{L}(X)} \leqslant \sum_{k=n+1}^{m} \|L^k\|_{\mathscr{L}(X)}.$$

One checks easily by induction that $\|L^{k+1}\|_{\mathscr{L}(X)} \leqslant \|L\|_{\mathscr{L}(X)} \, \|L^k\|_{\mathscr{L}(X)}$ for any $k \geqslant 0$ ***(Check this)***. Thus,

$$\|L^k\|_{\mathscr{L}(X)} \leqslant \|L\|^k_{\mathscr{L}(X)} \qquad \forall k \geqslant 0.$$

In particular, we see that

$$\|A_n - A_m\|_{\mathscr{L}(X)} \leqslant \sum_{k=n+1}^{m} \|L\|^k_{\mathscr{L}(X)} \leqslant \frac{\|L\|^{n+1}_{\mathscr{L}(X)}}{1 - \|L\|_{\mathscr{L}(X)}} \qquad \forall m > n \geqslant 0.$$

In particular, since $\|L\|_{\mathscr{L}(X)} < 1$, one sees that, for any $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$\|A_n - A_m\|_{\mathscr{L}(X)} < \varepsilon \qquad \forall m > n \geqslant N.$$

Then, $(A_n)_n$ is a Cauchy sequence in $\mathscr{L}(X)$. According to the previous Theorem, $(A_n)_n$ converges to some linear mapping, say, $A \in \mathscr{L}(X)$. Let us now prove that $I - L$ is necessarily invertible and that $A = (I - L)^{-1}$. For any $n \geqslant 0$, it holds that

$$(I - L) \circ A_n = A_n \circ (I - L) = I - L^{n+1}.$$

In particular,

$$\|(I - L) \circ A_n - I\|_{\mathscr{L}(X)} \leqslant \|L\|^{n+1}_{\mathscr{L}(X)}$$

and, letting $n \to \infty$, one sees that $(I - L) \circ A_n$ converges to I in $\mathscr{L}(X)$. By continuity, this means that

$$(I - L) \circ A = I.$$

In the same way, we have $A \circ (I - L) = I$. This shows that $(I - L)$ is invertible and that A is its inverse ***(Explain why)***. ∎

Remark 2.1.2 The series $\sum_{n=0}^{\infty} L^n$ defining $(I - L)^{-1}$ is called the Neumann's series. It is of course reminiscent to the well-known formula, valid for $0 < x < 1$ which asserts that

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1 - x} = (1 - x)^{-1}.$$

Exercise 2.3 Let $(X, \|\cdot\|)$ be a Banach space and let $(Y, \|\cdot\|_Y)$ be a normed vector space. We introduce the set

$$\mathcal{U} := \left\{A \in \mathscr{L}(X,Y);\ A\ :\ X \to Y \text{ is a bijection and } A^{-1} \in \mathscr{L}(Y;X)\right\}.$$

1. Prove that, if $A \in \mathcal{U}$ and $B \in \mathscr{L}(X,Y)$ is such that

$$\|B - A\|_{\mathscr{L}(X,Y)} < \frac{1}{\|A^{-1}\|_{\mathscr{L}(Y,X)}}$$

then $I + A^{-1}(B-A) \in \mathscr{L}(X)$ is invertible and deduce that

$$B = A\left(I + A^{-1}(B-A)\right) \in \mathscr{L}(X,Y)$$

is invertible.
2. Deduce from this that $\mathcal{U}$ is open in the normed vector space $\mathscr{L}(X,Y), \|\cdot\|_{\mathscr{L}(X,Y)}$.

■

A very special class of space of the form $\mathscr{L}(X,Y)$ is obtained whenever $Y = \mathbb{R}$. One has then the following definition

Definition 2.1.3 If $(X, \|\cdot\|)$ is a normed space, we define the ***dual space*** of X, denoted by $X^\star$ the space of all continuous and linear mappings $\Phi\ :\ X \to \mathbb{R}$, i.e.

$$X^\star = \mathscr{L}(X, \mathbb{R}).$$

We also denote

$$\|\Phi\|_\star = \|\Phi\|_{\mathscr{L}(X,\mathbb{R})} = \sup_{\|x\| \leqslant 1} |\Phi(x)| \qquad \forall \Phi \in X^\star.$$

One has the direct consequence of the above Proposition

Corollary 2.1.13 If $(X, \|\cdot\|)$ is a normed vector space, then $(X^\star, \|\cdot\|_\star)$ is a Banach space.

Exercise 2.4 Let $p > 1$ be given and let $X = \ell^p(\mathbb{N})$ be endowed with its usual $\|\cdot\|_p$ norm. Let q be the conjugate exponent of p given by $1/q + 1/p = 1$.
1. For any $\boldsymbol{y} = \{y_k\}_k \in \ell^q(\mathbb{N})$, prove that the mapping

$$\Phi\ :\ \boldsymbol{x} = \{x_k\}_k \in \ell^p(\mathbb{N}) \mapsto \Phi(\boldsymbol{x}) = \sum_{k=1}^{\infty} x_k y_k \in \mathbb{R} \tag{2.3}$$

belongs to $X^\star$ and prove that $\|\Phi\|_\star = \|\boldsymbol{y}\|_q$.
2. Let now $\Phi \in X^\star$ be given. Prove that there exists a unique $\boldsymbol{y} = \{y_k\}_k \in \ell^q(\mathbb{N})$ such that Φ is given by (4.1).

This exercise shows that the dual of $\ell^p(\mathbb{N})$ can be identified with $\ell^q(\mathbb{N})$ (there is an isonormed bijection between $\ell^p(\mathbb{N})^\star$ and $\ell^q(\mathbb{N})$). ■

The above Exercise can be actually generalize to abstract L^p spaces:

Theorem 2.1.14 — Riesz representation theorem. Given $1 < p < \infty$ and $\Phi \in (L^p(S,\mu))^*$, there exists a unique $g \in L^q(S,\mu)$ (with $\frac{1}{p}+\frac{1}{q}=1$) such that

$$\Phi(f) = \int_S fg\,\mathrm{d}\mu \qquad \forall f \in L^p(S,\mu).$$

Moreover,

$$\|\Phi\|_* = \|g\|_q.$$

If (S,Σ,μ) is σ-finite, then the result is still true for $p=1$.

Proof. If $g \in L^q(S,\mu)$, the mapping $\Phi_g : f \in L^p(S,\mu) \mapsto \int_S fg\mathrm{d}\mu \in \mathbb{R}$ is clearly linear and continuous thanks to Holder's inequality. Namely,

$$\Phi_g \in (L^p(S,\mu))^* \qquad \text{and} \qquad \|\Phi_g\|_* \leqslant \|g\|_q.$$

Moreover, taking $f_0 = \lambda|g|^{q-2}g$ with $\lambda = \|g\|_q^{1-q}$. one sees that $\|f_0\|_p = \lambda\|g\|_q^{q-1} = 1$ and

$$\Phi_g(f_0) = \lambda\|g\|_q^q = \|g\|_q$$

Check all these details. Therefore, $\|\Phi_g\|_* = \|g\|_q$.

Now, given $\Phi \in (L^p(S,\mu))^*$, we have to show that $\Phi = \Phi_g$ for some (unique) $g \in L^q(S,\mu)$. We give the proof only in the simplest case when $\mu(S) < \infty$ and for $1 < p < \infty$.

For any $A \in \Sigma$ the mapping $f = \mathbf{1}_A$ belongs to $L^p(S,\mu)$. We define then

$$\nu(A) = \Phi(\mathbf{1}_A) \qquad \forall A \in \Sigma.$$

One sees without major difficulty that $\nu(\cdot)$ is a measure on Σ [1] If $\mu(A)=0$, then $\mathbf{1}_A = 0$ (μ-a.e.) and, since Φ is linear $\Phi(0)=0$, i.e. $\Phi(\mathbf{1}_A)=0$. In other words

$$\mu(A) = 0 \implies \nu(A) = 0.$$

This shows that ν is absolutely continuous with respect to the measure μ. According to Radon-Nikodym Theorem, there exists $g : S \to \mathbb{R}$, $g \in L^1(S,\mu)$ such that $\nu = g\mu$, i.e.

$$\nu(A) = \int_A g\mathrm{d}\mu \qquad \forall A \in \Sigma$$

or, equivalently,

$$\Phi(\mathbf{1}_A) = \int_A g\mathrm{d}\mu = \int_S g\mathbf{1}_A\mathrm{d}\mu \qquad \forall A \in \Sigma.$$

The linearity of Φ shows that

$$\Phi(f) = \int_S fg\mathrm{d}\mu$$

for any nonnegative simple function f. As usual, one can extend this to any $f \in L^p(S,\mu)$ using the fact that Φ is continuous on $L^p(S,\mu)$. In other words

$$\Phi(f) = \int_S fg\mathrm{d}\mu \qquad \forall f \in L^p(S,\mu).$$

[1] Pay attention here: we use a notion of measure which extends the usual one since clearly ν is not necessarily nonnegative. However, one can get convinced easily that ν can be split as $\nu = \nu^+ - \nu^-$ where $\nu^\pm$ are nonnegative measure. In particular, the use of Radon-Nikodym theorem is justified for both ν^+ and ν^-.

It remains to prove that $g \in L^q(S,\mu)$. This comes from the fact that there exists $C > 0$ such that $|\Phi(f)| \leqslant C\|f\|_p$ for any $f \in L^p(S,\mu)$. Indeed, since g is integrable, one knows that there exists a sequence of simple function $(\varphi_n)_n$ such that $\varphi_n(s) \to |g(s)|$ and $0 \leqslant \varphi_n(s) \leqslant |g(s)|$ for μ-almost any $s \in S$. Then, $(\varphi_n^q)_n$ is a sequence of simple function that converges pointwise to $|g|^p$ with $0 \leqslant \varphi_n^q \leqslant |g|^q$ for any $n \in \mathbb{N}$. To prove that $|g|^q \in L^1(S,\mu)$, it is enough thanks to Fatou's Lemma to show that

$$\int_S |\varphi_n|^q \mathrm{d}\mu \leqslant C^q \qquad \forall n \in \mathbb{N}. \tag{2.4}$$

This simply comes from the fact that $\varphi_n^q = \varphi_n\, \varphi_n^{q-1} \leqslant |g|\varphi_n^{q-1} = g f_n$ with

$$f_n = \mathrm{sign}(g)\, \varphi_n^{q-1}.$$

Notice that $f_n \in L^p(S,\mu)$ (since φ_n is simple) and $\Phi(f_n) = \int_S g f_n \mathrm{d}\mu \leqslant C\|f_n\|_p$ exactly reads as

$$\int_S \varphi_n^q \mathrm{d}\mu \leqslant C\|f_n\|_p = C\left(\int_S |\varphi_n|^{p(q-1)}\mathrm{d}\mu\right)^{1/p} = C\left(\int_S |\varphi_n|^q \mathrm{d}\mu\right)^{1/p}.$$

This exactly means that $\|\varphi_n\|_q \leqslant C$ which is (2.4). Therefore, $g \in L^q(S,\mu)$ with $\|g\|_q \leqslant C$. ∎

Remark 2.1.3 Consider the mapping

$$T\ :\ g \in L^q(S,\mu) \mapsto T(g) = \Phi_g \in (L^p(S,\mu))^*.$$

The mapping T is linear and, since $\|T(g)\|_* = \|g\|_q$, one sees that T is injective (and continuous, i.e. $T \in \mathscr{L}(L^q(S,\mu)\,, (L^p(S,\mu))^*)$.) Riesz-representation theorem actually asserts that T is surjective. This shows that $(L^p(S,\mu))^*$ is isonormed to $L^q(S,\mu)$ which allows the following identification

$$(L^p(S,\mu))^* \simeq L^q(S,\mu).$$

Notice that the this identification is true for $p = 1$ if (S,Σ,μ) is σ-finite while it is not true for $p = \infty$.

Remark 2.1.4 The above characterization, and the fact that T is an isometry, allows also to compute in an alternative way the norm $\|g\|_q$ for some given $1 < q \leqslant \infty$ as

$$\|g\|_q = \sup\{\|fg\|_1\,,\ f \in L^p(S,\mu)\,,\ \|f\|_p \leqslant 1\}.$$

2.2 Simple consequences of completeness – Banach-Cacciopoli-Picard Theorem

We begin with the first important consequence of completeness (compare the result to Lemma 2.1.9) which extends to general Banach spaces the well-known property of real series which asserts that absolutely converging series are converging:

Proposition 2.2.1 Let $(X, \|\cdot\|)$ be a Banach space and let $\{x_n\}_n \subset X$ be such that

$$\sum_{n=1}^{\infty} \|x_n\| < \infty.$$

Then, the series $\sum_{n=1}^{\infty} x_n$ converges in X, i.e. there exists $x \in X$ such that

$$\lim_{N\to\infty} \left\| \sum_{n=1}^{N} x_n - x \right\| = 0.$$

Proof. Consider the sequence

$$u_N = \sum_{n=1}^{N} x_n \qquad N \in \mathbb{N}.$$

It is easy to check that, for $N > M$,

$$\|u_N - u_M\| = \left\| \sum_{n=M+1}^{N} x_n \right\| \leqslant \sum_{n=M+1}^{N} \|x_n\|$$

according to the triangle inequality. Since the series $\sum_n \|x_n\| < \infty$, one has

$$\lim_{M\to\infty} \sup_N \sum_{n=M+1}^{N} \|x_n\| = 0$$

which easily implies that $\{u_N\}_N$ is a Cauchy sequence in X. Being X complete, the sequence converges to some $x \in X$ and this proves the result. ∎

We present here another application of the completeness. We begin with the following definition

Definition 2.2.1 Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be two normed spaces and $k \in (0,1)$. A function $f : X \to Y$ is said to be a k-contraction mapping if

$$\|f(x) - f(y)\|_Y \leqslant k\,\|x - y\|_X \qquad \forall x, y \in X \times X.$$

Remark 2.2.1 It is obvious that any k-contraction mapping is continuous. This is clear that a k-contraction mapping is an application which "contracts" distances: the images of x and y through f are closer to each other than x, y.

One has then the following fundamental result

Theorem 2.2.2 — Banach fixed-point theorem. Let $(X, \|\cdot\|)$ be a *complete* normed space and let $f : X \to X$ be a k-contraction mapping, $k \in (0,1)$. Then, there exists a **unique** fixed point $\boldsymbol{a} \in X$ for f, i.e. there is a unique $\boldsymbol{a} \in X$ such that

$$f(\boldsymbol{a}) = \boldsymbol{a}.$$

Moreover, for any $x_0 \in X$, the sequence $(x_n)_{n\geqslant 0}$ defined inductively by $f(x_n) = x_{n+1}$, $n \in \mathbb{N}$ is converging to $\boldsymbol{a}$.

✍ *Proof.* We first prove that f can admit only *one* fixed point. Assume $\boldsymbol{a}, \boldsymbol{b} \in X$ are fixed-point points of f, i.e. $f(\boldsymbol{a}) = \boldsymbol{a}$ and $f(\boldsymbol{b}) = \boldsymbol{b}$. Since f is a k-contraction mapping one has

$$\|f(\boldsymbol{a}) - f(\boldsymbol{b})\| \leqslant k\|\boldsymbol{a} - \boldsymbol{b}\|$$

which reads $\|\boldsymbol{a} - \boldsymbol{b}\| \leqslant k\|\boldsymbol{a} - \boldsymbol{b}\|$. Since $k > 0$ and $\|\boldsymbol{a}, \boldsymbol{b}\| \geqslant 0$, one sees that necessarily $\|\boldsymbol{a} - \boldsymbol{b}\| = 0$, i.e. $\boldsymbol{a} = \boldsymbol{b}$.

To prove now the existence of some fixed-point, one proves actually that the sequence defined in the above statement converges to some fixed-point of f. Let then $x_0 \in X$ be given and define inductively

$$x_{n+1} = f(x_n).$$

Since f is a k-contraction mapping, one has $\|x_{n+1} - x_n\| \leqslant k\|x_n - x_{n-1}\|$ for any $n \geqslant 1$ and we deduce easily that

$$\|x_{n+1} - x_n\| \leqslant k^n \|x_1 - x_0\| \qquad \forall n \in \mathbb{N}.$$

If $m > n \geqslant 1$ are given, one deduces from the above inequality together with the triangle inequality that

$$\|x_n - x_m\| \leqslant \sum_{j=n}^{m-1} \|x_j - x_{j+1}\| \leqslant \|x_1 - x_0\| \sum_{j=n}^{m-1} k^j. \tag{2.5}$$

Since the geometric series $\sum_{j=1}^{\infty} k^j$ is convergent, one has $\lim_{n,m\to\infty} \sum_{j=n}^{m-1} k^j = 0$ and therefore the sequence $(x_n)_n$ is a Cauchy sequence in X. Since X is complete, $(x_n)_n$ converges to some limit $\boldsymbol{a} \in X$. Moreover, being k-contracting, f is continuous so that the sequence $(f(x_n))_n$ converges to $f(\boldsymbol{a})$. Since $x_{n+1} = f(x_n)$ one has $\boldsymbol{a} = f(\boldsymbol{a})$ and the result is proven. ■

Remark 2.2.2 Notice that the above inequality (2.5) provides actually the convergence rate of $(x_n)_n$ to $\boldsymbol{a}$. Indeed, taking the limit $m \to \infty$ in (2.5) yields

$$\|x_n - \boldsymbol{a}\| \leqslant \|x_1 - x_0\| \sum_{j=n}^{\infty} k^j = \frac{k^n}{1-k} \|x_1 - x_0\| s \qquad \forall n \in \mathbb{N}.$$

■ **Example 2.4** Let $\varphi : [0,1] \to [0,1]$ be a continuous function such that there is $x_0 \in [0,1]$ with $\varphi(x_0) \neq 1$. For any $\alpha \in \mathbb{R}$, there exists a unique continuously differentiable function $f : [0,1] \to \mathbb{R}$ such that

$$f(0) = \alpha, \qquad f'(x) = f(\varphi(x)) \qquad \forall x \in [0,1]. \tag{2.6}$$

Indeed, introduce the set $X = \mathscr{C}([0,1])$ of continuous functions over $\mathbb{R}$ endowed with the supremum norm $\|\cdot\|_\infty$. Let then $T : X \to X$ be defined as $T(f) = g$ where

$$g(x) = T(f)(x) = \alpha + \int_0^x f(\varphi(t))\mathrm{d}t.$$

One notices that (2.6) is equivalent to $T(f) = f$ which amounts to look for a fixed-point of T in X. To apply Theorem 2.2.2, one route would be to check that T is a k-contraction mapping in X for some $k \in (0, 1)$. Unfortunately, this is not the case. However, one will prove that $T \circ T$ is a k-contraction mapping. Therefore, Theorem 2.2.2 would imply that there is some unique $g \in X$ such that

$$T \circ T(g) = g.$$

Applying again T to that identity, one sees that $h = T(g)$ is another fixed-point of $T \circ T$: indeed, $T \circ T(h) = T \circ T \circ T(g) = T(g) = h$. By uniqueness, $h = g$. Therefore, $T(g) = g$ and g is a fixed-point of T. The uniqueness follows from the fact that any fixed-point of T is also a fixed-point of $T \circ T$ and there is only one such fixed-point.

Let us therefore prove that $T^2 := T \circ T$ is a k-contraction. Let $f_1, f_2 \in X$. Set $g_1 = T(f_1)$, $g_2 = T(f_2)$, $h_1 = T(g_1)$ and $h_2 = T(g_2)$. For any $x \in [0, 1]$, one has by definition

$$|g_1(x) - g_2(x)| = \left| \int_0^x f_1(\varphi(t))\mathrm{d}t - \int_0^x f_2(\varphi(t))\mathrm{d}t \right| \leqslant \int_0^x |f_1(\varphi(t)) - f_2(\varphi(t))| \, \mathrm{d}t.$$

Recalling that d_∞ is the supremum distance, one gets

$$|g_1(x) - g_2(x)| \leqslant \int_0^x \|f_1 - f_2\|_\infty \mathrm{d}t = x \, \|f_1 - f_2\|_\infty \qquad \forall x \in [0, 1]. \tag{2.7}$$

Then, as above,

$$|h_1(x) - h_2(x)| \leqslant \int_0^x |g_1(\varphi(t)) - g_2(\varphi(t))| \, \mathrm{d}t \qquad \forall x \in [0, 1].$$

Applying (2.7) with $x = \varphi(t)$, $t \in [0, 1]$, one sees that $|g_1(\varphi(t)) - g_2(\varphi(t))| \leqslant \varphi(t) \, \|f_1 - f_2\|_\infty$ for any $t \geqslant 0$, which yields

$$|h_1(x) - h_2(x)| \leqslant \|f_1 - f_2\|_\infty \int_0^x \varphi(t)\mathrm{d}t \qquad \forall x \in [0, 1].$$

Taking the supremum over x we get

$$\mathrm{d}_\infty(h_1, h_2) \leqslant k \, \|f_1 - f_2\|_\infty \qquad \forall f_1, f_2 \in X$$

where $k = \displaystyle\int_0^1 \varphi(t)\mathrm{d}t$. Since φ is nonnegative, bounded by 1 but not identically equal to 1, one checks that $k < 1$. This shows that $T \circ T$ is a k-contraction mapping and the result follows. ■

Exercise 2.5 As suggested by the above Example, prove the following version of Banach fixed-point theorem: Let $(X, \|\cdot\|)$ be a *complete* normed space and let $f : X \to X$ be such that some iterate of f is a k-contraction mapping, $k \in (0, 1)$ (i.e. there is $p \in \mathbb{N}$ such that $f^{(p)} = \underbrace{f \circ f \circ f \circ \ldots \circ f}_{p \text{ times}}$ is a k-contraction mapping). Then, there exists a **unique** fixed point $\boldsymbol{a} \in X$ for f. ■

2.3 Fundamental properties of Banach spaces

We collect here several important properties of Banach spaces. We begin with the most fundamental with which is the Baire property:

Definition 2.3.1 A normed space $(X, \|\cdot\|)$ is said to have the Baire property if the intersection of any sequence of dense open sets of X is dense in X, i.e. for any $(U_n)_n$ open subsets of X with $\overline{U_n} = X$ for any $n \in \mathbb{N}$, it holds

$$\overline{\bigcap_n U_n} = X.$$

Remark 2.3.1 The Baire property is clearly equivalent to the following property: the union of any sequence of closed subsets with empty interior has an empty interior. ***(Explain why!)***. In particular, the theorem is often used in the following form: if $(C_n)_n$ is a sequence of closed subsets such that

$$\mathrm{Int}\left(\bigcup_{n\in\mathbb{N}} C_n\right) = X$$

then there exists some $n_0 \in \mathbb{N}$ such that $\mathrm{Int}(C_{n_0}) \neq \varnothing$.

Then, one has

Theorem 2.3.1 — Baire Theorem. Any ***complete*** normed space $(X, \|\cdot\|)$ has the Baire property.

✍ *Proof.* Let $(U_n)_n$ be a sequence of open subsets of $(X, \|\cdot\|)$ with $\overline{U_n} = X$ for any $n \in \mathbb{N}$ and let $A = \bigcap_{n\in\mathbb{N}} U_n$. To prove that A is dense in X, we shall use Lemma 1.1.7 and proves that

$$A \cap \mathcal{O} \neq \varnothing \qquad \text{for any non-empty open subset } \mathcal{O} \subset X.$$

Let us then pick a non-empty open subset $\mathcal{O}$ of X. Pick any $x_0 \in \mathcal{O}$ and let $r_0 > 0$ such that

$$B_c(x_0, r_0) \subset \mathcal{O}.$$

Since U_1 is dense, one has $B(x_0, r_0) \cap U_1$ is open and non-empty. This allows to pick $x_1 \in B(x_0, r_0) \cap U_1$ and $r_1 > 0$ such that

$$B_c(x_1, r_1) \subset B(x_0, r_0) \cap U_1 \qquad \text{with} \quad 0 < r_1 < \frac{r_0}{2}.$$

We iterate this reasoning and construct a sequence $(x_n)_n \subset X$ and $(r_n)_n \subset (0, \infty)$ such that

$$B_c(x_{n+1}, r_{n+1}) \subset B(x_n, r_n) \cap U_{n+1} \qquad \text{with} \quad 0 < r_{n+1} < \frac{r_n}{2} \qquad \forall n \in \mathbb{N}.$$

In particular, one sees that $r_n < \frac{r_0}{2^n}$ so that the sequence $(x_n)_n$ is a Cauchy sequence (see Exercise 2.1). Since X is complete, there is $x \in X$ such that $\lim_n x_n = x$. Since $x_{n+k} \in B(x_n, r_n)$ for any $n \in \mathbb{N}$ and any $k \in \mathbb{N}$, taking the limit as $k \to \infty$, we get that

$$x \in \overline{B(x_n, r_n)} \subset B_c(x_n, r_n) \qquad \forall n \in \mathbb{N}.$$

In particular, $x \in \mathcal{O} \cap A$ which is non-empty and this proves the result. ∎

An important consequence of Baire Theorem is the following well-known fact (notice that the usual proof – using Cantor diagonalisation argument – is much more complicated than the one using Baire Theorem):

Corollary 2.3.2 $\mathbb{R}$ is not countable.

Proof. Recall that $(\mathbb{R}, |\cdot|)$ is a complete normed space. We argue by contradiction assuming that $\mathbb{R}$ is countable, i.e. there is a sequence $\{x_n\}_n \subset \mathbb{R}$ such that

$$\mathbb{R} = \bigcup_{n\in\mathbb{N}} \{x_n\}.$$

Since the singleton $\{x_n\}$ is closed in $(\mathbb{R}, |\cdot|)$ with empty interior, Baire Theorem implies that

$$\mathrm{Int}(\mathbb{R}) = \varnothing$$

which is impossible. Therefore, $\mathbb{R}$ is not countable. ■

Baire Theorem enjoys a lot of interesting consequences: the fist one is the following surprising result:

Theorem 2.3.3 — Banach-Steinhaus Theorem – Uniform Boundedness Principle. Let $(X, \|\cdot\|_X)$ a Banach space and let $(Y, \|\cdot\|_Y)$ be a normed space. Let $(T_i)_{i\in I} \subset \mathscr{L}(X,Y)$ be a given collection of continuous linear applications. Assume that, for any $x \in X$, there exists $M_x > 0$ such that

$$\sup_{i\in I} \|T_i(x)\|_Y \leqslant M_x. \tag{2.8}$$

Then, there exists $M > 0$ such that, for any $x \in X$ and any $i \in I$, it holds

$$\|T_i(x)\|_Y \leqslant M\|x\|_X$$

i.e. $\sup_{i\in I} \|T_i\|_{\mathscr{L}(X,Y)} \leqslant M < \infty$.

Remark 2.3.2 The Uniform Boundedness Principle can be reformulated as follows: Let $(X, \|\cdot\|_X)$ a Banach space and let $(Y, \|\cdot\|_Y)$ be a normed space. Let $\mathcal{F} \subset \mathscr{L}(X,Y)$ be a given collection of continuous linear applications (here above $\mathcal{F} = (T_i)_{i\in I}$). Then the following are equivalent:

- **(Pointwise boundedness)** For every $x \in X$, the set $\{T(x)\ ;\ T \in \mathcal{F}\}$ is bounded in Y, i.e.
$$\sup_{T\in\mathcal{F}} \|T(x)\|_Y = M_x < \infty, \qquad \forall x \in X;$$
- **(Uniform boundedness)** The operator norms $\{\|T\|_{\mathscr{L}(X,Y)}\ ;\ T \in \mathcal{F}\}$ are bounded, i.e.
$$\sup_{T\in\mathcal{F}} \|T\|_{\mathscr{L}(X,Y)} = M < \infty.$$

✍ *Proof.* For every $n \in \mathbb{N}$ set

$$X_n = \{x \in X\ ;\ \|T_i(x)\|_Y \leqslant n\ \forall i \in I\,\}.$$

Since for any $i \in I$, T_i is continuous, one sees that X_n is closed as the intersection of closed subsets of X. Moreover, according to (2.8),

$$X = \bigcup_n X_n$$

(Get convinced by this). It follows from Baire Theorem that there exists $n_0 \in \mathbb{N}$ such that $\mathrm{Int}(X_{n_0}) \neq \varnothing$. Pick then $x_0 \in X$ and $r > 0$ so that $B(x_0, r) \subset X_{n_0}$. By definition, it holds

$$\|T_i(x_0 + rz)\|_Y \leqslant n_0 \qquad \forall i \in I\,;\, \forall z \in B(0,1)$$

(Explain why). By linearity, we get

$$\|T_i(z)\|_Y \leqslant \frac{1}{r}\left(\|T_i(x_0+rz)\|_Y + \|T_i(x_0)\|_Y\right) \leqslant 2\frac{n_0}{r} \qquad \forall i \in I\,;\, \forall z \in B(0,1).$$

This clearly gives the result with $M = 2\frac{n_0}{r}$. *(Again, explain why)*. ■

■ **Example 2.5** Let us explain here the necessity of the completeness assumption. Namely, consider $X = c_{00}(\mathbb{N})$ the set of all real sequences with finite support (i.e., only finitely many terms are non-zero) endowed with the norm $\|\boldsymbol{x}\|_\infty = \sup_n |x_n|$, $\boldsymbol{x} = (x_n)_n \in X$. We already saw in Example 2.2 (see more precisely Exercise 2.2) that $(X, \|\cdot\|)$ is a normed space which is not complete. Consider then $Y = \mathbb{R}$ endowed with its usual norm and, for $n \in \mathbb{N}$, let

$$T_n(\boldsymbol{x}) = \sum_{k=1}^n x_k, \qquad \forall \boldsymbol{x} = \{x_k\}_k \in X.$$

For any $\boldsymbol{x} \in X$, there exists $N \geqslant 1$ such that $x_k = 0$ for all $k \geqslant N$ and it holds

$$|T_n(\boldsymbol{x})| \leqslant \sum_{k=1}^n |x_k| \leqslant \sum_{k=1}^N |x_k| < \infty$$

i.e. $\sup_n |T_n(\boldsymbol{x})| < \infty$ for any $\boldsymbol{x} \in X$. Consider now the sequence $\{\boldsymbol{x}^{(n)}\}_n$ in X defined as $\boldsymbol{x}^{(n)} = \{x_k^{(n)}\}_k$ with $x_k = 1$ for $k \leqslant n$ and $x_k = 0$ for $k > n$. Then,

$$|T_n(\boldsymbol{x}^{(n)})| = n \qquad \forall n \in \mathbb{N}$$

while $\|\boldsymbol{x}^{(n)}\|_\infty = 1$ for any $n \in \mathbb{N}$. Therefore,

$$\sup_{\|\boldsymbol{x}\|_\infty \leqslant 1} |T_n(\boldsymbol{x})| \geqslant n$$

and $\sup_n \|T_n\|_{\mathscr{L}(X,\mathbb{R})} = \infty$. ■

Notation: Given a vector space X, for any subset $A \subset X$ and any $x \in X$, $\lambda \in \mathbb{R} \setminus \{0\}$ we set

$$\{x\} + A = \{z \in X \;\; z = x + y \quad y \in A\} = \{z \in X\,;\; z - x \in A\}$$
$$\text{and} \quad \lambda A = \{z \in X\,;\, z = \lambda y,\; y \in A\} = \{z \in X\,;\, \lambda^{-1} z \in A\}.$$

We have the following fact which is easy to check: if A is convex and $x \in A$ then

$$\{x\} + A \subset 2A.$$

Indeed, given $y \in A$, since $x \in A$ and A is convex, $\frac{x+y}{2} \in A$ and

$$x + y = 2\left(\frac{x+y}{2}\right) \in 2A.$$

Another consequence of Baire Theorem is the following

Theorem 2.3.4 — Open mapping theorem. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be two Banach spaces and let

$$T \,:\, X \to Y$$

be a continuous linear application which is surjective (i.e. $T \in \mathscr{L}(X,Y)$ is onto). Then, T maps open sets of X into open sets of Y, i.e. if $U \subset X$ is an open set then $T(U)$ is an open set of Y.

Proof. In the proof, we use the following notations: given $r > 0$ and $x \in X$, we will denote by $\boldsymbol{B}(x,r) \subset X$ the open ball *centered at* x with radius $r > 0$, i.e.

$$\boldsymbol{B}(x,r) = \{y \in X \;;\; \|x - y\|_X < r\}$$

and, will simply write

$$\boldsymbol{B}_r = \boldsymbol{B}(0_X, r)$$

whenever the center of the ball is 0_X. We will instead denote balls in Y as $\mathbb{B}(y,r)$, i.e. for any $y \in Y$,

$$\mathbb{B}(y,r) = \{z \in Y \;;\; \|z - y\|_Y < r\}.$$

(i) Reducing the problem to $T(\boldsymbol{B}_1)$. We begin with the following observation: it is enough to prove that $T(\boldsymbol{B}_1)$ contains an open ball of Y centered at 0_Y. Indeed, assume there exists $\varrho > 0$ such that

$$\mathbb{B}(0_Y, \varrho) \subset T(\boldsymbol{B}_1) \tag{2.9}$$

and let $U \subset X$ be an open set. To deduce that $T(U)$ is an open subset of Y from (2.9), we pick $y \in T(U)$. By definition, there is $x \in U$ such that $y = T(x)$. Being U open, there exists $\varepsilon > 0$ such that $\boldsymbol{B}(x,\varepsilon) \subset U$ which can be written equivalently as

$$\{x\} + \boldsymbol{B}_\varepsilon = \{x\} + \varepsilon \boldsymbol{B}_1 \subset U.$$

Using then the linearity of T (in particular that $T(0_X) = 0_Y$) and (2.9), we deduce ***(Check the details)*** that

$$\mathbb{B}(y, \varrho\,\varepsilon) = \{y\} + \varepsilon\mathbb{B}(0_Y, \varrho) \subset \{y\} + \varepsilon T(\boldsymbol{B}_1) = T\left(\{x\} + \varepsilon \boldsymbol{B}_1\right) \subset T(U)$$

which proves that $T(U)$ is open.

Let us there focus now on the proof of (2.9). We split the argument into several additional steps:

(ii) The closure $\overline{T(\boldsymbol{B}_1)}$ contains an open ball of Y.

To prove this, we notice that, since T is surjective, for any $y \in Y$, there is $x \in X$ with $T(x) = y$. Since there is $n \in \mathbb{N}$ such that $\|x\|_X < n$, i.e. $x \in \boldsymbol{B}_n$ and therefore $y \in T(\boldsymbol{B}_n) \subset \overline{T(\boldsymbol{B}_n)}$ for some $n \in \mathbb{N}$, i.e.

$$Y = \bigcup_n \overline{T(\boldsymbol{B}_n)}.$$

Since $\overline{T(\boldsymbol{B}_n)}$ is closed (this is why we needed to introduce the closure here!) and Y is a Banach space, we deduce from Baire Theorem that there exists $n_0 \in \mathbb{N}$ such that

$$\mathrm{Int}\left(\overline{T(\boldsymbol{B}_{n_0})}\right) \neq \varnothing.$$

As easily seen,

$$\overline{T(\boldsymbol{B}_{n_0})} = n_0\overline{T(\boldsymbol{B}_1)}$$

so that

$$\mathrm{Int}\left(\overline{T(\boldsymbol{B}_1)}\right) \neq \varnothing$$

which exactly means that $\overline{T(\boldsymbol{B}_1)}$ is containing some open ball of Y.

(iii) $\overline{T(\boldsymbol{B}_1)}$ *contains an open ball of* Y centered *at* 0_Y.

From point *(ii)*, there is some $y \in \overline{T(\boldsymbol{B}_1)}$ and $r > 0$ such that

$$\mathbb{B}(y, 2r) \subset \overline{T(\boldsymbol{B}_1)}.$$

Notice that $y \in \overline{T(\boldsymbol{B}_1)}$ and, using the fact that $\boldsymbol{B}_1$ is symmetric and $T \in \mathscr{L}(X, Y)$, $-y \in \overline{T(\boldsymbol{B}_1)}$. Observe then that

$$\mathbb{B}(0_Y, 2r) = \{-y\} + \mathbb{B}(y, 2r) \subset \{-y\} + \overline{T(\boldsymbol{B}_1)} \subset 2\overline{T(\boldsymbol{B}_1)}$$

we deduce easily that

$$\mathbb{B}(0_Y, r) \subset \overline{T(\boldsymbol{B}_1)}. \tag{2.10}$$

(iv) Removing the closure: proof of (2.9).

Let us now deduce (2.9) from (2.10). We pick $\varrho = \frac{r}{2}$ and observe that, for $y \in \mathbb{B}(0_Y, \frac{r}{2})$, one has

$$2y \in \mathbb{B}(0_Y, r) \subset \overline{T(\boldsymbol{B}_1)}$$

which means that, for any $\varepsilon > 0$, there is $z \in T(\boldsymbol{B}_1)$ such that

$$\|2y - z\|_Y < \varepsilon.$$

In particular, with $\varepsilon = \frac{r}{2}$, there is $x_1 \in \boldsymbol{B}_1$ such that $z = T(x_1)$ and

$$\|2y - T(x_1)\|_Y < \frac{r}{2}.$$

Now, this means that

$$2\,(2y - T(x_1)) \in \mathbb{B}(0_Y, r) \subset \overline{T(\boldsymbol{B}_1)}$$

so, with the same reasoning, there exists $x_2 \in \boldsymbol{B}_1$ such that

$$\|2\,(2y - T(x_1)) - T(x_2)\|_Y < \frac{r}{2}$$

or, equivalently,

$$\left\| y - T\left(\frac{1}{2}x_1\right) - T\left(\frac{1}{4}x_2\right) \right\|_Y < \frac{r}{8}.$$

Introducing $z_1 = \frac{1}{2}x_1$, $z_2 = \frac{1}{4}x_2$, we see that

$$\|y - T(z_1)\|_Y < \frac{r}{4}, \qquad \|y - T(z_1) - T(z_2)\|_Y < \frac{r}{8}.$$

Repeating the same argument, we construct easily by induction a sequence $(z_n)_n \subset X$ such that

$$\|z_n\|_X < \frac{1}{2^n} \qquad \text{and} \qquad \left\| y - \sum_{k=1}^{n} T(z_k) \right\|_Y < \frac{r}{2^{n+1}} \qquad \forall n \in \mathbb{N}.$$

One deduces from the first property that the sequence $(\boldsymbol{x}_n)_n \subset X$ given by $\boldsymbol{x}_n = \sum_{k=1}^{n} z_k$ is a Cauchy sequence in X. Since X is a Banach space, it converges to $\boldsymbol{x} \in X$. Moreover

$$\|\boldsymbol{x}\|_X < \sum_{n=1}^{\infty} \frac{1}{2^n} = 1$$

and the second property (together with the continuity of T) implies that

$$y = \lim_n T(\boldsymbol{x}_n) = T(\boldsymbol{x}).$$

In other words, we proved that for any $y \in Y$ with $\|y\|_Y < \frac{r}{2}$, one can find $x \in \boldsymbol{B}_1$ such that $y = T(\boldsymbol{x})$ which proves the claim. ■

Remark 2.3.3 The Open Mapping Theorem asserts that, if $T \in \mathscr{L}(X, Y)$ is surjective then it is an open mapping. Actually, the converse is true: if $T \in \mathscr{L}(X, Y)$ is an open mapping, then T is surjective. Indeed, since T is open, $T(X)$ is an open set of Y and it contains a ball $\mathbb{B}(z, r)$ with $z \in T(X)$, $r > 0$. It is easy to deduce then from the linearity of T that

$$\mathbb{B}(0_Y, r) \subset T(X)$$

and then using the homogeneity of the norm and the fact that T is linear, we get $Y \subset T(X)$, i.e. T is surjective.

The Open Mapping Theorem could be reformulated like this: if $T \in \mathscr{L}(X, Y)$ is surjective, there is a constant $C > 0$ such that, for any $y \in Y$, there is a solution $x \in X$ to the linear equation

$$T(x) = y$$

which obeys the bound $\|x\|_X < \frac{1}{r}\,\|y\|_Y$ where $r \in (0, \varrho)$ and $\varrho > 0$ is the constant in (2.9). Indeed, for $y = 0_Y$, we can choose $x = 0_X$. If $y \neq 0_Y$, for any $0 < r < \varrho$, let

$$z_r := \frac{r}{\|y\|_Y} y.$$

One has $z_r \in \mathbb{B}(0_Y, r) \subset \mathbb{B}(0_Y, \varrho)$ so that, according to the Open Mapping Theorem, there is $x_r \in \boldsymbol{B}_1$ such that

$$T(x_r) = z_r.$$

But then, setting $x = \frac{\|y\|_Y}{r} x_r$ one has $T(x) = y$. Now

$$\|x\|_X \leqslant \frac{\|y\|_Y}{r} \|x\|_X < \frac{1}{r} \|y\|_Y$$

since $x_r \in \boldsymbol{B}_1$. Therefore, for any $r \in (0, \varrho)$ we proved that there is a solution x to $T(x) = y$ and

$$\|x\|_X \leqslant \frac{1}{r} \|y\|_Y.$$

In particular, we can deduce the following:

Corollary 2.3.5 Let $(X_i, \|\cdot\|_i)$ $(i = 1, 2)$ be two Banach spaces and let $T : X_1 \to X_2$ be a continuous linear application which is bijective, i.e., injective (one-to-one) and surjective (onto). Then , its inverse T^{-1} is also continuous, i.e. $T^{-1} \in \mathscr{L}(X_2, X_1)$.

Proof. Since T is surjective, according to the open mapping theorem, one sees that, if $x \in X_1$ is chosen so that $\|T(x)\|_2 < c$ then $\|x\|_1 < 1$. By homogeneity of the norm (see the previous reasoning) and the uniqueness of the solution we see that, for any $r \in (0, c)$,

$$\|x\|_1 \leqslant \frac{1}{r} \|T(x)\|_2, \qquad \forall x \in X_1$$

and, letting $r \to c$, we get

$$\|x\|_1 \leqslant \frac{1}{c} \|T(x)\|_2 \qquad \forall x \in X_1$$

and this means exactly, setting for instance $x = T^{-1}(y)$ that $\|T^{-1}y\|_1 \leqslant \frac{1}{c}\|y\|_2$, for any $y \in X_2$. As well-known, this proves that $T^{-1} \in \mathscr{L}(X_2, X_1)$. ■

An important consequence of the above *Open Mapping Theorem* is the following:

Proposition 2.3.6 Let $(X, \|\cdot\|)$ be a Banach space and let $\|\cdot\|_0$ be a norm on X such that

1. $(X, \|\cdot\|_0)$ is a Banach space;
2. there exists $C_0 > 0$ such that

$$\|x\|_0 \leqslant C_0 \|x\| \qquad \forall x \in X_0.$$

Then, the two norms $\|\cdot\|$ and $\|\cdot\|_0$ are equivalent.

Proof. We only need to prove that there exists $C_1 > 0$ such that

$$\|x\| \leqslant C_1 \|x\|_0 \qquad \forall x \in X.$$

For simplicity, we will write Y for the normed space $Y = (X, \|\cdot\|_0)$ and simply X for $(X, \|\cdot\|)$. By assumptions, both spaces Y and X are Banach spaces. Moreover, from 2., the identity mapping

$$\imath : x \in X \mapsto x \in Y$$

belongs to $\mathscr{L}(X,Y)$ ***(explain this)***. Since the identity mapping $\imath$ is clearly invertible with inverse

$$\imath^{-1} \,:\, x \in Y \mapsto x \in X$$

one deduces from the previous corollary that $\imath^{-1} \in \mathscr{L}(Y,X)$ i.e. there is $C_1 = \|\imath^{-1}\|_{\mathrm{op}}$ such that

$$\|\imath^{-1}(x)\| = \|x\| \leqslant C_1\|x\|_0 \qquad \forall x \in X$$

which proves the result. ■

An equivalent formulation of the open mapping theorem is the following:

Theorem 2.3.7 — Closed Graph Theorem. Let $(X_i, \|\cdot\|_i)$ $(i=1,2)$ be two Banach spaces and let $T : X_1 \to X_2$ be a linear application. Assume that the graph of T

$$\mathcal{G}(T) = \{(x_1, T(x_1) \in X_1 \times X_2, \qquad x_1 \in X_1\}$$

is closed in $X_1 \times X_2$ (endowed with the norm $\|(x_1,x_2)\|_{\max} = \max(\|x_1\|_1, \|x_2\|_2)$). Then, $T \in \mathscr{L}(X_1,X_2)$.

✍ *Proof.* Consider, on X_1 the norm

$$\|x\|_T = \|x\|_1 + \|T(x)\|_2, \qquad x \in X_1.$$

One checks that $\|\cdot\|_T$ is indeed a norm. Moreover, since $\mathcal{G}(T)$ is closed in $X_1 \times X_2$, one can check that $(X_1, \|\cdot\|_T)$ is a Banach space ***(Check this)***. Moreover, one clearly has $\|x\|_1 \leqslant \|x\|_T$ for any $x \in X_1$. Then, the previous Proposition asserts that the norms $\|\cdot\|_1$ and $\|\cdot\|_T$ are equivalent norms on X_1, so that there is $c > 0$ such that $\|x\|_T \leqslant c\|x\|_1$ for any $x \in X_1$. This is enough to prove the continuity of T. ■

2.4 Problems

Exercise 2.6 — Extension of Bounded Operators. . Let Y be a dense subspace of a normed space $(X, \|\cdot\|_X)$, and let $(Z, \|\cdot\|_Z)$ be a Banach space. Suppose that $L : Y \to Z$ is linear and continuous. Show that there exists a unique bounded linear operator $\tilde{L} \in \mathscr{L}(X,Z)$ whose restriction to Y is L. Prove that

$$\|\tilde{L}\|_{\mathscr{L}(X,Z)} = \|L\|_{\mathscr{L}(Y,Z)}.$$

■

Exercise 2.7 Let $(X, \|\cdot\|)$ be a normed space. Assume that for any $\{x_n\}_n \subset X$ such that

$$\sum_{n=1}^{\infty} \|x_n\| < \infty$$

then the series $\sum_{n=1}^{\infty} x_n$ converges in X. Prove that $(X, \|\cdot\|)$ is a Banach space. ■

Exercise 2.8 Let $(X, \|\cdot\|)$ be a Banach space. We denote by $(X^\star, \|\cdot\|_*)$ its dual space.

Let $Y \subset X$ be a given subset such that $\Phi(Y)$ is bounded in $\mathbb{R}$ for any $\Phi \in X^\star$. Prove that Y is bounded in X. ■

Exercise 2.9 Let $(X, \|\cdot\|)$ be a Banach space and let $T \in \mathscr{L}(X)$. We denote with T^n the n-th iterated of T (i.e. $T^0 = I, T^1 = T, T^2 = T \circ T$ and so on). Assume that, for any $x \in X$, there is $n \in \mathbb{N}$ such that $T^n(x) = 0$. Prove that there exists $n \in \mathbb{N}$ such that

$$T^n = 0.$$

(*Hint: Use Baire Theorem and Exercise 1.7 in Chapter 1*). ■

Exercise 2.10 Let X, Y be two Banach spaces. Let $\{T_n\}_n \subset \mathscr{L}(X,Y)$ be such that, for every $x \in X$, $\{T_n(x)\}_n \subset Y$ converges to a limit denoted by $T(x)$. Deduce from Banach-Steinhaus Theorem that

1. $\sup_n \|T_n\|_{\mathscr{L}(X,Y)} < \infty$,
2. $T \in \mathscr{L}(X,Y)$
3. $\|T\|_{\mathscr{L}(X,Y)} \leqslant \liminf_n \|T_n\|_{\mathscr{L}(X,Y)}$.

■

2.5 Complements

2.5.1 Compact operators in Banach spaces

In the sequel, $(X, \|\cdot\|)$ is a given Banach space.

Definition 2.5.1 Let $A \in \mathscr{L}(X)$. We say that A is a compact operator if the image of $B_c(0,1)$ through A is contained in a compact set of X, i.e. for any sequence $(\boldsymbol{x}_n)_n \subset X$ with $\|\boldsymbol{x}_n\| \leqslant 1$, the sequence $(A(\boldsymbol{x}_n))_n$ admits a subsequence which converges. We shall denote by $\mathscr{K}(X)$ the collection of all compact operators in X.

Remark 2.5.1 Since A is linear, it is clear that A is compact if and only if for any *bounded* sequence $(\boldsymbol{x}_n)_n$, the sequence $(A(\boldsymbol{x}_n))_n$ admits a subsequence which converges.

Definition 2.5.2 Let $A \in \mathscr{L}(X)$. We say that A is of finite rank if the image of A, $R(A) = \{A\boldsymbol{x}\,,\ \boldsymbol{x} \in X\}$ is a linear subspace of finite dimension of X.

■ **Example 2.6** Any finite rank operator is compact. ■

One has the following simple property whose proof is a simple **Exercise**:

Lemma 2.5.1 If $A \in \mathscr{K}(X)$ and $B \in \mathscr{L}(X)$ then $B \circ A$ and $A \circ B$ belong to $\mathscr{K}(X)$.

Proposition 2.5.2 Let $(A_n)_n \subset \mathscr{K}(X)$ and let $A \in \mathscr{L}(X)$ be given with

$$\lim_{n\to\infty} \|A_n - A\|_{\mathscr{L}(X)} = 0.$$

Then, $A \in \mathscr{K}(X)$.

Proof. Let $(\boldsymbol{x}_n)_n$ be a sequence in X with $\|\boldsymbol{x}_n\| \leqslant 1$ for any $n \in \mathbb{N}$. We should extract a subsequence $(\boldsymbol{x}_{\varphi(n)})_n$ such that $(A(\boldsymbol{x}_{\varphi(n)}))_n$ converges. Since A_1 is compact, one can extract from $(\boldsymbol{x}_n)_n$ a subsequence $(\boldsymbol{x}_{1,n})_n$ such that $(A_1(\boldsymbol{x}_{1,n}))_n$ converges. Since A_2 is compact, one can extract from $(\boldsymbol{x}_{1,n})_n$ a subsequence $(\boldsymbol{x}_{2,n})_n$ such that $(A_2(\boldsymbol{x}_{2,n}))_n$ converges. We construct then inductively, for any $k \geqslant 1$, a subsequence $(\boldsymbol{x}_{k+1,n})_n$ of $(\boldsymbol{x}_{k,n})_n$ such that $(A_{k+1}(x_{k+1,n}))_n$ converges. We define then the diagonal subsequence $(\boldsymbol{x}_{\varphi(n)})_n$ with $\boldsymbol{x}_{\varphi(n)} = \boldsymbol{x}_{n,n}$. Let us prove that $(A(\boldsymbol{x}_{\varphi(n)}))_n$ is a Cauchy sequence. For all $k, n, m \in \mathbb{N}$ one has

$$\|A(\boldsymbol{x}_{\varphi(m)}) - A(\boldsymbol{x}_{\varphi(n)})\| \leqslant \|A(\boldsymbol{x}_{\varphi(m)}) - A_k(\boldsymbol{x}_{\varphi(m)})\| + \\ \|A_k(\boldsymbol{x}_{\varphi(m)}) - A_k(\boldsymbol{x}_{\varphi(n)})\| + \|A_k(\boldsymbol{x}_{\varphi(n)}) - A(\boldsymbol{x}_{\varphi(n)})\|.$$

Now, for any $\varepsilon > 0$, pick $k \geqslant 1$ such that $\|A - A_k\|_{\mathscr{L}(H)} \leqslant \frac{\varepsilon}{3}$. Since $(A_k(\boldsymbol{x}_{\varphi(n)}))_n$ converges, it is a Cauchy sequence so that there exists $N = N_k \geqslant 1$ such that

$$\|A_k(\boldsymbol{x}_{\varphi(m)}) - A_k(\boldsymbol{x}_{\varphi(n)})\| \leqslant \frac{\varepsilon}{3} \qquad \forall n, m \geqslant N_k.$$

Then, from the above inequality, one sees that

$$\|A(\boldsymbol{x}_{\varphi(m)}) - A(\boldsymbol{x}_{\varphi(n)})\| \leqslant \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon \qquad \forall n, m \geqslant N_k.$$

This shows that $(A(\boldsymbol{x}_{\varphi(n)}))_n$ is a Cauchy sequence and since X is complete, $(A(\boldsymbol{x}_{\varphi(n)}))_n$ converges which proves the result. ■

Corollary 2.5.3 Let $(A_n)_n$ be a sequence of finite-rank operators and let $A \in \mathscr{L}(X)$ such that

$$\lim_{n\to\infty} \|A_n - A\|_{\mathscr{L}(X)} = 0$$

then $A \in \mathscr{K}(X)$.

2.5.2 More properties of the dual space/Hahn-Banach Theorem

We recall several notions concerning partial order relation: let P be a set with a (partial) order relation "$\leqslant$". We say that a subset $Q \subset P$ is totally ordered if for any pair $(a, b) \in Q$ either $a \leqslant b$ or $b \leqslant a$. Let $Q \subset P$; we say that $c \in P$ is an upper bound for Q if $a \leqslant c$ for every $a \in Q$. We say that $m \in P$ is a maximal element of P if there is no element $x \in P$ such that $m \leqslant x$, except for $x = m$. Note that a maximal element of P need not be an upper bound for P. We say that P is inductive if every totally ordered subset $Q \subset P$ has an upper bound. We have the following

Theorem 2.5.4 — Zorn's lemma. Every non empty ordered set that is inductive has a maximal element.

Theorem 2.5.5 — Helly, Hahn-Banach analytic form. Let X be a vector space over $\mathbb{R}$. Let $p : X \to \mathbb{R}$ be a function satisfying

1. $p(\lambda x) = \lambda p(x)$ for any $x \in X$ and any $\lambda > 0$.
2. $p(x + y) \leqslant p(x) + p(y)$ for any $x, y \in X$.

Let $Y \subset X$ be a linear subspace and $g : Y \to \mathbb{R}$ a linear function such that

3. $g(x) \leqslant p(x)$ for any $x \in Y$.

Then, there exists a linear function $f : X \to \mathbb{R}$ such that

$$f(x) = g(x) \qquad \forall x \in Y$$

(We say then that f extends g to X) and

$$f(x) \leqslant p(x) \quad \forall x \in X.$$

Proof. Consider the set

$$\mathcal{P} = \Big\{ h \ : \ D(h) \to X \ , \ D(h) \text{ is a linear subspace of } X, \ h \text{ is linear, } Y \subset D(h)$$
$$h(x) = g(x) \quad \forall x \in Y \qquad \text{and} \qquad h(x) \leqslant p(x) \quad \forall x \in D(h). \Big\}$$

It is clear that $\mathcal{P}$ is nonempty since $g \in \mathcal{P}$ with $D(g) = Y$. On $\mathcal{P}$ we define the order relation

$$h_1 \leqslant h_2$$

if and only if

$$D(h_1) \subset D(h_2) \qquad \text{and} \qquad h_2(x) = h_1(x) \qquad \forall x \in D(h_1).$$

We claim that $\mathcal{P}$ is inductive. Indeed, let $\mathcal{Q} \subset \mathcal{P}$ be a totally ordered subset; we write $\mathcal{Q}$ as $\mathcal{Q} = (h_i)_{i \in I}$ and we set

$$D(h) = \bigcup_{i \in I} D(h_i) \qquad h(x) = h_i(x) \qquad \text{if } x \in D(h_i) \quad \text{for some } i \in I.$$

We need first to prove that $h \ : \ D(h) \to \mathbb{R}$ is well-defined. Indeed, if $x \in D(h_i) \cap D(h_j)$ then $h(x) = h_i(x)$ and $h(x) = h_j(x)$. But, since $\mathcal{Q}$ is totally ordered, one has $h_i \leqslant h_j$ (or $h_j \leqslant h_i$). From the choice of the order relation, one sees then that $h_j(x) = h_i(x)$ so that $h(x)$ is well-defined. The fact that h is linear is clear. Moreover, it is clear that $Y \subset D(h)$. It is easy then to check that h is an upper bound for $\mathcal{Q}$. Then, according to Zorn's lemma, $\mathcal{P}$ has a maximal element, say $f \in \mathcal{P}$. The result will be proven if we can prove that $D(f) = X$. Suppose, by contradiction, that $D(f) \neq X$. Let $x_0 \notin D(f)$ and set

$$D(h) = D(f) + x_0\mathbb{R} = \{y \in X \ ; \ y = x + tx_0, \ x \in D(f), \, t \in \mathbb{R}\}$$

and, for any $y = x + tx_0 \in D(f)$, set

$$h(y) = f(x) + t\alpha$$

where the constant $\alpha \in \mathbb{R}$ will be chosen in such a way that $h \in \mathcal{P}$, i.e.

$$f(x) + t\alpha \leqslant p(x + tx_0) \qquad \forall x \in Y, \forall t \in \mathbb{R}.$$

According to *1)*, it suffices to check that

$$f(x) + \alpha \leqslant p(x + x_0) \qquad \text{and} \qquad f(x) - \alpha \leqslant p(x - x_0) \quad \forall x \in D(f).$$

In other words, we must find some α satisfying

$$\sup_{y \in D(f)} \left(f(y) - p(y - x_0)\right) \leqslant \alpha \leqslant \inf_{x \in D(f)} \left(p(x + x_0) - f(x)\right).$$

Such an α necessarily exists since

$$f(y) - p(y - x_0) \leqslant p(x + x_0) - f(x) \qquad \forall x, y \in D(f).$$

With this choice of α, one sees that $h \in \mathcal{P}$ and $f \leqslant h$; but this is impossible, since f is maximal and $h \neq f$. ■

A first simple consequence of the above Theorem is the following

Corollary 2.5.6 Let $(X, \|\cdot\|)$ be a normed space and let $Y \subset X$ be a linear subspace. If $g : Y \to \mathbb{R}$ is a continuous linear mapping, then there exists $f \in X^\star$ that extends g and such that

$$\|f\|_\star = \|g\|_{Y^\star}.$$

Proof. The proof is a direct application of the above Theorem with the choice $p(x) = \|g\|_{Y^\star}\,\|x\|$ for any $x \in X$. ■

Corollary 2.5.7 Let $(X, \|\cdot\|)$ be a normed space. For every $x_0 \in X$ there exists $\Phi_0 \in X^\star$ such that

$$\|\Phi_0\|_* = \|x_0\| \qquad \text{and} \qquad \Phi_0(x_0) = \|x_0\|^2.$$

Proof. Consider $Y = \mathrm{Span}(x_0)$ and $g(tx_0) = t\|x_0\|^2$ for any $t \in \mathbb{R}$. Then, $g \in Y^\star$ and $\|g\|_{Y^\star} = \|x_0\|$. The previous corollary gives then the result. ■

Corollary 2.5.8 Let $(X, \|\cdot\|)$ be a normed space. For every $x \in X$ we have

$$\|x\| = \sup_{\substack{\Phi \in X^* \\ \|\Phi\|_\star \leqslant 1}} |\Phi(x)|.$$

Proof. We may always assume that $x \neq 0$. It is clear that

$$\sup_{\substack{\Phi \in X^* \\ \|\Phi\|_\star \leqslant 1}} |\Phi(x)| \leqslant \|x\|.$$

On the other hand, we know from the previous Corollary that there is some $\Phi_0 \in X^\star$ such that $\|\Phi_0\|_\star = \|x\|$ and $\Phi_0(x) = \|x\|^2$. Setting then $\Phi_1 = \dfrac{1}{\|x\|}\Phi_0$ one has $\|\Phi_1\|_\star = 1$ and $\Phi_1(x) = \|x\|$ which proves the result. ■

3. Inner product spaces and Hilbert spaces

3.1 Definitions and Elementary Properties.

3.1.1 General properties

We start with the definition of inner product

Definition 3.1.1 Let H be a given vector space. An inner product $\langle \cdot, \cdot \rangle$ is a mapping from $H \times H$ with value in $\mathbb{R}$ with the following properties:

1. $\langle x, y \rangle = \langle y, x \rangle$ for any $x, y \in H$ *(symmetry)*;
2. $\langle x, \alpha y + \beta z \rangle = \alpha \langle x, y \rangle + \beta \langle x, z \rangle$ for any $x, y, z \in H$, $\alpha, \beta \in \mathbb{R}$ *(bilinearity)*;
3. $\langle x, x \rangle \geqslant 0$ for any $x \in H$;
4. $\langle x, x \rangle = 0$ if and only if $x = 0$.

If H is endowed with an inner product $\langle \cdot, \cdot \rangle$, we say that $(H, \langle \cdot, \cdot \rangle)$ is an inner product space.

■ **Example 3.1** If $H = \mathbb{R}^N$, $N \geqslant 2$, the euclidean product

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^{n} x_i y_i \qquad \boldsymbol{x} = (x_1, \ldots, x_N), \boldsymbol{y} = (y_1, \ldots, y_N) \in \mathbb{R}^N$$

defines an inner product on $\mathbb{R}^N$. ■

■ **Example 3.2** Recall the definition of $\ell^p(\mathbb{N})$ we introduced in **Chapter 2**. For $p = 2$, one has

$$\ell^2(\mathbb{N}) = \{\boldsymbol{x} = (x_k)_k \subset \mathbb{R} \; ; \; \sum_{k=1}^{\infty} |x_k|^2 < \infty\}.$$

The space $\ell^2(\mathbb{N})$ is endowed with the norm

$$\|\boldsymbol{x}\|_2 = \sqrt{\sum_{k=1}^{\infty} |x_k|^2} \qquad \forall \boldsymbol{x} = (x_k)_k \in \ell^2(\mathbb{N}).$$

Introduce

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{k=1}^{\infty} x_k\, y_k \qquad \forall \boldsymbol{x} = (x_k)_k,\ \boldsymbol{y} = (y_k)_k \in \ell^2(\mathbb{N}).$$

Then one can check as an **Exercise** that $\langle \cdot, \cdot \rangle$ is an inner product on $\ell^2(\mathbb{N})$ and that

$$\|\boldsymbol{x}\|_2^2 = \langle \boldsymbol{x}, \boldsymbol{x} \rangle \qquad \forall \boldsymbol{x} \in \ell^2(\mathbb{N}).$$

Notice that $|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| < \infty$ for any $\boldsymbol{x}, \boldsymbol{y} \in \ell^2(\mathbb{N})$ by virtue of Hölder inequality ***(Check this!)***. ∎

One has the following obvious observation

Proposition 3.1.1 If $(H, \langle \cdot, \cdot \rangle)$ is an inner product space, then it is a norm space with the norm given by

$$\|x\| = \sqrt{\langle x, x \rangle} \qquad \forall x \in X.$$

The norm $\|\cdot\|$ is called the norm induced by the inner product $\langle \cdot, \cdot \rangle$ (or sometimes called the inner norm). It satisfies

$$\|x \pm y\|^2 = \|x\|^2 + \|y\|^2 \pm 2\langle x, y \rangle \qquad \forall x, y \in H$$

and the ***parallelogram law***

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2 \qquad \forall x, y \in H. \tag{3.1}$$

One has the fundamental

Proposition 3.1.2 — Cauchy-Schwarz inequality. If $(H, \langle \cdot, \cdot \rangle)$ is an inner product space and $\|\cdot\|$ denotes the inner norm, then, for any $x, y \in H$ it holds

$$|\langle x, y \rangle| \leqslant \|x\|\,\|y\| \qquad \forall x, y \in H.$$

Proof. The proof of this general inequality is exactly the one already encountered for the classical euclidean product in $\mathbb{R}^N$ or mimicking the proof of Hölder inequality for the peculiar case $p = q = 2$. We recall it for convenience. Given $x, y \in H$, consider the mapping $g(t) = \|x + ty\|^2$ for any $t \in \mathbb{R}$. It is nonnegative: $g(t) \geqslant 0$ for any $t \in \mathbb{R}$. Moreover, using the properties of the inner product

$$g(t) = t^2\,\|y\|^2 + 2t\,\langle x, y \rangle + \|y\|^2 \qquad \forall x, y \in H.$$

In particular, being a quadratic function of t which remains nonnegative, one should have that the associate discriminant is nonpositive ***(recall why)*** and therefore

$$4\,\langle x, y \rangle^2 - 4\|x\|^2\|y\|^2 \leqslant 0$$

which is exactly Cauchy-Schwarz inequality. ∎

Notice that the above parallelogram law completely characterizes the norm induced by the inner product. Indeed, assume a normed space $(H, \|\cdot\|)$ is given. How to decide whether there is some inner product $\langle \cdot, \cdot \rangle$ for which $\|\cdot\|$ is the associated norm ? The answer is exactly the parallelogram law; if the norm $\|\cdot\|$ satisfies (3.1) then the norm is induced by an inner product which is actually given by the polarization identity

$$\langle x, y \rangle = \frac{1}{4}\left(\|x + y\|^2 - \|x - y\|^2\right) \qquad \forall x, y \in H.$$

The fact that such an identity defines an inner product is not an easy task but is left as an **Exercise**.

■ **Example 3.3** Introduce the space $\mathscr{C}([0,1])$ with the norm $\|\cdot\|_1$ as defined in **Chapter 1**. Is $\mathscr{C}([0,1])$ an inner product space for this norm ? This amounts to check whether the norm $\|\cdot\|_1$ can be induced by an inner product. This is not the case since $\|\cdot\|_1$ does not satisfy the parallelogram law. To get convinced by this, consider for instance the functions $f(x) = 1$ and $g(x) = 2x$ for any $x \in [0,1]$. Then one has

$$\|f\|_1 = 1, \qquad \|g\|_1 = 1, \qquad \|f-g\|_1 = \frac{1}{2}, \qquad \|f+g\|_1 = 2$$

(check this!) so that (3.1) is violated by f, g. ■

Exercise 3.1 Check that the space $\ell^1(\mathbb{N})$ defined in **Chapter 1**, is not an inner product space for the norm $\|\cdot\|_1$. Actually, $\ell^p(\mathbb{N})$ is an inner product space if and only if $p = 2$. ■

Exercise 3.2 Check that the space $\mathscr{C}([0,1])$ endowed with the norm $\|\cdot\|_\infty$ is not an inner product space. ■

Since an inner product space is a normed space, we can define the notion of converging sequence and has the following

Proposition 3.1.3 Let $(H, \langle\cdot,\cdot\rangle)$ be an inner product space. Let $(x_n)_n$, $(y_n)_n$ be two converging sequences with respective limits x, y (for the inner norm). Then

$$\lim_{n\to\infty} \langle x_n, y_n\rangle = \langle x, y\rangle .$$

Proof. The proof is an easy **Exercise**. ■

Exercise 3.3 Let H be an inner product space with inner product $\langle\cdot,\cdot\rangle$ and let $u, v \in H$. Prove that, if $\langle x, u\rangle = \langle x, v\rangle$ for any $x \in H$, then $u = v$. ■

3.1.2 Orthogonality

The interest of inner product is that it allows to define a notion of orthogonality:

Definition 3.1.2 Let $(H, \langle\cdot,\cdot\rangle)$ be a given inner product space. Two vectors $x, y \in H$ are said to be orthogonal if $\langle x, y\rangle = 0$. Given two linear subspaces $M, N \subset H$, we say that M and N are orthogonal subspaces, and denote $M \perp N$ if $\langle x, y\rangle = 0$ for any $x \in M$ and any $y \in N$.

Remark 3.1.1 Notice that, in virtue of Cauchy-Schwarz inequality, the null element $0 \in H$ is orthogonal to any $x \in H$ (since $|\langle x, 0\rangle| \leqslant \|0\|\,\|x\| = 0$).

We also introduce the concept of orthonormal family

Definition 3.1.3 Let $(H, \langle\cdot,\cdot\rangle)$ be a inner product space. A finite family $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_N\} \subset$

H is said to be orthonormal if

$$\|e_k\| = 1 \qquad \forall k = 1, \ldots, N \qquad \text{and} \qquad \langle e_n, e_m \rangle = 0 \qquad \forall n \neq m.$$

Remark 3.1.2 We shall write $\langle e_n, e_m \rangle = \delta_{n,m}$ for any $n, m \in \mathbb{N}$ where $\delta_{n,m}$ is the Kronecker symbol $\delta_{n,m} = 1$ if $n = m$ and $\delta_{n,m} = 0$ otherwise.

With the analogy of finite dimensional space, one has

Proposition 3.1.4 — Gram-Schmidt Procedure. Let $(H, \langle \cdot, \cdot \rangle)$ be an inner product space.

$a)$ Any orthonormal set $\{e_1, \ldots, e_N\} \subset H$ is linearly independent;

$b)$ Given a linearly independent subset $\{v_1, \ldots, v_N\}$ of H. Set $S = \text{span}(v_1, \ldots, v_N)$. Then, there is an orthonormal basis $\{e_1, \ldots, e_N\}$ of S.

✍ *Proof.* We prove first point *a)*. Let $\alpha_1, \ldots, \alpha_N \in \mathbb{R}$ such that

$$\sum_{k=1}^{N} \alpha_k \, e_k = 0.$$

We have to prove that $\alpha_j = 0$ for any j. To do so, for a given $j \in \{1, \ldots, N\}$, it is enough to compute the inner product of the above sum with e_j:

$$0 = \left\langle e_j, \sum_{k=1}^{N} \alpha_k \, e_k \right\rangle = \sum_{k=1}^{N} \alpha_k \, \langle e_j, e_k \rangle = \alpha_j$$

thanks to the orthonormality property ***(check this!)***.

For the point $b)$, we construct the orthonormal basis through ***Gram-Schmidt procedure***. We argue by induction over N. Assume first $N = 1$. Then, $v_1 \neq 0$ and setting $e_1 = \frac{1}{\|v_1\|} v_1$ we have the result. Assume the result is true for some integer $N \geqslant 1$ and let $\{v_1, \ldots, v_{N+1}\}$ be a linearly independent subset of H. Let $\{e_1, \ldots, e_N\}$ be an orthonormal basis of $\text{span}(v_1, \ldots, v_N)$. One has to construct e_{N+1} such that $\{e_1, \ldots, e_N, e_{N+1}\}$ is an orthonormal basis of $\text{span}(v_1, \ldots, v_N, v_{N+1})$. Since $\{v_1, \ldots, v_{N+1}\}$ is a linearly independent familly, $v_{N+1} \notin \text{span}(v_1, \ldots, v_N) = \text{span}(e_1, \ldots, e_N)$. Set

$$h_{N+1} = v_{N+1} - \sum_{k=1}^{N} \langle v_{N+1}, e_k \rangle \, e_k.$$

By construction, $h_{N+1} \in \text{span}(v_1, \ldots, v_{N+1})$ and $h_{N+1} \neq 0$. Moreover

$$\langle h_{N+1}, e_n \rangle = 0 \qquad \forall n \in \{1, \ldots, N\}$$

(check this). Set then $e_{N+1} = h_{N+1}/\|h_{N+1}\|$, it holds $\|e_{N+1}\| = 1$ and $\{e_1, \ldots, e_{N+1}\}$ is an orthonormal family such that $\text{span}(e_1, \ldots, e_{N+1}) \subset \text{span}(v_1, \ldots, v_{N+1})$. Since $\{e_1, \ldots, e_{N+1}\}$ is a linearly independent subset from *a)*, these two linear subspaces have the same dimension so $\text{span}(e_1, \ldots, e_{N+1}) = \text{span}(v_1, \ldots, v_{N+1})$. This proves the result for any $N \in \mathbb{N}$. ■

Exercise 3.4 Let $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_N\}$ be an orthonormal family of H and let $\alpha_1, \ldots, \alpha_N \in \mathbb{R}^N$. Show

$$\left\|\sum_{k=1}^{N} \alpha_k \boldsymbol{e}_k\right\|^2 = \sum_{k=1}^{N} |\alpha_k|^2.$$

Definition 3.1.4 Let $(H, \langle \cdot, \cdot \rangle)$ be an inner space and let M be a subset of H. The orthogonal complement of M is the set

$$M^\perp = \{x \in H\,;\, \langle x, u\rangle = 0\, \forall u \in M\}.$$

In particular, one sees that $M \perp M^\perp$.

Remark 3.1.3 One sees that $M \perp M^\perp$. If $M = \varnothing$ then $M^\perp = H$. Note however that $M^\perp$ is not the set-theoretic complement of M, i.e. $M^\perp \neq M^c$.

Example 3.4 If $H = \mathbb{R}^N$ and $M = \mathrm{span}(\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_k)$ for $k < N$ and $\boldsymbol{e}_i$ are the vectors of the canonical basis of $\mathbb{R}^N$. Then, $M^\perp = \mathrm{span}(\boldsymbol{e}_{k+1}, \ldots, \boldsymbol{e}_N)$. ***(Check this!)***

The properties of the orthogonal complement are listed in the following

Proposition 3.1.5 Let $(H, \langle \cdot, \cdot \rangle)$ is an inner product space and let $M \subset H$. Then,

1. $0 \in M^\perp$.
2. If $0 \in M$ then $M \cap M^\perp = \{0\}$; otherwise $M \cap M^\perp = \varnothing$.
3. $\{0\}^\perp = H$ and $H^\perp = \{0\}$.
4. If M is a non-empty open subset of H, then $M^\perp = \{0\}$.
5. If $N \subset M$ then $M^\perp \subset N^\perp$.
6. $M^\perp$ is a closed linear subspace of H.
7. $M \subset (M^\perp)^\perp$.

Proof. We leave the proof as an **Exercise**. The only non trivial proof is the one of point *4*. ∎

If M is a linear subspace of H, its orthogonal complement can be characterize as follow

Proposition 3.1.6 Let M be a linear subspace of an inner product space $(H, \langle \cdot, \cdot \rangle)$. Then

$$x \in M^\perp \iff \|x - y\| \geqslant \|y\| \qquad \forall y \in M.$$

Proof. Recall that, for any $x, y \in H$ and any $\alpha \in \mathbb{R}$, it holds

$$\|x - \alpha y\|^2 = \|x\|^2 + \alpha^2 \|y\|^2 - 2\alpha \langle x, y\rangle\,. \tag{3.2}$$

If $x \in M^\perp$ and $y \in M$, one gets therefore

$$\|x - \alpha y\|^2 = \|x\|^2 + \alpha^2 \|y\|^2 \qquad \forall \alpha \in \mathbb{R}$$

and, in particular, for $\alpha = 1$, one sees that $\|x - y\| \geqslant \|y\|$. This proves the first implication.

Assume on the contrary that, $x \in H$ is such that $\|x-y\| \geqslant \|y\|$ for any $y \in M$. Since M is a subspace, one sees that $\alpha y \in M$ for any $\alpha \in \mathbb{R}$ and therefore $\|x-\alpha y\|^2 \geqslant \alpha^2\|y\|^2$. From (3.2), it holds then

$$\|x\|^2 - 2\alpha \langle x, y\rangle \geqslant 0 \qquad \forall \alpha \in \mathbb{R}.$$

If $x=0$, this is clearly satisfied and $x \in M^\perp$. If $x \neq 0$, assume that $x \notin M^\perp$, then there is $y \in M$ such that $\langle x, y\rangle \neq 0$. Letting then $\alpha \to \pm\infty$ in the above inequality yields a contradiction. Thus, $x \in M^\perp$. ■

Exercise 3.5 Repeat the above proof to show that actually, under the assumptions of the above Proposition, it holds

$$x \in M^\perp \iff \|x-y\| \geqslant \|x\| \qquad \forall y \in M.$$

■

3.2 Hilbert spaces and Projection Theorem

As for normed spaces, whenever an inner product space is complete, we call it with another name !

Definition 3.2.1 Let $(H, \langle\cdot,\cdot\rangle)$ be an inner product space and let $\|\cdot\|$ be the associated inner norm. If $(H, \|\cdot\|)$ is complete, then $(H, \langle\cdot,\cdot\rangle)$ is called a ***Hilbert space***.

■ **Example 3.5** Every finite-dimensional inner space is a Hilbert space. The space $\ell^2(\mathbb{N})$ with the inner product defined in the previous section is a Hilbert space. ■

■ **Example 3.6** Given $I=[a,b]$, set $H = \mathscr{C}(I)$ with inner product

$$\langle f, g\rangle = \int_a^b f(t)g(t)dt, \qquad f, g \in H.$$

Then, $(H, \langle\cdot,\cdot\rangle)$ is an inner product space but is ***not*** a Hilbert space. ■

One has the following, whose proof is immediate:

Lemma 3.2.1 If $(H, \langle\cdot,\cdot\rangle)$ is a Hilbert space and $Y \subset H$ is a linear subspace of H. Then, Y is a Hilbert space if and only if Y is closed in H.

We recall the definition of convex set (in a general vector space):

Definition 3.2.2 Let X be a given vector space. A subset $C \subset X$ is convex if, for any $x, y \in C$ one has

$$tx + (1-t)y \in C \qquad \forall t \in [0,1].$$

Remark 3.2.1 A convex set is such that, when it contains two points, it contains the segment joining them. Of course, if the vector space X is a normed space then any (closed) ball is convex.

Remark 3.2.2 If X is a given vector space and $V \subset X$ is a linear subspace of X then V is convex.

One has the following

Theorem 3.2.2 — Projection over closed convex subset. Let $(H, \langle\cdot,\cdot\rangle)$ be a given Hilbert space and let $K \subset H$ be ***closed and convex***. Then for every $f \in H$ there exists a unique element $u \in K$ such that

$$\|f-u\| = \min_{v\in K}\|f-v\| = \operatorname{dist}(f,K). \tag{3.3}$$

Moreover, u is characterized by the property

$$u \in K \qquad \text{and} \qquad \langle f-u, w-u\rangle \leqslant 0 \qquad \forall w \in K. \tag{3.4}$$

Proof. Let $(v_n)_n$ be a minimizing sequence for (3.3), i.e. $v_n \in K$ for any $n \in \mathbb{N}$ and

$$\lim_{n\to\infty}\|f-v_n\| = \inf_{v\in K}\|f-v\|.$$

Set for simplicity $D := \inf_{v\in K}\|f-v\|$ and $D_n = \|f-v_n\|$. We claim that $(v_n)_n$ is a Cauchy sequence. Indeed, given $n,m \in \mathbb{N}$, the parallelogram law (3.1) applied with $x = f - v_n$ and $y = f - v_m$ leads to

$$\begin{aligned}\|v_n-v_m\|^2 &= 2\left(\|f-v_n\|^2+\|f-v_m\|^2\right) - \|2f-(v_n+v_m)\| \\ &= 2\left(D_n^2+D_m^2\right) - 4\left\|f-\frac{v_n+v_m}{2}\right\|^2.\end{aligned}$$

Now, since K is convex, for any n,m, $\frac{v_n+v_m}{2} \in K$ so that $\|f-\frac{v_n+v_m}{2}\|^2 \geqslant D^2$. Therefore,

$$\|v_n-v_m\|^2 \leqslant 2\left(D_n^2+D_m\right)^2 - 4D^2.$$

Letting n,m go to infinity, since $D_n \to D$ and $D_m \to D$ one sees that $\lim_{n,m\to\infty}\|v_n - v_m\|^2 = 0$, i.e. $(v_n)_n$ is a Cauchy sequence. Since H is complete, it converges to some limit $u \in H$ and, since K is closed, $u \in K$. In particular, $\|f-u\| = D$ which gives the existence of some $u \in K$ satisfying (3.3).

Before proving uniqueness, we show the equivalence between (3.3) and (3.4). Assume that $u \in K$ satisfies (3.3) and let $w \in K$. We have

$$v = (1-t)u + tw \in K \quad \forall t \in [0,1].$$

Therefore,

$$\|f-u\| \leqslant \|f-v\| = \|(f-u)-t(w-u)\| \qquad \forall t \in [0,1].$$

Taking the square and expanding the second norm, we get

$$\|f-u\|^2 \leqslant \|f-u\|^2 + t^2\|w-u\|^2 - 2t\langle f-u, w-u\rangle \qquad \forall t \in [0,1].$$

or equivalently

$$\langle f-u, w-u\rangle \leqslant \frac{t}{2}\|w-u\|^2 \qquad \forall t \in (0,1].$$

Letting now $t \to 0$, we see that u satisfies (3.4). Conversely, assume that u satisfies (3.4). Then, for any $w \in K$ we have ***(check this)***

$$\|u-f\|^2 - \|w-f\|^2 = 2\langle f-u, w-u\rangle - \|u-w\|^2 \leqslant 0$$

which implies (3.3).

Let us now prove the uniqueness of u. Assume u_1 and u_2 satisfy (3.4). We have

$$\langle f-u_1, v-u_1\rangle \leqslant 0 \quad \forall v\in K \quad \text{and } \langle f-u_2, w-u_2\rangle \qquad \forall w\in K.$$

Choosing $v=u_2$ and $w=u_1$ in the above inequalities yields

$$\langle f-u_1, u_2-u_1\rangle \leqslant 0 \quad \text{and } \langle f-u_2, u_1-u_2\rangle \leqslant 0$$

which, after adding both inequalities, gives $\|u_1-u_2\|^2 \leqslant 0$, i.e. $u_1=u_2$. ■

Definition 3.2.3 Given a closed convex subset $K\subset H$ of a Hilbert space $(H,\langle\cdot,\cdot\rangle)$ and $f\in H$, the unique $u\in K$ satisfying (3.3) is called the *projection* of f over K and is denoted by

$$u=\mathbf{P}_K f.$$

The following asserts that $\mathbf{P}_K$ is a contraction:

Proposition 3.2.3 Let $(H,\langle\cdot,\cdot\rangle)$ be a Hilbert space and let $K\subset H$ be a closed and convex subset. Then,

$$\|\mathbf{P}_K f-\mathbf{P}_K g\| \leqslant \|f-g\| \qquad \forall f,g\in H.$$

✍ *Proof.* Let $f,g\in H$ be given. Set $u_1=\mathbf{P}_K f$ and $u_2=\mathbf{P}_K g$. According to the characterization (3.4) we have

$$\langle f-u_1, v-u_1\rangle \leqslant 0 \quad \forall v\in K \quad \text{and } \langle g-u_2, w-u_2\rangle \qquad \forall w\in K.$$

Choosing $v=u_2$ and $w=u_1$ and adding the corresponding inequalities, we obtain

$$\|u_1-u_2\|^2 \leqslant \langle f-g, u_1-u_2\rangle$$

which, together with Cauchy-Schwarz inequality yields $\|u_1-u_2\|\leqslant\|f-g\|$ ***(Check this)***. ■

In the particular case in which K is a linear subspace, one can be more precise:

Corollary 3.2.4 Let $(H,\langle\cdot,\cdot\rangle)$ be a Hilbert space. Assume that $M\subset H$ is a closed linear subspace and let $f\in H$. Then $u=\mathbf{P}_M f$ is characterized by

$$u\in M \qquad \text{and} \qquad \langle f-u, z\rangle = 0 \qquad \forall z\in M. \tag{3.5}$$

Moreover, $\mathbf{P}_M$ is a linear mapping, $\mathbf{P}_M\in\mathscr{L}(H)$, called the orthogonal projection over M.

✍ *Proof.* By (3.4) we already have $\langle f-u, w-u\rangle\leqslant 0$ for any $w\in M$. In particular, given $z\in M$, since $w=z+u\in M$ one has

$$\langle f-u, z\rangle \leqslant 0.$$

Since $-z\in M$, one has $\langle f-u,z\rangle\geqslant 0$ and (3.5) follows. The fact that (3.5) implies (3.4) is obvious.

Let us now prove that $\mathbf{P}_M$ is linear. To do so, let $f_1,f_2\in H$ with $u_1=\mathbf{P}_M f_1, u_2=\mathbf{P}_M f_2$. For $\alpha_1,\alpha_2\in\mathbb{R}$, let $f=\alpha_1 f_1+\alpha_2 f_2$ and $u=\mathbf{P}_M(f)$. To show that $\mathbf{P}_M$ is linear, we need to prove that $u = \alpha_1 u_2 + \alpha_2 u_2$. Recall that u is *characterized* by (3.5). Therefore, to prove the result, it is enough to show that $\alpha_1 u_1 + \alpha_2 u_2$ also satisfies (3.5) and then, by uniqueness, we will deduce $\alpha_1 u_1 + \alpha_2 u_2 = u$. First, since M is a linear subspace and $u_1, u_2 \in M$, $\alpha_1 u_1 + \alpha_2 u_2 \in M$. Moreover, for $z \in M$,

$$\begin{aligned}\langle f - (\alpha_1 u_1 + \alpha_2 u_2), z\rangle &= \langle \alpha_1 f_1 + \alpha_2 f_2 - (\alpha_1 u_1 + \alpha_2 u_2), z\rangle \\ &= \alpha_1 \langle f_1 - u_1, z\rangle + \alpha_2 \langle f_2 - u_2, z\rangle \\ &= 0\end{aligned}$$

since, by (3.5) is satisfied by (f_1, u_1) and (f_2, u_2) i.e. $\langle f_1 - u_1, z\rangle = \langle f_2 - u_2, z\rangle = 0$. Thus, $\alpha_1 u_1 + \alpha_2 u_2$ satisfies (3.5) and by uniqueness, $\alpha_1 u_1 + \alpha_2 u_2 = \mathbf{P}_M(f)$. This shows the linearity of $\mathbf{P}_M$. The fact that then $\mathbf{P}_M$ is continuous is an obvious consequence of the above Proposition. ■

Remark 3.2.3 The fact that $\mathbf{P}_M$ is a projection means that $\mathbf{P}_M^2 = \mathbf{P}_M \circ \mathbf{P}_M = \mathbf{P}_M$ ***(Check this last property)***. Notice also that (3.5) means that

$$f - \mathbf{P}_M f \in M^\perp.$$

In other words, the image of $\mathbf{I} - \mathbf{P}_M$ is a subspace of $M^\perp$. It turns out that it exactly is $M^\perp$. Indeed, if $f \in M^\perp$ then, one can write

$$f = f - \mathbf{P}_M(f) + \mathbf{P}_M(f)$$

but $f - \mathbf{P}_M(f) \in M^\perp$ and $\mathbf{P}_M(f) \in M$. Thus,

$$\langle f - \mathbf{P}_M(f), \mathbf{P}_M(f)\rangle = 0.$$

Since $\langle f, \mathbf{P}_M(f)\rangle = 0$, this implies that $\langle \mathbf{P}_M(f), \mathbf{P}_M(f)\rangle = 0$, i.e. $\mathbf{P}_M(f) = 0$ and $f = f - \mathbf{P}_M(f)$ belongs to the image of $\mathbf{I} - \mathbf{P}_M$.

Corollary 3.2.5 Let $(H, \langle\cdot,\cdot\rangle)$ be a Hilbert space. Assume that $M \subset H$ is a closed linear subspace. For any $f \in H$, there exist a unique $u \in M$ and a unique $v \in M^\perp$ such that $f = u + v$. Moreover,

$$\|f\|^2 = \|u\|^2 + \|v\|^2.$$

In particular, $(M^\perp)^\perp = M$.

Proof. The proof is an easy **Exercise** which follows the line of the previous Remark (actually, $u = \mathbf{P}_M f$ and $v = \mathbf{P}_{M^\perp} f$). For the second part of the statement, one already knows that $M \subset (M^\perp)^\perp$ (see Prop. 3.1.5). If the inclusion were a strict inclusion, from the previous splitting, there would be some $x \in H$, $x \neq 0$ such that $x \in M^\perp \cap (M^\perp)^\perp$. But then $\langle x, x\rangle = 0$ i.e. $x = 0$ which is a contradiction. ■

Remark 3.2.4 The above result can be reformulated as $\mathbf{P}_M + \mathbf{P}_{M^\perp} = \mathbf{I}$.

■ **Example 3.7** The simplest example of closed linear subpsace in a given Hilbert space H is the one of a ***finite dimension*** linear subspace (recall that, from **Chapter 2** any finite dimensional space is closed). Assume M to be spanned by $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$, we can assume without loss of generality that $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ is an orthonormal family (see Proposition 3.1.4), i.e.

$$\langle \boldsymbol{e}_n, \boldsymbol{e}_m \rangle = \delta_{n,m}.$$

In this case, the orthogonal projection is given by

$$\mathbf{P}_M(x) = \sum_{j=1}^{n} \langle x, \boldsymbol{e}_j \rangle \, \boldsymbol{e}_j \qquad \forall x \in H.$$

To prove this, it suffices to show that $x - \mathbf{P}_M(x) \in M^\perp$ which amounts to prove that $x - \mathbf{P}_M(x) \perp \boldsymbol{e}_k$ for any $k = 1, \dots, n$. ***Check that this is indeed the case.*** In particular, since $\mathbf{P}_M$ is surjective on M, the above expression of $\mathbf{P}_M$ shows that

$$x = \sum_{j=1}^{n} \langle x, \boldsymbol{e}_j \rangle \, \boldsymbol{e}_j \qquad \forall x \in M. \tag{3.6}$$

In other words, in any finite dimensional Hilbert space M with orthonormal basis $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$, one has the splitting (3.6). ∎

Fundamental example: $L^2(S, \mu)$ and the construction of the conditional expectation

Recall that – on a given measure space $(S, \mathcal{F}, \mu)$ the space $L^2(S, \mu)$ has been defined in Chapter 2 and is defined as the class of equivalence of measurable functions $f : S \to \mathbb{R}$ such that

$$\int_S |f|^2 \mathrm{d}\mu < \infty.$$

One can define a norm on $L^2(S, \mu)$ by

$$\|f\|_2 = \left(\int_S |f|^2 \mathrm{d}\mu \right)^{1/2}, \qquad f \in L^2(S, \mu)$$

and, for that norm $(L^2(S, \mu), \|\cdot\|_2)$ is a Banach space (Riesz-Fisher). As a consequence, one has the following

Theorem 3.2.6 The space $L^2(S, \mu)$ endowed with the scalar product

$$\langle f, g \rangle = \int_S f \, g \, \mathrm{d}\mu, \qquad \forall f, g \in L^2(S, \mu)$$

is an Hilbert space.

Proof. It is clear that $\langle \cdot, \cdot \rangle$ is an inner product on $L^2(S, \mu)$ (notice that $\langle f, g \rangle$ is finite thanks to Holder's inequality) and that the induced norm is $\|\cdot\|_2$. According to the previous result, $(L^2(S, \mu), \|\cdot\|_2)$ is complete so that $(L^2(S, \mu), \langle \cdot, \cdot \rangle)$ is a Hilbert space. ∎

We provide here a construction of the conditional expectation based on a suitable application of the projection theorem in the Hilbert space $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Other constructions are possible based on the Radon-Nikodym Theorem.

Theorem 3.2.7 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability triple and let X be a real random variable with $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Then, there exists a random variable Y such that

1. Y is $\mathcal{G}$-measurable.
2. $\mathbf{E}(|Y|) < \infty$.
3. For any set $A \in \mathcal{G}$ we have $\mathbf{E}(Y\mathbf{1}_A) = \mathbf{E}(X\mathbf{1}_A)$ i.e.

$$\int_A Y \,\mathrm{d}\mathbb{P} = \int_A X \,\mathrm{d}\mathbb{P} \qquad \forall A \in \mathcal{G}. \tag{3.7}$$

Moreover, if $\tilde{Y}$ is another random variable with the above properties, then $\mathbb{P}[Y = \tilde{Y}] = 1$. In such a case, Y is called a version of the conditional expectation of X given $\mathcal{G}$ and write $Y = \mathbf{E}(X|\mathcal{G})$.

Proof. We divide the proof in various steps assuming first various stronger assumptions on X.

Step 0: uniqueness. We begin with proving that, if $\tilde{Y}, Y$ are two random variables statifying the properties of the theorem, then $\mathbb{P}[Y = \tilde{Y}] = 1$. We argue by contradiction and assume that $\mathbb{P}(Y \neq \tilde{Y}) \neq 0$. Up to switching the role of Y and $\tilde{Y}$, we may assume $\mathbb{P}(Y > \tilde{Y}) \neq 0$. Then, since

$$\{Y > \tilde{Y}\} = \bigcup_{n \geqslant 1} \left\{ Y \geqslant \tilde{Y} + \frac{1}{n} \right\}$$

one sees that there is $n_0 \in \mathbb{N}$ such that $\mathbb{P}(Y \geqslant \tilde{Y} + 1/n_0) \neq 0$. Since $Y - \tilde{Y}$ is $\mathcal{G}$-measurable, $A := \{Y \geqslant \tilde{Y} + 1/n_0\} \in \mathcal{G}$. But then

$$\int_A Y \mathrm{d}\mathbb{P} = \int_A \tilde{Y} \mathrm{d}\mathbb{P} = \int_A X \mathrm{d}\mathbb{P}$$

which is a contradiction since

$$\int_A Y \mathrm{d}\mathbb{P} - \int_A \tilde{Y} \mathrm{d}\mathbb{P} \geqslant \frac{1}{n_0} \mathbb{P}(A)$$

(Explain this). Therefore, $Y = \tilde{Y}$ $\mathbb{P}$-a.e.

Step 1: the case $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. We assume here that $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ and use the projection theorem. The set $K = L^2(\Omega, \mathcal{G}, \mathbb{P})$ is a linear subspace of $H = L^2(\Omega, \mathcal{F}, \mathbb{P})$ which is closed (since any limit of $\mathcal{G}$-measurable functions is still $\mathcal{G}$-measurable). Since H is a Hilbert space, the projection theorem asserts that there exists $Y = \mathbf{P}_K(X)$ which is such that

$$\|X - Y\|_2 = \min\{\|X - Z\|_2\,,\ Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})\}$$

and is characterized by the fact that

$$Y \in L^2(\Omega, \mathcal{G}, \mathbb{P}) \qquad \text{and} \qquad \langle X - Y, Z\rangle = 0 \qquad \forall Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})$$

where $\langle \cdot, \cdot \rangle$ is the scalar product of $L^2(\Omega, \mathcal{F}, \mathbb{P})$, i.e.

$$\int_\Omega (X - Y)\, Z \mathrm{d}\mathbb{P} = 0 \qquad \forall L^2(\Omega, \mathcal{G}, \mathbb{P}).$$

In particular, given $A \in \mathcal{G}$, since $Z = \mathbf{1}_A \in L^2(\Omega, \mathcal{G}, \mathbb{P})$, it holds that

$$\int_A (X - Y)\mathrm{d}\mathbb{P} = 0,$$

i.e. Y satisfies (3.7). We proved the existence of $\mathbf{E}(X|\mathcal{G}) = Y$ whenever $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ and notice that, in this case, $Y \in L^2(\Omega, \mathcal{G}, \mathbb{P})$.

Step 2: the case $X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ and X nonnegative. If $X \geqslant 0$ is bounded, then, since $\mathbb{P}$ is a probability measure $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ and Step 1 ensures the existence of $Y = \mathbf{E}(X|\mathcal{G})$ with moreover $Y \in L^2(\Omega, \mathcal{G}, \mathbb{P})$. Let us prove moreover that $Y \geqslant 0$ $\mathbb{P}$-a.e. We argue by contradiction and assume $\mathbb{P}(Y < 0) \neq 0$. Then,

$$\{Y < 0\} = \bigcup_{n \geqslant 1} \left\{Y \leqslant -\frac{1}{n}\right\}$$

and, since $\mathbb{P}(Y < 0) \neq 0$, there is $n_0 \in \mathbb{N}$ such that $\mathbb{P}(Y \leqslant -\frac{1}{n_0}) \neq 0$. Setting then $A = \{Y \leqslant -\frac{1}{n_0}\}$. Since Y is $\mathcal{G}$-measurable, $A \in \mathcal{G}$ and, from (3.7),

$$\int_A X\mathrm{d}\mathbb{P} = \int_A Y\mathrm{d}\mathbb{P}.$$

Now, since $X \geqslant 0$, the left-hand-side is nonnegative while, by definition of A, $\int_A Y\mathrm{d}\mathbb{P} \leqslant -\frac{1}{n_0}\mathbb{P}(A) < 0$. This is a contradiction !

Step 3: the case $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, X nonnegative. Assume now $X \geqslant 0$ $\mathbb{P}$-a.e. is integrable. For any $n \geqslant 1$, set

$$X_n = \min(X, n).$$

Then, $X_n \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ is nonnegative and, from the previous step, there is $Y_n = \mathbf{E}(X_n|\mathcal{G})$ with moreover $Y_n \geqslant 0$ $\mathbb{P}$-a.e. for any $n \in \mathbb{N}$. Notice that, for $\mathbb{P}$-a.e. $\omega \in \Omega$, $(X_n(\omega))_n$ is an increasing sequence with $X(\omega) = \lim_n X_n(\omega)$. Set then

$$Y(\omega) = \limsup_n Y(\omega) \qquad \forall \omega \in \Omega.$$

It is well-known that Y is $\mathcal{G}$-measurable and nonnegative. Let us prove that $Y \in L^1(\Omega, \mathcal{G}, \mathbb{P})$ is the conditional expectation of X. One sees that, for any $n \in \mathbb{N}$, $Z_n := X_{n+1} - X_n \geqslant 0$ $\mathbb{P}$-a.e. and $Z_n \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ so that, using the Step 2 again, the conditional expectation $\mathbf{E}(Z_n|\mathcal{G})$ is nonnegative. Using the uniqueness of conditional expectation (Step 0), one checks easily that

$$\mathbf{E}(Z_n|\mathcal{G}) = Y_{n+1} - Y_n$$

so that $(Y_n)_n$ is increasing. Using the monotone convergence theorem, one sees then that, for any $A \in \mathcal{G}$

$$\int_A Y\mathrm{d}\mathbb{P} = \lim_n \int_A Y_n\mathrm{d}\mathbb{P} = \lim_n \int_A X_n\mathrm{d}\mathbb{P} = \int_A X\mathrm{d}\mathbb{P} < \infty$$

so that Y satisfies (3.7). In the particular case $A = \Omega$ (which belongs to $\mathcal{G}$), one sees that

$$\int_\Omega Y\mathrm{d}\mathbb{P} < \infty$$

so that $Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$.

Step 4: the general case $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Splitting X into positive and negative parts: $X = X^+ - X^-$, one sees from the previous step that

$$Y^+ := \mathbf{E}(X^+|\mathcal{G}) \qquad \text{and} \qquad Y^- = \mathbf{E}(X^-|\mathcal{G})$$

are well-defined random variables $\mathcal{G}$-measurable, both integrable. Set then $Y = Y^+ - Y^-$, one has $Y \in L^1(\Omega, \mathcal{G}, \mathbb{P})$ and the linearity of integral readily gives that Y satisfies (3.7). This proves the result. ■

3.3 The dual space of a Hilbert space

We recall that, for a given Banach space $(X, \|\cdot\|)$, the dual of X, denoted by $X^\star$ is the space of continuous linear applications from X to $\mathbb{R}$, i.e.

$$X^\star = \mathscr{L}(X, \mathbb{R}).$$

Endowed with the norm $\|\Phi\|_\star = \sup_{\|x\|\leqslant 1} |\Phi(x)| = \|\Phi\|_{\mathscr{L}(X,\mathbb{R})}$, $X^\star$ is a Banach space. We recall the following easy case: let $X = \mathbb{R}^N$ and let $\Phi : \mathbb{R}^N \to \mathbb{R}$ be linear. Then, there exists $\boldsymbol{v} \in \mathbb{R}^N$ such that

$$\Phi(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{v} \rangle \qquad \forall \boldsymbol{x} \in \mathbb{R}^N.$$

Indeed, consider the canonical basis $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_N\}$ of $\mathbb{R}^N$. Any $\boldsymbol{x} = (x_1, \ldots, x_N) \in \mathbb{R}^N$ can be written as

$$\boldsymbol{x} = \sum_{k=1}^N x_k \boldsymbol{e}_k$$

and therefore

$$\Phi(\boldsymbol{x}) = \sum_{k=1}^N x_k \Phi(\boldsymbol{e}_k).$$

Denote $\boldsymbol{v} = (\Phi(\boldsymbol{e}_1, \ldots, \Phi(\boldsymbol{e}_N))$, one sees that

$$\Phi(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{v} \rangle \qquad \forall \boldsymbol{x} \in \mathbb{R}^N.$$

This shows that the dual space of $\mathbb{R}^N$ can be identified with $\mathbb{R}^N$ through the identification $\Phi \leftrightarrow \boldsymbol{v}$. We shall see that this generalizes to any Hilbert space of infinite dimension:

Theorem 3.3.1 — Riesz-Fréchet representation theorem. Let $(H, \langle\cdot,\cdot\rangle)$ be a Hilbert space. Given any $\Phi \in H^\star$ there exists a unique $f \in H$ such that

$$\Phi(u) = \langle f, u \rangle \qquad \forall u \in H \qquad \text{and} \qquad \|f\| = \|\Phi\|_\star.$$

✍ *Proof.* Let $M = \Phi^{-1}(\{0\}) = \{u \in H\ ;\ \Phi(u) = 0\}$. Since Φ is linear and continuous, M is a closed subspace of H. We may always assume that $M \neq H$ (otherwise $\Phi \equiv 0$ and the conclusion is obvious with $f = 0$). We claim that there exists some element $g \in H$ such that

$$\|g\| = 1 \qquad \text{and} \qquad \langle g, v \rangle = 0 \qquad \forall v \in M. \tag{3.8}$$

Indeed, let $g_0 \in H$ with $g_0 \notin M$. Set $g_1 = \mathbf{P}_M g_0$ Since $g_1 \in M$ and $g_0 \notin M$, $\|g_0 - g_1\| \neq 0$. Then

$$g = \frac{1}{\|g_0 - g_1\|}(g_0 - g_1) \in H$$

satisfies (3.8) according to the Projection Theorem. Notice that $g \notin M$ so that $\Phi(g) \neq 0$. Given $u \in H$, set then

$$v = u - \lambda g \qquad \text{with} \quad \lambda = \frac{\Phi(u)}{\Phi(g)}.$$

One has $\Phi(v) = \Phi(u) - \lambda\Phi(g) = 0$, i.e. $v \in M$ and $\langle g, v\rangle = 0$. In other words,

$$\langle g, u\rangle = \lambda \langle g, g\rangle = \lambda,$$

i.e. $\Phi(u) = \Phi(g)\langle g, u\rangle$. Setting $f = \Phi(g)g$, one sees that

$$\Phi(u) = \langle f, u\rangle \qquad \forall u \in H.$$

Moreover, by Cauchy-Schwarz inequality $|\Phi(u)| \leqslant \|f\|\,\|u\|$ for any $u \in H$, so that $\|\Phi\|_* \leqslant \|f\|$. Moreover, $|\Phi(f)| = \|f\|^2$ which shows that $\|\Phi\|_* = \|f\|$ ***(Explain why)***. The fact that such a f is unique comes from the fact that $\|f\|^2 = |\Phi(f)|$. If they were two such elements f_1, f_2, then $\Phi(f_1 - f_2) = 0$ which would get $f_1 = f_2$. ■

3.4 Hilbert bases

We first begin with the following

> **Proposition 3.4.1** Any infinite-dimensional inner product space $(H, \langle\cdot,\cdot\rangle)$ contains an orthonormal sequence, i.e. there is $(\boldsymbol{e}_n)_n \subset H$ such that
>
> $$\langle \boldsymbol{e}_n, \boldsymbol{e}_m\rangle = \delta_{n,m} \qquad \forall n, m \in \mathbb{N}.$$

Proof. The proof combines the Gram-Schmidt procedure we saw in Proposition 3.1.4 together with the proof of Riesz-Theorem seen in **Chapter 2**. Indeed, arguing exactly as in the proof of Riesz Theorem, we can construct an infinite family $(\boldsymbol{x}_n)_n \subset H$ such that $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ is linearly independent for any n. Then, apllying inductively the Gram-Schmidt procedure, we can construct a family $(\boldsymbol{e}_n)_n$ which is orthonormal from $(\boldsymbol{x}_n)_n$. Details are omitted. ■

■ **Example 3.8** If $H = \ell^2(\mathbb{N})$, the sequence $(\boldsymbol{e}^n)_n$ given by

$$\boldsymbol{e}^n = (\boldsymbol{e}^n_k)_k \qquad \text{with} \quad \boldsymbol{e}^n_k = \delta_{n,k} \qquad \forall k \geqslant 1, n \geqslant 1$$

is an orthonormal sequence of H. ■

For a general orthonormal sequence $(\boldsymbol{e}_n)_n$ in an infinite-dimensional inner product space H and an element $x \in H$, the obvious generalization of the identity (3.6) would of course be

$$x = \sum_{n=1}^{\infty} \langle x, \boldsymbol{e}_n\rangle\ \boldsymbol{e}_n$$

but of course, several problems occur here: first, does the above series converge ? Second, if it converges, is the limit truly x ? The following two lemmas give positive answer to these questions:

Lemma 3.4.2 Let $(H, \langle \cdot, \cdot \rangle)$ be a given Hilbert space and let $(\boldsymbol{e}_n)_n$ be a given orthonormal sequence of H. Then, for any $x \in H$, the series $\sum_{n=1}^{\infty} |\langle x, \boldsymbol{e}_n \rangle|^2$ converges (in $\mathbb{R}$) and

$$\sum_{n=1}^{\infty} |\langle x, \boldsymbol{e}_n \rangle|^2 \leqslant \|x\|^2 \qquad \text{(Bessel's inequality).}$$

Proof. For a given $x \in H$ and $n \in \mathbb{N}$, set

$$S_n = \sum_{k=1}^{n} \langle x, \boldsymbol{e}_k \rangle \, \boldsymbol{e}_k \in H.$$

Notice that $S_n = \mathbf{P}_n(x)$ where $\mathbf{P}_n$ is the orthogonal projection on the finite dimensional subspace $E_n = \mathrm{span}(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$. One checks easily that

$$\|x - S_n\|^2 = \langle x - S_n, x - S_n \rangle = \|x\|^2 - 2\langle x, S_n \rangle + \|S_n\|^2 = \|x\|^2 - \sum_{k=1}^{n} |\langle x, \boldsymbol{e}_n \rangle|^2 .$$

In particular, for any $n \in \mathbb{N}$

$$\sum_{k=1}^{n} |\langle x, \boldsymbol{e}_k \rangle|^2 \leqslant \|x\|^2$$

and, since the real sequence $(\sum_{k=1}^{n} |\langle x, \boldsymbol{e}_k \rangle|^2)_n$ is increasing, it converges to a limit $\sum_{k=1}^{\infty} |\langle x, \boldsymbol{e}_n \rangle|^2$ which satisfies Bessel's inequality. ∎

Lemma 3.4.3 Let $(H, \langle \cdot, \cdot \rangle)$ be a given Hilbert space and let $(\boldsymbol{v}_n)_n$ be an sequence in H such that

$$\langle \boldsymbol{v}_n, \boldsymbol{v}_m \rangle = 0 \qquad \forall n \neq m \qquad \text{and} \qquad \sum_{n=1}^{\infty} \|\boldsymbol{v}_n\|^2 < \infty.$$

Then, setting $S_n = \sum_{k=1}^{n} \boldsymbol{v}_k \in H$, $n \geqslant 1$, it holds that $(S_n)_n \subset H$ is a convergent sequence whose limit S is such that

$$\|S\|^2 = \sum_{k=1}^{\infty} \|\boldsymbol{v}_k\|^2.$$

Proof. It is very easy to check that

$$\|S_n - S_m\|^2 = \sum_{k=m+1}^{n} \|\boldsymbol{v}_k\|^2 \qquad \forall n > m.$$

Therefore, $(S_n)_n$ is a Cauchy sequence and, since H is a Hilbert space, it converges to some S. Since moreover

$$\|S_n\|^2 = \sum_{k=1}^{n} \|\boldsymbol{v}_k\|^2 \qquad \forall n \geqslant 1$$

we see that, at the limit, the conclusion follows. ∎

With these two results in hands, one has the following:

Theorem 3.4.4 Let $(H, \langle \cdot, \cdot \rangle)$ be a given Hilbert space and let $(\boldsymbol{e}_n)_n$ be a given orthonormal sequence of H. If $(\alpha_n)_n \subset \mathbb{R}$, then the series $\sum_{n=1}^{\infty} \alpha_n \boldsymbol{e}_n$ converges in H if and only if

$$\sum_{n=1}^{\infty} |\alpha_n|^2 < \infty.$$

In that case, one has

$$\left\| \sum_{n=1}^{\infty} \alpha_n \boldsymbol{e}_n \right\|^2 = \sum_{n=1}^{\infty} |\alpha_n|^2.$$

Proof. If $\sum_n |\alpha_n|^2 < \infty$, setting $\boldsymbol{v}_n = \alpha_n \boldsymbol{e}_n$ for any $n \in \mathbb{N}$, Lemma 3.4.3 proves the result. Assume conversely that the series $\sum_n \alpha_n \boldsymbol{e}_n$ converges in H. Set

$$x = \sum_{n=1}^{\infty} \alpha_n \boldsymbol{e}_n.$$

Then, one sees easily that, for any $k \in \mathbb{N}$, $\alpha_k = \langle x, \boldsymbol{e}_k \rangle$ and Lemma 3.4.2 shows that $\sum_{n=1}^{\infty} |\alpha_n|^2 < \infty$. ■

One has then the following

Theorem 3.4.5 Let $(H, \langle \cdot, \cdot \rangle)$ be a given Hilbert space and let $(\boldsymbol{e}_n)_n$ be a given orthonormal sequence of H. The following are equivalent:

$a)$ $\overline{\text{span}\{\boldsymbol{e}_n \,,\, n \in \mathbb{N}\}} = H$;

$b)$ for any $x \in H$ it holds

$$x = \sum_{n=1}^{\infty} \langle x, \boldsymbol{e}_n \rangle \, \boldsymbol{e}_n.$$

$c)$ for any $x \in H$, *Bessel-Parseval identity* holds

$$\|x\|^2 = \sum_{n=1}^{\infty} | \langle x, \boldsymbol{e}_n \rangle |^2 ;$$

$d)$ $\{\boldsymbol{e}_n \,,\, n \in \mathbb{N}\}^{\perp} = \{0\}$.

Proof. d) $\implies$ b). Given $x \in H$, set $z = \sum_{n=1}^{\infty} \langle x, \boldsymbol{e}_n \rangle \, \boldsymbol{e}_n$. We know from Lemma 3.4.2 and the previous theorem that the series converges and defines an element in H. One has moreover

$$\langle z, \boldsymbol{e}_k \rangle = \langle x, \boldsymbol{e}_k \rangle \qquad \forall k \in \mathbb{N}.$$

Set then $y = x - z$, it holds $\langle y, \boldsymbol{e}_k \rangle = 0$ for any $k \in \mathbb{N}$, i.e. $y \in \{\boldsymbol{e}_n \,,\, n \in \mathbb{N}\}^{\perp}$. Thus, $y = 0$ and $x = z = \sum_{n=1}^{\infty} \langle x, \boldsymbol{e}_n \rangle \, \boldsymbol{e}_n$.

b) $\implies$ c) directly comes from the previous Theorem.

b) $\implies$ a) is obvious since $\sum_{k=1}^{n} \langle x, \boldsymbol{e}_k \rangle \, \boldsymbol{e}_k \in \text{span}(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ for any $n \in \mathbb{N}$ so that $x = \sum_{n=1}^{\infty} \langle x, \boldsymbol{e}_n \rangle \, \boldsymbol{e}_n$ belongs to $\overline{\text{span}\{\boldsymbol{e}_n \,,\, n \in \mathbb{N}\}}$.

a) $\implies$ d) Let $y \in \{\boldsymbol{e}_n \,,\, n \in \mathbb{N}\}^{\perp}$, i.e. $\langle y, \boldsymbol{e}_n \rangle = 0$ for any $n \in \mathbb{N}$. It is clear then that $\boldsymbol{e}_n \in \{y\}^{\perp}$ for any $n \in \mathbb{N}$ but $\{y\}^{\perp}$ is a closed linear subspace of H. Thus, from the closedness,

$$\overline{\text{span}\{\boldsymbol{e}_n \,,\, n \in \mathbb{N}\}} \subset \{y\}^{\perp}$$

and, since $\overline{\text{span}\{\boldsymbol{e}_n\,,\,n \in \mathbb{N}\}} = H$, we get $H \subset \{y\}^\perp$ which, in particular, implies $\langle y, y\rangle = 0$ i.e. $y = 0$.

c) $\implies$ *d)* If $\langle x, \boldsymbol{e}_n\rangle = 0$ for all $n \in \mathbb{N}$,, from *c)* it is clear that $\|x\|^2 = 0$ which means $x = 0$ and proves *d)*. ∎

Definition 3.4.1 Let $(H, \langle\cdot,\cdot\rangle)$ be a given Hilbert space and let $(\boldsymbol{e}_n)_n$ be a given orthonormal sequence of H. If the equivalent conditions of the previous Theorem hold, we say that $(\boldsymbol{e}_n)_n$ is an orthonormal basis or *Hilbert basis* of H.

Definition 3.4.2 A given normed space $(X, \|\cdot\|)$ is said to be ***separable*** if it contains a countable dense subset.

■ **Example 3.9** $\mathbb{R}$ endowed with its usual norm is separable since $\mathbb{Q}$ is a dense and countable subset of $\mathbb{R}$. ■

■ **Example 3.10** If $(H, \langle\cdot,\cdot\rangle)$ is a given Hilbert space and $(\boldsymbol{e}_n)_n$ is a Hilbert basis of H, then H is separable. Indeed, consider

$$E = \text{span}_{\mathbb{Q}}\{\boldsymbol{e}_n\,,\,n \in \mathbb{N}\}$$

be the space of all finite linear combinations of the $\boldsymbol{e}_k$ with *rational* coefficients. It is clear that E is countable. Since $\mathbb{Q}$ is dense in $\mathbb{R}$ one sees that

$$\overline{E} = \overline{\text{span}\{\boldsymbol{e}_n\,,\,n \in \mathbb{N}\}} = H$$

so that H contains a dense and countable subset. ■

The following provides a reciprocal to the previous example:

Theorem 3.4.6 If $(H, \langle\cdot,\cdot\rangle)$ is a separable Hilbert space, then there exists an orthonormal basis $(\boldsymbol{e}_n)_n$ of H.

Proof. Let $(\boldsymbol{v}_n)_n$ be a countable dense subset of H. Let F_k denote the linear space spanned by $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$. The sequence $(F_k)_k$ is a nondecreasing sequence of finite-dimensional spaces such that $\bigcup_{k=1}^\infty F_k$ is dense in H. Pick any unit vector $\boldsymbol{e}_1 \in F_1$. If $F_2 \neq F_1$ there is some vector $\boldsymbol{e}_2 \in F_2$ such that $\{\boldsymbol{e}_1, \boldsymbol{e}_2\}$ is an orthonormal basis of F_2. Repeating the same construction, one obtains an orthonormal basis of H. ∎

Proposition 3.4.7 If $(H, \langle\cdot,\cdot\rangle)$ is a separable Hilbert space and $M \subset H$ is a closed linear subspace, then $(M, \langle\cdot,\cdot\rangle)$ is separable.

Proof. Let $E = \{\boldsymbol{x}_n\,,\,n \in \mathbb{N}\}$ be a countable and dense subset of H. We denote by

$$F = \mathbf{P}_M(E) = \{\mathbf{P}_M(x_n)\,,\,n \in \mathbb{N}\}.$$

Clearly, $F \subset M$ is countable. We prove that F is dense in M. Let $y \in M$. For any $\varepsilon > 0$, there exists $n \in \mathbb{N}$ such that $\|y - \boldsymbol{x}_n\| \leqslant \varepsilon$ (since E is dense in H. Recall that $\mathbf{P}_M$ is a contraction so that

$$\|\mathbf{P}_M(y) - \mathbf{P}_M(\boldsymbol{x}_n)\| \leqslant \varepsilon$$

and, since $\mathbf{P}_M(y) = y$, we see that $\|y - \mathbf{P}_M(\boldsymbol{x}_n)\| \leqslant \varepsilon$, i.e. F dense in M. ∎

3.5 Introduction to the weak topology

We recall that, if $(H, \langle\cdot,\cdot\rangle)$ is an infinite-dimensional Hilbert space, the closed unit ball $B_c(0,1) = \{x \in H\,;\, \|x\| \leqslant 1\}$ is not compact. In particular, a bounded sequence in H does not necessarily admits a subsequence which converges. We shall see a way to overcome this problem. Notice that the notion introduce here can be generalized to any Banach space. In all the sequel, $(H, \langle\cdot,\cdot\rangle)$ is a fixed Hilbert space.

Definition 3.5.1 Let $(x_n)_n \subset H$ be a given sequence. We say that $(x_n)_n$ converges weakly to $x \in H$ and writes $\mathrm{w} - \lim_n x_n = x$ if

$$\lim_n \langle x_n, y\rangle = \langle x, y\rangle \qquad \forall y \in H.$$

Remark 3.5.1 The weak convergence is of course equivalent to

$$\lim_n \Phi(x_n) = \Phi(x) \qquad \forall \Phi \in H^\star.$$

In a general Banach space X, this is the correct notion of weak convergence.

A first observation is the following

Proposition 3.5.1 If $(x_n)_n$ converges to x (for the norm) then $(x_n)_n$ converges weakly to x.

Proof. The proof is a simple application of Cauchy-Schwarz inequality: given $y \in H$

$$|\,\langle x_n, y\rangle - \langle x, y\rangle\,| = |\,\langle x_n - x, y\rangle\,| \leqslant \|y\|\,\|x_n - x\|$$

which proves the result. ■

■ **Example 3.11** Let $(\boldsymbol{e}_n)_n$ be a Hilbert basis of H. Then, $(\boldsymbol{e}_n)_n$ converges weakly to 0. Indeed, if $y \in H$, according to Bessel-Parseval identity, the series

$$\sum_{n=1}^{\infty} |\,\langle y, \boldsymbol{e}_n\rangle\,|^2 < \infty$$

in particular, $\lim_n |\,\langle y, \boldsymbol{e}_n\rangle\,|^2 = 0$ which gives the result. ■

The above example illustrates some delicate point: the sets which are closed for the usual convergence (we say the strong convergence) are not necessarily closed for the weak convergence: in some separable Hilbert space, the unit sphere $\{x \in H\,,\, \|x\| = 1\}$ is closed for the strong convergence but not for the weak convergence since the sequence $(\boldsymbol{e}_n)_n$ belongs to the sphere but its weak limit is zero which is not in the sphere !!! Nevertheless, one can still prove that, if a sequence admits a weak limit, such a weak limit is unique **(Exercise)**. Moreover, one has

Proposition 3.5.2 Let $(x_n)_n$ be a sequence in H which converges weakly to x. Then

$$\|x\| \leqslant \liminf_n \|x_n\|.$$

Proof. Notice that

$$|\,\langle x_n, x\rangle\,| \leqslant \|x\|\,\|x_n\| \qquad \forall n \in \mathbb{N}$$

taking the limit as $n \to \infty$, by definition the left-hand-side converges to $\langle x, x\rangle = \|x\|^2$ which gives the iniequality
$$\|x\|^2 \leqslant \|x\| \liminf_n \|x_n\|$$
and proves the result. ■

One has then the fundamental result

Theorem 3.5.3 — Banach-Alaoglu. Let $(x_n)_n$ be a bounded sequence of H. Then, $(x_n)_n$ admits a subsequence which converges weakly in H.

Proof. The proof introduces here the fundamental *diagonal procedure* which allows to extract a common converging subsequence from an infinite number of sequences (see also Prop. 2.5.2 in **Chapter 2**). Let $(x_n)_n$ be a bounded sequence in H. Then, for any $k \in \mathbb{N}$, the sequence $(\langle x_n, x_k\rangle)_n$ is a bounded sequence of $\mathbb{R}$. It admits a subsequence which converges. Then, we first construct a subsequence $(x_{1,n})_n$ such that $(\langle x_{1,n}, x_1\rangle)_n$ converges in $\mathbb{R}$. From that sequence, one extract a subsequence $(x_{2,n})_n$ such that $(\langle x_{2,n}, x_2\rangle)_n$ converges. Inductively, for any $k \geqslant 1$, one can extract a subsequence $(x_{k+1,n})_n$ of $(x_{k,n})_n$ such that $(\langle x_{k+1,n}, x_{k+1}\rangle)_n$ converges in $\mathbb{R}$. We define then the diagonal subsequence $(x_{\varphi(n)})_n$ given by $x_{\varphi(n)} = x_{n,n}$ for any $n \in \mathbb{N}$. The interest of this diagonal subsequence is that it shares *all the properties of the previous sequences*. Set then
$$E = \text{span}\{x_n\,,\, n \in \mathbb{N}\}$$
which is the space of all the *finite* linear combination of the x_n. Using the linearity of the innner product, one sees that, for any $y \in E$, it holds that $(\langle y, x_{\varphi(n)}\rangle)_n$ converges. Then, it also holds that $(\langle y, x_{\varphi(n)}\rangle)_n$ converges for any $y \in \overline{E}$. We denote by $A(y)$ the limit. Then, the mapping $A : \overline{E} \to \mathbb{R}$ is clearly linear and, since $(x_n)_n$ is bounded, one has
$$|\langle y\,,\, x_{\varphi(n)}\rangle| \leqslant C\,\|y\| \qquad \forall n$$
so that
$$|A(y)| \leqslant C\|y\| \qquad \forall y \in \overline{E}$$
ie. $A \in \mathscr{L}(\overline{E}, \mathbb{R})$, $A \in \overline{E}^\star$. According to the Riesz Representation Theorem, there is therefore $x \in \overline{E}$ such that
$$A(y) = \langle y, x\rangle \qquad \forall y \in \overline{E}$$
i.e.
$$\lim_n \langle y, x_{\varphi(n)}\rangle = \langle y, x\rangle \qquad \forall y \in \overline{E}.$$
Since $H = \overline{E} + \overline{E}^\perp$ and $x_{\varphi(n)} \in \overline{E}$ for any $n \in \mathbb{N}$ one sees that the above convergence extends to any $y \in H$ and the sequence $(x_{\varphi(n)})_n$ converges weakly to x. ■

Remark 3.5.2 A sequence $(x_n)_n$ which weakly converges in H is necessarily bounded. This is an easy consequence of Banach-Steinhaus Theorem: indeed, for any $n \in \mathbb{N}$, one can define the linear continuous mapping
$$T_n(y) = \langle y, x_n\rangle \qquad y \in H$$

and, since $(T_n(y))_n$ converges, it is bounded, i.e. there is $M_y > 0$ such that

$$|T_n(y)| \leqslant M_y \qquad \forall n \in \mathbb{N}.$$

Then, according to Banach-Steinhaus Theorem, there is $M > 0$ such that

$$|T_n(y)| \leqslant M\,\|y\| \qquad \forall y \in H.$$

In particular, for $y = x_n$, this gives $\|x_n\|^2 \leqslant M\,\|x_n\|$ and the sequence is bounded.

3.6 Problems

Exercise 3.6 Let $(H, \langle\cdot,\cdot\rangle)$ be a Hilbert space and let $A \in \mathscr{L}(H)$. Assume that there is $c > 0$ such that

$$\langle A(x), x\rangle \geqslant c\|x\|^2 \qquad \forall x \in H.$$

Define

$$\mathrm{Ker}(A) = \{x \in H\,,\ A(x) = 0\} \qquad \text{and} \qquad \mathrm{Im}(A) = \{y \in H\,,\ \exists x \in H\,,\ y = A(x)\}.$$

1. Prove that $\mathrm{Ker}(A) = \{0\}$.
2. Prove that $\mathrm{Im}(A)$ is closed in H (Hint: If $(A(x_n))_n \subset \mathrm{Im}(A)$ is convergent, prove that $(x_n)_n$ is a Cauchy sequence).
3. Prove that $\mathrm{Im}(A)^\perp = \{0\}$ and deduce from it that $\mathrm{Im}(A)$ is dense in H.
4. Deduce from this that A is a bijection and that its inverse $A^{-1} \in \mathscr{L}(H)$.

■

Exercise 3.7 Let $(H, \langle\cdot,\cdot\rangle)$ be a given Hilbert space and let $T \in \mathscr{L}(H)$.
1. For any $y \in H$, consider the mapping $F_y\ :\ x \in H \mapsto \langle T(x), y\rangle \in \mathbb{R}$. Prove that $F_y \in H^\star$ for any $y \in H$.
2. Deduce from this, using Riesz-representation theorem, that there exists $T^\star\ :\ y \in H \mapsto T^\star(y) \in H$ such that

$$\langle T(x), y\rangle = \langle x, T^\star(y)\rangle\,.$$

3. Prove that $T^\star \in \mathscr{L}(H)$ and $\|T^\star\|_{\mathscr{L}(H)} = \|T\|_{\mathscr{L}(H)}$.
4. Prove that, if $T_1, T_2 \in \mathscr{L}(H)$, then $(\alpha T_1 + T_2)^\star = \alpha\,T_1^\star + T_2$ for any $\alpha \in \mathbb{R}$.
5. If $T \in \mathscr{L}(H)$ is invertible with inverse T^{-1}, prove that $(T^{-1})^\star = (T^\star)^{-1}$.
6. If M is a closed linear subspace of H and $\mathbf{P}_M$ is the projection operator on M, prove that $\mathbf{P}_M^\star = \mathbf{P}_M$.

The operator $T^\star$ is called the adjoint of T.

■

Exercise 3.8 Let $(H, \langle\cdot,\cdot\rangle)$ be a given Hilbert space and let $\mathbf{P}$ be a linear application such that $\mathbf{P}^2 = \mathbf{P}$ (i.e. $\mathbf{P}(\mathbf{P}(u)) = \mathbf{P}(u)$ for any $u \in H$).
1. With the notation of Exercise 3.6, prove that

$$H = \mathrm{Ker}(\mathbf{P}) + \mathrm{Im}(\mathbf{P})$$

and that $\mathrm{Im}(\mathbf{P}) = \mathrm{Ker}(\mathbf{P} - \mathbf{I})$ where $\mathbf{I}$ is the identity in H.
2. Assume moreover that $\|\mathbf{P}(x)\| \leqslant \|x\|$ for all $x \in H$. Prove that $\mathrm{Ker}(\mathbf{P})$ and $\mathrm{Im}(\mathbf{P})$ are two closed linear subspaces of H with $\mathrm{Ker}(\mathbf{P})^{\perp} \subset \mathrm{Im}(\mathbf{P})$.
3. Prove that $\mathrm{Im}(\mathbf{P}) = \mathrm{Ker}(\mathbf{P})^{\perp}$ and deduce from that $\mathbf{P}$ is the orthogonal projection on $\mathrm{Im}(\mathbf{P})$.

■

Exercise 3.9 Let $(H_1, \langle\cdot,\cdot\rangle_1)$ and $(H_2, \langle\cdot,\cdot\rangle_2)$ be two Hilbert spaces with respective norms $\|\cdot\|_1$ and $\|\cdot\|_2$. Let $\Phi \;:\; H_1 \to H_2$ be a linear application. Prove that the following are equivalent:
1. $\|\Phi(x)\|_2 = \|x\|_1$ for all $x \in H_1$;
2. $\langle\Phi(x), \Phi(y)\rangle_2 = \langle x, y\rangle_2$ for all $x, y \in H_2$.

■

Exercise 3.10 Let $H = \ell^2(\mathbb{N})$ be the Hilbert space of square-summable sequences with inner product

$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \sum_{k=1}^{\infty} x_k y_k \qquad \forall \boldsymbol{x} = (x_k)_k,\ \boldsymbol{y} = (y_k)_k \in H.$$

Let $\boldsymbol{a} = (a_k)_k \in \ell^{\infty}(\mathbb{N})$, i.e. $\sup_{k\geqslant} |a_k| = \|\boldsymbol{a}\|_{\infty} < \infty$, and let introduce the linear operator

$$T_{\boldsymbol{a}} \;:\; \boldsymbol{x} = (x_k)_k \in H \mapsto T_{\boldsymbol{a}}(\boldsymbol{x}) = (a_k x_k)_k \in H.$$

1. Prove that $T_{\boldsymbol{a}} \in \mathscr{L}(H)$ and determine its adjoint $T_{\boldsymbol{a}}^{\star}$.
2. Prove that the operator $T_{\boldsymbol{a}}$ is invertible if and only if the sequence $\boldsymbol{a}$ is such that $a_k \neq 0$ for any $k \in \mathbb{N}$ and $(\frac{1}{a_k})_k \in \ell^{\infty}(\mathbb{N})$.

■

Exercise 3.11 Let $H = \mathscr{C}([-1,1])$ be the space of continuous functions over $[-1,1]$ and let

$$\langle f, g\rangle = \int_{-1}^{1} g(t) f(t) dt \qquad \forall f, g \in H.$$

1. Show that $(H, \langle\cdot,\cdot\rangle)$ is an inner product space but not a Hilbert space.
2. Construct a sequence of polynomial functions $(P_n)_n \in H$ such that $\deg(P_n) = n$ and $P_n(1) = 1$ for any $n \in \mathbb{N}$ with moreover

$$P_k \perp P_n \qquad \forall k \neq n.$$

■

Exercise 3.12 Let $(H, \langle\cdot,\cdot\rangle)$ be a given Hilbert space and let $\boldsymbol{a} \;:\; H \times H \to \mathbb{R}$ bilinear for which there are $c > 0$, $\alpha > 0$ such that

$$|\boldsymbol{a}(u,v)| \leqslant c\|u\|\|v\| \qquad \text{and} \qquad \boldsymbol{a}(u,u) \geqslant \alpha\|u\|^2 \qquad \forall u, v \in H.$$

Let $K \subset H$ be a closed convex subset of H and $\Phi \in H^{\star}$.

1. Prove that there exists $f \in H$ and $A \in \mathscr{L}(H)$ such that

$$\boldsymbol{a}(u,v) = \langle A(u), v\rangle \qquad \text{and} \qquad \Phi(u) = \langle f, u\rangle \qquad u, v \in H.$$

2. Let $\varrho > 0$. For any $v \in K$ we set

$$S(v) = \mathbf{P}_K(\varrho f - \varrho A(v) + v)$$

   where $\mathbf{P}_K$ is the projection over K. Prove that there is $\varrho > 0$ and $k \in (0,1)$ such that $S : K \to H$ is a k-contraction.
3. Prove therefore that there is $u \in K$ such that

$$u = \mathbf{P}_K(\varrho f - \varrho A(u) + u)$$

   and that $\langle A(u), v - u\rangle \geqslant \langle f, v - u\rangle$ for any $v \in K$.
4. Deduce from this that there is $u \in K$ such that

$$\boldsymbol{a}(u, v-u) \geqslant \Phi(v-u) \qquad \forall v \in K.$$

■

Exercise 3.13 Let $(H_1, \langle\cdot,\cdot\rangle_1)$ and $(H_2, \langle\cdot,\cdot\rangle_2)$ be two Hilbert spaces with respective norms $\|\cdot\|_1$ and $\|\cdot\|_2$ and let $T \in \mathscr{L}(H_1, H_2)$ be such that for any sequence $(u_n)_n \subset H_1$ which converges *weakly* to 0, it holds that $(T(u_n))_n$ converges strongly to 0 (in H_2). Prove that $T(B)$ is a relatively compact subset of H_2 where B is the unit ball of H_1. ■

4. Complements about Lebesgue spaces

We complement here the first study of the Lebesgue spaces initiated in **Chapter 2**. We begin with general additional inequalities

4.1 Useful inequalities

4.1.1 Jensen's inequality

We provide here a useful general inequality, valid in probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Recall that, given an open interval $I \subset \mathbb{R}$ a function

$$\Phi \ : \ I \to \mathbb{R}$$

is said to be convex on I if and only if, for any $x, y \in I$

$$\Phi(\theta x + (1-\theta)y) \leqslant \theta\Phi(x) + (1-\theta)\Phi(y) \qquad \forall \theta \in [0,1].$$

One sees that this readily extends to any convex combination, i.e. given $(p_i)_{i=1,\ldots,n}$ be nonnegative with $\sum_{i=1}^n p_i = 1$, then

$$\Phi\left(\sum_{i=1}^n p_i x_i\right) \leqslant \sum_{i=1}^n p_i \Phi(x_i) \qquad \forall (x_1, \ldots, x_n) \in I^n.$$

Remark 4.1.1 In the language of probability, if X is a discrete random variable on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking a finite number of values $x_1, \ldots, x_n$ with

$$\mathbb{P}(X = x_i) = p_i \qquad \forall i = 1, \ldots, n.$$

The above definition exactly means that, for any convex function $\Phi \ : \ \mathbb{R} \to \mathbb{R}$ it holds

$$\Phi(\mathbb{E}(X)) \leqslant \mathbb{E}(\Phi(X)) \tag{4.1}$$

where $\mathbb{E}$ is the expectation with respect to the probability measure $\mathbb{P}$. ***Explain why***.

Our scope here is to prove that (4.1) generalizes to any random variable. Before this, let us investigate briefly additional properties of convex functions. First, one has the following

Lemma 4.1.1 Let $\Phi : I \to \mathbb{R}$ be a convex function on some open interval $I \subset \mathbb{R}$. Then, for any $x_0 \in I$, there is an affine function $\ell_0(x) = a_0 x + b_0$ such that

$$\ell(x_0) = \Phi(x_0) \qquad \text{and} \qquad \ell(y) \leqslant \Phi(y) \qquad \forall y \in I.$$

Proof. Let $x_0 \in I$ be given. For any $h > 0$ (small enough so that $x_0 \pm h \in I$), applying the definition of convexity with $\theta = 1/2$ one finds that

$$\frac{\Phi(x_0) - \Phi(x_0 - h)}{h} \leqslant \frac{\Phi(x_0 + h) - \Phi(x_0)}{h}.$$

Moreover, for any $h_1 > h_2$, applying the definition of convexity with $\theta = h2/h1$ to points x_0 and $x_0 - h_1$ gives

$$\frac{\Phi(x_0) - \Phi(x_0 - h_1)}{h_1} \leqslant \frac{\Phi(x_0) - \Phi(x_0 - h_2)}{h_2}$$

and

$$\frac{\Phi(x_0 + h_1) - \Phi(x_0)}{h_1} \geqslant \frac{\Phi(x_0 + h_2) - \Phi(x_0)}{h_2}.$$

Therefore, both the mapping

$$h \in (0, \delta) \mapsto \frac{\Phi(x_0 + h) - \Phi(x_0)}{h} \qquad \text{and} \qquad h \in (0, \delta) \mapsto \frac{\Phi(x_0 - h) - \Phi(x_0)}{h}$$

are monotone. In particular, the following limit

$$\Phi'_+(x_0) = \lim_{h \to 0^+} \frac{\Phi(x_0 + h) - \Phi(x_0)}{h} \qquad \text{and} \qquad \Phi'_-(x_0) = \lim_{h \to 0^-} \frac{\Phi(x_0 - h) - \Phi(x_0)}{h}$$

exist and satisfy $\Phi'_+(x_0) \geqslant \Phi'_-(x_0)$. Taking any $\lambda \in [\Phi'_-(x_0), \Phi'_+(x_0)]$, the mapping

$$\ell_0(y) = f(y) + \lambda(y - x_0)$$

satisfies the desired request. ∎

A concrete way to check that a function is convex is the following:

Lemma 4.1.2 Let $I \subset \mathbb{R}$ be a given open interval and let $\Phi : I \to \mathbb{R}$ be differentiable over I. If the derivative Φ' is increasing on I then Φ is convex on I. In particular, if Φ is twice-differentiable on I and $\Phi''(x) \geqslant 0$ for any $x \in I$ then Φ is convex on I.

Proof. Suppose Φ' is increasing on I and let $x_0, x_1 \in I$, say $x_0 < x_1$ and let $\theta \in [0, 1]$. Set $x = \theta x_0 + (1 - \theta) x_1$. Applying the mean-value theorem, there exist $c, d \in I$ with $x_0 < c < x < d < x_1$ such that

$$\begin{aligned}
\Phi(x) - \Phi(x_0) &= (x - x_0)\Phi'(c) = \theta(x_1 - x_0)\Phi'(c) \\
\Phi(x_1) - \Phi(x) &= (x_1 - x)\Phi'(d) = (1 - \theta)(x_1 - x)\Phi'(d).
\end{aligned}$$

Multiplying the first equation by $(\theta - 1)$ and the second by θ, and adding, we find that

$$\theta\Phi(x_1) + (1-\theta)\Phi(x_0) - \Phi(x) = \theta(1-\theta)\left(\Phi'(c) - \Phi'(d)\right)$$

and, since $\Phi'(d) \geqslant \Phi'(c)$, we find that

$$\Phi(x) \leqslant \theta\Phi(x_1) + (1-\theta)\Phi(x_0)$$

which proves the convexity of Φ on I. The second part of the result is clear since, whenever Φ'' is nonnegative on I, Φ' is increasing on I. ■

▪ **Example 4.1** Thanks to this practical criterion, one sees that $\Phi(x) = \exp(x)$ is convex on $\mathbb{R}$, $\Phi(x) = x^p$ is convex on $\mathbb{R}^+$ for any $p > 1$ and $\Phi(x) = -x\log x$ is convex on $(0,\infty)$. ▪

Theorem 4.1.3 — Jensen's inequality. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X : \Omega \to \mathbb{R}$ be a given integrable random variable. Let $\Phi : I \to \mathbb{R}$ be a convex function on some open interval $I \subset \mathbb{R}$ such that $\mathbb{P}(X \in I) = 1$ and

$$\int_\Omega |\Phi(X)|\mathrm{d}\mathbb{P} < \infty$$

then

$$\Phi(\mathbb{E}(X)) \leqslant \mathbb{E}(\Phi(X)).$$

Proof. First of all, since $\mathbb{P}(X \in I) = 1$, $x_0 = \mathbb{E}(X) \in I$. Apply then the above Lemma, there is $\ell_0(x) = ax + b$ such that

$$\ell(\mathbb{E}(X)) = \Phi(\mathbb{E}(X))$$

and $\ell(y) \leqslant \Phi(y)$ for all $y \in I$. In particular, for $y = X(\omega)$, it holds $\ell(X(\omega)) \leqslant \Phi(X(\omega))$ for $\mathbb{P}$-almost every $\omega \in \Omega$. Integrating with respect to ω we see that

$$\int_\Omega \ell(X(\omega))\mathrm{d}\mathbb{P}(\omega) \leqslant \int_\Omega \Phi(X(\omega))\mathrm{d}\mathbb{P}(\omega)$$

and, since ℓ is affine

$$\int_\Omega \ell(X(\omega))\mathrm{d}\mathbb{P}(\omega) = a\mathbb{E}(X) + b = \ell(\mathbb{E}(X)).$$

To summarize,

$$\ell(\mathbb{E}(X)) \leqslant \mathbb{E}(\Phi(X)) \qquad \text{and} \qquad \ell(\mathbb{E}(X)) = \Phi(\mathbb{E}(X))$$

which gives the result. ■

Remark 4.1.2 Assume that the distribution of X has a density f with respect to the Lebesgue measure. Then, Jensen's inequality reads

$$\Phi\left(\int_\mathbb{R} xf(x)\mathrm{d}x\right) \leqslant \int_\mathbb{R} \Phi(x)f(x)\mathrm{d}x$$

for any convex function Φ such that $\int_\mathbb{R} |\Phi(x)|f(x)\mathrm{d}x < \infty$.

Remark 4.1.3 If $\Phi(t) = t^p$, $p \geqslant 1$, one sees that Jensen's inequality (applied to $|X|$) reads

$$\|X\|_1 \leqslant \|X\|_p$$

Explain why.

Exercise 4.1 Let $f, g : \mathbb{R} \to \mathbb{R}$ be two positive measurable mappings such that

$$\int_{\mathbb{R}} f(x)\mathrm{d}x = 1 = \int_{\mathbb{R}} g(x)\mathrm{d}x$$

Prove that

$$H(f|g) := \int_{\mathbb{R}} f(x)\log\left(\frac{f(x)}{g(x)}\right)\mathrm{d}x \geqslant 0.$$

(*Hint: introduce the function* $h = \frac{f}{g}$ *and prove that* $H(f|g) = -\int_{\mathbb{R}} \Phi(h(x))g(x)\mathrm{d}x$ *where* $\Phi(s) = -s\log s$ *for any* $s > 0$.) ■

Exercise 4.2 Deduce from the previous exercise that, if X is a random variable with positive density f and variance σ^2, then its entropy

$$H(X) = -\int_{\mathbb{R}} f(x)\log f(x)\mathrm{d}x$$

satisfies

$$H(X) \leqslant \frac{1}{2}\left(\log(2\pi\sigma^2) + 1\right).$$

In particular, for given variance σ^2, the normal distribution has the maximum entropy. (*Hint: introduce the density* g *of the normal distribution with the same mean and variance as* X *and compare* $H(X)$ *to* $H(f|g)$.) ■

4.1.2 Minkowski's integral inequality

Let us consider here two measure spaces (S_1, Σ_1, μ_1) and (S_2, Σ_2, μ_2). One has the following

Theorem 4.1.4 — Minkowski's integral inequality. Let $F : S_1 \times S_2 \to \mathbb{R}$ be measurable. Then, for any $1 \leqslant p < \infty$, it holds:

$$\left[\int_{S_2}\left|\int_{S_1} F(x,y)\,\mathrm{d}\mu_1(x)\right|^p \mathrm{d}\mu_2(y)\right]^{1/p} \leqslant \int_{S_1}\left(\int_{S_2} |F(x,y)|^p\,\mathrm{d}\mu_2(y)\right)^{1/p}\mathrm{d}\mu_1(x).$$

Proof. We can assume, without loss of generality that $F : S_1 \times S_2 \to \mathbb{R}$ is nonnegative. If $p = 1$, the conclusion is simply Fubini's theorem so let us assume $1 < p < \infty$. Let $\frac{1}{q} + \frac{1}{p} = 1$. We denote by I the left-hand-side of the above inequality. For μ_2-almost every $y \in S_2$, set

$$G(y) = \left(\int_{S_1} F(x,y)\mathrm{d}\mu_1(x)\right)^{p-1}.$$

Then,

$$\|G\|^q_{L^q(S_2,\mu_2)} = \int_{S_2} |G(y)|^q \mathrm{d}\mu_2(y) = \int_{S_2} \left(\int_{S_1} F(x,y)\mathrm{d}\mu_1(x)\right)^{q(p-1)} \mathrm{d}\mu_2(y)$$
$$= \int_{S_2} \left(\int_{S_1} F(x,y)\mathrm{d}\mu_1(x)\right)^{p} \mathrm{d}\mu_2(y)$$

i.e.

$$\|G\|^q_q = I^p.$$

One also has

$$I^p = \int_{S_2} \left(\int_{S_1} F(x,y)\mathrm{d}\mu_1(x)\right)^{p} \mathrm{d}\mu_2(y) = \int_{S_2} G(y) \left(\int_{S_1} F(x,y)\mathrm{d}\mu_1(x)\right) \mathrm{d}\mu_2(y)$$

by definition of G. According to Fubini's theorem, since all functions are nonnegative, we deduce that

$$I^p = \int_{S_2\times S_1} F(x,y)G(y)\mathrm{d}\mu_1(x)\mathrm{d}\mu_2(y) = \int_{S_1} \left(\int_{S_2} F(x,y)G(y)\mathrm{d}\mu_2(y)\right) \mathrm{d}\mu_1(x).$$

Now, for μ_1-almost every $x \in S_1$, one estimates the integral $\int_{S_2} F(x,y)G(y)\mathrm{d}\mu_2(y)$ using Holder's inequality, namely

$$\int_{S_2} F(x,y)G(y)\mathrm{d}\mu_2(y) \leqslant \|G\|_{L^q(S_2,\mu_2)} \left(\int_{S_2} F(x,y)^p\mathrm{d}\mu_2(y)\right)^{\frac{1}{p}}.$$

Then, integrating with respect to $x \in S_1$

$$I^p \leqslant \|G\|_{L^q(S_2,\mu_2)} \int_{S_1} \left(\int_{S_2} F(x,y)^p\mathrm{d}\mu_2(y)\right)^{\frac{1}{p}} \mathrm{d}\mu_1(x).$$

Since $\|G\|^q_{L^q(S_2,\mu_2)} = I^p$ one gets

$$I^p \leqslant I^{\frac{p}{q}} \int_{S_1} \left(\int_{S_2} F(x,y)^p\mathrm{d}\mu_2(y)\right)^{\frac{1}{p}} \mathrm{d}\mu_1(x)$$

which gives

$$I \leqslant \int_{S_1} \left(\int_{S_2} F(x,y)^p\mathrm{d}\mu_2(y)\right)^{\frac{1}{p}} \mathrm{d}\mu_1(x)$$

as desired. ■

Remark 4.1.4 The above inequality can be rewritten as

$$\left\|\int_{S_1} F(x,\cdot)\mathrm{d}\mu_1(x)\right\|_{L^p(S_2,\mu_2)} \leqslant \int_{S_1} \|F(x,\cdot)\|_{L^p(S_2,\mu_2)}\mathrm{d}\mu_1(x).$$

In particular, if $f,g \in L^p(S_2,\mu_2)$ and $S_1 = \{1,2\}$ is endowed with the counting measure $\mathrm{d}\mu_1 = \delta_1 + \delta_2$, one can define

$$F(1,y) = f(y), \quad F(2,y) = g(y)$$

and the above inequality becomes

$$\|f+g\|_p \leqslant \|f\|_p + \|g\|_p$$

which is what we already called Minkowski's inequality (this is the triangle inequality for the $\|\cdot\|_p$ norm).

Remark 4.1.5 ***(Integral of vector valued-functions)*** Let $f : \mathbb{R}^N \to \mathbb{R}^2$ be a *vector-valued function*, i.e. $f(x) \in \mathbb{R}^2$ for all $x \in \mathbb{R}^N$. We denote $|\cdot|_2$ the euclidean norm of $\mathbb{R}^2$, i.e. given $\boldsymbol{y} = (y_1, y_2) \in \mathbb{R}^2$, $|\boldsymbol{y}|_2 = \sqrt{y_1^2 + y_2^2}$. Clearly, one can define two functions $f_1 : \mathbb{R}^N \to \mathbb{R}$ and $f_2 : \mathbb{R}^N \to \mathbb{R}$ as follows

$$f(x) = (f_1(x), f_2(x)) \qquad \forall x \in \mathbb{R}^N.$$

We say then that $f \in L^1(\mathbb{R}^N, \mathbb{R}^2)$ if $f_1, f_2 \in L^1(\mathbb{R}^N)$ and denote

$$\int_{\mathbb{R}^N} f(x)\mathrm{d}x = \left(\int_{\mathbb{R}} f_1(x)\mathrm{d}x, \int_{\mathbb{R}^N} f_2(x)\mathrm{d}x\right) \in \mathbb{R}^2.$$

Notice that, in such a case, $\int_{\mathbb{R}^N} f(x)\mathrm{d}x \in \mathbb{R}^2$. We denote $|\cdot|_2$ the euclidean norm of $\mathbb{R}^2$, i.e. given $\boldsymbol{y} = (y_1, y_2) \in \mathbb{R}^2$, $|\boldsymbol{y}|_2 = \sqrt{y_1^2 + y_2^2}$. In particular, one can compute the euclidean norm of $\int_{\mathbb{R}^2} f(x)\mathrm{d}x$:

$$\left|\int_{\mathbb{R}^N} f(x)\mathrm{d}x\right|_2 = \left(\left(\int_{\mathbb{R}^N} f_1(x)\mathrm{d}x\right)^2 + \left(\int_{\mathbb{R}^N} f_2(x)\mathrm{d}x\right)^2\right)^{1/2}.$$

What this then the link between $\left|\int_{\mathbb{R}^N} f(x)\mathrm{d}x\right|_2$ and $\int_{\mathbb{R}^N} |f(x)|_2\mathrm{d}x$? The answer is given by the following

$$\left|\int_{\mathbb{R}^N} f(x)\mathrm{d}x\right|_2 \leqslant \int_{\mathbb{R}^N} |f(x)|_2\mathrm{d}x.$$

Indeed, as in the previous remark, one can define $S_1 = \mathbb{R}^N$ (with the lebesgue measure) and $S_2 = \{1, 2\}$ endowed with the counting measure $\mu_2 = \delta_1 + \delta_2$. Then, defining $F : S_1 \times S_2 \to \mathbb{R}$ by

$$F(x,y) = \begin{cases} f_1(x) & \text{if } y = 1 \\ f_2(x) & \text{if } y = 2. \end{cases}$$

one gets that

$$\begin{aligned}\left|\int_{\mathbb{R}^N} f(x)\mathrm{d}x\right|_2 &= \left(\left(\int_{\mathbb{R}^N} F(x,1)\mathrm{d}x\right)^2 + \left(\int_{\mathbb{R}^N} F(x,2)\mathrm{d}x\right)^2\right)^{1/2} \\ &= \left(\int_{S_2} \left(\int_{\mathbb{R}^N} F(x,y)\mathrm{d}x\right)^2 \mathrm{d}\mu_2(y)\right)^{1/2}.\end{aligned}$$

Therefore, Minkowski's integral inequality yields

$$\left|\int_{\mathbb{R}^N} f(x)\mathrm{d}x\right|_2 \leqslant \int_{\mathbb{R}^N}\left(\int_{S_2} F(x,y)^2\mathrm{d}\mu_2(y)\right)^{1/2}\mathrm{d}x$$
$$= \int_{\mathbb{R}^N}\left(F(x,1)^2+F(x,2)^2\right)^{1/2}\mathrm{d}x = \int_{\mathbb{R}^N}\left(f_1(x)^2+f_2(x)\right)^{1/2}\mathrm{d}x$$
$$= \int_{\mathbb{R}^N}|f(x)|_2\mathrm{d}x.$$

Of course, the above generalizes easily to vector-valued functions $f : \mathbb{R}^N \to \mathbb{R}^d$, $d \geqslant 2$.

From now on, we will assume in the rest of this Chapter that

$$(S,\Sigma,\mu) = (\mathbb{R}^N, \mathscr{B}(\mathbb{R}^N), \mathfrak{m})$$

where $\mathscr{B}(\mathbb{R}^N)$ is the Borel σ-algebra over $\mathbb{R}^N$ and $\mathfrak{m}$ is the Lebesgue measure over $\mathbb{R}^N$. Of course, for applications, this is one fundamental peculiar case and we shall study in more details the space $L^p(S,\mu)$ in this case. We simply denote

$$L^p(\mathbb{R}^N) = L^p(\mathbb{R}^N, \mathscr{B}(\mathbb{R}^N), \mathfrak{m}), \qquad \forall 1 \leqslant p \leqslant \infty$$

and will simply denote integral with respect to the measure $\mathfrak{m}$ as follows

$$\int_{\mathbb{R}^N} f\mathrm{d}\mathfrak{m} = \int_{\mathbb{R}^N} f\mathrm{d}x, \qquad f \in L^1(\mathbb{R}^N).$$

4.2 Integral depending on a parameter

We present briefly here two applications of the dominated convergence theorem to the study of integrals depending on a parameter. Let us expose the general framework: let I be a given open interval of $\mathbb{R}$ and let $\Omega \subset \mathbb{R}^N$ be an open subset (I and Ω are non empty). We introduce a function

$$f : I \times \Omega \to \mathbb{R}$$

such that, for all $t \in I$, the function $f(t,\cdot) \in L^1(\Omega)$ (where, as above, $f(t,\cdot)$ for given t is a shorthand notation to the mapping $x \mapsto f(t,x)$). For all $t \in I$, we can define then

$$F(t) := \int_\Omega f(t,x)\mathrm{d}x. \tag{4.2}$$

The regularity properties of F can be deduced, under natural hypothesis, from those of f thanks to the Dominated convergence theorem. Let us start with the continuity:

Theorem 4.2.1 — Continuity of integral depending on a parameter. Under the above assumptions, assume moreover that, for almost every $x \in \Omega$, the mapping $t \mapsto f(t,x)$ is continuous at some $t_0 \in I$. Assume moreover that there is $\Phi \in L^1(\Omega)$ such that

$$|f(t,x)| \leqslant \Phi(x) \qquad \forall t \in I\,, \quad \text{for a.e. } x \in \Omega.$$

Then, $F : I \to \Omega$ defined by (4.2) is continuous at t_0, i.e.

$$\lim_{t\to t_0} F(t) = \lim_{t\to t_0} \int_\Omega f(t,x)\mathrm{d}x = \int_\Omega \lim_{t\to t_0} f(t,x)\mathrm{d}x = F(t_0).$$

Proof. The proof is a direct application of the dominated convergence theorem. ■

Regarding the differentiability properties of F one has the following

Theorem 4.2.2 — Differentiability of integral depending on a parameter. Under the above assumptions, assume moreover that, for almost every $x \in \Omega$, the mapping $t \mapsto f(t,x)$ is differentiable on I. Assume moreover that there is $\Phi \in L^1(\Omega)$ such that

$$\left|\frac{\partial}{\partial t} f(t,x)\right| \leqslant \Phi(x) \qquad \forall t \in I\,, \quad \text{for a.e. } x \in \Omega.$$

Then, $F : I \to \Omega$ defined by (4.2) is differentiable on I and

$$F'(t) = \int_\Omega \frac{\partial}{\partial t} f(t,x)\mathrm{d}x \qquad \forall t \in I.$$

If moreover, for almost every $x \in \Omega$, the mapping $t \mapsto f(t,x)$ belongs to $\mathscr{C}^1(I)$, then $F \in \mathscr{C}^1(I)$.

Proof. Let $t \in I$ be given and let $(t_n)_n \subset I$ be a sequence which converges to $t \in I$ (with $t_n \neq t$ for any n). We consider, for any $n \in \mathbb{N}$ the sequence

$$\frac{F(t_n) - F(t)}{t_n - t} = \int_\Omega \frac{f(t_n,x) - f(t,x)}{t_n - t}\mathrm{d}x.$$

By assumption, there is some $\mathscr{Z} \subset \Omega$ such that $\mathfrak{m}(\mathscr{Z}) = 0$ and

$$\lim_{n\to\infty} \frac{f(t_n,x) - f(t,x)}{t_n - t} = \frac{\partial}{\partial t} f(t,x) \qquad \forall x \in \Omega \setminus \mathscr{Z}$$

and, by virtue of the mean value theorem

$$\left|\frac{f(t_n,x) - f(t,x)}{t_n - t}\right| \leqslant \Phi(x) \qquad \forall x \in \Omega \setminus \mathscr{Z}.$$

Thanks to the dominated convergence theorem, one has

$$\lim_{n\to\infty} \int_\Omega \frac{f(t_n,x) - f(t,x)}{t_n - t}\mathrm{d}x = \int_\Omega \frac{\partial}{\partial t} f(t,x)\mathrm{d}x$$

i.e. F is differentiable at t with

$$F'(t) = \int_\Omega \frac{\partial}{\partial t} f(t,x)\mathrm{d}x.$$

This proves the first part of the result. Assume now that there is $\mathscr{Z}' \subset \Omega$ with $\mathfrak{m}(\mathscr{Z}') = 0$ such that $f(\cdot,x) \in \mathscr{C}^1(I)$ for any $x \in \Omega \setminus \mathscr{Z}'$. Let then $t \in I$ and $(t_n)_n \subset I$ be a sequence which converges to $t \in I$. Then, for any $x \in \Omega \setminus \mathscr{Z}'$ one has

$$\lim_{n\to\infty} \frac{\partial}{\partial t} f(t_n,x) = \frac{\partial}{\partial t} f(t,x)$$

with

$$\left|\frac{\partial}{\partial t} f(t_n, x)\right| \leqslant \Phi(x) \qquad x \in \Omega \setminus \mathcal{Z}'.$$

Then, according to the dominated convergence theorem

$$\lim_{n\to\infty} F'(t_n) = F'(t)$$

which proves that $F \in \mathscr{C}(I)$. ■

4.3 Convolution

We begin by recalling the following property of the Lebesgue measure $\mathfrak{m}$

Lemma 4.3.1 — Borel regularity. A subset $A \subset \mathbb{R}^N$ is Lebesgue measurable if and only if, for any $\varepsilon > 0$ there is an open set G and a closed set F such that $F \subset A \subset G$ and

$$\mathfrak{m}(G \setminus F) < \varepsilon.$$

If $\mathfrak{m}(A) < \infty$, then F may be chosen to be compact.

Introduce the following

Definition 4.3.1 We denote by $\mathscr{C}_c(\mathbb{R}^N)$ the space of all continuous functions on $\mathbb{R}^N$ with compact support, i.e.,

$$\mathscr{C}_c(\mathbb{R}^N) = \{f \in \mathscr{C}(\mathbb{R}^N)\,,\, \exists K \subset \mathbb{R}^N \text{ compact such that } f(x) = 0\, \forall x \notin K\}.$$

Then, one has the following density result that we admit without a proof:

Theorem 4.3.2 For any $1 \leqslant p < \infty$, the space $\mathscr{C}_c(\mathbb{R}^N)$ is dense in $L^p(\mathbb{R}^N)$, i.e., for any $f \in L^p(\mathbb{R}^N)$ and any $\varepsilon > 0$, there is $g \in \mathscr{C}_c(\mathbb{R}^N)$ such that $\|f - g\|_p \leqslant \varepsilon$.

Proof. We give the proof only in the case $p = 1$. Note first that by the dominated convergence theorem

$$\lim_{R\to\infty} \|f - f\mathbf{1}_{B(0,R)}\|_1 = 0$$

we can assume that $f \in L^1(\mathbb{R}^N)$ has compact support (i.e. vanishes outside some compact subset). Decomposing $f = f^+ - f^-$ into positive and negative part, we can also assume that f is nonnegative. Then there is an increasing sequence of compactly supported simple functions that converge to f pointwise and hence, by the monotone convergence theorem, it also converges in $L^1(\mathbb{R}^N)$. Since every simple function is a finite linear combination of characteristic functions, it is sufficient to prove the result for the characteristic function $\mathbf{1}_A$ of a bounded, measurable set $A \subset \mathbb{R}^N$. Given $\varepsilon > 0$, by the Borel regularity of Lebesgue measure, there exists a bounded open set $G \subset \mathbb{R}^N$ and a compact set $K \subset \mathbb{R}^N$ such that $K \subset A \subset G$ and $\mathfrak{m}(G \setminus K) < \varepsilon$. Let now $g \in \mathscr{C}_c(\mathbb{R}^N)$ be such that $g = 1$ on K, $g = 0$ on $\mathbb{R}^N \setminus G$ and $0 \leqslant g \leqslant 1$ (such a function exists thanks to Urysohn Lemma, see Chapter I). We then have that

$$\|\mathbf{1}_A - g\|_1 = \int_{G\setminus K} |g - \mathbf{1}_A| \mathrm{d}x \leqslant \mathfrak{m}(G \setminus K) \leqslant \varepsilon$$

which proves the result.

The extension to the case $p > 1$ is not difficult and is based simply on the fact that the set of simple functions belonging to $L^p(\mathbb{R}^N)$ is a dense subset of $L^p(\mathbb{R}^N)$ ***(Check this)*** and then to argue as before. ■

Recall the Lebesgue measure is invariant by translation, i.e.

$$\mathfrak{m}(A) = \mathfrak{m}(A + x) \qquad \forall A \in \mathscr{B}(\mathbb{R}^N), \quad x \in \mathbb{R}^N$$

(where $A + x = \{y + x\,,\, y \in A\}$). This allows to define for instance the translation operator on $L^p(\mathbb{R}^N)$: for a given $y \in \mathbb{R}^N$, let $\tau_y \,:\, f \in L^p(\mathbb{R}^N) \mapsto \tau_y f \in L^p(\mathbb{R}^N)$ be defined by

$$\tau_y f(x) = f(x - y) \qquad \forall x \in \mathbb{R}^N.$$

The invariance of the Lebesgue measure under translation shows that $\|\tau_y f\|_p = \|f\|_p$ for any $y \in \mathbb{R}^N$, i.e.

$$\int_{\mathbb{R}^N} |f(x-y)|^p \mathrm{d}x = \int_{\mathbb{R}^N} |f(x)| \mathrm{d}x$$

(Check that this is a simple consequence of the invariance under translation). Actually, one has

> **Proposition 4.3.3** For any $f \in L^p(\mathbb{R}^N)$ $(1 \leqslant p < \infty)$, one has
>
> $$\lim_{y \to 0} \|\tau_y f - f\|_p = 0.$$

✍ *Proof.* Let $f \in L^p(\mathbb{R}^N)$ be given and let $\varepsilon > 0$ be fixed. From Theorem 4.3.2, there is $g \in \mathscr{C}_c(\mathbb{R}^N)$ such that

$$\|f - g\|_p \leqslant \varepsilon.$$

In particular, using the triangle inequality, for any $y \in \mathbb{R}^N$:

$$\|\tau_y f - f\|_p \leqslant \|\tau_y f - \tau_y g\|_p + \|\tau_y g - g\|_p + \|g - f\|_p \leqslant 2\varepsilon + \|\tau_y g - g\|_p \tag{4.3}$$

where we used that $\|\tau_y g - \tau_y f\|_p = \|f - g\|_p \leqslant \varepsilon$. Thus, it is enough to prove that

$$\lim_{y \to 0} \|\tau_y g - g\|_p = 0.$$

Recall that g is continuous with compact support. In particular, there is $R > 0$, such that $g(x) = 0$ for any $\|x\| > R$. For any $y \in \mathbb{R}^N$ with $\|y\| < 1$, one sees then that $\tau_y g(x) = 0$ for any $\|x\| \geqslant R + 1$. In other words, for any $y \in B(0,1)$, $\tau_y g = g = 0$ on $B(0, R+1)^c$. Thus,

$$\|\tau_y g - g\|_p^p = \int_{B(0,R+1)} |g(x-y) - g(x)|^p \mathrm{d}x.$$

Since g is continuous, one sees that, for any $x \in \mathbb{R}^N$, $\lim_{y \to 0} |g(x-y) - g(x)|^p = 0$. Moreover, $|g(x-y) - g(x)|^p \leqslant 2^p \|g\|_\infty^p$ for any $x \in B(0, R+1)$ where $\|g\|_\infty = \sup_{x \in \mathbb{R}^N} |g(x)|$ is finite since g is continuous over a compact and vanishes outside ***(Explain this)***. Since the measure of $B(0, R+1)$ is finite, one sees that the mapping $x \mapsto |g(x-y) - g(x)|^p$ is dominated by an integrable function and, from the dominated convergence theorem,

$$\lim_{y \to 0} \int_{B(0,R+1)} |g(x-y) - g(x)|^p \mathrm{d}x = 0.$$

This shows that $\lim_{y\to 0} \|\tau_y g - g\|_p = 0$ and achieves the proof. Indeed, applying the definition of limit with the above choice of ε, we see there is $\delta > 0$ such that, for $\|y\| \leqslant \delta$, it holds $\|\tau_y g - g\|_p \leqslant \varepsilon$. Inserting this into (4.3), we finally proved that, for any $\varepsilon > 0$, there is $\delta > 0$ such that, for $\|y\| \leqslant \delta$ one has $\|\tau_y f - f\|_p \leqslant 3\varepsilon$. By definition, this means that $\lim_{y\to 0} \|\tau_y f - f\|_p = 0$. ∎

Remark 4.3.1 In the above proof, it was also possible to invoke the fact that g is uniformly continuous on $B(0, R+1)$ from which there is $\delta > 0$ such that

$$\|y\| < \delta \implies |g(x-y) - g(x)|^p < \varepsilon \qquad \forall x \in B(0, R+1).$$

Still using the invariance of $\mathfrak{m}$ under translation, the following convolution will enjoy nice properties:

Theorem 4.3.4 — Young convolution inequality. Let $f \in L^1(\mathbb{R}^N)$ and let $g \in L^p(\mathbb{R}^N)$ with $1 \leqslant p \leqslant \infty$. Then for $\mathfrak{m}$-a.e. $x \in \mathbb{R}^N$ the function $y \mapsto f(x-y)g(y)$ is integrable on $\mathbb{R}^N$ and we define

$$f * g(x) = \int_{\mathbb{R}^N} f(x-y)g(y)\mathrm{d}y \qquad \mathfrak{m} - \text{a.e. } x \in \mathbb{R}^N.$$

In addition, $f * g \in L^p(\mathbb{R}^N)$ and

$$\|f * g\|_p \leqslant \|f\|_1 \|g\|_p.$$

✍ *Proof.* If $p = \infty$, the conclusion is clear. Assume now $p = 1$. For any $x, y \in \mathbb{R}^N$, set $F(x,y) = f(x-y)g(y)$. For $\mathfrak{m}$-a.e. $y \in \mathbb{R}^N$ we have

$$\int_{\mathbb{R}^N} |F(x,y)|\mathrm{d}x = |g(y)| \int_{\mathbb{R}^N} |f(x-y)|\mathrm{d}x = |g(y)|\|f\|_1 < \infty$$

(here again we exactly used the fact that the Lebesgue measure is invariant by translation), so that

$$\int_{\mathbb{R}^N} \mathrm{d}y \int_{\mathbb{R}^N} |F(x,y)|\mathrm{d}x = \|f\|_1 \int_{\mathbb{R}^N} |g(y)|\mathrm{d}y = \|g\|_1 \|f\|_1.$$

We deduce from Tonelli's theorem that $F \in L^1(\mathbb{R}^N \times \mathbb{R}^N, \mathfrak{m} \otimes \mathfrak{m})$. Applying now Fubini's theorem, we see that

$$\int_{\mathbb{R}^N} |F(x,y)|\mathrm{d}y < \infty \qquad \text{for a.e.} \quad x \in \mathbb{R}^N$$

and

$$\int_{\mathbb{R}^N} \mathrm{d}x \int_{\mathbb{R}^N} |F(x,y)|\mathrm{d}y = \int_{\mathbb{R}^N} \mathrm{d}y \int_{\mathbb{R}^N} |F(x,y)|\mathrm{d}x = \|f\|_1 \|g\|_1.$$

Since $|f * g(x)| \leqslant \int_{\mathbb{R}^N} |F(x,y)|\mathrm{d}y$, we get the conclusion whenever $p = 1$.

Assume now $1 < p < \infty$. From the previous step, we know that for a.e. fixed $x \in \mathbb{R}^N$ the function $y \mapsto |f(x-y)|\,|g(y)|^p$ is integrable on $\mathbb{R}^N$. Define

$$h_1 \ : \ y \mapsto |f(x-y)|^{1/p}|g(y)| \qquad \text{and} \qquad h_2 \ : \ y \mapsto |f(x-y)|^{\frac{1}{q}}$$

where $1 = \frac{1}{p} + \frac{1}{q}$, one has therefore

$$h_1 \in L^p(\mathbb{R}^N) \qquad \text{and} \qquad h_2 \in L^q(\mathbb{R}^N)$$

with $\|h_2\|_q^q = \|f\|_1$ while, from the previous step

$$\|h_1\|_p = \left(\int_{\mathbb{R}^N} |f(x-y)|\, |g(y)|^p \mathrm{d}y \right)^{\frac{1}{p}} = \left(\left(|f| * |g|^p \right)(x) \right)^{1/p}$$

(notice that $\|h_1\|_p$ actually depends on x). Therefore, according to Hölder inequality, $h_1 h_2 \in L^1(\mathbb{R}^N)$ with $\|h_1 h_2\|_1 \leqslant \|h_1\|_p \|h_2\|_q$. Now

$$h_1(y) h_2(y) = |f(x-y)||g(y)|$$

since $1/p + 1/q = 1$ and we get

$$\int_{\mathbb{R}^N} |f(x-y)||g(y)| \mathrm{d}y \leqslant \|h_1\|_p \|h_2\|_q$$

so that

$$|f * g(x)| \leqslant \|f\|_1^{1/q} \left(\left(|f| * |g|^p \right)(x) \right)^{1/p} \qquad \text{for a.e. } x \in \mathbb{R}^N$$

and, after integrating the p-th power of the above inequality

$$\|f * g\|_p^p \leqslant \|f\|_1^{p/q} \| \, |f| * |g|^p \|_1 \leqslant \|f\|_1^{p/q} \|f\|_1 \| |g|^p \|_1$$

where we used the first step again. Now, since $\| |g|^p \|_1 = \|g\|_p^p$ we easily get the conclusion. ■

Remark 4.3.2 Notice that $f * g(x) = g * f(x)$ for a.e. $x \in \mathbb{R}^N$ and any $f \in L^1(\mathbb{R}^N)$, $g \in L^p(\mathbb{R}^N)$, i.e the convolution product is commutative. It is also clear, from the linearity of integral, that such a product is also associative with respect to the addition. We can prove however that there is no neutral element for such a product.

Exercise 4.3 Let $f_i(x) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{(x-\mu_i)^2}{2\sigma_i^2} \right)$, $x \in \mathbb{R}$, $i = 1, 2$ where $\sigma_i > 0$, $\mu_i \in \mathbb{R}$ $(i = 1, 2)$. Prove that

$$f_1 * f_2(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right) \qquad \forall x \in \mathbb{R}.$$

where $\mu = \mu_1 + \mu_2$ and $\sigma^2 = \sigma_1^2 + \sigma_2^2$. Compute, for any $1 \leqslant p < \infty$, $\|f_1 * f_2\|_p$ and compare it to $\|f_1\|_1 \|f_2\|_p$. ■

4.3.1 Application to Probability

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a given probability space and let $X, Y\ \Omega \to \mathbb{R}^N$ be two random variables with density f_X and f_Y respectively. We have the following

Proposition 4.3.5 — *Independence and convolution: first version.* If X, Y are independent, then $X+Y$ has a density f_{X+Y} given by

$$f_{X+Y}(x) = f_X * f_Y(x) \qquad x \in \mathbb{R}^N.$$

Proof. According to the above result ($p=1$), notice that $f_X * f_Y$ is well-defined and nonnegative, moreover $f_X * f_Y \in L^1(\mathbb{R}^N)$ and

$$\|f_X * f_Y\|_1 \leqslant \|f_X\|_1 \, \|f_Y\|_1 = 1.$$

Notice that a careful reading of the above proof for $p=1$ one actually has $\|f_X * f_Y\|_1 = \|f_X\|_1 \, \|f_Y\|_1 = 1$ since f_X, f_Y are both nonnegative. Then, $f_X * f_Y$ is indeed the density f_Z of a random variable $Z : \Omega \to \mathbb{R}^N$. Let us show $Z = X+Y$. For a given $x \in \mathbb{R}^N$, set

$$A = \{(u,v) \in \mathbb{R}^N \times \mathbb{R}^N \, ; \, u_i + v_i \leqslant x_i \quad \forall i = 1, \ldots, N\}.$$

Let $\mathbb{P}_{(X,Y)}$ be the law of (X,Y). Since (X,Y) are independent, $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$ has the density $f_X(u) f_Y(v)$ with respect to the Lebesgue measure on $\mathbb{R}^N \times \mathbb{R}^N$. For given $w \in \mathbb{R}^N$, we denote by

$$I(w) = \{u \in \mathbb{R}^N \, , \, u_i \leqslant w_i \qquad \forall i = 1, \ldots, N\}.$$

The repartition function of $X+Y$ is defined as

$$F_{X+Y}(x) = \mathbb{P}(X+Y \in I(x)) \qquad \forall x \in \mathbb{R}^N.$$

Simple application of Fubini's theorem shows that

$$\begin{aligned}
F_{X+Y}(x) = \mathbb{P}\left[(X,Y) \in A\right] = \int_{\mathbb{R}^N \times \mathbb{R}^N} \mathbf{1}_A(u,v) \mathrm{d}\mathbb{P}_{(X,Y)}(u,v) \\
= \int_{\mathbb{R}^N \times \mathbb{R}^N} \mathbf{1}_A(u,v) f_X(u) f_Y(v) \mathrm{d}\mathfrak{m}(u) \mathrm{d}\mathfrak{m}(v) \\
= \int_{\mathbb{R}^N} \left(\int_{I(x-v)} f_X(u) \mathrm{d}\mathfrak{m}(u) \right) f_Y(v) \mathrm{d}\mathfrak{m}(v).
\end{aligned}$$

Now, for given $v \in \mathbb{R}^N$, one has $\int_{I(x-v)} f_X(u) \mathrm{d}\mathfrak{m}(u) = \int_{I(x)} f_X(u-v) \mathrm{d}\mathfrak{m}(u)$ so that

$$\begin{aligned}
F_{X+Y}(x) = \int_{\mathbb{R}^N} \left(\int_{I(x)} f_X(u-v) \mathrm{d}\mathfrak{m}(u) \right) f_Y(v) \mathrm{d}\mathfrak{m}(v) \\
= \int_{I(x)} \left(\int_{R^N} f_X(u-v) f_Y(v) \mathrm{d}\mathfrak{m}(v) \right) \mathrm{d}\mathfrak{m}(u)
\end{aligned}$$

i.e.

$$F_{X+Y}(x) = \int_{I(x)} (f_X \times f_Y)(u) \mathrm{d}\mathfrak{m}(u) = \int_{I(x)} f_Z(u) \mathrm{d}\mathfrak{m}(u)$$

which shows that $F_{X+Y}(x) = F_Z(x)$ and therefore $X+Y = Z$. ∎

4.3.2 Extension and consequences of Young's convolution inequality

The above convolution inequality can be generalized as follows

Theorem 4.3.6 — Generalized Young's inequality. Let $1 \leqslant p, q, r \leqslant \infty$ be such that

$$1/r = 1/p + 1/q - 1.$$

Then, for any $f \in L^p(\mathbb{R}^N)$ and $g \in L^q(\mathbb{R}^N)$ we have that

$$f * g \in L^r(\mathbb{R}^N) \qquad \text{and} \qquad \|f * g\|_r \leqslant \|f\|_p \|g\|_q.$$

Proof. Let $x \in \mathbb{R}^N$ be given. Let us estimate $|f * g(x)|$. One has

$$\begin{aligned}|f * g(x)| &\leqslant \int_{\mathbb{R}^N} |f(x-y)||g(y)|\mathrm{d}y = \int_{\mathbb{R}^N} |f(x-y)|^{1-\frac{p}{r}+\frac{p}{r}}|g(y)|^{1-\frac{q}{r}+\frac{q}{r}}\mathrm{d}y \\ &= \int_{\mathbb{R}^N} (|f(x-y)|)^{1-\frac{p}{r}} (|g(y)|)^{1-\frac{q}{r}} \left(|f(x-y)|^p \, |g(y)|^q\right)^{\frac{1}{r}} \mathrm{d}y.\end{aligned}$$

Set, for fixed $x \in \mathbb{R}^N$

$$h_1(y) = |f(x-y)|^{1-\frac{p}{r}} = |f(x-y)|^{\frac{r-p}{r}}, \qquad h_2(y) = |g(y)|^{\frac{r-q}{r}}$$

and

$$h_3(y) = \left(|f(x-y)|^p \, |g(y)|^q\right)^{\frac{1}{r}}.$$

Since $f \in L^p(\mathbb{R}^N)$, one sees that $h_1 \in L^a(\mathbb{R}^N)$ with $a\frac{r-p}{r} = p$, i.e.

$$h_1 \in L^a(\mathbb{R}^N) \quad \text{ for } \quad \frac{1}{a} = \frac{r-p}{rp} = \frac{1}{p} - \frac{1}{r}$$

with moreover

$$\|h_1\|_a = \left(\int_{\mathbb{R}^N} |f(x-y)|^p \mathrm{d}y\right)^{\frac{1}{a}} = \|f\|_p^{\frac{r-p}{r}}.$$

In the same way, since $g \in L^q(\mathbb{R}^N)$, one has $h_2 \in L^b(\mathbb{R}^N)$ with $b\frac{r-q}{r} = q$, i.e.

$$h_2 \in L^b(\mathbb{R}^N) \quad \text{ for } \quad \frac{1}{b} = \frac{r-q}{rq} = \frac{1}{q} - \frac{1}{r}$$

with moreover

$$\|h_2\|_b = \left(\int_{\mathbb{R}^N} |g(y)|^q \mathrm{d}y\right)^{\frac{1}{b}} = \|g\|_q^{\frac{r-q}{r}}.$$

Now, setting $c = r$, one sees that $h_3 \in L^c(\mathbb{R}^N)$. Indeed,

$$\int_{\mathbb{R}^N} |h_3(y)|^r \mathrm{d}y = \int_{\mathbb{R}^N} |f(x-y)|^p \, |g(y)|^q \mathrm{d}y = \left(|f|^p * |g|^q\right)(x)$$

which is well-defined according to Young's inequality with $|f|^p \in L^1(\mathbb{R}^N)$ and $|g|^q \in L^1(\mathbb{R}^N)$.

Notice that

$$\frac{1}{a} + \frac{1}{b} + \frac{1}{c} = \frac{1}{p} + \frac{1}{q} - \frac{1}{r} = 1$$

so that, according to the generalized Holder inequality

$$|f * g(x)| \leqslant \int_{\mathbb{R}^N} h_1(y)h_2(y)h_3(y)\mathrm{d}y \leqslant \|h_1\|_a \, \|h_2\|_b \, \|h_3\|_c \qquad \forall x \in \mathbb{R}^N.$$

According to the previous computations, it holds

$$|f * g(x)| \leqslant \|f\|_p^{\frac{r-p}{r}} \|g\|_q^{\frac{r-q}{r}} \left(|f|^p * |g|^q(x)\right)^{\frac{1}{r}} \qquad x \in \mathbb{R}^N.$$

Therefore, taking the r-th power of this inequality and integrating with respect to x, it holds

$$\|f * g\|_r^r \leqslant \|f\|_p^{r-p} \|g\|_q^{r-q} \||f|^p * |g|^q\|_1$$

and, since $|f|^p, |g|^q \in L^1(\mathbb{R}^N)$, Young's convolution inequality states that

$$\||f|^p * |g|^q\|_1 \leqslant \||f|^p\|_1 \||g|^q\|_1 = \|f\|_p^p \|g\|_q^q.$$

In other words

$$\|f * g\|_r^r \leqslant \|f\|_p^{r-p} \|g\|_q^{r-q} \|f\|_p^p \|g\|_q^q = \|f\|_p^r \|g\|_q^r$$

which is the desired result. ■

Exercise 4.4 For any $t > 0$, define the ***Heat kernel***

$$G_t(x) = \frac{1}{(4\pi t)^{N/2}} \exp\left(-\frac{\|x\|^2}{4t}\right) \qquad \forall x \in \mathbb{R}^N.$$

Prove that $G_t \in L^q(\mathbb{R}^N)$ for any $1 \leqslant q \leqslant \infty$ and compute $\|G_t\|_q$. Given $f \in L^p(\mathbb{R}^N)$ we define

$$u_t(x) = G_t * f(x) \qquad \forall x \in \mathbb{R}^N.$$

Show that

$$\|u_t\|_r \leqslant \frac{\|f\|_p}{(4\pi t)^{\frac{N}{2}\left(\frac{1}{p}-\frac{1}{r}\right)}} \qquad \forall 1 \leqslant p \leqslant r \leqslant \infty.$$

In the special case in which $r = \infty$, we have the following useful consequence:

Corollary 4.3.7 Let $1 \leqslant p \leqslant \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$ be given. If $f \in L^p(\mathbb{R}^N)$ and $g \in L^q(\mathbb{R}^N)$ then

$$f * g \in L^\infty(\mathbb{R}^N) \qquad \text{and} \qquad \|f * g\|_\infty \leqslant \|f\|_p \|g\|_q.$$

The above result can be improved since, in the above situation, the convolution $f * g$ is actually (uniformly) continuous:

Proposition 4.3.8 Let $1 \leqslant p \leqslant \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$ be given. If $f \in L^p(\mathbb{R}^N)$ and $g \in L^q(\mathbb{R}^N)$ then $f * g$ is uniformly continous over $\mathbb{R}^N$.

Proof. To prove the uniform continuity, one notices first that convolution commutes with translations, i.e.

$$\tau_y(f * g) = \tau_y f * g = f * \tau_y g \qquad \forall y \in \mathbb{R}^N$$

(Check this). Since $1/p + 1/q = 1$, one has that $p < \infty$ or $q < \infty$. Let us assume $p < \infty$. Assume also $\|g\|_q \neq 0$ (otherwise there is nothing to check, $f * g = 0$). Then, from the previous observation

$$\|\tau_y(f * g) - (f * g)\|_\infty = \|\tau_y f * g - f * g\|_\infty = \|(\tau_y f - f) * g\|_\infty$$

and, according to the previous Corollary,

$$\|\tau_y(f*g)-(f*g)\|_\infty \leqslant \|\tau_y f - f\|_p \|g\|_q.$$

Now, according to Proposition 4.3.3, for any $\varepsilon>0$, there is $\delta>0$ such that

$$\|\tau_y f - f\|_p < \varepsilon/\|g\|_q \qquad \forall \|y\|<\delta$$

which means that

$$\|\tau_y(f*g)-(f*g)\|_\infty < \varepsilon \qquad \forall \|y\|<\delta.$$

This is exactly the definition of the uniform continuity of $f*g$ over $\mathbb{R}^N$. If $p=\infty$ and $q<\infty$, the proof is exactly the same. ■

Exercise 4.5 If $f,g\in\mathscr{C}_c(\mathbb{R}^N)$ prove that $f*g\in\mathscr{C}_c(\mathbb{R}^N)$. Using this, if $f\in L^p(\mathbb{R}^N)$ and $g\in L^q(\mathbb{R}^N)$ with $\frac{1}{p}+\frac{1}{q}=1$ and both $p,q<\infty$, prove that

$$\lim_{\|x\|\to\infty} f*g(x)=0.$$

Hint: use the fact that, since $p,q<\infty$, for any $\varepsilon\in(0,1)$, there are $f_\varepsilon, g_\varepsilon\in\mathscr{C}_c(\mathbb{R}^N)$ such that $\|f_\varepsilon*g_\varepsilon - f*g\|_\infty \leqslant \varepsilon\left(1+\|f\|_p+\|g\|_q\right)$. ■

The importance of convolution is illustrated by the following result :

Theorem 4.3.9 Let $1\leqslant p<\infty$. Let $(\varrho_n)_n$ be a sequence of functions such that, for any $n\geqslant 1$:

1. ϱ_n is infinitely differentiable on $\mathbb{R}^N$, i.e. $\varrho_n\in\mathscr{C}^\infty(\mathbb{R}^N)$.
2. $\varrho_n(x)\geqslant 0$ for any $x\in\mathbb{R}^N$.
3. $\varrho_n(x)=0$ for any $x\in\mathbb{R}^N$ with $\|x\|>\frac{1}{n}$.
4. $\displaystyle\int_{\mathbb{R}^N}\varrho_n(x)\mathrm{d}x=1$.

Then, for any $f\in L^p(\mathbb{R}^N)$, it holds that

$$\varrho_n * f\in\mathscr{C}^\infty(\mathbb{R}^N), \qquad \forall n\in\mathbb{N},$$

and

$$\lim_{n\to\infty}\|\varrho_n*f-f\|_p=0.$$

In particular, $\mathscr{C}^\infty(\mathbb{R}^N)$ is dense in $L^p(\mathbb{R}^N)$.

Remark 4.3.3 A sequence $(\varrho_n)_n$ with the above properties is called a sequence of ***mollifiers*** (also called *approximations to the identity*). Notice that such a sequence exists. Indeed, consider the single function $\varrho\in\mathscr{C}_c^\infty(\mathbb{R}^N)$ with $\varrho(x)=0$ for $\|x\|>1$, ϱ non identically zero [a] then, for $n\geqslant 1$, setting

$$\varrho_n(x)=\frac{n^N}{\|\varrho\|_1}\varrho(n\,x), \qquad x\in\mathbb{R}^N$$

then $(\varrho_n)_n$ is a sequence of mollifiers.

[a]For instance, take

$$\varrho(x) = \begin{cases} \exp\left(\frac{1}{\|x\|^2-1}\right) & \forall \|x\| < 1 \\ 0 & \forall \|x\| \geqslant 1. \end{cases}$$

Proof. Let $n \in \mathbb{N}$ be fixed. Notice that, since ϱ_n has compact support, $\varrho_n \in L^q(\mathbb{R}^N)$ and, from Prop. 4.3.8, $\varrho_n * f$ is uniformly continous for any $n \in \mathbb{N}$. Let now prove that $\varrho_n * f$ is infinitely differentiable. We focus mainly on first-order derivatives. Let $i \in \{1, \dots, N\}$. One wishes to compute the partial derivative of $\varrho_n * f$ with respect to the i-th variable x_i. One has

$$\varrho_n * f(x) = \int_{\mathbb{R}^N} \varrho_n(x-y) f(y) \mathrm{d}y$$

with, ϱ_n differentiable. Since ϱ_n is compactly supported on $B_c(0, \frac{1}{n})$, the same occur for $\dfrac{\partial}{\partial x_i}\varrho_n$ and, since $\varrho_n \in \mathscr{C}^\infty$, one has $\dfrac{\partial}{\partial x_i}\varrho_n$ is bounded on $B_c(0, \frac{1}{n})$, i.e.

$$\frac{\partial}{\partial x_i}\varrho_n \in \bigcap_{1 \leqslant q \leqslant \infty} L^q(\mathbb{R}^N).$$

Then, using Theorem 4.2.2 ***(Check the assumptions are satisfied)***, one sees that

$$\frac{\partial}{\partial x_i}(\varrho_n * f)(x) = \int_{\mathbb{R}^N} \frac{\partial}{\partial x_i}\left(\varrho_n(x-y) f(y)\right) \mathrm{d}y$$

i.e.

$$\frac{\partial}{\partial x_i}(\varrho_n * f)(x) = \int_{\mathbb{R}^N} f(y) \frac{\partial}{\partial x_i}\varrho_n(x-y) \mathrm{d}y.$$

which proves that

$$\frac{\partial}{\partial x_i}(\varrho_n * f)(x) = (f * \frac{\partial}{\partial x_i}\varrho_n)(x) \qquad \forall x \in \mathbb{R}^N, \quad i = 1, \dots, N.$$

By induction, one proves easily that, for any $k \geqslant 1$, one has

$$\frac{\partial^k}{\partial x_{i_1} \dots \partial x_{i_k}}(\varrho_n * f)(x) = \left(\frac{\partial^k}{\partial x_{i_1} \dots \partial x_{i_k}}\varrho_n * f\right)(x) \qquad \forall x \in \mathbb{R}^N.$$

This shows that $f * \varrho_n \in \mathscr{C}^\infty(\mathbb{R}^N)$. Let us now prove that $\lim_{n\to\infty} \|\varrho_n * f - f\|_p = 0$. One has

$$|\varrho_n * f(x) - f(x)| = \left| \int_{\mathbb{R}^N} \left(f(x-y) - f(x)\right) \varrho_n(y) \mathrm{d}y \right|$$

since $\int_{\mathbb{R}^N} \varrho_n(y)\mathrm{d}y = 1$. Therefore

$$|\varrho_n * f(x) - f(x)| \leqslant \int_{\mathbb{R}^N} |f(x-y) - f(x)| \, \varrho_n(y) \mathrm{d}y.$$

Using Minkowski's integral inequality (see Section 4.1), one has

$$\|\varrho_n * f - f\|_p \leqslant \left(\int_{\mathbb{R}^N}\left[\int_{\mathbb{R}^N}|f(x-y)-f(x)|\,\varrho_n(y)\mathrm{d}y\right]^p \mathrm{d}x\right)^{1/p}$$
$$\leqslant \int_{\mathbb{R}^N}|\varrho_n(y)|\left[\int_{\mathbb{R}^N}|f(x-y)-f(x)|^p\,\mathrm{d}x\right]^{\frac{1}{p}}\mathrm{d}y$$
$$\leqslant \int_{\mathbb{R}^N}|\varrho_n(y)|\|\tau_y f - f\|_p \mathrm{d}y = \int_{B(0,\frac{1}{n})}\varrho_n(y)\|\tau_y f - f\|_p \mathrm{d}y.$$

Given $\varepsilon > 0$, one knows that there exists $\delta > 0$ such that $\|\tau_y f - f\|_p < \varepsilon$ whenever $\|y\| < \delta$. Then, for $n > \frac{1}{\delta}$, one has

$$\int_{B(0,\frac{1}{n})}\varrho_n(y)\|\tau_y f - f\|_p \mathrm{d}y \leqslant \varepsilon \int_{B(0,\frac{1}{n})}\varrho_n(y)\mathrm{d}y = \varepsilon$$

which proves that $\lim_{n\to\infty}\|\varrho_n * f - f\|_p = 0$. ■

4.4 Complements : Compact subsets of $L^p(\mathbb{R}^N)$

The following theorem and its corollary characterize compact subset of L^p spaces:

Theorem 4.4.1 — Kolmogorov-Riesz-Fréchet. Let $\mathcal{E}$ be a bounded set in $L^p(\mathbb{R}^N)$ with $1 \leqslant p < \infty$. Assume that

$$\limsup_{h\to 0}\, \sup_{f\in\mathcal{E}} \|\tau_h f - f\|_p = 0$$

$$\forall \varepsilon > 0, \exists \delta > 0 \text{ such that } \|\tau_h f - f\|_p < \varepsilon \qquad \forall f \in \mathcal{E}\,;\, \forall h \in \mathbb{R}^N,\ \|h\| < \delta. \tag{4.4}$$

Then, for any measurable set $\Omega \subset \mathbb{R}^N$ with $\mathfrak{m}(\Omega) < \infty$, the set $\mathcal{E}_{|\Omega}$ is relatively compact in $L^p(\mathbb{R}^N)$ where $\mathcal{E}_{|\Omega}$ is the set of all restrictions of function of $\mathcal{E}$ on Ω.

Proof. We divide the proof into 4 steps introducing first a sequence of mollifiers $(\varrho_n)_n$. Let $\varepsilon > 0$ and $\delta > 0$ be given satisfying (4.4).

Step 1. We prove that, for

$$\|(\varrho_n * f) - f\|_p \leqslant \varepsilon \qquad \forall f \in \mathcal{E}, \quad \forall n > 1/\delta. \tag{4.5}$$

Indeed, let $f \in \mathcal{E}$ and $n \in \mathbb{N}$ be fixed. Since $\int_{\mathbb{R}^N}\varrho_n(y)\mathrm{d}y = 1$, one sees that, for any $x \in \mathbb{R}^N$

$$|(\varrho_n * f)(x) - f(x)| \leqslant \int_{\mathbb{R}^N}|f(x-y)-f(x)|\varrho_n(y)\mathrm{d}y$$
$$\leqslant \left(\int_{\mathbb{R}^N}|f(x-y)-f(x)|^p\varrho_n(y)\mathrm{d}y\right)^{1/p}$$
$$= \left(\int_{B(0,\frac{1}{n})}|f(x-y)-f(x)|^p\varrho_n(y)\mathrm{d}y\right)^{1/p}$$

thanks to Holder's inequality ***(Check this !!)***. Taking the p-th power and integrating over $\mathbb{R}^N$ we get

$$\|(\varrho_n * f) - f\|_p^p \leqslant \int_{\mathbb{R}^N} \mathrm{d}x \int_{B(0,\frac{1}{n})} |f(x-y) - f(x)|^p \varrho_n(y)\mathrm{d}y$$

so that, thanks to Fubini theorem

$$\|(\varrho_n * f) - f\|_p^p \leqslant \int_{B(0,\frac{1}{n})} \varrho_n(y)\mathrm{d}y \int_{\mathbb{R}^N} |f(x-y) - f(y)|^p \mathrm{d}y = \int_{B(0,\frac{1}{n})} \varrho_n(y)\mathrm{d}y \|\tau_y f - f\|_p^p$$

so that, if $1/n < \delta$, then

$$\|(\varrho_n * f) - f\|_p^p \leqslant \varepsilon^p \int_{B(0,\frac{1}{n})} \varrho_n(y)\mathrm{d}y = \varepsilon^p$$

which is (4.5).

Step 2. For any $n \in \mathbb{N}$, there is $C_n > 0$ such that

$$\|(\varrho_n * f) - f\|_\infty \leqslant C_n \|f\|_p \qquad \forall f \in \mathcal{E}$$

and

$$|\varrho_n * f(x) - \varrho_n * f(y)| \leqslant C_n \|f\|_p \, \|x-y\| \qquad \forall f \in \mathcal{E}\,;\, x, y \in \mathbb{R}^N.$$

The first inequality is easily proven thanks to Young's inequality with $C_n = \|\varrho_n\|_q$ $(1/q + 1/p = 1)$. To prove the second inequality, one applies the first one to the mollifier $\nabla \varrho_n$ (the gradient of ϱ_n). Indeed, recall that, for any $i = 1, \dots, N$

$$\frac{\partial}{\partial x_i}(\varrho_n * f) = (\frac{\partial}{\partial x_i}\varrho_n * f)$$

so that, thanks to Young's inequality

$$\left\| \frac{\partial}{\partial x_i}(\varrho_n * f) \right\|_\infty \leqslant \left\| \frac{\partial}{\partial x_i}\varrho_n \right\|_q \|f\|_p \qquad \forall f \in \mathcal{E}$$

where $1/p + 1/q = 1$. Taking $C_n = \max_i(\|\varrho_n\|_q, \left\| \frac{\partial}{\partial x_i}\varrho_n \right\|_q)$, one sees from the mean value theorem that

$$|\varrho_n * f(x) - \varrho_n * f(y)| \leqslant C_n \|f\|_p \|x-y\| \qquad \forall f \in \mathcal{E}\,;\, x, y \in \mathbb{R}^N.$$

Step 3. For any $\kappa > \varepsilon$ and any $\Omega \subset \mathbb{R}^N$ of finite measure, there is a bounded measurable subset $A \subset \Omega$ such that

$$\|f\|_{L^p(\Omega \setminus A)} < \kappa \qquad \forall f \in \mathcal{E}. \tag{4.6}$$

Indeed, we write

$$\|f\|_{L^p(\Omega\setminus A)} \leqslant \|f - (\varrho_n * f)\|_p + \|\varrho_n * f\|_{L^p(\Omega\setminus A)} \qquad \forall n \in \mathbb{N}$$

From the previous step, one sees that

$$\|\varrho_n * f\|_{L^p(\Omega\setminus A)} \leqslant \|\varrho_n * f\|_\infty \, \mathfrak{m}(\Omega \setminus A) \leqslant C_n \, \|f\|_p \mathfrak{m}(\Omega \setminus A)$$

so that, using (4.5) and choosing $n > 1/\delta$

$$\|f\|_{L^p(\Omega\setminus A)} \leqslant \varepsilon + C_n \|f\|_p \mathfrak{m}(\Omega \setminus A)$$

and it suffices to choose A so that $\mathfrak{m}(\Omega \setminus A)$ is small enough.

Step 4. Given $\Omega \subset \mathbb{R}^N$ with $\mathfrak{m}(\Omega) < \infty$, we prove the relative compactness of $\mathcal{E}_{|\Omega}$. According to Bolzano-Weierstrass Theorem (see Chapter I), it suffices to prove that $\mathcal{E}_{|\Omega}$ is totally bounded. Given $\varepsilon > 0$, we fix $A \subset \Omega$ such that (4.6) holds with $\kappa = 2\varepsilon$. Fix then $\delta > 0$ such that (4.4) holds and let $n > 1/\delta$ be fixed. Consider the set

$$\mathcal{H} = \{(\varrho_n * f)|_{\overline{A}},\ \ f \in \mathcal{E}\}.$$

We know that $\mathcal{H} \subset \mathscr{C}(\overline{A})$ and, according Step 2, $\mathcal{H}$ is an equicontinuous family of $\mathscr{C}(\overline{A})$ which is pointwise bounded. According to Ascoli-Arzelà Theorem, $\mathcal{H}$ is relatively compact in $\mathscr{C}(\overline{A})$. Since $\|\cdot\|_p \leqslant \mathfrak{m}(\overline{A})^{1/p}\|\cdot\|_\infty$ and $\mathscr{C}(\overline{A})$ is dense in $L^p(\overline{A})$, one checks easily ***(Check it)*** that $\mathcal{H}$ is relatively compact in $L^p(\overline{A})$. In particular, it is totally bounded, i.e. there is a finite number of continuous functions $h_1, \ldots, h_k \in \mathcal{C}(\overline{A})$ such that

$$\mathcal{H} \subset \bigcup_{i=1}^{\infty} B_p(h_i, \varepsilon)$$

where $B_p(h_i, \varepsilon) = \{h \in L^p(\overline{A})\ ;\ \|h - h_i\|_p \leqslant \varepsilon\}$. Set then

$$g_i = h_i \mathbf{1}_{\overline{A}} \in L^p(\Omega)$$

and $B_{p,\Omega}(g_i, 4\varepsilon) = \{g \in L^p(\Omega)\,,\ \|g - g_i\|_p \leqslant 4\varepsilon\}$. We prove that

$$\mathcal{E}_{|\Omega} \subset \bigcup_{i=1}^{k} B_{p,\Omega}(g_i, 4\varepsilon).$$

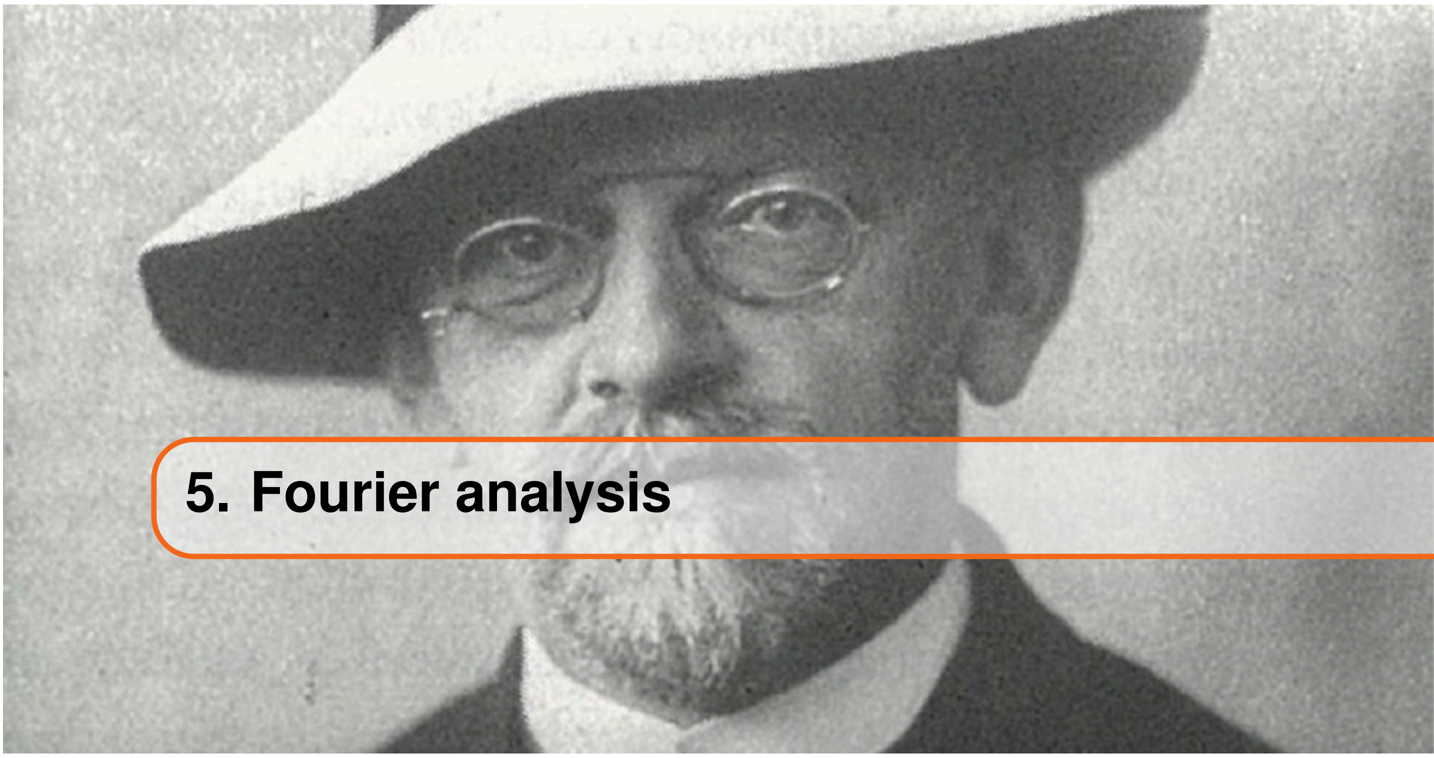
Indeed, given $f \in \mathcal{E}$, there is some $i = 1, \ldots, k$ such that $\|(\varrho_n * f) - h_i\|_{L^p(\overline{A}} \leqslant \varepsilon$. Since

$$\|f - g_i\|^p_{L^p(\Omega)} = \int_{\Omega\setminus\overline{A}} |f|^p \mathrm{d}x + \int_{\overline{A}} |f - h_i|^p \mathrm{d}x$$

we see from (4.6) that

$$\begin{aligned}\|f - g_i\|^p_{L^p(\Omega)} &\leqslant 2\varepsilon + \int_{\overline{A}} |f - h_i|^p \mathrm{d}x \\ &\leqslant 2\varepsilon + \int_{\overline{A}} |f - (\varrho_n * f)|^p \mathrm{d}x + \int_{\overline{A}} |(\varrho_n * f) - h_i|^p \mathrm{d}x \leqslant 4\varepsilon\end{aligned}$$

according to Step 1. This proves the result. ∎



5. Fourier analysis

We denote here $\mathrm{d}x$ the Lebesgue measure $\mathrm{d}\mathfrak{m}$ over $\mathbb{R}^d$ and, as in **Chapter 4**, we denote by

$$L^p(\mathbb{R}^d), \qquad 1 \leqslant p \leqslant \infty$$

the Lebesgue spaces with respect to the Lebesgue measure. The only difference with respect to the previous Chapter is that we shall consider now functions

$$f \,:\, \mathbb{R}^d \to \mathbb{C}$$

i.e. f takes *complex* values. For this reason, we will often write $L^p(\mathbb{R}^d, \mathbb{C})$ to keep track of the fact that we are dealing with complex-valued function. This is no loss of generality and everything we proved in **Chapter 4** still applies here since any function $f \,:\, \mathbb{R}^d \to \mathbb{C}$ can be written as

$$f = \mathrm{Re}f + i\mathrm{Im}f$$

where $\mathrm{Re}f$ and $\mathrm{Im}f$ are real-valued functions. Then, f is said to be integrable over $\mathbb{R}^d$ if and only if $\mathrm{Re}f$ and $\mathrm{Im}f$ are integrable over $\mathbb{R}^d$ and we have then

$$\int_{\mathbb{R}^d} f(x)\mathrm{d}x = \int_{\mathbb{R}^d} \mathrm{Re}f(x)\mathrm{d}x + i\int_{\mathbb{R}^d} \mathrm{Im}f(x)\mathrm{d}x.$$

More generally, we will have $f \in L^p(\mathbb{R}^d, \mathbb{C})$ if and only if both $\mathrm{Re}f$ and $\mathrm{Im}f$ belong to $L^p(\mathbb{R}^d, \mathbb{R})$. Notice that $L^p(\mathbb{R}^d, \mathbb{C})$ is a $\mathbb{C}$-vector space. Moreover, if $|f|$ is the modulus of f,

$$|f(x)| = \sqrt{\mathrm{Re}f(x)^2 + \mathrm{Im}f(x)^2} \qquad x \in \mathbb{R}^d$$

then one sees that

$$f \in L^p(\mathbb{R}^d, \mathbb{C}) \Longleftrightarrow |f| \in L^p(\mathbb{R}^d, \mathbb{R})$$

and

$$\|f\|_p = \|\,|f|\,\|_p.$$

For a given function $f \in L^p(\mathbb{R}^d, \mathbb{C})$, we indicate by $\overline{f}$ its complex conjugate:

$$f = \mathrm{Re}f + i\mathrm{Im}f \implies \overline{f} = \mathrm{Re}f - i\mathrm{Im}f$$

so that

$$|f|^2 = f\,\overline{f}.$$

Notice that a consequence of Minkowski's integral inequality is that, for $f : \mathbb{R}^d \to \mathbb{C}$

$$\left|\int_{\mathbb{R}^d} f(x)\mathrm{d}x\right| \leqslant \int_{\mathbb{R}^d} |f(x)|\mathrm{d}x$$

(see Remark 4.1.5).

5.1 Fourier transform in $L^1(\mathbb{R}^d)$

With this preamble, we define the Fourier transform on $L^1(\mathbb{R}^d) = L^1(\mathbb{R}^d, \mathbb{C})$ as follows:

Definition 5.1.1 Given $f \in L^1(\mathbb{R}^d, \mathbb{C})$, the Fourier transform of f, denoted by $\mathcal{F}(f)$ or $\widehat{f}$ is defined by

$$\mathcal{F}(f)(\xi) = \widehat{f}(\xi) = \int_{\mathbb{R}^d} f(x)\mathbf{e}^{-i\xi\cdot x}\mathrm{d}x \qquad \forall \xi \in \mathbb{R}^d$$

where $\xi \cdot x = \langle \xi, x\rangle = \sum_{j=1}^d \xi_j x_j$ is the standard euclidean inner space on $\mathbb{R}^d$.

Notice that, since

$$\left|f(x)\mathbf{e}^{-i\xi\cdot x}\right| \leqslant |f(x)| \qquad \forall x \in \mathbb{R}^d, \xi \in \mathbb{R}^d$$

the above integral is well-defined for any $\xi \in \mathbb{R}^d$.

5.1.1 Main properties – Riemann-Lebesgue Theorem

Let us list several simple properties of Fourier transform:

Proposition 5.1.1 Let $f \in L^1(\mathbb{R}^d, \mathbb{C})$ be given.

1. For any $\xi \in \mathbb{R}^d$
$$\mathcal{F}(\overline{f})(\xi) = \overline{\mathcal{F}(f)(-\xi)}.$$
2. For any $h \in \mathbb{R}^d$, set
$$\tau_h f(\cdot) = f(\cdot - h).$$
Then,
$$\mathcal{F}(\tau_h f)(\xi) = \mathbf{e}^{-i\xi\cdot h}\mathcal{F}(f)(\xi) \qquad \xi \in \mathbb{R}^d.$$
3. Given $h \in \mathbb{R}^d$, set
$$g(x) = \mathbf{e}^{ih\cdot x}f(x) \qquad x \in \mathbb{R}^d$$
then
$$\mathcal{F}(g)(\xi) = \mathcal{F}(f)(\xi - h) = \tau_h\mathcal{F}(f)(\xi) \qquad \forall \xi \in \mathbb{R}^d.$$
4. For $\lambda > 0$, set $g(x) = f(x/\lambda)$. Then
$$\mathcal{F}(g)(\xi) = \lambda^d\mathcal{F}(f)(\lambda\xi) \qquad \forall \xi \in \mathbb{R}^d.$$

5. If $g \in L^1(\mathbb{R}^d)$ and $f * g$ denotes the convolution of f with g then

$$\mathcal{F}(f * g)(\xi) = \mathcal{F}(f)(\xi)\,\mathcal{F}(g)(\xi) \qquad \forall \xi \in \mathbb{R}^d. \tag{5.1}$$

Proof. The proof of the first 4 properties are an easy **Exercise**. Let us prove (5.1). Recall that, since $f, g \in L^1(\mathbb{R}^d, \mathbb{C})$, Young's integral inequality ensures that $f * g \in L^1(\mathbb{R}^d)$. We can therefore compute $\mathcal{F}(f * g)$. Given $\xi \in \mathbb{R}^d$ one has, thanks to Fubini's theorem,

$$\begin{aligned}
\mathcal{F}(f * g)(\xi) &= \int_{\mathbb{R}^d} \mathrm{e}^{-i\xi\cdot x} f * g(x)\mathrm{d}x = \int_{\mathbb{R}^d} \mathrm{e}^{-i\xi\cdot x}\left(\int_{\mathbb{R}^d} f(x-y)g(y)\mathrm{d}y\right)\mathrm{d}x \\
&= \int_{\mathbb{R}^d} g(y)\left(\int_{\mathbb{R}^d} \mathrm{e}^{-i\xi\cdot x} f(x-y)\mathrm{d}x\right)\mathrm{d}y = \int_{\mathbb{R}^d} g(y)\left(\int_{\mathbb{R}^d} \mathrm{e}^{-i\xi\cdot x}\tau_y f(x)\mathrm{d}x\right)\mathrm{d}y.
\end{aligned}$$

Using the second point, one has

$$\int_{\mathbb{R}^d} \mathrm{e}^{-i\xi\cdot x}\tau_y f(x)\mathrm{d}x = \mathcal{F}(\tau_y f)(\xi) = \mathrm{e}^{-iy\cdot\xi}\mathcal{F}(f)(\xi)$$

so that

$$\begin{aligned}
\mathcal{F}(f * g)(\xi) &= \int_{\mathbb{R}^d} g(y)\mathrm{e}^{-iy\cdot\xi}\mathcal{F}(f)(\xi)\mathrm{d}y \\
&= \mathcal{F}(f)(\xi)\int_{\mathbb{R}^d} g(y)\mathrm{e}^{-iy\cdot\xi}\mathrm{d}y = \mathcal{F}(f)(\xi)\mathcal{F}(g)(\xi)
\end{aligned}$$

which proves the result. ■

We can notice an important difference between f and its Fourier transform $\mathcal{F}(f)$: the element $f \in L^1(\mathbb{R}^d, \mathbb{C})$ is actually an equivalent class of functions which are all identified to a single function defined almost every where on $\mathbb{R}^d$ while $\mathcal{F}(f)$ is a function defined at each ξ of $\mathbb{R}^d$. Actually, one has the following

Theorem 5.1.2 — Riemann-Lebesgue Theorem. For any $f \in L^1(\mathbb{R}^d, \mathbb{C})$, the Fourier transform $\mathcal{F}(f)$ is continuous on $\mathbb{R}^d$ (i.e. $\mathcal{F}(f) \in \mathscr{C}(\mathbb{R}^d)$) with moreover

$$\lim_{|\xi|\to\infty} \mathcal{F}(f)(\xi) = 0 \tag{5.2}$$

and

$$|\mathcal{F}(f)(\xi)| \leqslant \|f\|_1 \qquad \forall \xi \in \mathbb{R}^d.$$

✍ *Proof.* For any $\xi \in \mathbb{R}^d$, one has

$$|\mathcal{F}(f)(\xi)| \leqslant \int_{\mathbb{R}^d} \left|f(x)\mathrm{e}^{-i\xi\cdot x}\right|\mathrm{d}x \leqslant \int_{\mathbb{R}^d} |f(x)|\mathrm{d}x = \|f\|_1 \quad \forall \xi \in \mathbb{R}^d.$$

Let us show the continuity of $\mathcal{F}(f)$. Pick $\xi \in \mathbb{R}^d$ and let $(\xi_n)_n \subset \mathbb{R}^d$ be a sequence which converges to ξ. Introduce then, for any $n \in \mathbb{N}$, the functions

$$g_n(x) = f(x)\mathrm{e}^{-i\xi_n\cdot x} \qquad \text{and} \qquad g(x) = f(x)\mathrm{e}^{-i\xi\cdot x} \qquad \text{for a.e. } x \in \mathbb{R}^d.$$

Clearly, $g_n(x) \to g(x)$ for a.e. $x \in \mathbb{R}^d$ and

$$|g_n(x)| \leqslant |f(x)| \qquad \text{for a.e. } x \in \mathbb{R}^d, \quad n \in \mathbb{N}$$

with $f \in L^1(\mathbb{R}^d)$. Then, according to the dominated convergence theorem

$$\lim_{n\to\infty}\int_{\mathbb{R}^d} g_n(x)\mathrm{d}x = \int_{\mathbb{R}^d} g(x)\mathrm{d}x$$

which exactly means

$$\lim_{n\to\infty}\mathcal{F}(f)(\xi_n) = \mathcal{F}(f)(\xi).$$

This shows that $\mathcal{F}(f) \in \mathscr{C}(\mathbb{R}^d)$. Let us now prove (5.2). For $\xi \neq 0$, since $\mathbf{e}^{-i\pi} = -1$ one has

$$\mathcal{F}(f)(\xi) = -\int_{\mathbb{R}^d} f(x)\mathbf{e}^{-i\xi\cdot\left(x+\frac{\pi}{\|\xi\|^2}\xi\right)}\mathrm{d}x$$

and, using the change of variable $y = x + \frac{\pi}{\|\xi\|^2}\xi$, we see that

$$\mathcal{F}(f)(\xi) = -\int_{\mathbb{R}^d} f\left(y - \frac{\pi}{\|\xi\|^2}\xi\right)\mathbf{e}^{-i\xi\cdot y}\mathrm{d}y.$$

Summing this with the actual definition of $\mathcal{F}(f)$ gives

$$\mathcal{F}(f)(\xi) = \frac{1}{2}\int_{\mathbb{R}^d}\left(f(x) - f\left(x - \frac{\pi}{\|\xi\|^2}\xi\right)\right)\mathbf{e}^{-i\xi\cdot x}\mathrm{d}x$$

i.e., setting $\boldsymbol{a} = \boldsymbol{a}(\xi) = \frac{\pi}{\|\xi\|^2}\xi$ one has

$$\mathcal{F}(f)(\xi) = \frac{1}{2}\int_{\mathbb{R}^d}\left(f(x) - \tau_{\boldsymbol{a}}f(x)\right)\mathbf{e}^{-i\xi\cdot x}\mathrm{d}x$$

where $\tau_{\boldsymbol{a}}$ denote the translation operator $(\tau_{\boldsymbol{a}}f)(x) = f(x - \boldsymbol{a})$. In particular

$$|\mathcal{F}(f)(\xi)| \leqslant \frac{1}{2}\|f - \tau_{\boldsymbol{a}}f\|_1.$$

Since $\lim_{|\xi|\to\infty}\boldsymbol{a} = 0$, using the continuity of translation in $L^1(\mathbb{R}^d)$ (see **Chapter 4**), we get (5.2). ■

■ **Example 5.1** A simple consequence of the Riemann-Lebesgue Lemma is the following: for any Borel subset $A \subset \mathbb{R}$ with finite Lebesgue measure $\lambda(A) < \infty$, it holds

$$\lim_{n\to\infty}\int_A \cos^2(nx)\mathrm{d}x = \frac{\lambda(A)}{2}.$$

Indeed, since $\cos^2(nx) = \frac{1+\cos(2nx)}{2}$ and $\mathbf{1}_A$ is integrable over $\mathbb{R}$ (because $\lambda(A) < \infty$), from Riemann-Lebesgue Theorem

$$\lim_{n\to\infty}\int_A \cos(2nx)\mathrm{d}x = \lim_{n\to\infty}\mathrm{Re}\left(\int_{\mathbb{R}}\mathbf{e}^{-2inx}\mathbf{1}_A(x)\mathrm{d}x\right) = \lim_{n\to\infty}\mathrm{Re}\left(\mathcal{F}(\mathbf{1}_A)(2n)\right) = 0$$

from which the result follows. ■

Actually, assuming further decay on f at infinity, we can prove that $\mathcal{F}(f)$ is even more regular. Namely, one has the following

Theorem 5.1.3 Let $f \in L^1(\mathbb{R}^d, \mathbb{C})$ be given and let $\mathcal{F}(f)$ denote its Fourier transform.

1. Assume that $x \mapsto \|x\| f(x)$ belongs to $L^1(\mathbb{R}^d, \mathbb{C})$, i.e.

$$\int_{\mathbb{R}^d} \|x\| |f(x)| \mathrm{d}x < \infty$$

then $\mathcal{F}(f)$ is continuously differentiable on $\mathbb{R}^d$ and

$$\frac{\partial}{\partial \xi_k} \mathcal{F}(f)(\xi) = -i \mathcal{F}(x_k f)(\xi) \qquad \forall \xi \in \mathbb{R}^d, \ \ \forall k = 1, \ldots, d.$$

2. If $f \in \mathscr{C}^1(\mathbb{R}^d)$ with

$$\partial_k f := \frac{\partial f}{\partial x_k} \in L^1(\mathbb{R}^d, \mathbb{C}) \qquad \forall k = 1, \ldots, d$$

then

$$\mathcal{F}(\partial_k f) = i \xi_k \mathcal{F}(f)(\xi) \qquad \forall \xi \in \mathbb{R}^d, \ \ \forall k = 1, \ldots, d.$$

Proof. Let $Z \subset \mathbb{R}^d$ be a measurable set with zero Lebesgue measure such that $|f(x)| < \infty$ for any $x \in \mathbb{R}^d \setminus Z$. We define the function $\Phi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$ by

$$\Phi(\xi, x) := \mathbf{e}^{-i\xi \cdot x} f(x) \qquad \forall (\xi, x) \in \mathbb{R}^d \times \mathbb{R}^d.$$

Then, for any $x \in \mathbb{R}^d \setminus Z$, one has $\Phi(\cdot, x) \in \mathscr{C}^1(\mathbb{R}^d, \mathbb{C})$ [1] while, for any $\xi \in \mathbb{R}^d$, $\Phi(\xi, \cdot) \in L^1(\mathbb{R}^d, \mathbb{C})$. Moreover,

$$\left| \frac{\partial}{\partial \xi_k} \Phi(\xi, x) \right| = |-i x_k \Phi(\xi, x)| \leqslant |x_k| \, |f(x)| \leqslant \|x\| \, |f(x)|$$

so that $\dfrac{\partial}{\partial \xi_k} \Phi(\cdot, x)$ is dominated over $\mathbb{R}^d$ by a L^1-function. Using the general result about the differentiability of integrals depending on a parameter (see **Chapter 4**) we get that the mapping $\mathcal{F}(f) : \xi \mapsto \int_{\mathbb{R}^d} \Phi(\xi, x) \mathrm{d}x$ belongs to $\mathscr{C}^1(\mathbb{R}^d)$ with

$$\frac{\partial}{\partial \xi_k} \mathcal{F}(f)(\xi) = \int_{\mathbb{R}^d} \frac{\partial}{\partial \xi_k} \Phi(\xi, x) \mathrm{d}x = -i \int_{\mathbb{R}^d} x_k f(x) \mathbf{e}^{-i\xi \cdot x} \mathrm{d}x \qquad \forall k = 1, \ldots, d.$$

This proves the first point. To simplify the notations, for the second point, we shall only consider the derivative with respect to the first variable and, for any $x \in \mathbb{R}^d$, we write $x = (x_1, y)$ with $y \in \mathbb{R}^{d-1}$ and $\xi = (\xi_1, \zeta)$. Fix $a < b$ and consider the integral

$$\int_a^b \mathbf{e}^{-i\xi \cdot x} \frac{\partial}{\partial x_1} f(x) \mathrm{d}x_1$$

Integration by part yields

$$\int_a^b \mathbf{e}^{-i\xi \cdot x} \frac{\partial}{\partial x_1} f(x) \mathrm{d}x_1 = \mathbf{e}^{-i\xi \cdot x} f(x) \Big|_{x_1 = a}^{x_1 = b} + i \xi_1 \int_a^b \mathbf{e}^{-i\xi \cdot x} f(x) \mathrm{d}x_1.$$

[1] Here, as usual for a function of two variables $\Phi(\cdot, x)$ is a shorthand notation for, given x fixed, the mapping $\xi \mapsto \Phi(\xi, x)$.

Since both f and $\frac{\partial}{\partial x_1} f$ are integrable, one can invoke Fubini's theorem to prove that

$$\int_{\mathbb{R}^{d-1}} \left(\int_a^b \mathbf{e}^{-i\xi\cdot x} \frac{\partial}{\partial x_1} f(x) \mathrm{d}x_1 \right) \mathrm{d}y = \int_{\mathbb{R}^d} \mathbf{e}^{-i\xi\cdot x} \frac{\partial}{\partial x_1} f(x) \mathbf{1}_{[a,b]}(x_1) \mathrm{d}x$$

and

$$\int_{\mathbb{R}^{d-1}} \left(\int_a^b \mathbf{e}^{-i\xi\cdot x} f(x) \mathrm{d}x_1 \right) \mathrm{d}y = \int_{\mathbb{R}^d} \mathbf{e}^{-i\xi\cdot x} f(x) \mathbf{1}_{[a,b]}(x_1) \mathrm{d}x$$

so that

$$\int_{\mathbb{R}^d} \mathbf{e}^{-i\xi\cdot x} \frac{\partial}{\partial x_1} f(x) \mathbf{1}_{[a,b]}(x_1) \mathrm{d}x = i\xi_1 \int_{\mathbb{R}^d} \mathbf{e}^{-i\xi\cdot x} f(x) \mathbf{1}_{[a,b]}(x_1) \mathrm{d}x + \int_{\mathbb{R}^{d-1}} \mathbf{e}^{-i\zeta\cdot y} \left(\mathbf{e}^{-i\xi_1 b} f(b,y) - \mathbf{e}^{-i\xi_1 a} f(a,y) \right) \mathrm{d}y. \quad (5.3)$$

Since $\mathbf{1}_{[a,b]}(x_1)$ converges to 1 as $a \to -\infty$, $b \to \infty$ and both f, $\frac{\partial}{\partial x_1} f$ are L^1 functions, we have, from the dominated convergence theorem that

$$\lim_{\substack{a\to-\infty \\ b\to\infty}} \int_{\mathbb{R}^{d-1}} \left(\int_a^b \mathbf{e}^{-i\xi\cdot x} \frac{\partial}{\partial x_1} f(x) \mathrm{d}x_1 \right) \mathrm{d}y = \int_{\mathbb{R}^d} \mathbf{e}^{-i\xi\cdot x} \frac{\partial}{\partial x_1} f(x) \mathrm{d}x$$

and

$$\lim_{\substack{a\to-\infty \\ b\to\infty}} \int_{\mathbb{R}^{d-1}} \left(\int_a^b \mathbf{e}^{-i\xi\cdot x} f(x) \mathrm{d}x_1 \right) \mathrm{d}y = \int_{\mathbb{R}^d} \mathbf{e}^{-i\xi\cdot x} f(x) \mathrm{d}x$$

Since the mapping

$$x_1 \mapsto \int_{\mathbb{R}^d} \mathbf{e}^{-i\xi\cdot x} f(x_1, y) \mathrm{d}y$$

is integrable, there is $(a_n)_n$ converging to $-\infty$ and $(b_n)_n$ converging to $+\infty$ such that

$$\lim_n \int_{\mathbb{R}^d} \mathbf{e}^{-i\xi\cdot x} f(a_n, y) \mathrm{d}y = \lim_{n\to\infty} \int_{\mathbb{R}^d} \mathbf{e}^{-i\xi\cdot x} f(b_n, y) \mathrm{d}y = 0$$

and we see then that, applying (5.3) with $a = a_n$ and $b = b_n$ that, at the limit,

$$\int_{\mathbb{R}^d} \mathbf{e}^{-i\xi\cdot x} \frac{\partial}{\partial x_1} f(x) \mathrm{d}x = i\xi_1 \int_{\mathbb{R}^d} \mathbf{e}^{-i\xi\cdot x} f(x) \mathrm{d}x$$

which proves the result. ■

Remark 5.1.1 (f **differentiable** $\implies$ $\widehat{f}$ **decays**) From the above results, if $f \in \mathscr{C}^1(\mathbb{R}^d) \cap L^1(\mathbb{R}^d, \mathbb{C})$ and $\partial_k f \in L^1(\mathbb{R}^d, \mathbb{C})$ for all $k = 1, \ldots, d$, then

$$\left| \widehat{f}(\xi) \right| \leqslant \frac{1}{1 + \|\xi\|} \left(\|f\|_1 + \sum_{k=1}^d \|\partial_k f\|_1 \right) \qquad \forall \xi \in \mathbb{R}^d$$

Indeed, since $\|\xi\| \leqslant \sum_{k=1}^d |\xi_k|$, it holds

$$(1 + \|\xi\|)|\widehat{f}(\xi)| + \sum_{k=1}^d |\xi_k \widehat{f}(\xi)| = |\widehat{f}(\xi)| + \sum_{k=1}^d |\widehat{\partial_k f}(\xi)|$$

and the result follows since $|\widehat{f}(\xi)| \leqslant \|f\|_1$ and $|\widehat{\partial_k f}(\xi)| \leqslant \|\partial_k f\|_1$.

Of course, the above theorem can be generalized to any order. We introduce the following notations: for any $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$ we denote

$$|\boldsymbol{\alpha}| = \alpha_1 + \ldots + \alpha_d, \qquad \boldsymbol{\partial}^{\boldsymbol{\alpha}} = \frac{\partial^{|\boldsymbol{\alpha}|}}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}$$

and

$$\boldsymbol{\xi}^{\boldsymbol{\alpha}} = \xi_1^{\alpha_1} \cdots \xi_d^{\alpha_d} = \prod_{j=1}^d \xi_j^{\alpha_j} \in \mathbb{R} \qquad \forall \xi = (\xi_1, \ldots, \xi_d).$$

This means for example that, if $f : \mathbb{R}^3 \to \mathbb{C}$ is given then

$$\frac{\partial^5 f}{\partial x_1^2 \partial x_2 \partial x_3^2}(\boldsymbol{x}) = \boldsymbol{\partial}^{\boldsymbol{\alpha}} f(\boldsymbol{x}) \quad \text{with} \qquad \boldsymbol{\alpha} = (2,1,2), \quad |\boldsymbol{\alpha}| = 5.$$

In the same way $\dfrac{\partial^4 f}{\partial x_2^3 \partial x_3}(\boldsymbol{x}) = \boldsymbol{\partial}^{\boldsymbol{\alpha}} f(\boldsymbol{x})$ with now $\boldsymbol{\alpha} = (0,3,1)$ and $|\boldsymbol{\alpha}| = 4$.

Then, a simple induction argument shows that the previous theorem can be generalized to get

Corollary 5.1.4 Let $f \in L^1(\mathbb{R}^d, \mathbb{C})$ be given and let $\mathcal{F}(f)$ denote its Fourier transform. Let $k \in \mathbb{N}$ be given.

1. Assume that $x \mapsto \|x\|^k f(x)$ belongs to $L^1(\mathbb{R}^d, \mathbb{C})$ then $\mathcal{F}(f) \in \mathscr{C}^k(\mathbb{R}^d)$ and
$$\boldsymbol{\partial}^{\boldsymbol{\alpha}} \mathcal{F}(f)(\xi) = (-i)^k \mathcal{F}(\boldsymbol{x}^{\boldsymbol{\alpha}} f)(\xi) \qquad \forall \xi \in \mathbb{R}^d, \ \ \boldsymbol{\alpha} \in \mathbb{N}^d \text{ with } |\boldsymbol{\alpha}| = k.$$
2. If $f \in \mathscr{C}^k(\mathbb{R}^d)$ with
$$\boldsymbol{\partial}^{\boldsymbol{\alpha}} f \in L^1(\mathbb{R}^d, \mathbb{C}) \qquad \forall \boldsymbol{\alpha} \in \mathbb{N}^d \quad \text{with} \quad |\boldsymbol{\alpha}| = k$$
then
$$\mathcal{F}\left(\boldsymbol{\partial}^{\boldsymbol{\alpha}} f\right) = i^k \boldsymbol{\xi}^{\boldsymbol{\alpha}} \mathcal{F}(f)(\xi) \qquad \forall \xi \in \mathbb{R}^d.$$

Remark 5.1.2 As before, one can deduce that, if $f \in \mathscr{C}^k(\mathbb{R}^d)$ with $\boldsymbol{\partial}^{\boldsymbol{\alpha}} f \in L^1(\mathbb{R}^d, \mathbb{C})$ as soon as $|\boldsymbol{\alpha}| = k$, then there is $C > 0$ (depending on $\|f\|_1$ and $\|\boldsymbol{\partial}^{\boldsymbol{\alpha}} f\|_1$) such that

$$|\widehat{f}(\xi)| \leqslant \frac{C}{(1 + \|\xi\|)^k} \qquad \forall \xi \in \mathbb{R}^d.$$

■ **Example 5.2** Define the Laplace operator Δ which, for a given $f \in \mathscr{C}^2(\mathbb{R}^d)$, is defined as [2]

$$\Delta f(x) = \sum_{i=1}^d \frac{\partial^2}{\partial x_i^2} f(x)$$

If $f \in \mathscr{C}^2(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ is such that $\Delta f \in L^1(\mathbb{R}^d)$, then one checks easily that

$$\mathcal{F}(\Delta f)(\xi) = -\|\xi\|^2 \mathcal{F}(f)(\xi) \qquad \forall \xi \in \mathbb{R}^d.$$

[2] Notice that, if $\mathrm{Hess}(f)$ denotes the Hessian matrix of f whose entries are $\left(\frac{\partial^2}{\partial x_i \partial x_j} f\right)_{i,j}$ then Δf is exactly the trace of $\mathrm{Hess}(f)$.

■

■ **Example 5.3** Assume $d = 1$ and let

$$f(x) = \frac{1}{1+x^2} \qquad \forall x \in \mathbb{R}.$$

Notice that $f \in L^1(\mathbb{R})$. Moreover, for any $m \in \mathbb{N}$, the derivative $\partial_m f = \frac{\mathrm{d}^m}{\mathrm{d}x^m} f$ is such that

$$\partial_m f(x) = \sum_{j=[m/2]}^{m} c_{m,j} \frac{x^{2j-m}}{(1+x^2)^{j+1}} \qquad x \in \mathbb{R}$$

where $c_{m,j} \in \mathbb{R}$ are given coefficients and $[m/2]$ denote the integral part of $m/2$ ***(Check this)***. Thus

$$\partial_m f \in L^1(\mathbb{R}) \qquad \forall m \in \mathbb{N}$$

and therefore, for any $N \in \mathbb{N}$, there is $C_N > 0$ such that

$$|\widehat{f}(\xi)| \leqslant \frac{C_N}{(1+|\xi|)^N} \qquad \forall \xi \in \mathbb{R}.$$

Actually, one can prove (with method coming from complex analysis) that

$$\widehat{f}(\xi) = \pi \exp(-|\xi|) \qquad \forall \xi \in \mathbb{R}.$$

■

■ **Example 5.4** Assume still $d = 1$ and let

$$f(x) = \exp(-|x|), \qquad x \in \mathbb{R}.$$

Then, for any $k \in \mathbb{N}$, $x \mapsto |x|^k f(x)$ is integrable ***(Check this)***. Therefore, for any $k \in \mathbb{N}$, $\widehat{f} \in \mathscr{C}^k(\mathbb{R})$. Notice that, for all $\xi \in \mathbb{R}$,

$$\begin{aligned}\widehat{f}(\xi) = \int_{\mathbb{R}} \mathbf{e}^{-i\xi\, x} \exp(-|x|)\mathrm{d}x &= \int_{-\infty}^{0} \mathbf{e}^{-i\xi\, x + x}\mathrm{d}x + \int_{0}^{\infty} \mathbf{e}^{-i\xi\, x - x}\mathrm{d}x \\ &= \int_{-\infty}^{0} \mathbf{e}^{x(1-i\xi)}\mathrm{d}x + \int_{0}^{\infty} \mathbf{e}^{-x(1+\xi)}\mathrm{d}x\end{aligned}$$

so that [3]

$$\widehat{f}(\xi) = \frac{1}{1-i\xi} + \frac{1}{1+i\xi} = \frac{(1-i\xi)+(1+i\xi)}{(1-i\xi)(1+i\xi)} = \frac{2}{1+\xi^2} \qquad \forall \xi \in \mathbb{R}.$$

■

[3] Notice that here we compute the integral of complex exponential as we usually do for real exponential. This is licit since actually the complex exponential is such that, for any $\alpha \in \mathbb{R}$ and $z \in \mathbb{C}$, $\frac{\mathrm{d}}{\mathrm{d}\alpha}\mathbf{e}^{\alpha\, z} = \alpha \mathbf{e}^{z}$. This allows to see $\mathbf{e}^{x(1-i\xi)}$ as the primitive of $\frac{1}{1-i\xi}\frac{\mathrm{d}}{\mathrm{d}x}\mathbf{e}^{x(1-i\xi)}$.

5.1.2 Example of computations

For any $a > 0$, we call the Gaussian density over $\mathbb{R}^d$ with covariance matrix $\mathbf{I}_N$ the mapping

$$\mathbb{G}_a(x) = \frac{1}{(2\pi a)^{d/2}} \exp\left(-\frac{\|x\|^2}{2a}\right) \qquad \forall x \in \mathbb{R}^d.$$

Notice that $\mathbb{G}_a \in \mathscr{C}^1(\mathbb{R}^d)$ and

$$\int_{\mathbb{R}^d} \mathbb{G}_a(x) \mathrm{d}x = 1 \qquad \forall a > 0.$$

Proposition 5.1.5 — Gaussian Fourier transform. For any $a > 0$, one has

$$\mathcal{F}(\mathbb{G}_a)(\xi) = \exp\left(-a\|\xi\|^2/2\right) \qquad \forall \xi \in \mathbb{R}^d$$

i.exs.

$$\mathcal{F}(\mathbb{G}_a) = \left(\frac{2\pi}{a}\right)^{N/2} \mathbb{G}_{1/a}(\xi).$$

Proof. We begin with the case $d = 1$. For any $a > 0$, $\mathbb{G}_a \in \mathscr{C}^1(\mathbb{R})$ and satisfies

$$a\mathbb{G}_a'(x) + x\mathbb{G}_a(x) = 0 \qquad \forall x \in \mathbb{R}. \tag{5.4}$$

Set $h_a(x) = x\mathbb{G}_a(x)$. Since $\mathbb{G}_a'$, h_a and $\mathbb{G}_a$ are all integrable, one deduces fro the previous theorem that $\mathcal{F}(\mathbb{G}_a) \in \mathscr{C}^1(\mathbb{R})$ and

$$\mathcal{F}(\mathbb{G}_a')(\xi) = i\xi\mathcal{F}(\mathbb{G}_a)(\xi) \qquad \text{and} \qquad \mathcal{F}(\mathbb{G}_a)'(\xi) = -i\mathcal{F}(h_a)(\xi).$$

Notice that this last identity also reads as $\mathcal{F}(h_a) = i\mathcal{F}(\mathbb{G}_a)'$. One deduces from this and (5.4) that

$$a\xi\mathcal{F}(\mathbb{G}_a)(\xi) + \mathcal{F}(h_a)(\xi) = 0.$$

The Fourier transform $f = \mathcal{F}(\mathbb{G}_a)$ satisfies then the ODE

$$f'(\xi) = -a\xi f(\xi)$$

and a general solution is $f(\xi) = f(0)\exp(-a\xi^2/2)$. Since moreover

$$f(0) = \mathcal{F}(\mathbb{G}_a)(0) = \int_{\mathbb{R}} \mathbb{G}_a(x)dx = 1$$

we see that

$$\mathcal{F}(\mathbb{G}_a)(\xi) = \exp\left(-a\xi^2/2\right) \qquad \forall \xi \in \mathbb{R}.$$

This proves the result for $d = 1$. For $d > 1$, one denotes simply the $1D$-gaussian

$$g_a(x_k) = \frac{1}{\sqrt{2\pi a}} \exp\left(ax_k^2/2a\right) \qquad \forall x_k \in \mathbb{R}$$

and

$$\mathbb{G}_a(x) = \prod_{k=1}^{d} g_a(x_k) \qquad \forall x = (x_1, \ldots, x_N) \in \mathbb{R}^d.$$

Then, thanks to Fubini's theorem, we get

$$\begin{aligned}\mathcal{F}(\mathbb{G}_a)(\xi) &= \int_{\mathbb{R}^d} \prod_{k=1}^{d} \mathbf{e}^{-i\xi_k x_k} g_a(x_k)\mathrm{d}x \\ &= \prod_{k=1}^{d} \int_{\mathbb{R}} \mathbf{e}^{-i\xi_k x_k} g_a(x_k)\mathrm{d}x_k = \prod_{k=1}^{d} \mathcal{F}(g_a)(\xi_k)\end{aligned}$$

where $\mathcal{F}(g_a)$ is the $1D$ Fourier transform of g_a, which, from the first part of the proof, is

$$\mathcal{F}(g_a)(\xi_k) = \exp\left(-a\xi_k^2/2\right)$$

from which we get the result. ■

5.1.3 Inversion formula

We establish now the fact that the Fourier transform of f *completely characterizes* f:

Theorem 5.1.6 Assume that $f \in L^1(\mathbb{R}^d, \mathbb{C})$ is such that $\widehat{f} = \mathcal{F}(f) \in L^1(\mathbb{R}^d, \mathbb{C})$. Then, for almost every $x \in \mathbb{R}^d$ it holds

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathbf{e}^{i\xi\cdot x} \widehat{f}(\xi)\mathrm{d}\xi.$$

Before giving the rigorous proof of the result, let us begin with some heuristic ideas: the inversion formula we aim to establish is

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathbf{e}^{i\xi\cdot x} \left(\int_{\mathbb{R}^d} \mathbf{e}^{-i\xi\cdot y} f(y)\mathrm{d}y\right) \mathrm{d}\xi.$$

The first natural idea would be to exchange the order of integration between y and ξ which would yields

$$f(x) = \int_{\mathbb{R}^d} f(y) \left(\frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathbf{e}^{i\xi\cdot(x-y)}\mathrm{d}\xi\right) \mathrm{d}y$$

or, in other words

$$f(x) = \int_{\mathbb{R}^d} K(x-y)f(y)\mathrm{d}y = (f * K)(x)$$

with

$$K(z) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathbf{e}^{i\xi\cdot z}\mathrm{d}\xi \qquad z \in \mathbb{R}^d.$$

Unfortunately, that reasoning does not lead anywhere ! First, the expression of K is not explicit and, actually, in the Lebesgue integral framework, it does not make any sense !!! Indeed, for any $z \in \mathbb{R}^d$, the mapping $\Phi_z \,:\, \xi \mapsto \mathbf{e}^{i\xi\cdot z}$ is not integrable since

$$|\Phi_z(\xi)| = 1$$

so that

$$\int_{\mathbb{R}^d} |\Phi_z(\xi)|\mathrm{d}\xi = \int_{\mathbb{R}^d} \mathrm{d}\xi = \infty.$$

Second, the exchange of order of integration here above cannot be justified by Fubini's theorem. Indeed, for given $x \in \mathbb{R}^d$, the function

$$(y, \xi) \mapsto \mathbf{e}^{i\xi\cdot(x-y)} f(y)$$

is not integrable over $\mathbb{R}^d \times \mathbb{R}^d$ since

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} |\mathbf{e}^{i\xi\cdot(x-y)} f(y)| \mathrm{d}\xi \mathrm{d}y = \int_{\mathbb{R}^d\times\mathbb{R}^d} |f(y)| \mathrm{d}\xi \mathrm{d}y = \|f\|_1 \int_{\mathbb{R}^d} \mathrm{d}\xi = +\infty$$

unless $f(y) = 0$ for almost every $y \in \mathbb{R}^d$.

At first sight therefore, this approach is not applicable. However, everything will be made rigorous if we modify a bit the function K as follows: let

$$K_\varepsilon(z) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathbf{e}^{-i\xi\cdot z} \exp\left(-\frac{1}{2}\varepsilon^2\|\xi\|^2\right) \mathrm{d}\xi \qquad \forall z \in \mathbb{R}^d \tag{5.5}$$

which, for $\varepsilon > 0$ sufficiently small is expected to behave like $K(z)$.

Let us give a rigorous proof of the result. We begin with the following

Lemma 5.1.7 Given $\varepsilon > 0$, let K_ε be defined by (5.5). Then, for all $f \in L^1(\mathbb{R}^d, \mathbb{C})$ it holds

$$\lim_{\varepsilon\to 0^+} \|K_\varepsilon * f - f\|_1 = 0. \tag{5.6}$$

Proof. Given $\varepsilon > 0$, from Proposition 5.1.5, since

$$\exp\left(-\frac{1}{2}\varepsilon^2\|\xi\|^2\right) = \left(\frac{2\pi}{\varepsilon^2}\right)^{d/2} \mathbb{G}_{1/\varepsilon^2}(\xi)$$

one has

$$K_\varepsilon(z) = \frac{1}{(2\pi\varepsilon^2)^{d/2}} \mathcal{F}(\mathbb{G}_{1/\varepsilon^2})(z) = \mathbb{G}_{\varepsilon^2}(z) \qquad \forall z \in \mathbb{R}^d. \tag{5.7}$$

In particular, $K_\varepsilon \in L^1(\mathbb{R}^d, \mathbb{C})$ so that $K_\varepsilon * f$ is well defined and belongs to $L^1(\mathbb{R}^d, \mathbb{C})$ according to Young's convolution inequality. Let us prove (5.6). Set

$$I_\varepsilon := \int_{\mathbb{R}^d} K_\varepsilon(x-y) f(y) \mathrm{d}y = \int_{\mathbb{R}^d} \mathbb{G}_{\varepsilon^2}(x-y) f(y) \mathrm{d}y = \int_{\mathbb{R}^d} \mathbb{G}_{\varepsilon^2}(y) f(x-y) \mathrm{d}y$$

i.e.

$$I_\varepsilon = \frac{1}{(2\pi\varepsilon^2)^{d/2}} \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2\varepsilon^2}\|y\|^2\right) f(x-y) \mathrm{d}y.$$

Setting $z = \frac{y}{\varepsilon}$, $\mathrm{d}z = \frac{1}{\varepsilon^d}\mathrm{d}y$ so that

$$I_\varepsilon = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}\|z\|^2\right) f(x-\varepsilon z) \mathrm{d}z = \int_{\mathbb{R}^d} \mathbb{G}_1(z) f(x-\varepsilon z) \mathrm{d}z.$$

Since $\int_{\mathbb{R}^d} \mathbb{G}_1(z) \mathrm{d}z = 1$ one has

$$I_\varepsilon - f(x) = \int_{\mathbb{R}^d} \mathbb{G}_1(z) \left(f(x-\varepsilon z) - f(x)\right) \mathrm{d}z$$

so that

$$\int_{\mathbb{R}^d}\left|\int_{\mathbb{R}^d} K_\varepsilon(x-y)f(y)\mathrm{d}y - f(x)\right|\mathrm{d}x \leqslant \int_{\mathbb{R}^d}\mathrm{d}x\int_{\mathbb{R}^d}\mathbb{G}_1(z)\,|f(x-\varepsilon z)-f(x)|\,\mathrm{d}z$$
$$= \int_{\mathbb{R}^d}\mathbb{G}_1(z)\left(\int_{\mathbb{R}^d}|f(x-\varepsilon z)-f(x)|\mathrm{d}x\right)\mathrm{d}x.$$

Moreover, for almost every $z\in\mathbb{R}^d$

$$\lim_{\varepsilon\to 0^+}\int_{\mathbb{R}^d}|f(x-\varepsilon z)-f(x)|\mathrm{d}x = 0$$

and

$$\sup_{\varepsilon>0}\int_{\mathbb{R}^d}|f(x-\varepsilon z)-f(x)|\mathrm{d}x \leqslant 2\|f\|_1$$

so that, according to the dominated convergence theorem

$$\lim_{\varepsilon\to 0^+}\int_{\mathbb{R}^d}\mathbb{G}_1(z)\left(\int_{\mathbb{R}^d}|f(x-\varepsilon z)-f(x)|\mathrm{d}x\right)\mathrm{d}x = 0$$

which is the desired result. ■

✍ *Proof of Theorem 5.1.6.* According to the previous lemma, since the convergence occurs in $L^1(\mathbb{R}^d,\mathbb{C})$, we can assume that

$$\lim_{\varepsilon\to 0^+} K_\varepsilon * f(x) = f(x) \qquad \text{for a.e. } x\in\mathbb{R}^d. \tag{5.8}$$

The proof consists then in expressing $K_\varepsilon * f(x)$ in order to prove that

$$\lim_{\varepsilon\to 0^+} K_\varepsilon * f(x) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\mathbf{e}^{i\xi\cdot x}\widehat{f}(\xi)\mathrm{d}\xi \tag{5.9}$$

and invoke the uniqueness of the limit to obtain the inversion formula. Recall that $K_\varepsilon\in L^1(\mathbb{R}^d,\mathbb{C})$ for any $\varepsilon>0$ and, from (5.7), one has $K_\varepsilon(z)=K_\varepsilon(-z)$ $\forall z\in\mathbb{R}^d$. Now, for any $\varepsilon>0$, the use of Fubini's theorem becomes now licit to get, for given $x\in\mathbb{R}^d$

$$\int_{\mathbb{R}^d}K_\varepsilon(x-y)f(y)\mathrm{d}y = \int_{\mathbb{R}^d}f(y)\left(\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\mathbf{e}^{i\xi\cdot(x-y)}\exp\left(-\frac{1}{2}\varepsilon^2\|\xi\|^2\right)\mathrm{d}\xi\right)\mathrm{d}y$$
$$= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\mathbf{e}^{i\xi\cdot x}\exp\left(-\frac{1}{2}\varepsilon^2\|\xi\|^2\right)\left(\int_{\mathbb{R}^d}\mathbf{e}^{-i\xi\cdot y}f(y)\mathrm{d}y\right)\mathrm{d}\xi$$
$$= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\mathbf{e}^{i\xi\cdot x}\exp\left(-\frac{1}{2}\varepsilon^2\|\xi\|^2\right)\widehat{f}(\xi)\mathrm{d}\xi \tag{5.10}$$

where now the exchange of the integral order is licit since the mapping

$$(y,\xi)\mapsto \mathbf{e}^{i\xi\cdot(x-y)}\exp\left(-\frac{1}{2}\varepsilon^2\|\xi\|^2\right)f(y)$$

is integrable over $\mathbb{R}^d\times\mathbb{R}^d$ ***(Check this)***.

Let us now pass to the limit as $\varepsilon \to 0^+$ in the last identity. For any $x \in \mathbb{R}^d$ and almost every $\xi \in \mathbb{R}^d$ one has

$$\lim_{\varepsilon \to 0^+} \mathbf{e}^{i\xi \cdot x} \exp\left(-\frac{1}{2}\varepsilon^2 \|\xi\|^2\right) \widehat{f}(\xi) = \mathbf{e}^{i\xi \cdot x} \widehat{f}(\xi)$$

and

$$\sup_{\varepsilon > 0} \left| \mathbf{e}^{i\xi \cdot x} \exp\left(-\frac{1}{2}\varepsilon^2 \|\xi\|^2\right) \widehat{f}(\xi) \right| \leqslant |\widehat{f}(\xi)| \in L^1(\mathbb{R}^d, \mathrm{d}\xi)$$

so that, according to the dominated convergence theorem, we obtain

$$\lim_{\varepsilon \to 0^+} \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathbf{e}^{i\xi \cdot x} \exp\left(-\frac{1}{2}\varepsilon^2 \|\xi\|^2\right) \widehat{f}(\xi) \mathrm{d}\xi = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathbf{e}^{i\xi \cdot x} \widehat{f}(\xi) \mathrm{d}\xi \qquad \forall x \in \mathbb{R}^d.$$

which is exactly (5.9). Combining (5.9) with (5.8) gives the inversion formula. ∎

Remark 5.1.3 The choice of K_ε was not the unique possible choice in the previous proof. In particular, it is instructive in dimension $d = 1$ to consider for instance

$$H_\varepsilon(z) = \int_{\mathbb{R}} \exp(-\varepsilon\,|\xi|) \mathbf{e}^{i\xi\, z} \mathrm{d}\xi, \qquad \forall z \in \mathbb{R}.$$

Then, using Example 5.3, one can prove that

$$H_\varepsilon(z) = \sqrt{\frac{2}{\pi}} \frac{\varepsilon}{\varepsilon^2 + z^2} \qquad \forall z \in \mathbb{R}.$$

Moreover, one can prove as before that

$$\lim_{\varepsilon \to 0^+} \|H_\varepsilon * f - f\|_1 = 0 \qquad \forall f \in \mathbb{R}$$

while

$$\lim_{\varepsilon 0^+} H_\varepsilon * f(x) = \int_{\mathbb{R}} \widehat{f}(\xi) \mathbf{e}^{i\xi x} \mathrm{d}\xi \qquad \text{for a.e. } x \in \mathbb{R}.$$

This provides an alternative proof of the inversion formula.

The Fourier inversion formula has the following important consequence:

Corollary 5.1.8 If $f \in L^1(\mathbb{R}^d, \mathbb{C})$ is such that $\mathcal{F}(f)(\xi) = 0$ for any $\xi \in \mathbb{R}^d$ then $f(x) = 0$ for almost every $x \in \mathbb{R}^d$.

Remark 5.1.4 If $f, g \in L^1(\mathbb{R}^d, \mathbb{C})$ are such that $\mathcal{F}(f)(\xi) = \mathcal{F}(g)(\xi)$ for all $\xi \in \mathbb{R}^d$ then

$$\mathcal{F}(f - g)(\xi) = 0 \qquad \forall \xi \in \mathbb{R}^d$$

and the previous corollary asserts that $f = g$ almost everywhere (i.e. f, g are the same $L^1(\mathbb{R}^d, \mathbb{C})$ "function"). This exactly means that the Fourier transform of f characterizes f.

Remark 5.1.5 Let us comment a bit the limit (5.6). Recall that, for any $\varepsilon > 0$,

$$K_\varepsilon(z) = \mathbb{G}_{\varepsilon^2}(z)$$

so, up to set now $\delta = \varepsilon$ one has

$$\lim_{\delta\to 0} \|\mathbb{G}_\delta * f - f\|_1 = 0 \tag{5.11}$$

where $\mathbb{G}_\delta$ is the Gaussian probability density:

$$\mathbb{G}_\delta(x) = \frac{1}{(2\pi\delta)^{d/2}} \exp\left(-\frac{\|x\|^2}{2\delta}\right) \qquad \forall x \in \mathbb{R}^d.$$

In particular, for $\delta = 2t$, $t > 0$, setting $G_t = \mathbb{G}_{2t}$ the Heat Kernel defined in **Chapter 4**, EXERCISE 3.10 and

$$u_t = G_t * f$$

one sees that

$$\lim_{t\to 0^+} \|u_t - f\|_1 = 0.$$

Remark 5.1.6 We go on with the previous remark and still denote, for $f \in L^1(\mathbb{R}^d, \mathbb{R})$

$$u_t(x) = G_t * f(x) = \mathbb{G}_{2t} * f(x), \qquad \forall t \geqslant 0.$$

Assume that $f \in \mathscr{C}^2_c(\mathbb{R}^d)$. We know that then $u_t \in \mathscr{C}^2(\mathbb{R}^d)$. We shall compute the Laplace operator of u_t

$$\Delta u_t(x) = \sum_{i=1}^d \frac{\partial^2}{\partial x_i^2} u_t(x).$$

If we assume that $\Delta u_t \in L^1(\mathbb{R}^d)$ and compute its Fourier transform, as already seen, one has

$$\mathcal{F}(\Delta u_t)(\xi) = -\|\xi\|^2 \mathcal{F}(u_t)(\xi) \qquad \forall \xi \in \mathbb{R}^d$$

Moreover,

$$\mathcal{F}(u_t)(\xi) = \mathcal{F}(\mathbb{G}_{2t})(\xi)\mathcal{F}(f)(\xi) = \exp(-t\|\xi\|^2)\mathcal{F}(f)(\xi).$$

In particular, one sees that the mapping $t \mapsto \mathcal{F}(u_t)(\xi)$ is differentiable for any ξ and

$$\frac{\partial}{\partial t}\mathcal{F}(u_t)(\xi) = -\|\xi\|^2 \mathcal{F}(f)(\xi).$$

Consequently,

$$\frac{\partial}{\partial t}\mathcal{F}(u_t)(\xi) = \mathcal{F}(\Delta u_t)(\xi) \qquad \forall \xi \in \mathbb{R}^d.$$

Since clearly $\frac{\partial}{\partial t}\mathcal{F}(u_t) = \mathcal{F}(\frac{\partial}{\partial t} u_t)$ one sees that

$$\mathcal{F}\left(\frac{\partial}{\partial t} u_t\right) = \mathcal{F}(\Delta u_t)$$

and therefore, by injectivity

$$\frac{\partial}{\partial t} u_t(x) = \Delta u_t(x) \qquad \forall t > 0, \qquad \text{a.e.} x \in \mathbb{R}^d.$$

Since moreover $\lim_{t\to 0^+} u_t = f$ (in L^1), we see that u_t solves the *Heat equation*

$$\begin{cases} \frac{\partial}{\partial t} u_t(x) = \Delta u_t(x) \\ u_0(x) = f(x) \end{cases}$$

5.2 Fourier-Plancherel transform in $L^2(\mathbb{R}^d)$

Introduce the space

$$\mathscr{C}_0(\mathbb{R}^d) = \{f \in \mathscr{C}(\mathbb{R}^d, \mathbb{C})\,;\ \lim_{\|y\|\to\infty} |f(y)| = 0\}$$

endowed with the norm $\|f\|_\infty = \sup_{y\in\mathbb{R}^d} |f(y)|$. Notice that $\|f\|_\infty < \infty$ for any $f \in \mathscr{C}_0(\mathbb{R}^d)$ ***(Explain why)***. It is easy to check that $(\mathscr{C}_0(\mathbb{R}^d), \|\cdot\|_\infty)$ is a Banach space. Define then the linear mapping

$$\mathcal{F}\ :\ f \in L^1(\mathbb{R}^d, \mathbb{C}) \mapsto \mathcal{F}(f) \in \mathscr{C}_0(\mathbb{R}^d).$$

Notice that $\mathcal{F}$ is well-defined since, by Riemann-Lebesgue theorem, $\mathcal{F}(f) \in \mathscr{C}_0(\mathbb{R}^d)$ for any $f \in L^1(\mathbb{R}^d)$. Moreover, since $\|\mathcal{F}(f)\|_\infty \leqslant \|f\|_1$ for any $f \in L^1(\mathbb{R}^d)$, it holds that

$$\mathcal{F} \in \mathscr{L}(L^1(\mathbb{R}^d, \mathbb{C})\,;\, \mathscr{C}_0(\mathbb{R}^d)).$$

Theorem 5.1.6 actually states that $\mathcal{F}$ is injective.

Using the density of $L^1(\mathbb{R}^d, \mathbb{C}) \cap L^2(\mathbb{R}^d, \mathbb{C})$ in $L^2(\mathbb{R}^d, \mathbb{C})$, one can extend the previous operator to define the Fourier transform in $L^2(\mathbb{R}^d, \mathbb{C})$. Notice that the integral

$$\int_{\mathbb{R}^d} \mathrm{e}^{-i\xi\cdot x} f(x)\mathrm{d}x$$

does not usually make sense if $f \in L^2(\mathbb{R}^d, \mathbb{C})$!! The definition of the Fourier transform in $L^2(\mathbb{R}^d, \mathbb{C})$ will therefore extend the one we gave in $L^1(\mathbb{R}^d, \mathbb{C})$ but will not coincide with it.

Observe that $L^2(\mathbb{R}^d, \mathbb{C})$ is a Hilbert space endowed with the *complex* inner product

$$\langle f, g\rangle = \int_{\mathbb{R}^d} \overline{g}(x)\, f(x)\mathrm{d}x.$$

This inner product is a *hermitian* bilinear application over $L^2(\mathbb{R}^d, \mathbb{C})$ since it satisfies the following:

1. $\langle f, g\rangle = \overline{\langle g, f\rangle}$, $f, g \in L^2(\mathbb{R}^d, \mathbb{C})$;
2. Linearity in the first argument:

$$\langle af, g\rangle = a\langle f, g\rangle \qquad \text{and} \qquad \langle f + g, h\rangle = \langle f, h\rangle + \langle f, h\rangle, \qquad f, g, h \in L^2(\mathbb{R}^d, \mathbb{C})$$

3. Positive-definiteness: given $f \in L^2(\mathbb{R}^d, \mathbb{C})$ one has

$$\langle f, f\rangle \geqslant 0 \qquad \text{and} \qquad \langle f, f\rangle = 0 \Rightarrow f = 0 \quad \text{a.e.}.$$

Check these properties.

We first extend the convergence established in (5.11) to $L^p(\mathbb{R}^d, \mathbb{C})$, $1 \leqslant p < \infty$:

Lemma 5.2.1 Given $f \in L^p(\mathbb{R}^d, \mathbb{C})$ $(1 \leqslant p < \infty)$ it holds

$$\lim_{\delta \to 0^+} \|\mathbb{G}_\delta * f - f\|_p = 0.$$

Proof. The proof mimics the one seen in Theorem 4.3.9 and uses Minkowski's integral inequality. Indeed, for a given $\delta > 0$, since $\int_{\mathbb{R}^d} \mathbb{G}_\delta(x)\mathrm{d}x = 1$ one has

$$\begin{aligned}
\|\mathbb{G}_\delta * f - f\|_p &\leqslant \left(\int_{\mathbb{R}^d} \left[\int_{\mathbb{R}^d} |f(x-y) - f(x)|\, \mathbb{G}_\delta(y)\mathrm{d}y \right]^p \mathrm{d}x \right)^{1/p} \\
&\leqslant \int_{\mathbb{R}^d} \mathbb{G}_\delta(y)| \left[\int_{\mathbb{R}^d} |f(x-y) - f(x)|^p \, \mathrm{d}x \right]^{\frac{1}{p}} \mathrm{d}y \leqslant \int_{\mathbb{R}^d} \mathbb{G}_\delta(y) \|\tau_y f - f\|_p \mathrm{d}y.
\end{aligned}$$

We know that, since $f \in L^p(\mathbb{R}^d, \mathbb{C})$

$$\lim_{y \to 0} \|\tau_y f - f\|_p = 0$$

i.e. for any $\varepsilon > 0$, there exists $\eta > 0$ such that $\|\tau_y f - f\|_p < \varepsilon$ whenever $\|y\| < \eta$. One has then

$$\int_{\mathbb{R}^d} \mathbb{G}_\delta(y) \|\tau_y f - f\|_p \mathrm{d}y \leqslant \varepsilon \int_{B(0,\eta)} \mathbb{G}_\delta(y)\mathrm{d}y + \int_{\|y\| \geqslant \eta} \mathbb{G}_\delta(y) \|\tau_y f - f\|_p \mathrm{d}y.$$

Since, for any $y \in \mathbb{R}^d$, $\|\tau_y f - f\|_p \leqslant \|\tau_y f\|_p + \|f\|_p = 2\|f\|_p$ one gets

$$\|\mathbb{G}_\delta * f - f\|_2 \leqslant \varepsilon \int_{B(0,\eta)} \mathbb{G}_\delta(y)\mathrm{d}y + 2\|f\|_p \int_{\|y\| \geqslant r} \mathbb{G}_\delta(y)\mathrm{d}y \leqslant \varepsilon + 2\|f\|_p \int_{\|y\| \geqslant \eta} \mathbb{G}_\delta(y)\mathrm{d}y.$$

If we are able to prove that

$$\lim_{\delta \to 0^+} \int_{\|y\| \geqslant \eta} \mathbb{G}_\delta(y)\mathrm{d}y = 0 \qquad \forall \eta > 0 \tag{5.12}$$

then we get the result. Let us now prove (5.12). Given $\eta > 0$, one has

$$A_\delta := \int_{\|y\| \geqslant \eta} \mathbb{G}_\delta(y)\mathrm{d}y = \frac{1}{(2\pi\delta)^{d/2}} \int_{\|y\| \geqslant \eta} \exp\left(-\frac{\|y\|^2}{2\delta} \right) \mathrm{d}y.$$

Setting $z = \frac{y}{\sqrt{\delta}}$, one has $\mathrm{d}z = \frac{1}{\delta^{d/2}}\mathrm{d}y$ and

$$A_\delta = \frac{1}{(2\pi)^{d/2}} \int_{\|z\| \geqslant \eta/\sqrt{\delta}} \exp(-\|z\|^2/2)\mathrm{d}z = \int_{\mathbb{R}^d} \mathbb{G}_1(z) \mathbf{1}_{\{\|z\| \geqslant \eta/\sqrt{\delta}\}} \mathrm{d}z.$$

A simple use of the dominated convergence theorem proves then that $\lim_{\delta \to 0} A_\delta = 0$. ■

The key ingredient to extend the Fourier transform to $L^2(\mathbb{R}^d, \mathbb{C})$ is the following result which applies to $f \in L^1(\mathbb{R}^d, \mathbb{C}) \cap L^2(\mathbb{R}^d, \mathbb{C})$:

Proposition 5.2.2 — Plancherel. Assume that $f \in L^1(\mathbb{R}^d, \mathbb{C}) \cap L^2(\mathbb{R}^d, \mathbb{C})$. Then its Fourier transform $\mathcal{F}(f)$ belongs to $L^2(\mathbb{R}^d, \mathbb{C})$ and

$$\|\mathcal{F}(f)\|_2 = (2\pi)^{d/2} \|f\|_2.$$

✍ *Proof.* Since $\mathcal{F}(f) \in L^\infty(\mathbb{R}^d, \mathbb{C})$, for any $\delta > 0$,

$$I_\delta = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} |\mathcal{F}(f)(\xi)|^2 \mathbf{e}^{-\delta\|\xi\|^2/2} \mathrm{d}\xi < \infty. \tag{5.13}$$

Moreover, the mapping

$$\Phi \;:\; (x, y, \xi) \in \mathbb{R}^d \mapsto \Phi(x, y, \xi) = \frac{1}{(2\pi)^d} f(y)\overline{f}(x)\mathbf{e}^{-\delta\|\xi\|^2/2}$$

belongs to $L^1(\mathbb{R}^{3d}, \mathbb{C})$ ***(Check this)*** so

$$J_\delta = \int_{\mathbb{R}^{3d}} \Phi(x, y, \xi)\mathbf{e}^{i\xi\cdot(x-y)} \mathrm{d}x\mathrm{d}y\mathrm{d}\xi$$

is well defined and one sees easily that, from Fubini's theorem

$$\begin{aligned}
J_\delta &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathbf{e}^{-\delta\|\xi\|^2/2} \left(\int_{\mathbb{R}^d} \overline{f}(x)\mathbf{e}^{i\xi\cdot x} \mathrm{d}x \right) \left(\int_{\mathbb{R}^d} f(y)\mathbf{e}^{-i\xi\cdot y} \mathrm{d}y \right) \mathrm{d}\xi \\
&= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathbf{e}^{-\delta\|\xi\|^2/2} \left(\mathcal{F}(\overline{f})(-\xi)\right) \left(\mathcal{F}(f)(\xi)\right. \mathrm{d}\xi \\
&= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathbf{e}^{-\delta\|\xi\|^2/2} \overline{\mathcal{F}(f)(\xi)}\mathcal{F}(f)(\xi) \mathrm{d}\xi
\end{aligned}$$

where we used Proposition 5.1.1. Therefore

$$I_\delta = J_\delta = \int_{\mathbb{R}^{3d}} \Phi(x, y, \xi) \mathrm{d}x\mathrm{d}y\mathrm{d}\xi.$$

Using again Fubini's theorem one also has

$$I_\delta = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \overline{f}(x) \left(\int_{\mathbb{R}^d} f(y) \left[\int_{\mathbb{R}^d} \mathbf{e}^{i\xi\cdot(x-y)}\mathbf{e}^{-\delta\|\xi\|^2/2} \mathrm{d}\xi \right] \mathrm{d}y \right) \mathrm{d}x$$

According to the inversion formula, the integral between the square brackets is $\mathbb{G}_\delta(x - y)$ (see also Proposition 5.1.5). Thus,

$$I_\delta = \int_{\mathbb{R}^d} \overline{f}(x) \left(\int_{\mathbb{R}^d} f(y)\mathbb{G}_\delta(x - y)\mathrm{d}y \right) \mathrm{d}x = \int_{\mathbb{R}^d} \overline{f}(x) \left(\mathbb{G}_\delta * f\right)(x)\mathrm{d}x.$$

Therefore,

$$I_\delta - \|f\|_2^2 = \int_{\mathbb{R}^d} \overline{f}(x) \left[\left(\mathbb{G}_\delta * f\right)(x) - f(x) \right] \mathrm{d}x$$

so that, according to Holder's inequality

$$\left| I_\delta - \|f\|_2^2 \right| \leqslant \|f\|_2 \|\mathbb{G}_\delta * f - f\|_2.$$

Using the fact that $\lim_{\delta\to 0^+} \|\mathbb{G}_\delta * f - f\|_2 = 0$ one gets that

$$\lim_{\delta\to 0^+} I_\delta = \|f\|_2^2.$$

In particular, from the definition (5.13), one sees that

$$\sup_{\delta} \int_{\mathbb{R}^d} |\mathcal{F}(f)(\xi)|^2 \exp(-\delta\|\xi\|^2/2)\mathrm{d}\xi < \infty$$

and, since $\lim_{\delta\to 0} |\mathcal{F}(f)(\xi)|^2 \exp(-\delta\|\xi\|^2/2) = |\mathcal{F}(f)(\xi)|^2$ for any ξ, one can use the monotone convergence theorem ***(Explain why)*** to get that $\mathcal{F}(f) \in L^2(\mathbb{R}^d, \mathbb{C})$ and

$$\frac{1}{(2\pi)^d}\|\mathcal{F}(f)\|_2^2 = \lim_{\delta\to 0^+} I_\delta = \|f\|_2^2$$

which proves the result. ■

With this we can extend the Fourier transform to $L^2(\mathbb{R}^d, \mathbb{C})$:

Theorem 5.2.3 — Plancherel Theorem. The mapping $\mathcal{F} : L^1(\mathbb{R}^d,\mathbb{C}) \cap L^2(\mathbb{R}^d,\mathbb{C}) \to L^2(\mathbb{R}^d,\mathbb{C})$ admits a unique extension, that we still denote, $\mathcal{F}$ to $L^2(\mathbb{R}^d,\mathbb{C})$, i.e.

$$\mathcal{F} : L^2(\mathbb{R}^d,\mathbb{C}) \to L^2(\mathbb{R}^d,\mathbb{C})$$

which is such that, for any $f, g \in L^2(\mathbb{R}^d,\mathbb{C})$ the following ***Parseval formula*** holds

$$\begin{aligned}\langle f, g\rangle &= \int_{\mathbb{R}^d} \overline{g}(x) f(x)\mathrm{d}x = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} \overline{\mathcal{F}(g)(\xi)}\mathcal{F}(f)(\xi)\mathrm{d}\xi \\ &= \frac{1}{(2\pi)^d}\langle \mathcal{F}(f), \mathcal{F}(g)\rangle .\end{aligned} \tag{5.14}$$

In particular, $(2\pi)^{-d/2}\mathcal{F} : L^2(\mathbb{R}^d,\mathbb{C}) \to L^2(\mathbb{R}^d,\mathbb{C})$ is an isometry, i.e. ***Plancherel formula*** holds

$$\|\mathcal{F}(f)\|_2 = (2\pi)^{d/2}\|f\|_2 \qquad \forall f \in L^2(\mathbb{R}^d,\mathbb{C}).$$

The mapping $\mathcal{F} : L^2(\mathbb{R}^d,\mathbb{C}) \to L^2(\mathbb{R}^d,\mathbb{C})$ is called the ***Fourier-Plancherel transform***.

✍ *Proof.* Since $L^1(\mathbb{R}^d,\mathbb{C}) \cap L^2(\mathbb{R}^d,\mathbb{C})$ is dense in $L^2(\mathbb{R}^d,\mathbb{C})$ (it contains $\mathscr{C}_c(\mathbb{R}^d)$), given $f \in L^2(\mathbb{R}^d,\mathbb{C})$, there exists a sequence $(f_n)_n \in L^1(\mathbb{R}^d,\mathbb{C}) \cap L^2(\mathbb{R}^d,\mathbb{C})$ such that

$$\lim_{n\to\infty} \|f_n - f\|_2 = 0.$$

Then, for any $n \in \mathbb{N}$, according to Plancherel Proposition, $\mathcal{F}(f_n)$ is well-defined and belongs to $L^2(\mathbb{R}^d,\mathbb{C})$. Moreover

$$\|\mathcal{F}(f_n)\|_2 = (2\pi)^{d/2}\|f_n\|_2 \qquad \forall n \in \mathbb{N}.$$

In particular, one sees that the sequence $(\mathcal{F}(f_n)_n)$ is a Cauchy sequence in $L^2(\mathbb{R}^d,\mathbb{C})$ since

$$\|\mathcal{F}(f_n) - \mathcal{F}(f_m)\|_2 = (2\pi)^{d/2}\|f_n - f_m\|_2 \qquad \forall n, m \in \mathbb{N}.$$

Therefore, $(\mathcal{F}(f_n)_n)_n$ converges to some limit $h \in L^2(\mathbb{R}^d)$. Let us prove that this limit does not depend on the choice of the approximating sequence but only on f: indeed, if $(g_n)_n$ is another sequence in $L^1(\mathbb{R}^d,\mathbb{C}) \cap L^2(\mathbb{R}^d,\mathbb{C})$ such that

$$\lim_{n\to\infty} \|g_n - f\|_2 = 0$$

then $(\mathcal{F}(g_n)_n)$ converges to some limit, say $h_2 \in L^2(\mathbb{R}^d)$. But,

$$\|\mathcal{F}(g_n) - \mathcal{F}(f_n)\|_2 = (2\pi)^{d/2}\|f_n - g_n\|_2 \leqslant \|f_n - f\|_2 + \|g_n - f\|_2$$

so $\lim_n \|\mathcal{F}(g_n) - \mathcal{F}(f_n)\|_2 = 0$ and $h = h_2$. The limit depends only on f and we denote it $\mathcal{F}(f)$ [4]. Of course,

$$\|\mathcal{F}(f)\|_2 = (2\pi)^{d/2}\|f\|_2 \qquad \forall f \in L^2(\mathbb{R}^d, \mathbb{C}).$$

Using polarization identity for hermitian inner product ***(Check carefully the formula)***:

$$\langle f, g\rangle = \frac{1}{2}\left[\|f+g\|_2^2 - i\|g + if\|_2^2 - (1-i)\|f\|_2^2 - (1-i)\|g\|_2^2\right]$$

one gets easily (5.14). ■

Again, we insist on the fact that, strictly speaking, if $f \in L^2(\mathbb{R}^d, \mathbb{C})$ then

$$\mathcal{F}(f)(\xi) \neq \int_{\mathbb{R}^d} \mathbf{e}^{-i\xi\cdot x} f(x)\mathrm{d}x.$$

However, for any $n \in \mathbb{N}$, if one defines

$$\widehat{f}_n(\xi) = \int_{B(0,n)} \mathbf{e}^{-i\xi\cdot x} f(x)\mathrm{d}x$$

then, since $\mathbf{1}_{B(0,n)} f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, one has

$$\mathcal{F}(f) = \lim_{n\to} \widehat{f}_n$$

where the limit is meant in $L^2(\mathbb{R}^d)$. Given $f \in L^2(\mathbb{R}^d)$, set then

$$\mathcal{F}_{-1} f(\xi) := \frac{1}{(2\pi)^d}\mathcal{F}(f)(-\xi).$$

The Fourier inversion formula still holds in that case:

Theorem 5.2.4 Let $f \in L^2(\mathbb{R}^d, \mathbb{C})$ and let $\mathcal{F}(f)$ denotes its Fourier-Plancherel transform. Then,

$$f = \mathcal{F}_{-1}(\mathcal{F}(f)).$$

In other words, the mapping

$$\mathcal{F} \;:\; L^2(\mathbb{R}^d, \mathbb{C}) \to L^2(\mathbb{R}^d, \mathbb{C})$$

is bijective and its inverse is $\mathcal{F}_{-1}$: $\mathcal{F}_{-1} = \mathcal{F}^{-1}$. In particular, for any $f \in L^2(\mathbb{R}^d, \mathbb{C})$:

$$f = \frac{1}{(2\pi)^d}\overline{\mathcal{F}(\overline{f})}.$$

[4] pay attention, it is a notation and does not mean that $\mathcal{F}(f)$ is given by the integral $\mathcal{F}(f) = \int_{\mathbb{R}^d} \mathbf{e}^{-i\xi\cdot x} f(x)\mathrm{d}x$ which does not make necessarily sense !!

Proof. Let $(f_n)_n \subset L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ be such that

$$\lim_n \|f_n - f\|_2 = 0 \qquad \text{and} \qquad \lim_{n\to\infty} \|\widehat{f}_n - \mathcal{F}(f)\|_2 = 0. \tag{5.15}$$

Given $\varepsilon > 0$, we saw (see the proof of Theorem 5.1.6, Eq. (5.10)) that

$$\mathbb{G}_{\varepsilon^2} * f_n(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathbf{e}^{i\xi\cdot x} \mathbf{e}^{-\frac{\varepsilon^2}{2}\|\xi\|^2} \widehat{f}_n(\xi)\mathrm{d}\xi \qquad x \in \mathbb{R}^d.$$

Since $\mathbb{G}_{\varepsilon^2} \in {}^1(\mathbb{R}^d, \mathbb{C})$ has unit mass and $\xi \mapsto \mathbf{e}^{-\frac{\varepsilon^2}{2}\|\xi\|^2}$ belongs to $L^2(\mathbb{R}^d)$, one deduces from (5.15) that

$$\lim_n \|\mathbb{G}_{\varepsilon^2} * f_n - \mathbb{G}_{\varepsilon^2} * f\|_2 = 0$$

while, for any $x \in \mathbb{R}^d$:

$$\lim_n \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathbf{e}^{i\xi\cdot x} \mathbf{e}^{-\frac{\varepsilon^2}{2}\|\xi\|^2} \widehat{f}_n(\xi)\mathrm{d}\xi = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathbf{e}^{i\xi\cdot x} \mathbf{e}^{-\frac{\varepsilon^2}{2}\|\xi\|^2} \mathcal{F}(f)(\xi)\mathrm{d}\xi.$$

In particular, since $\mathbb{G}_{\varepsilon^2} * f_n(x) \to \mathbb{G}_{\varepsilon^2} * f(x)$ for a.e. $x \in \mathbb{R}^d$, one has

$$\mathbb{G}_{\varepsilon^2} * f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathbf{e}^{i\xi\cdot x} \mathbf{e}^{-\frac{\varepsilon^2}{2}\|\xi\|^2} \mathcal{F}(f)(\xi)\mathrm{d}\xi \qquad \text{for a.e. } x \in \mathbb{R}^d$$

i.e.

$$\mathbb{G}_{\varepsilon^2} * f = \mathcal{F}^{-1}\left(\mathbf{e}^{-\frac{\varepsilon^2}{2}\|\cdot\|^2} \mathcal{F}(f)\right).$$

Since $\lim_{\varepsilon\to 0} \|\mathbb{G}_{\varepsilon^2} * f - f\|_2 = 0$ while, by dominated convergence theorem

$$\lim_{\varepsilon\to 0} \|\mathbf{e}^{-\frac{\varepsilon^2}{2}\|\cdot\|^2} \mathcal{F}(f) - \mathcal{F}(f)\|_2 = 0$$

we get, thanks to Plancherel theorem

$$\lim_{\varepsilon\to 0} \|\mathcal{F}^{-1}(\mathbf{e}^{-\frac{\varepsilon^2}{2}\|\cdot\|^2} \mathcal{F}(f)) - \mathcal{F}^{-1}(\mathcal{F}(f))\|_2 = 0$$

from which the result follows. ■

To summarize the results we saw so far:

Theorem 5.2.5 The Fourier transform $f \in L^1(\mathbb{R}^d, \mathbb{C}) \cap L^2(\mathbb{R}^d, \mathbb{C}) \mapsto \widehat{f} \in L^2(\mathbb{R}^d, \mathbb{C})$ has an extension

$$\mathcal{F} \,:\, L^2(\mathbb{R}^d, \mathbb{C}) \to L^2(\mathbb{R}^d, \mathbb{C})$$

with the following properties:

1. $(2\pi)^{-d/2}\mathcal{F}$ is an isometry mapping of $L^2(\mathbb{R}^d, \mathbb{C}) \to L^2(\mathbb{R}^d, \mathbb{C})$.
2. $\mathcal{F}$ is a continuous bijection of $L^2(\mathbb{R}^d, \mathbb{C}))$ to $L^2(\mathbb{R}^d, \mathbb{C})$.
3. The inverse of $\mathcal{F}$ is given by

$$\mathcal{F}^{-1}(f) = \mathcal{F}_1(f) = \frac{1}{(2\pi)^d} \overline{\mathcal{F}(\overline{f})}.$$

One has also the following

Proposition 5.2.6 Let $f \in L^1(\mathbb{R}^d)$ and $g \in L^2(\mathbb{R}^d)$ be given. Then,

$$\mathcal{F}(f * g) = \mathcal{F}(f)\,\mathcal{F}(g).$$

✍ *Proof.* Notice that, since $f \in L^1(\mathbb{R}^d)$ and $g \in L^2(\mathbb{R}^d)$, according to Young's inequality

$$f * g \in L^2(\mathbb{R}^d)$$

so that $\mathcal{F}(f * g)$ is well-defined. Then, given $(g_n)_n \subset L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ such that

$$\lim_n \|g_n - g\|_2 = 0 \qquad \text{and} \qquad \lim_n \|\mathcal{F}(g_n) - \mathcal{F}(g)\|_2 = 0$$

one has $f * g_n \in L^1(\mathbb{R}^d, \mathbb{C}) \cap L^2(\mathbb{R}^d, \mathbb{C})$ so $\mathcal{F}(f * g_n)$ is the L^1-Fourier transform and

$$\mathcal{F}(f * g_n) = \mathcal{F}(f)\mathcal{F}(g_n) \qquad \forall n \in \mathbb{N}.$$

Now, because $\mathcal{F}(g_n) \to \mathcal{F}(g)$ in $L^2(\mathbb{R}^d)$, and

$$\|\mathcal{F}(f)\mathcal{F}(g_n) - \mathcal{F}(f)\mathcal{F}(g)\|_2 \leqslant \|\mathcal{F}(f)|_\infty \|\mathcal{F}(g_n) - \mathcal{F}(g)\|_2 \leqslant \|f\|_1 \|\mathcal{F}(g_n) - \mathcal{F}(g)\|_2$$

one deduces that

$$\lim_{n\to\infty} \|\mathcal{F}(f * g_n) - \mathcal{F}(g)\mathcal{F}(g)\|_2 = 0.$$

To conclude, one simply observes that, thanks to Young's inequality,

$$\lim_{n\to\infty} \|f * g_n - f * g\|_2 = 0$$

since $\|f * g_n - f * g\|_2 \leqslant \|f\|_1 \|g_n - g\|_2$ and therefore $\{f * g_n\}_n$ is a sequence in $L^1(\mathbb{R}^d, \mathbb{C}) \cap L^2(\mathbb{R}^d, \mathbb{C})$ which approximates $f * g$ in $L^2(\mathbb{R}^d, \mathbb{C})$ so that, by definition of Fourier-Plancher transform,

$$\lim_{n\to\infty} \|\mathcal{F}(f * g_n) - \mathcal{F}(f * g)\|_2 = 0.$$

By uniqueness of the limit, one has $\mathcal{F}(f * g) = \mathcal{F}(f)\mathcal{F}(g)$. ■

Pay attention that the Fourier-Plancherel transform in $L^2(\mathbb{R}^d)$ does not transforms a L^2-function into a continuous one, and, *a fortiori*, would not transform a L^2-function with enough decay into a differentiable one. We actually need another version of differentiation.

5.2.1 An alternative Hilbertian construction of Fourier-Plancherel transform

We propose here an alternative construction of the Fourier-Plancherel transform in $L^2(\mathbb{R}^d)$. This method is due to N. Wiener (1933) and is based on Hilbertian method. For simplicity, we will focus on the simplest case $d = 1$. The idea is to construct a suitable orthonormal basis $\{\boldsymbol{e}_n\,;\, n \in \mathbb{N}\}$ of $L^2(\mathbb{R})$ and to define, for a given $f \in L^2(\mathbb{R})$ written as

$$f = \sum_{n\in\mathbb{N}} \langle f, \boldsymbol{e}_n\rangle\, \boldsymbol{e}_n$$

the Fourier transform $\mathcal{F}(f)$ by its action on the basis $\boldsymbol{e}_n$. It will be somehow simpler to start with working on the Gaussian L^2-space

$$H = L^2(\gamma) = L^2(\mathbb{R}, \mathrm{d}\gamma)$$

where $\mathrm{d}\gamma$ is the Gaussian measure

$$\mathrm{d}\gamma(A) = \frac{1}{\sqrt{2\pi}} \int_A \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x, \qquad \forall A \in \mathcal{B}(\mathbb{R}).$$

Of course γ is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and we will make no distinction between the measure $\mathrm{d}\gamma$ and its density with respect to the Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$:

$$\gamma(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \qquad x \in \mathbb{R}.$$

Of course, the space H is a Hilbert space when endowed with the inner product

$$\langle f, g\rangle_\gamma = \int_{\mathbb{R}} f\,\overline{g} \mathrm{d}\gamma = \int_{\mathbb{R}} f(x)\overline{g(x)}\gamma(x)\mathrm{d}x, \qquad f, g \in H.$$

In the sequel, we set

$$\mathbb{N}_0 = \mathbb{N} \cup \{0\}.$$

We define now the Hermite polynomials on $\mathbb{R}$ as follows:

Definition 5.2.1 We define $H_0(x) = 1$ for any $x \in \mathbb{R}$ and, for a given $n \in \mathbb{N}$, we define

$$H_n(x) = (-1)^n \frac{1}{\gamma(x)} \frac{\mathrm{d}^n}{\mathrm{d}x^n}\gamma(x), \qquad x \in \mathbb{R}.$$

The family $\{H_n\ ;\ n \in \mathbb{N}_0\}$ is called the family of *Hermite polynomials* over $\mathbb{R}$.

Exercise 5.1 Show that, for any $n \in \mathbb{N}$,

$$H_{n+1}(x) = xH_n(x) - H_n'(x).$$

By induction, deduce from this expression that $H_n' = nH_{n-1}$, $\forall n \geqslant 1$. ■

One can check easily that

$$H_1(x) = x, \qquad H_2(x) = x^2 - 1, \qquad H_3(x) = x^3 - 3x, \ldots.$$

More generally, H_n is a polynomial function of degree n whose dominant term is x^n.

Remark 5.2.1 A simple use of integration by parts shows that, for any $n \in \mathbb{N}$, one has

$$\int_{\mathbb{R}} f^{(n)} \mathrm{d}\gamma = \int_{\mathbb{R}} f\, H_n \mathrm{d}\gamma, \qquad \forall f \in \mathscr{C}_c^\infty(\mathbb{R}) \tag{5.16}$$

where $f^{(n)}$ is the n-th derivative of f, $f^{(n)} = \frac{\mathrm{d}^n}{\mathrm{d}x^n} f$, $n \in \mathbb{N}$.

One has the following

Lemma 5.2.7 For any $n, m \in \mathbb{N}$, one has

$$\langle H_n, H_m\rangle_\gamma = \begin{cases} n! & \text{if } n = m, \\ 0 & \text{if } n \neq m. \end{cases}$$

Moreover, $\gamma H_n \in L^1(\mathbb{R})$ and

$$\mathcal{F}(\gamma\, H_n)(\xi) = (-i\xi)^n \sqrt{2\pi}\gamma(\xi). \tag{5.17}$$

Proof. The proof is easily made by induction. Assume for simplicity $n \geqslant m$. The easiest way to check the result is to use (5.16) with $f = H_m$ (even if $f \notin \mathscr{C}_c^\infty(\mathbb{R})$, (5.16) is still valid for $f = H_m$ since both sides of the identity are well-defined). One has then

$$\langle H_m, H_n\rangle_\gamma = \int_{\mathbb{R}} H_m\, H_n \mathrm{d}\gamma = \int_{\mathbb{R}} \frac{\mathrm{d}^n}{\mathrm{d}x^n} H_m \mathrm{d}\gamma$$

and, of course, if $n > m$, then since H_m is of degree n, one has

$$\frac{\mathrm{d}^n}{\mathrm{d}x^n} H_m(x) = 0$$

which shows that, for $n > m$, $\langle H_m, H_n\rangle_\gamma = 0$. For $n = m$, since H_n is of degree n with $H_n(x) = x^n + \sum_{k=0}^{n-1} x^k$, one has

$$\frac{\mathrm{d}^n}{\mathrm{d}x^n} H_n(x) = \frac{\mathrm{d}^n}{\mathrm{d}x^n} x^n = n!$$

i.e.

$$\langle H_n, H_n\rangle_\gamma = \int_{\mathbb{R}} n! \mathrm{d}\gamma = n!.$$

This proves the result. For the computation of the Fourier transform, one sees that, by definition of H_n:

$$\int_{\mathbb{R}} \mathbf{e}^{-ix\xi} \gamma(x) H_n(x) \mathrm{d}x = (-1)^n \int_{\mathbb{R}} \mathbf{e}^{-ix\xi} \frac{\mathrm{d}^n}{\mathrm{d}x^n} \gamma(x) \mathrm{d}x$$

i.e.

$$\mathcal{F}(\gamma\, H_n) = (-1)^n \mathcal{F}\left(\frac{\mathrm{d}^n}{\mathrm{d}x^n}\gamma\right).$$

Using the well-known property of the Fourier transform (in L^1), we have

$$\mathcal{F}\left(\frac{\mathrm{d}^n}{\mathrm{d}x^n}\gamma\right) = (-i\xi)^n \mathcal{F}(\gamma)$$

and, using the fact that $\widehat{\gamma}(\xi) = \sqrt{2\pi}\gamma(\xi)$, one gets (5.17). ∎

Remark 5.2.2 By the injectivity of the Fourier transform, (5.17) can actually serve as a *definition* of the Hermite polynomials H_n.

Introducing then

$$h_n = \frac{1}{n!} H_n, \qquad n \in \mathbb{N}_0$$

one has the following

Proposition 5.2.8 The family $\{h_n \,;\, n \in \mathbb{N}_0\}$ is a *Hilbert basis* of $H = L^2(\gamma)$.

Proof. The previous Lemma shows that $\{h_n \,;\, n \in \mathbb{N}_0\}$ is an orthonormal family of $L^2(\gamma)$. To show that it is a Hilbert basis, one needs to prove that, if

$$\langle f, H_n\rangle_\gamma = 0 \qquad \forall n \in \mathbb{N}_0$$

then $f = 0$ (see Theorem 3.4.5, point *d)*). Since $\mathrm{Span}\left(\{h_n\,,\, n \in \mathbb{N}_0\}\right) = \mathrm{Span}\left(\{x^n\,,\, n \in \mathbb{N}_0\}\right)$, one sees easily that it is equivalent to prove that

$$\langle f, x^n\rangle_\gamma = 0 \qquad \forall n \in \mathbb{N}_0 \Longrightarrow f = 0.$$

Let us then fix $f \in L^2(\gamma)$ such that $\langle f, x^n\rangle_\gamma = 0$ for any $n \in \mathbb{N}$. This implies clearly that, for any $\xi \in \mathbb{R}$

$$\sum_{k=0}^{n} \int_{\mathbb{R}} f(x)\gamma(x)\, \frac{(i\xi\, x)^k}{k!} \mathrm{d}x = 0 \qquad \forall n \in \mathbb{N}.$$

Notice that, for any $\xi \in \mathbb{R}$ and $n \in \mathbb{N}$,

$$\left| f(x)\gamma(x) \sum_{k=0}^{n} \frac{(-i\xi\, x)^k}{k!} \right| \leqslant |f(x)|\gamma(x) \sum_{k=0}^{n} \frac{|\xi\, x|^k}{k!} \leqslant \frac{1}{\sqrt{2\pi}} |f(x)| \exp\left(|x\xi| - \frac{x^2}{2}\right),$$

and the mapping $x \mapsto \exp\left(|x\xi| - \frac{x^2}{2}\right)$ belongs to $L^2(\gamma)$ ***(Check this)*** which, by Cauchy-Schwarz inequality implies that $x \mapsto |f(x)| \exp\left(|x\xi| - \frac{x^2}{2}\right)$ is integrable over $\mathbb{R}$. Then, according to the dominated convergence theorem, for any $\xi \in \mathbb{R}$, one has

$$\lim_{n\to\infty} \sum_{k=0}^{n} \int_{\mathbb{R}} f(x)\gamma(x)\, \frac{(-i\xi\, x)^k}{k!} \mathrm{d}x = \int_{\mathbb{R}} f(x)\gamma(x) \mathbf{e}^{-ix\xi} \mathrm{d}x$$

so that

$$\int_{\mathbb{R}} f(x)\gamma(x) \mathbf{e}^{-ix\xi} \mathrm{d}x = 0 \qquad \forall \xi \in \mathbb{R}.$$

Because $f\gamma \in L^1(\mathbb{R})$ and has a zero Fourier transform, we get $f(x)\gamma(x) = 0$ for a.e. $x \in \mathbb{R}$ which proves the result. ∎

Turning back to the space $L^2(\mathbb{R};\mathbb{C})$, the above Proposition translates easily in the following

Proposition 5.2.9 Introduce

$$\boldsymbol{e}_n(x) = \sqrt{\gamma(x)}\, h_n(x) = \frac{1}{n!} \sqrt{\gamma(x)}\, H_n(x), \qquad x \in \mathbb{R}, \qquad n \in \mathbb{N}_0.$$

Then, $\{\boldsymbol{e}_n \,;\, n \in \mathbb{N}_0\}$ is a Hilbert basis of $L^2(\mathbb{R})$. In particular, any $f \in L^2(\mathbb{R};\mathbb{C})$ can be written as

$$f = \sum_{n \in \mathbb{N}_0} \langle f, \boldsymbol{e}_n\rangle\, \boldsymbol{e}_n$$

where the series converges in $L^2(\mathbb{R};\mathbb{C})$ with moreover

$$\|f\|_2^2 = \sum_{n\in\mathbb{N}_0} |\langle f, \boldsymbol{e}_n\rangle|^2 .$$

Proof. The proof simply comes from the fact that one can identify $L^2(\gamma)$ and $L^2(\mathbb{R})$ through the isometric mapping

$$\boldsymbol{\iota}\ :\ f \in L^2(\gamma) \longmapsto \sqrt{\gamma} f \in L^2(\mathbb{R}).$$

Notice that, with $\langle\cdot,\cdot\rangle$ the usual inner product of $L^2(\mathbb{R},\mathbb{C})$, one has

$$\langle \boldsymbol{\iota}(f), \boldsymbol{\iota}(g)\rangle = \langle f, g\rangle_\gamma \qquad \forall f, g \in L^2(\gamma).$$

Then, for any $n \in \mathbb{N}$, since $\boldsymbol{e}_n = \boldsymbol{\iota}(h_n)$ one gets that $\{\boldsymbol{e}_n\ ;\ n \in \mathbb{N}_0\}$ is a Hilbert bases of $L^2(\mathbb{R};\mathbb{C})$. The rest of the statement follows then from Theorem 3.4.5. ■

Exercise 5.2 Deduce from Exercise 5.1 that the derivative of $\boldsymbol{e}_n$ is

$$\boldsymbol{e}_n' + \frac{x}{2}\boldsymbol{e}_n = \boldsymbol{e}_{n-1}, \qquad n \geqslant 1$$

whereas $\boldsymbol{e}_n' - \frac{x}{2}\boldsymbol{e}_n = -(n+1)\boldsymbol{e}_{n+1}$, $(n \geqslant 0)$. ■

Notice that

$$\boldsymbol{e}_n \in L^1(\mathbb{R}) \cap L^2(\mathbb{R}) \qquad \forall n \in \mathbb{N}$$

as can be easily seen since

$$\int_{\mathbb{R}} |\boldsymbol{e}_n(x)| \mathrm{d}x = \frac{1}{(2\pi)^{\frac{1}{4}}} \int_{\mathbb{R}} \exp\left(-\frac{x^2}{4}\right) |H_n(x)|\, \mathrm{d}x$$

which is finite, H_n being a polynomial of degree n. As such one can compute the Fourier transform of $\boldsymbol{e}_n$ and one has

Lemma 5.2.10 For any $n \in \mathbb{N}$, one has

$$\widehat{\boldsymbol{e}_n}(\xi) = \sqrt{4\pi}(-i)^n \boldsymbol{e}_n(2\xi) \qquad \forall \xi \in \mathbb{R}.$$

Proof. From Exercise 5.2, one has

$$\mathcal{F}(\boldsymbol{e}_n')(\xi) - \frac{1}{2}\mathcal{F}(x\boldsymbol{e}_n) = -(n+1)\mathcal{F}(\boldsymbol{e}_{n+1}), \qquad n \geqslant 1.$$

Since $\mathcal{F}(\boldsymbol{e}_n')(\xi) = i\xi\widehat{\boldsymbol{e}_n}$ and $\mathcal{F}(x\boldsymbol{e}_n)(\xi) = i\frac{\mathrm{d}}{\mathrm{d}\xi}\widehat{\boldsymbol{e}_n}(\xi)$ one sees that the function $y_n(\xi) = \widehat{\boldsymbol{e}_n}(\xi)$ satisfies $i\xi y_n(\xi) - \frac{i}{2}y_n'(\xi) = -(n+1)y_{n+1}(\xi)$, or equivalently

$$y_n'(\xi) - 2\xi\, y_n(\xi) = -2i(n+1)y_{n+1}(\xi)$$

with, thanks to Proposition 5.1.5

$$y_0(\xi) = \mathcal{F}(\boldsymbol{e}_0) = \mathcal{F}(\sqrt{\gamma})(\xi) = \frac{1}{(2\pi)^{\frac{1}{4}}} \int_{\mathbb{R}} \mathbf{e}^{-ix\xi} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x = \frac{\sqrt{4\pi}}{(2\pi)^{\frac{1}{4}}} \exp\left(-\frac{\xi^2}{2}\right)$$

i.e.

$$y_0(\xi) = \sqrt{4\pi\gamma(2\xi)}.$$

Now, one notices easily thanks to Exercise 5.2 that the function

$$z_n(\xi) = (-i)^n\sqrt{4\pi}\boldsymbol{e}_n(2\xi)$$

satisfies also

$$z_n'(\xi) - 2\xi\, z_n(\xi) = -2i(n+1)z_{n+1}(\xi), \qquad n \geqslant 0$$

and

$$z_0(\xi) = \sqrt{4\pi}\boldsymbol{e}_0(2\xi) = \sqrt{4\pi\gamma(2\xi)} = y_0(\xi).$$

Therefore, the sequence (z_n) and (y_n) satisfies the same induction relation with $z_0 = y_0$ so that $z_n = y_n$ which is the desired result. ■

We have all in hands to give the following alternative definition of the Fourier-Plancherel transform on $L^2(\mathbb{R})$:

Theorem 5.2.11 For any $f \in L^2(\mathbb{R})$ given by

$$f = \sum_{n\in\mathbb{N}_0} \langle f, \boldsymbol{e}_n\rangle \boldsymbol{e}_n$$

define

$$\mathcal{F}(f) = \sum_{n\in\mathbb{N}_0} \langle f, \boldsymbol{e}_n\rangle \widehat{\boldsymbol{e}_n}.$$

Then, $\mathcal{F}$ satisfies Parseval formula

$$\langle f, g\rangle = \frac{1}{2\pi}\langle \mathcal{F}(f), \mathcal{F}(g)\rangle, \qquad \forall f, g \in L^2(\mathbb{R}).$$

In particular

$$\|\mathcal{F}(f)\|_2 = \sqrt{2\pi}\|f\|_2, \qquad \forall f \in L^2(\mathbb{R}).$$

Proof. Given $f \in L^2(\mathbb{R})$ given by

$$f = \sum_{n\in\mathbb{N}_0} \langle f, \boldsymbol{e}_n\rangle \boldsymbol{e}_n$$

we define

$$\mathcal{F}(f) = \sum_{n\in\mathbb{N}_0} \langle f, \boldsymbol{e}_n\rangle \widehat{\boldsymbol{e}_n} = \sqrt{4\pi} \sum_{n\in\mathbb{N}_0} (-i)^n \langle f, \boldsymbol{e}_n\rangle\, \boldsymbol{e}_n(2\cdot).$$

where the series is still convergent in $L^2(\mathbb{R})$ since

$$\sum_{n\in\mathbb{N}_0} |\langle f, \boldsymbol{e}_n\rangle|^2 < \infty.$$

By construction, given $f, g \in L^2(\mathbb{R})$

$$\langle \mathcal{F}(f), \mathcal{F}(g)\rangle = \sum_{n,m} \langle f, \boldsymbol{e}_n\rangle\langle g, \boldsymbol{e}_m\rangle\langle \widehat{\boldsymbol{e}_n}, \widehat{\boldsymbol{e}_m}\rangle$$

with, from the previous Lemma

$$\langle \widehat{\boldsymbol{e}_n}, \widehat{\boldsymbol{e}_m} \rangle = 4\pi(-i)^n i^m \int_{\mathbb{R}} \boldsymbol{e}_n(2\xi)\boldsymbol{e}_m(2\xi)\mathrm{d}\xi.$$

Making the change of variable $x = 2\xi$, $2\mathrm{d}\xi = \mathrm{d}x$ we got

$$\langle \widehat{\boldsymbol{e}_n}, \widehat{\boldsymbol{e}_m} \rangle = 2\pi(-i)^n i^m \langle \boldsymbol{e}_n, \boldsymbol{e}_m \rangle = 2\pi(-i)^n i^m \delta_{n,m} = 2\pi\delta_{n,m}$$

so that

$$\langle \mathcal{F}(f), \mathcal{F}(g) \rangle = 2\pi \sum_n \langle f, \boldsymbol{e}_n \rangle \langle g, \boldsymbol{e}_m \rangle = 2\pi \langle f, g \rangle.$$

So, Parseval identity is true. ■

5.3 Fourier transform of measure

5.3.1 Basic definition

Assume that μ is a finite measure over $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, i.e. $\|\mu\| = \mu(\mathbb{R}^d) < \infty$. Then, for any $\xi \in \mathbb{R}^d$, one can define

$$\widehat{\mu}(\xi) = \int_{\mathbb{R}^d} \mathbf{e}^{-i\xi\cdot x}\mu(\mathrm{d}x)$$

which is well-defined and satisfies

$$|\widehat{\mu}(\xi)| \leqslant \|\mu\| < \infty \qquad \forall \xi \in \mathbb{R}^d.$$

■ **Example 5.5** Of course, if μ is absolutely continuous with respect to the Lebesgue measure $\mathfrak{m}$ with density f then

$$\widehat{\mu}(\xi) = \mathcal{F}(f)(\xi) \qquad \forall \xi \in \mathbb{R}^d$$

and $\|\mu\| = \|f\|_1$. ■

■ **Example 5.6** If $\mu(\mathrm{d}x) = \delta_{\boldsymbol{a}}(\mathrm{d}x)$ then

$$\widehat{\mu}(\xi) = \mathbf{e}^{-i\xi\cdot\boldsymbol{a}} \qquad \forall \xi \in \mathbb{R}^d.$$

Notice then that, in such a case, $\widehat{\mu}$ is infinitely differentiable with

$$\frac{\partial^k}{\partial\xi_1^{\alpha_1}\ldots\partial\xi_d^{\alpha_d}}\widehat{\mu}(\xi) = (-i)^k a_1^{\alpha_1}\ldots a_d^{\alpha_d}\widehat{\mu}(\xi) \qquad \forall \xi \in \mathbb{R}^d$$

where $\boldsymbol{a} = (a_1, \ldots, a_d)$ and $(\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$ with $\sum_j \alpha_j = k$.
More generally, if

$$\mu(\mathrm{d}x) = \sum_{k\in I} \alpha_k \delta_{\boldsymbol{a}_k}(\mathrm{d}x)\,, \qquad I \subset \mathbb{N}\,,\ \alpha_k \in \mathbb{R}^+, \boldsymbol{a}_k \in \mathbb{R}^d, \quad k \in I$$

then

$$\widehat{\mu}(\xi) = \sum_{k\in I} \alpha_k \mathbf{e}^{-i\xi\cdot\boldsymbol{a}_k} \qquad \forall \xi \in \mathbb{R}^d.$$

■

5.3.2 Application to probability theory

We establish now some general results that apply to a large variety of problems and go beyond the problem we have in mind, i.e. the proof of the Central Limit Theorem. We begin with a definition

Definition 5.3.1 Let $(X_n)_n$ be a sequence of real random variables and let X be a real random variable over the same triple $(\Omega, \mathcal{F}, \mathbb{P})$. We say that $(X_n)_n$ converges weakly to X (as $n \to \infty$) and write $X_n \implies X$ if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \qquad \forall x \in \mathcal{C}_X$$

where F_{X_n} and F_X stand respectively for the distribution function of X_n and X and $\mathcal{C}_X$ is the set of continuity points of F_X, i.e.

$$\mathcal{C}_X = \left\{x_0 \in \mathbb{R}\,,\ \lim_{x\to x_0} F_X(x) = F_X(x_0)\right\} = \left\{x_0 \in \mathbb{R}\,,\ \lim_{\substack{x\to x_0\\ x<x_0}} F_X(x) = F_X(x_0)\right\}$$

where we recall that F_X is right-continuous on $\mathbb{R}$.

■ **Example 5.7** Let X have distribution function F_X. Then, setting $X_n = X + 1/n$ for any $n \geqslant 1$, the distribution function F_{X_n} of X_n is given by

$$F_{X_n}(x) = \mathbb{P}(X_n \leqslant x) = \mathbb{P}(X \leqslant x - 1/n) = F_X(x - 1/n) \qquad \forall x \in \mathbb{R}.$$

In particular, $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ if and only if $\lim_{n\to\infty} F_X(x-1/n) = \lim_{\substack{y\to x\\ y<x}} F_X(y)$ and this explains why, in our definition, we focused only on the continuity points of F_X. ■

We introduce now the concept of characteristic function.

Definition 5.3.2 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability triple and let X be a real random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. We define its characteristic function $\Phi_X \,:\, t \in \mathbb{R} \to \mathbb{C}$ by

$$\Phi_X(t) = \mathbb{E}(\mathrm{e}^{itX}) \qquad \forall t \in \mathbb{R}.$$

Remark 5.3.1 Writing $\mathbb{P}_X$ the law of X under $\mathbb{P}$, one has

$$\Phi_X(t) = \int_{\mathbb{R}} \mathrm{e}^{itx} \mathrm{d}\mathbb{P}_X(x).$$

In other words, $\Phi_X(t)$ is exactly the Fourier transform of $\mathrm{d}\mathbb{P}_X$ in $-t$: $\Phi_X(t) = \widehat{\mathbb{P}_X}(-t)$.

Definition 5.3.3 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability triple and let $X = (X_1, \ldots, X_n)$ be a random vector on $(\Omega, \mathcal{F}, \mathbb{P})$, i.e. X_i is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ for any $i = 1, \ldots, n$. Then, the characteristic function of $(X_1, \ldots, X_n)$ is defined as $\Phi_{(X_1,\ldots,X_n)} : \mathbb{R}^n \to \mathbb{C}$ by

$$\Phi_{(X_1,\ldots,X_n)}(t_1, \ldots, t_n) = \mathbb{E}\left(e^{i\sum_{k=1}^n t_k X_k}\right) \qquad \forall (t_1, \ldots, t_n) \in \mathbb{R}^n.$$

Exercise 5.3 — Bernoulli. Assume that X follow the Bernoulli distribution with parameter p. Then, for any $t \in \mathbb{R}$:

$$\begin{aligned}\Phi_X(t) = \mathbb{E}(e^{itX}) &= \int_{\mathbb{R}} e^{ixt} \mathrm{d}\mathbb{P}_X(x) \\ &= p \int_{\mathbb{R}} e^{ixt} \mathrm{d}\delta_1(x) + (1-p) \int_{\mathbb{R}} e^{ixt} \mathrm{d}\delta_0(x) = pe^{ipt} + (1-p).\end{aligned}$$

■

Exercise 5.4 — Binomial distribution. Assume $X \sim \mathcal{B}(n, p)$. Then, for any $t \in \mathbb{R}$:

$$\begin{aligned}\Phi_X(t) = \mathbb{E}(e^{itX}) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \int_{\mathbb{R}} e^{ixt} \mathrm{d}\delta_k(x) \\ &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} e^{ikt} \\ &= \sum_{k=0}^{n} \binom{n}{k} (pe^{it})^k (1-p)^{n-k} = \left(1 - p + pe^{it}\right)^n\end{aligned}$$

thanks to the binomial identity. ■

Exercise 5.5 — Geometrical distribution. Assume that $X \sim \mathcal{G}(p)$. Then, for any $t \in \mathbb{R}$:

$$\begin{aligned}\Phi_X(t) = \mathbb{E}(e^{itX}) &= \sum_{n=0}^{\infty} p(1-p)^n \int_{\mathbb{R}} e^{ixt} \mathrm{d}\delta_n(x) \\ &= \sum_{n=0}^{\infty} p(1-p)^n e^{int} = p \sum_{n=0}^{\infty} \left((1-p)\, e^{it}\right)^n = \frac{p}{1 - (1-p)\, e^{it}}.\end{aligned}$$

■

Exercise 5.6 — Poisson distribution. Let $\lambda \geqslant 0$ and let $X \sim \mathcal{P}(\lambda)$. For any $t \in \mathbb{R}$, one has

$$\begin{aligned}\Phi_X(t) = \mathbb{E}(e^{itX}) &= \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \int_{\mathbb{R}} e^{ixt} \mathrm{d}\delta_n(x) = \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} e^{int} \\ &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{\left(\lambda\, e^{it}\right)^n}{n!} = e^{-\lambda} e^{\lambda\, e^{it}} = \exp\left(\lambda \left(e^{it} - 1\right)\right).\end{aligned}$$

■

Exercise 5.7 — Normal distribution. Assume that $X \sim \mathcal{N}(\mu, \sigma^2)$. Recall that X admits a density function f_X given by:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \forall x \in \mathbb{R}$$

Then, for any $t \in \mathbb{R}$:

$$\Phi_X(t) = \mathbb{E}(e^{itX}) = \int_{\mathbb{R}} e^{ixt} f_X(t)\mathrm{d}x = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} e^{ixt-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x.$$

Then,

$$\Phi_X(t) = \exp(-i\mu t - \sigma^2 t^2/2) \qquad \forall t \in \mathbb{R}. \tag{5.18}$$

■

Exercise 5.8 — Exponential distribution. If X is a random variable with exponential distribution with parameter θ, then X admits a probability density function f_X given by:

$$f_X(x) = \theta\, e^{-\theta\, x} \mathbf{1}_{[0,\infty)}(x) \qquad \forall x \in \mathbb{R}.$$

Therefore, for any $t \in \mathbb{R}$,

$$\Phi_X(t) = \mathbb{E}(e^{itX}) = \int_{\mathbb{R}} e^{ixt} f_X(t)\mathrm{d}x = \theta \int_0^\infty e^{ixt} e^{-\theta\, x}\mathrm{d}x = \theta \int_0^\infty e^{x(it-\theta)}\mathrm{d}x.$$

Using classical property of (complex) integration, one sees that, as it is the case for real integration

$$\int_0^\infty e^{x(it-\theta)}\mathrm{d}x = \frac{1}{(it-\theta)} e^{x(it-\theta)}\Big|_{x=0}^{\infty}$$

and, because $\theta > 0$, one proves that $\lim_{x\to\infty} e^{x(it-\theta)} = 0$. Therefore,

$$\Phi_X(t) = \frac{\theta}{(\theta - it)} \qquad \forall t \in \mathbb{R}.$$

■

Exercise 5.9 Prove that, if Φ_X is the characteristic function of some random variable X, then the following hold:

1. $\Phi_X(0) = 1$.
2. $\Phi_X(-t) = \overline{\Phi_X(t)}$ for any $t \geqslant 0$ (where, for any $z \in \mathbb{C}$, $\overline{z}$ is the complex conjugate of z).
3. $|\Phi_X(t)| \leqslant 1$ for any $t \in \mathbb{R}$.
4. For any $a, b \in \mathbb{R}$ and any $t \in \mathbb{R}$: $\Phi_{aX+b}(t) = e^{itb}\Phi_X(at)$.

■

Exercise 5.10 If X, Y are two independent random variables with respective characteristic functions Φ_X and Φ_Y, prove that the characteristic function Φ_{X+Y} of $X + Y$ is given by the following

$$\Phi_{X+Y}(t) = \Phi_X(t)\,\Phi_Y(t) \qquad \forall t \in \mathbb{R}.$$

Recall that "independence means multiplication." ■

It is not difficult to convince ourselves that, if $X \in L^2$ then $t \mapsto \Phi_X(t)$ is twice differentiable with

$$\Phi'_X(t) = i\mathbb{E}(Xe^{itX}) \qquad \text{and} \qquad \Phi''_X(t) = -\mathbb{E}(X^2e^{iXt}) \qquad \forall t \in \mathbb{R}.$$

In particular,

$$\Phi'_X(0) = i\mathbb{E}(X) \qquad \text{and} \qquad \Phi''_X(0) = -\mathbb{E}(X^2).$$

The above property actually allows to compute in an easy way the various moments of a random variable X. Precisely, if a random variable X has moments up to k-th order, then the characteristic function Φ_X is k times continuously differentiable and

$$\mathbb{E}(X^k) = (-i)^k \Phi_X^{(k)}(0).$$

Actually, one can even be more precise:

Proposition 5.3.1 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability triple and let X be a real random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. If $\mathbb{E}(X) = \mu$ and $\mathbb{E}(X^2) = \sigma^2 < \infty$ then

$$\Phi_X(t) = 1 + it\mu - \frac{\sigma^2}{2}t^2 + t^2\delta(t^2) \qquad t \in \mathbb{R}$$

where the mapping $\delta : \mathbb{R} \to \mathbb{C}$ is such that $\lim_{h\to 0} \frac{\delta(h)}{h} = 0$.

The proof is reminiscent from Taylor formula and is omitted here. As far a weak convergence is concerned, the interest of the characteristic functions lies in the two following results (see [4]). First

Proposition 5.3.2 Let $(X_n)_n$ be a sequence of real random variables and let X be a given random variable. We denote by Φ_{X_n} the characteristic function of X_n for an n and by Φ_X the one of X. If $X_n \implies X$ as $n \to \infty$, then $\lim_n \Phi_{X_n}(t) = \Phi_X(t)$ for any $t \in \mathbb{R}$.

The counterpart is the most interesting one:

Theorem 5.3.3 Let $(X_n)_n$ be a sequence of real random variables and let Φ_{X_n} the characteristic function of X_n for an n. Assume that there exists a mapping $\Phi : \mathbb{R} \to \mathbb{C}$ that is continuous at $t = 0$ and such that

$$\lim_n \Phi_{X_n}(t) = \Phi(t) \qquad \forall t \in \mathbb{R}.$$

Then, there exists a random variable X with characteristic function Φ such that $X_n \implies X$ as $n \to \infty$.

5.3.3 Central Limit Theorem

We now turn our attention on the following problem: consider a sequence $(X_1, X_2, \ldots)$ of i.i.d. random variables with $X_1 \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ and define, for any $n \geqslant 1$

$$Y_n = \sum_{k=1}^{n} X_k$$

with $Y_0 = 0$. The main result is the following

Theorem 5.3.4 Let $(X_1, X_2, \ldots)$ be a sequence of i.i.d. random variables with $\mathbb{E}(X_k) = \mu$ and $\mathbf{Var}(X_k) = \sigma^2 \in (0, \infty)$ for any $k \geqslant 1$. If $Y_n = X_1 + \ldots + X_n$ then

$$\frac{Y_n - n\mu}{\sqrt{\sigma^2 n}} \Longrightarrow \mathcal{N}(0,1) \qquad \text{as} \quad n \to \infty$$

which means that the sequence of random variables $\left(\frac{Y_n - n\mu}{\sqrt{\sigma^2 n}}\right)_n$ converges weakly to some random variable X with $X \sim \mathcal{N}(0,1)$.

Proof. With the tools we introduced just before, the proof of this fundamental result becomes very easy ! Indeed, set

$$\overline{X}_k = X_k - \mu$$

so that $\mathbb{E}(\overline{X}_k) = 0$ and $\mathbf{Var}(\overline{X}_k) = \sigma^2$ for any $k \geqslant 1$. Set $\overline{Y}_n = \sum_{k=1}^n \overline{X}_k$. From Prop. 5.3.1, one

$$\Phi_{\overline{X}_k}(t) = 1 - \sigma^2 \frac{t^2}{2} + t^2 \delta(t^2)$$

where $\delta(h)/h \to 0$ as $h \to 0$ (notice that, since the $\overline{X}_i$ are identically distributed, $\delta(\cdot)$ does not depend on k). Now, from the independence of $(X_1, X_2, \ldots)$, the characteristic function Φ_n of $\frac{\overline{Y}_n}{\sqrt{\sigma^2 n}}$ is

$$\Phi_n(t) = \mathbb{E}(e^{it\frac{\overline{Y}_n}{\sqrt{\sigma^2 n}}}) = \left(\Phi_{X_1}\left(\frac{t}{\sqrt{\sigma^2 n}}\right)\right)^n = \left(1 - \sigma^2 \frac{t^2}{2\sigma^2 n} + \frac{t^2}{\sigma^2 n}\delta\left(\frac{t^2}{\sigma^2 n}\right)\right)^n$$

i.e.

$$\Phi_n(t) = \left(1 - \frac{t^2}{2n} + \frac{t^2}{\sigma^2 n}\delta\left(\frac{t^2}{\sigma^2 n}\right)\right)^n \qquad \forall n \in \mathbb{N}.$$

It is a classical exercises of complex analysis (see [4, Theorem 4.2, p. 112]) to prove that, if $(z_n)_n \subset \mathbb{C}$ is such that $z_n \to z \in \mathbb{C}$, then

$$\lim_n \left(1 + \frac{z_n}{n}\right)^n = \lim_n \left(1 + \frac{z}{n}\right)^n = e^z.$$

Let now $t \in \mathbb{R}$ be fixed. Applying this with $z_n = \frac{t^2}{2} + \frac{t^2}{\sigma^2}\delta\left(\frac{t^2}{\sigma^2 n}\right)$ with $z_n \to \frac{t^2}{2}$, one gets that

$$\lim_n \Phi_n(t) = e^{\frac{t^2}{2}}$$

With (5.18), one recognizes in the right hand side the characteristic function of the normal distribution $\mathcal{N}(0,1)$. Therefore, according to Prop. 5.3.2,

$$\frac{\overline{Y}_n}{\sqrt{\sigma^2 n}} \implies \mathcal{N}(0,1).$$

Turning back to Y_n we actually get $\dfrac{Y_n}{\sqrt{\sigma^2 n}} \implies \mathcal{N}(0,1)$ and the result is proven. ■

Remark 5.3.2 The Central Limit Theorem is maybe one of the most stunning result in probability theory. The most stunning thing is of course that the (weak) limit is the same for **any** sequence of i.i.d. This illustrates somehow the universal nature of the normal distribution and appears quite surprising. If you think of X_n to be uniformly distributed or X_n to be, say, a Poisson random variable , at the end, the approximation is always a normal distribution!

Remark 5.3.3 The Central Limit Theorem can be also reformulated as follows: let $(X_1, X_2, \ldots)$ be a sequence of i.i.d. random variables with $\mathbb{E}(X_k) = \mu$ and $\mathbf{Var}(X_k) = \sigma^2 \in (0,\infty)$ for any $k \geqslant 1$. If $Y_n = X_1 + \ldots + X_n$ then

$$\mathbb{P}\left(\frac{Y_n - n\mu}{\sqrt{\sigma^2 n}} \leqslant x\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt \qquad \forall x \in \mathbb{R}$$

when we turned back to our initial definition of weak convergence.

The following illustrating example is taken from [4]:

Exercise 5.11 — Coin flips. Let $(X_1, X_2, \ldots)$ be i.i.d. with $P(X_k = 0) = P(X_k = 1) = 1/2$ for any $k \geqslant 1$. We have in mind here a Bernoulli trial associated to coin flips and the event $\{X_k = 1\}$ indicates that a heads occurred on the k-th toss. Then,

$$Y_n = X_1 + \ldots + X_n$$

is the total number of heads at time n (i.e. after n tosses). One has

$$\mathbb{E}(X_k) = 1/2 \qquad \text{and} \qquad \mathbf{Var}(X_k) = 1/4.$$

Hence, the Central Limit Theorem tells us that

$$\mathbb{P}\left(\frac{Y_n - \frac{n}{2}}{\sqrt{n/4}} \leqslant x\right) \to \mathbb{P}(X \leqslant x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt \qquad \forall x \in \mathbb{R}$$

where $X \sim \mathcal{N}(0,1)$. Thus

$$\mathbb{P}\left(\frac{Y_n - \frac{n}{2}}{\sqrt{n/4}} \in [-2,2]\right) = \mathbb{P}\left(\frac{Y_n - \frac{n}{2}}{\sqrt{n/4}} \leqslant 2\right) - \mathbb{P}\left(\frac{Y_n - \frac{n}{2}}{\sqrt{n/4}} \leqslant -2\right) \simeq \mathbb{P}(X \in [-2,2]).$$

Any table of the normal distribution shows that

$$\mathbb{P}(X > 2) = 1 - .9773 = .0227$$

i.e. $\mathbb{P}(|X| \leqslant 2) = 1 - 2(.0227) = .9546$. In other words, if n is large enough

$$\mathbb{P}\left(\frac{Y_n - \frac{n}{2}}{\sqrt{n/4}} \in [-2,2]\right) \simeq 0.95$$

Since

$$\mathbb{P}\left(\frac{Y_n - \frac{n}{2}}{\sqrt{n/4}} \in [-2,2]\right) = \mathbb{P}\left((Y_n - \frac{n}{2}) \in [-\sqrt{n}, \sqrt{n}]\right),$$

taking for instance $n = 10\,000$, one can say that, approximatively, 95% of the time the number of heads will be between 4900 and 5100. ■

5.4 Laplace transform

Given a mapping

$$f \;:\; t \in \mathbb{R} \mapsto f(t) \in \mathbb{C}$$

which is identically equal to zero for $t < 0$, i.e. $f(t) = 0$ for $t < 0$, we define, for any $p \in \mathbb{R}$ for which the integral makes sense, the *Laplace transform* of f as

$$F(p) = \mathcal{L}[f](p) = \int_0^\infty f(t)e^{-pt}\mathrm{d}t. \tag{5.19}$$

Of course, this is an integral over $[0,\infty)$ and not over $\mathbb{R}$. This is somehow the main difference between Fourier transform and Laplace transform. One will say that the Laplace transform exists if the integral (5.19) exists, i.e. $e^{-pt}f(t) \in L^1([0,\infty)$. We denote with

$$L^1_{\mathrm{loc}}(\mathbb{R}^+)$$

the set of all measurable mappings $f \;:\; [0,\infty) \to \mathbb{C}$ such that

$$\int_0^T |f(t)|\mathrm{d}t < \infty \qquad\qquad \forall T > 0.$$

One has the following

Lemma 5.4.1 Let $f \in L^1_{\mathrm{loc}}(\mathbb{R}^+)$ and $p_0 \in \mathbb{R}$.

1. if $\displaystyle\int_0^\infty |f(t)|e^{-p_0 t}\mathrm{d}t < \infty$ then

$$\int_0^\infty |f(t)|e^{-pt}\mathrm{d}t < \infty \qquad \forall p > p_0.$$

2. if $t \mapsto e^{-p_0 t}f(t)$ is bounded in the vicinity of $+\infty$, i.e.

$$\limsup_{t\to\infty} e^{-p_0 t}|f(t)| < \infty$$

then

$$\int_0^\infty |f(t)|e^{-pt}\mathrm{d}t < \infty \quad \forall p > p_0.$$

Remark 5.4.1 The previous Lemma asserts only that

1. if $\mathcal{L}[f](p_0)$ exists, then $\mathcal{L}[f](p)$ exists for any $p > p_0$.
2. If $e^{-p_0 t}f(t)$ is bounded in the vicinity of $+\infty$, then $\mathcal{L}[f](p)$ exists for any $p > p_0$.

One denotes with $\mathcal{A}$ the set of all mapping $f \in L^1_{\mathrm{loc}}(\mathbb{R}^+)$ satisfying the point (2) of the previous Lemma for some $p_0 \in \mathbb{R}$.

Definition 5.4.1 if $f \in \mathcal{A}$, one defines the convergence abscissa of f, denoted $\omega(f)$, the real number

$$\omega(f) = \inf\{p \in \mathbb{R} \text{ such that } e^{-pt}f(t) \text{ is bounded in the vicinity of } +\infty\}.$$

One has then the following:

Proposition 5.4.2 Let $f \in \mathcal{A}$, then $\omega(f) < \infty$ and

1. for any $p > \omega(f)$, $\mathcal{L}[f](p)$ exists,
2. if $p < \omega(f)$,

$$\int_0^\infty e^{-pt}\,|f(t)|\,\mathrm{d}t = \infty.$$

The convergence abscissa $\omega(f)$ is therefore the smallest real number p_0 from which one can define the Laplace transform of f

Example 5.8 **a)** If $f(t) = H(t) = \mathbf{1}_{[0,\infty)}$, then

$$\mathcal{L}[f](p) = \int_0^\infty e^{-pt}\mathrm{d}t = \frac{1}{p} \qquad \forall p > 0,$$

et $\int_0^\infty e^{-pt}\mathrm{d}t = \infty\ \forall p \leqslant 0$. Therefore $\omega(H) = 0$.

b) If $f(t) = tH(t)$, then using integration by parts,

$$\mathcal{L}[f](p) = \int_0^\infty te^{-pt}\mathrm{d}t = \frac{1}{p^2}, \qquad \forall p > 0$$

whereas $\int_0^\infty te^{-pt}\mathrm{d}t = \infty$ for $p \leqslant 0$. Again $\omega(f) = 0$.

c) If $f(t) = \cos(t)\,H(t)$, then

$$\begin{aligned}\mathcal{L}[f](p) &= \int_0^\infty \cos(t)e^{-pt}\mathrm{d}t = \int_0^\infty \mathrm{Re}(e^{it})e^{-pt}\mathrm{d}t \\ &= \mathrm{Re}\left(\int_0^\infty e^{-(p-i)t}\mathrm{d}t\right) = \mathrm{Re}\left(\frac{1}{p-i}\right) = \frac{p}{p^2+1}, \qquad \forall p > 0\end{aligned}$$

and $\int_0^\infty |\cos(t)|\,e^{-pt}\mathrm{d}t = \infty$ for $p < 0$. ■

As it is the case for the Fourier transform, the Laplace transform enjoys several nice properties

Theorem 5.4.3 One has

1. **Linearity** If $f_1, f_2 \in \mathcal{A}$ and $\lambda_1, \lambda_2 \in \mathbb{C}$ then
$$\mathcal{L}\left[\lambda_1 f_1 + \lambda_2 f_2\right](p) = \lambda_1 \mathcal{L}[f_1](p) + \lambda_2 \mathcal{L}[f_2](p), \qquad \forall p > \max(\omega(f_1)\,;\omega(f_2)).$$
2. **Continuity:** For $f \in \mathcal{A}$, the mapping $p \mapsto \mathcal{L}[f](p)$is continuous over $(\,\omega(f), +\infty[$.
3. **Riemann-Lebesgue:** For any $f \in \mathcal{A}$,
$$\lim_{p\to+\infty} \mathcal{L}[f](p) = 0.$$
4. **Initial value theorem:** If $\lim_{t\to 0^+} f(t) = f(0^+)$ exists then
$$\lim_{p\to\infty} p\mathcal{L}[f](p) = f(0^+).$$
5. **Final value theorem:** If $\lim_{t\to\infty} f(t) = \ell$ exists, then $\omega(f) \leqslant 0$ and
$$\lim_{p\to 0} p\mathcal{L}[f](p) = \ell.$$

Proof. The proof is similar to that of the analogue properties of the Fourier transform. Let us just show the points (4) and (5). One assumes without loss of generality that $f(0^+) = 0$. Then, for any $\varepsilon > 0$, there is $\tau > 0$ such that $|f(t)| < \frac{\varepsilon}{2}$ for any $0 < t < \tau$. In this case,

$$\begin{aligned}|p\mathcal{L}[f](p)| &= \left|\int_0^\tau pf(t)e^{-pt}\mathrm{d}t + \int_\tau^\infty pf(t)e^{-pt}\mathrm{d}t\right| \leqslant p\frac{\varepsilon}{2}\int_0^\tau e^{-pt}\mathrm{d}t + \int_\tau^\infty p|f(t)|e^{-pt}\mathrm{d}t \\ &\leqslant \frac{\varepsilon}{2}\left(1 - e^{-p\tau}\right) + I(p)\end{aligned}$$

where
$$I(p) = \int_\tau^\infty p|f(t)|e^{-pt}\mathrm{d}t.$$

One only has to prove that, for p large enough it holds $I(p) \leqslant \frac{\varepsilon}{2}$. Let $\omega(f)$ be the convergence abscissa of f and let $\alpha > \omega(f)$. One has

$$I(p) = p\int_\tau^\infty |f(t)|e^{-\alpha t}e^{-(p-\alpha)t}\mathrm{d}t \leqslant pe^{-(p-\alpha)\tau}\int_\tau^\infty |f(t)|e^{-\alpha t}\mathrm{d}t \leqslant pe^{-(p-\alpha)\tau}\mathcal{L}[f](\alpha).$$

Since $\mathcal{L}[f](\alpha)$ is a given complex number with finite modulus (because $\alpha > \omega(f)$), one has
$$\lim_{p\to\infty} I(p) = 0$$
so that, for p large enough $I(p) \leqslant \frac{\varepsilon}{2}$. One obtains then
$$p\mathcal{L}[f](p) \leqslant \varepsilon, \qquad \text{for } p \text{ large enough}$$
i.e. $\lim_{p\to\infty} p\mathcal{L}[f](p) = 0$. If now $f(0^+) \neq 0$, one sets $g(t) = f(t) - f(0^+)H(t)$. One has then $\lim_{t\to 0^+} g(t) = 0$ and, from the preceding point $\lim_{p\to\infty} p\mathcal{L}[g](p) = 0$. Since $\mathcal{L}[f](p) = \mathcal{L}[g](p) + \frac{f(0^+)}{p}$, one has point (4). The point (5) follows the same proof. ■

For any $f \in \mathcal{A}$ and any $a \in \mathbb{R}\setminus\{0\}$, one denotes with f_a the mapping $f_a(t) = f(at)$ and $\tau_a f$ the translation
$$\tau_a f(t) = f(t-a), \qquad t \in \mathbb{R}.$$

One has then

Theorem 5.4.4 Let $f \in \mathcal{A}$. One has

$$\mathcal{L}[\tau_a f](p) = e^{-pa}\mathcal{L}(p), \qquad \forall p > \omega(f).$$

If $g(t) = e^{at}f(t)$ one has

$$\mathcal{L}[g](p) = \mathcal{L}[f](p-a) = \tau_a\left(\mathcal{L}[f]\right)(p), \qquad \forall p > \omega(f) + a.$$

Finally

$$\mathcal{L}[f_a](p) = \frac{1}{a}\mathcal{L}[f]\left(\frac{p}{a}\right)$$

for any $p > a\omega(f)$.

5.4.1 Differentiation and integration

One starts with the following

Theorem 5.4.5 Let $f \in \mathcal{A}$. Assume that f is differentiable over $\mathbb{R}^+ \setminus \{0\}$ with continuous derivative. Then,

$$\mathcal{L}[f'](p) = p\mathcal{L}[f](p) - f(0^+), \qquad \forall p > \omega(f).$$

Proof. One has

$$\mathcal{L}[f'](p) = \int_0^\infty f'(t)e^{-pt}\mathrm{d}t$$

and, by integration by parts,

$$\mathcal{L}[f'](p) = \left[f(t)e^{-pt}\right]_0^\infty + \int_0^\infty pf(t)e^{-pt}\mathrm{d}t$$

Since $f \in \mathcal{A}$, f decreases exponentially fast and $\lim_{t\to\infty} f(t)e^{-pt} = 0$ for any $p > \omega(f)$ and

$$\mathcal{L}[f'](p) = -\lim_{t\to 0^+} f(t) + p\int_0^\infty f(t)e^{-pt}\mathrm{d}t.$$

One has then $\mathcal{L}[f'](p) = \mathcal{L}[f](p) - f(0^+)$ $\forall p > \omega(f)$. ■

One deduces by induction the following

Corollary 5.4.6 If $f \in \mathcal{A}$ is of class $\mathcal{C}^n$ on $\mathbb{R}_+ \setminus \{0\}$ then,

$$\mathcal{L}[f^{(n)}](p) = p^n\mathcal{L}[f](p) - \left(p^{n-1}f(0^+) + p^{n-2}f'(0^+) + \cdots + f^{(n-1)}(0^+)\right), \qquad \forall p > \omega(f).$$

As for the Fourier transform one can recognize the original function from derivatives of $\mathcal{L}[f]$.

Theorem 5.4.7 Let $f \in \mathcal{A}$, then, for any $n \in \mathbb{N}$, the mapping $t \mapsto t^n f(t)$belongs to $\mathcal{A}$ and

$$\mathcal{L}[t^n f(t)](p) = (-1)^n \frac{\mathrm{d}^n}{\mathrm{d}p^n}\mathcal{L}[f](p), \qquad \forall p > \omega(f).$$

Remark 5.4.2 In particular, one sees that if $f \in \mathcal{A}$ then $\mathcal{L}[f]$ is of class $\mathcal{C}^\infty$.

Proof. It suffices to observe that $\frac{\mathrm{d}}{\mathrm{d}p}e^{-pt} = -te^{-pt}$ and therefore, by derivation of integral depending on a parameter,

$$\frac{\mathrm{d}}{\mathrm{d}p}\mathcal{L}[f](p) = \frac{\mathrm{d}}{\mathrm{d}p}\int_0^\infty f(t)e^{-pt}\mathrm{d}t = -\int_0^\infty tf(t)e^{-pt}\mathrm{d}t = -\mathcal{L}[tf(t)](p).$$

One proceeds in the same way for the higher order derivatives. ■

If $f, g \in \mathcal{A}$, then both $f$and g are vanishing on $\mathbb{R}_-$ and, in particular, for $t > 0$, one has $f(t-s)g(s) = 0$ whenever $s < 0$ or $t - s < 0$. Then, the convolution

$$[f * g](t) = \int_{\mathbb{R}} f(t-s)g(s)\mathrm{d}s = \int_0^\infty f(t-s)g(s)\mathrm{d}s = \int_0^t f(t-s)g(s)\mathrm{d}s,$$

i.e.

$$[f * g](t) = \int_0^t f(t-s)g(s)\mathrm{d}s = \int_0^t f(s)g(t-s)\mathrm{d}s, \qquad \forall t > 0.$$

One has, in analogy with the Fourier transform,

Theorem 5.4.8 Let $f, g \in \mathcal{A}$. Then

$$\mathcal{L}[f * g](p) = \mathcal{L}[f](p)\,\mathcal{L}[g](p), \qquad \forall p > \max(\omega(f), \omega(g)).$$

The proof is an easy ***Exercise***. Notice that, if g is identically equal to 1 over $\mathbb{R}^+$, i.e. $g = H$ then

$$[f * g](t) = [f * H](t) = \int_0^t f(s)\mathrm{d}s$$

is the primitive of f which vanishes at 0. Since $\mathcal{L}[H](p) = 1/p$ for any $p > 0$, one has

Corollary 5.4.9 Let $f \in \mathcal{A}$ with convergence abscissa $\omega(f)$. For any $p > \max(\omega(f), 0)$ one has

$$\mathcal{L}\left[\int_0^t f(s)\mathrm{d}s\right](p) = \frac{1}{p}\mathcal{L}[f](p).$$

For periodic functions, one has

Theorem 5.4.10 Let $f \in \mathcal{A}$ be a periodic function with period $T > 0$. Then $\omega(f) = 0$ and

$$\mathcal{L}[f](p) = \frac{1}{1-e^{-pT}}\int_0^T e^{-ps}f(s)\mathrm{d}s, \qquad \forall p > 0.$$

Proof. The proof is made by splitting $[0, \infty[$ in a infinite sequence of intervals, each of them with length T. Then, for any p,

$$\mathcal{L}[f](p) = \int_0^\infty e^{-pt}f(t)\mathrm{d}t = \sum_{k=0}^\infty \int_{kT}^{(k+1)T} e^{-pt}f(t)\mathrm{d}t.$$

For any $k \in \mathbb{N}$, the change of variables $t = s + kT$ with $s \in [0, T]$ leads

$$\mathcal{L}[f](p) = \sum_{k=0}^{\infty} e^{-pkT} \int_0^T e^{-ps} f(s + kT)\mathrm{d}s = \sum_{k=0}^{\infty} e^{-pkT} \int_0^T e^{-ps} f(s)\mathrm{d}s$$

where we used that $f(s + kT) = f(s)$ for any $s \in \mathbb{R}$ and any $k \in \mathbb{N}$. This last integral does not depend on p anymore and therefore, if $\sum_{k=0}^{\infty} e^{-pkT}$ converges then the Laplace transform of f is well-defined. Clearly, if $p \leqslant 0$, the series diverge whereas it converges for $p > 0$. This shows that $\omega(f) = 0$ and one deduces the result, ■

We end this section with the injective property of the Laplace transform

Theorem 5.4.11 Let $f_1, f_2 \in \mathcal{A}$. If there exists p^* such that $\mathcal{L}[f_1](p) = \mathcal{L}[f_2](p)$ for any $p \geqslant p^*$ then $f_1(t) = f_2(t)$ for almost every $t \in \mathbb{R}$.

5.5 Complements

5.5.1 Vector differentiation and Sobolev spaces

The goal of this section is to interpret the notion of derivative in terms of translation operators. For simplicity, we shall consider here only functions f taking real values (of course, their Fourier transform will then take complex values).

Definition 5.5.1 Given $1 \leqslant p < \infty$, $f \in L^p(\mathbb{R}^d)$ and $a \in \mathbb{R}^d$, we say that the derivative of f in the direction of a exists in the L^p-sense if there exists $g \in L^p$ such that

$$\lim_{\varepsilon \to 0} \|g + \tfrac{1}{\varepsilon}\left(\tau_{\varepsilon\, a} f - f\right)\|_p = 0$$

and we set then

$$g = D_a f.$$

In other words, the notion of derivative in L^p reads

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\left(f(\cdot - \varepsilon\, a) - f(\cdot)\right) = -D_a f = -g$$

where the convergence is meant in L^p. This clearly generalizes the notion of derivative in the classical sense (i.e. pointwise). We introduce then

$$W_1^p(\mathbb{R}^d) = \left\{ f \in L^p(\mathbb{R}^d)\,;\, D_a f \text{ exists for any } a \in \mathbb{R}^d \right\}.$$

If $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\}$ denotes the standard basis of $\mathbb{R}^d$, one sees that

$$W_1^p(\mathbb{R}^d) = \left\{ f \in L^p(\mathbb{R}^d)\,;\, D_{\boldsymbol{e}_k} f \text{ exists for any } k \in \{1, \ldots, d\} \right\}.$$

Proposition 5.5.1 Given $f \in W_1^1(\mathbb{R}^d)$, one has

$$\mathcal{F}(D_a f)(\xi) = i(a \cdot \xi)\mathcal{F}(f)(\xi) \qquad \forall \xi \in \mathbb{R}^d.$$

Proof. Since $D_a f \in L^1(\mathbb{R}^d)$, its Fourier transform is well-defined. Moreover, for any $\varepsilon > 0$,

$$\frac{1}{\varepsilon}\mathcal{F}(\tau_{\varepsilon\, a} f - f)(\xi) = \frac{\mathrm{e}^{-i\varepsilon a \cdot \xi} - 1}{\varepsilon}\mathcal{F}(f)(\xi) \qquad \forall \xi \in \mathbb{R}^d.$$

Since
$$\mathcal{F}(D_a f)(\xi) = -\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\mathcal{F}(\tau_{\varepsilon a} f - f)(\xi)$$
(Explain why), we get the result. ■

Corollary 5.5.2 If $f \in W_1^1(\mathbb{R}^d)$ then $\mathcal{F}(\xi) = o(\|\xi\|^{-1})$ as $\|\xi\| \to \infty$.

Proof. Since $\lim_{\|\xi\|\to\infty} \mathcal{F}(D_a f)(\xi) = 0$ according to Riemann-Lebesgue lemma, we get the result. ■

We introduce another notion of differentiation

Definition 5.5.2 — Weak derivative. Given $1 \leqslant p < \infty$, $f \in L^p(\mathbb{R}^d)$, we say that f admits a weak derivative in the k-th direction if there exists $u_k \in L^1_{loc}(\mathbb{R}^d)$ such that
$$\int_{\mathbb{R}^d} f\frac{\partial \varphi}{\partial x_k}\mathrm{d}x = -\int_{\mathbb{R}^d} u_k(x)\varphi(x)\mathrm{d}x \qquad \forall \varphi \in \mathscr{C}_c^\infty(\mathbb{R}^d)$$
and we write
$$\partial_k f = u_k.$$
Here $L^1_{loc}(\mathbb{R}^d)$ is the set of functions u such that $\int_K |u(x)|\mathrm{d}x < \infty$ for any compact subset K of $\mathbb{R}^d$.

The two notions of differentiation somehow coincide:

Theorem 5.5.3 Given $1 \leqslant p < \infty$ and $f \in L^p(\mathbb{R}^d)$ the following are equivalent:
1. $f \in W_1^p(\mathbb{R}^d)$
2. f is weakly differentiable in the k-th direction for any $k \in \{1, \dots, d\}$ and $\partial_k f \in L^p(\mathbb{R}^d)$.

Moreover, in that case $D_{\boldsymbol{e}_k} f = \partial_k f$.

Proof. We shall only prove the fact that $i) \implies ii)$. Given $h \in L^q(\mathbb{R}^d)$, $k \in \{1, \dots, d\}$ and $\varepsilon > 0$, one has
$$\int_{\mathbb{R}^d} \frac{1}{\varepsilon}\left(\tau_{\varepsilon \boldsymbol{e}_k} f - f\right) h\mathrm{d}x = \int_{\mathbb{R}^d} \frac{1}{\varepsilon}\left(\tau_{-\varepsilon \boldsymbol{e}_k} h - h\right) f\mathrm{d}x$$
(check this). In particular, if $h = \varphi \in \mathscr{C}^\infty$ one can pass to the limit in the left-hand-side since $\varphi \in L^q(\mathbb{R}^d)$ to get
$$-\int_{\mathbb{R}^d} D_{\mathrm{e}_k} f\varphi = \lim_{\varepsilon\to 0}\int_{\mathbb{R}^d}\frac{1}{\varepsilon}\left(\tau_{-\varepsilon \boldsymbol{e}_k}\varphi - \varphi\right) f\mathrm{d}x.$$
Moreover, since $\varphi \in \mathscr{C}^\infty$, for any $x \in \mathbb{R}^d$
$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(\tau_{-\varepsilon \boldsymbol{e}_k}\varphi(x) - \varphi(x)\right) f(x) = \frac{\partial \varphi}{\partial x_k}(x) f(x)$$
and actually the convergence is uniform on the compact K which is the support of φ. Therefore
$$\lim_{\varepsilon\to 0}\int_{\mathbb{R}^d}\frac{1}{\varepsilon}\left(\tau_{-\varepsilon \boldsymbol{e}_k}\varphi - \varphi\right) f\mathrm{d}x = \int_{\mathbb{R}^d}\frac{\partial \varphi}{\partial x_k}(x) f(x)\mathrm{d}x.$$

This proves that f is weakly-differentiable with $\partial_k f = D_{\mathbf{e}_k}$. In particular, $\partial_k f \in L^p(\mathbb{R}^d)$. The proof of the other implication is not too difficult and resorts on Taylor formula for $\varphi \in \mathscr{C}_c^\infty(\mathbb{R}^d)$, see [7]. ■

Corollary 5.5.4 Let $(f_n)_n \subset W_1^p(\mathbb{R}^d)$ $(1 \leqslant p < \infty)$ be such that

$$\lim_{n\to\infty} \|f_n - f\|_p = 0$$

and such that, for any $k \in \{1,\ldots,d\}$, $(D_{\mathbf{e}_k} f_n)_n$ converges in $L^p(\mathbb{R}^d)$. Then, $f \in W_1^p(\mathbb{R}^d)$ and

$$\lim_{n\to\infty} \|D_{\mathbf{e}_k} f_n - D_{\mathbf{e}_k} f\|_p = 0 \qquad \forall k = 1,\ldots,d.$$

Proof. Let $k \in \{1,\ldots,d\}$ be fixed. To prove the result, it is enough to show that f is weakly differentiable in the k-th direction. Let $\varphi \in \mathscr{C}_c^\infty(\mathbb{R}^d)$ be given. One has

$$\int_{\mathbb{R}^d} \varphi\, D_{\mathbf{e}_k} f_n \mathrm{d}x = -\int_{\mathbb{R}^d} f_n \frac{\partial \varphi}{\partial x_k} \mathrm{d}x \qquad \forall n \in \mathbb{N}. \tag{5.20}$$

Moreover, since $(D_{\mathbf{e}_k} f_n)$ converges in $L^p(\mathbb{R}^d)$ to some limit, say $u_k \in L^p(\mathbb{R}^d)$, one has for the left-hand-side:

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} \varphi\, D_{\mathbf{e}_k} f_n \mathrm{d}x = \int_{\mathbb{R}^d} \varphi\, u_k \mathrm{d}x$$

while, since $\lim_n \|f_n - f\|_p = 0$, the right-hand-side of (5.20) satifies

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} f_n \frac{\partial \varphi}{\partial x_k} \mathrm{d}x = \int_{\mathbb{R}^d} f \frac{\partial \varphi}{\partial x_k} \mathrm{d}x.$$

Therefore, according to (5.20)

$$\int_{\mathbb{R}^d} \varphi\, u_k \mathrm{d}x = -\int_{\mathbb{R}^d} f \frac{\partial \varphi}{\partial x_k} \mathrm{d}x.$$

This shows that f is weakly-differentiable in the k-th direction with $\partial_k f = u_k$. In particular, $\partial_k f \in L^p(\mathbb{R}^d)$ so that $f \in W_1^p(\mathbb{R}^d)$. We also proved that $\lim_{n\to\infty} \|\partial_k f - D_{\mathbf{e}_k} f_n\| = 0$ which proves the second part of the result. ■

Corollary 5.5.5 Introduce, for a given $1 \leqslant p < \infty$,

$$\|f\|_{W_1^p} = \|f\|_p + \sum_{k=1}^d \|D_{\mathbf{e}_k} f\|_p \qquad \forall f \in W_1^p(\mathbb{R}^d).$$

Then, $(W_1^p(\mathbb{R}^d), \|\cdot\|_{W_1^p})$ is a Banach space.

Proof. The fact that $W_1^p(\mathbb{R}^d)$ is a vector space is easy to check while the fact that $\|\cdot\|_{W_1^p}$ is a norm is left as an **Exercise**. Let us prove that $(W_1^p(\mathbb{R}^d), \|\cdot\|_{W_1^p})$ is complete. Let $(f_n)_n \subset W_1^p(\mathbb{R}^d)$ be a Cauchy sequence. Then, in particular, $(f_n)_n$ is a Cauchy sequence in $L^p(\mathbb{R}^d)$ and, for any $k \in \{1,\ldots,d\}$, $(D_{\mathbf{e}_k} t f_n)_n$ is a Cauchy sequence in $L^p(\mathbb{R}^d)$. In particular, there is $f \in L^p(\mathbb{R}^d)$ such that $\lim_n \|f_n - f\|_p = 0$ while $(D_{\mathbf{e}_k} f_n)_n$ converges for any k. According to the previous Corollary, $f \in W_1^p(\mathbb{R}^d)$ and $\lim_n \|D_{\mathbf{e}_k} f_n - D_{\mathbf{e}_k} f\|_p = 0$ for any k. Thus, $\lim_n \|f_n - f\|_{W_1^p} = 0$ which proves the result. ■

We can extend the above to higher-order derivatives:

Definition 5.5.3 Given $k \in \mathbb{N}$ and $1 \leqslant p < \infty$, we define

$$W_k^p(\mathbb{R}^d) = \{f \in W_1^p(\mathbb{R}^d)\ ;\ D_a f \in W_{k-1}^p(\mathbb{R}^d \quad \forall a \in \mathbb{R}^d\}$$

with the convention that $W_0^p(\mathbb{R}^d) = L^p(\mathbb{R}^d)$.

For $f \in W_k^p(\mathbb{R}^d)$, we define the inductively $D_a^k f$ by

$$D_a^k f = D_a(D^{k-1} f_a) \qquad k \geqslant 2.$$

Clearly one has

Proposition 5.5.6 Introduce, for a given $1 \leqslant p < \infty$ and $k \in \mathbb{N}$:

$$\|f\|_{W_k^p} = \|f\|_p + \sum_{\ell=1}^{k} \sum_{j=1}^{d} \|D_{\mathbf{e}_j}^{\ell} f\|_p \qquad \forall f \in W_k^p(\mathbb{R}^d).$$

Then, $(W_k^p(\mathbb{R}^d), \|\cdot\|_{W_k^p})$ is a Banach space.

We can extend Proposition 5.5.1 to $W_1^2(\mathbb{R}^d)$:

Proposition 5.5.7 Let $f \in W_1^2(\mathbb{R}^d)$. Then

$$\mathcal{F}(D_{\mathbf{e}_k} f)(\xi) = i\xi_k \mathcal{F}(f)(\xi) \qquad \text{for a.e. } \xi \in \mathbb{R}^d,\ \ k = 1, \ldots, d.$$

Proof. Let $f \in W_1^2(\mathbb{R}^d)$ be given. Since $f \in L^2(\mathbb{R}^d)$, its Fourier-Plancherel transform $\mathcal{F}(f) \in L^2(\mathbb{R}^d)$ is well-defined. Moreover

$$\mathcal{F}(\tau_{\varepsilon\,\mathbf{e}_k} f)(\xi) = \mathbf{e}^{-i\varepsilon\xi_k} \mathcal{F}(f)(\xi) \in L^2(\mathbb{R}^d)$$

for any $k = 1, \ldots, d$ and any $\varepsilon > 0$. Therefore, for any $k \in \{1, \ldots, d\}$:

$$\frac{1}{\varepsilon} \mathcal{F}\left(\tau_{\varepsilon \mathbf{e}_k} f - f\right) = \frac{\mathbf{e}^{-i\varepsilon\xi_k} - 1}{\varepsilon} \mathcal{F}(f)(\xi).$$

Since the left-hand side converges in $L^2(\mathbb{R}^d)$ to $-\mathcal{F}(D_{\mathbf{e}_k} f)$, the right-hand side also converges in L^2. Taking a subsequence (ε_n) is necessary, the convergences occurs for almost every ξ, i.e.

$$\lim_{\varepsilon \to 0^+} \frac{\mathbf{e}^{i\varepsilon\xi_k} - 1}{\varepsilon} \mathcal{F}(f)(\xi) = -\mathcal{F}(D_{\mathbf{e}_k} f)(\xi) \qquad \text{for a.e. } \xi \in \mathbb{R}^d.$$

Since, for almost every $\xi \in \mathbb{R}^d$

$$\lim_{\varepsilon \to 0^+} \frac{\mathbf{e}^{-i\varepsilon\xi_k} - 1}{\varepsilon} \mathcal{F}(f)(\xi) = -i\xi_k \mathcal{F}(f)(\xi)$$

we get the result. ∎

Theorem 5.5.8 Introduce

$$H^1(\mathbb{R}^d) = \left\{ f \in L^2(\mathbb{R}^d) \ ; \ \int_{\mathbb{R}^d} \left(1 + \|\xi\|^2\right) |\mathcal{F}(f)(\xi)|^2 \mathrm{d}\xi < \infty \right\}.$$

Then,

$$H^1(\mathbb{R}^d) = W_1^2(\mathbb{R}^d).$$

Proof. Assume $f \in W_1^2(\mathbb{R}^d)$ and let $h = \mathcal{F}(f)$ denote its Fourier-Plancherel transform. Recall that $h \in L^2(\mathbb{R}^d)$. According to the previous Proposition

$$\mathcal{F}(D_{\boldsymbol{e}_k} f)(\xi) = i\xi_k \mathcal{F}(f)(\xi)$$

holds for almost every $\xi \in \mathbb{R}^d$ and any $k = 1, \ldots, d$. Since $D_{\boldsymbol{e}_k} f \in L^2(\mathbb{R}^d)$, also its Fourier-Plancherel transform lies in $L^2(\mathbb{R}^d)$ and therefore $\xi \mapsto i\xi_k \mathcal{F}(f)(\xi)$ belongs to $L^2(\mathbb{R}^d)$, i.e.

$$\int_{\mathbb{R}^d} |i\xi_k \mathcal{F}(f)(\xi)|^2 \,\mathrm{d}\xi = \int_{\mathbb{R}^d} |\xi_k|^2 \, |\mathcal{F}(f)(\xi)|^2 \mathrm{d}\xi < \infty$$

for any $k \in \{1, \ldots, d\}$. Summing up all the indices k, we get

$$\int_{\mathbb{R}^d} \|\xi\|^2 \, |\mathcal{F}(f)(\xi)|^2 \mathrm{d}\xi < \infty$$

which, with $F(f) \in L^2(\mathbb{R}^d)$ gives

$$\int_{\mathbb{R}^d} \left(1 + \|\xi\|^2\right) |\mathcal{F}(f)(\xi)|^2 \mathrm{d}\xi < \infty$$

and shows that $f \in H^1(\mathbb{R}^d)$. Conversely, assume that $f \in H^1(\mathbb{R}^d)$. Let $\varphi \in \mathscr{C}_c^\infty(\mathbb{R}^d)$ be given. According to Plancherel Formula

$$\int_{\mathbb{R}^d} f \overline{\frac{\partial \varphi}{\partial x_k}} \mathrm{d}x = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathcal{F}(f)(\xi) \overline{\mathcal{F}\left(\frac{\partial \varphi}{\partial x_k}\right)(\xi)} \mathrm{d}\xi$$

with (see for instance Theorem 5.1.3)

$$\mathcal{F}\left(\frac{\partial \varphi}{\partial x_k}\right)(\xi) = i\xi_k \mathcal{F}(\varphi)(\xi).$$

Thus,

$$\int_{\mathbb{R}^d} f \overline{\frac{\partial \varphi}{\partial x_k}} \mathrm{d}x = -\frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} i\xi_k \mathcal{F}(f)(\xi) \overline{\mathcal{F}(\varphi)(\xi)} \mathrm{d}\xi. \tag{5.21}$$

Since $f \in H^1(\mathbb{R}^d)$, $\xi \mapsto i\xi_k \mathcal{F}(f)(\xi)$ belongs to $L^2(\mathbb{R}^d)$. Applying the inverse Fourier isomorphism $\mathcal{F}^{-1}$, one sees that there exists a function $u_k \in L^2(\mathbb{R}^d)$ such that

$$\mathcal{F}(u_k)(\xi) = i\xi_k \mathcal{F}(f)(\xi).$$

Then, from (5.21),

$$\int_{\mathbb{R}^d} f \overline{\frac{\partial \varphi}{\partial x_k}} \mathrm{d}x = -\frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathcal{F}(u_k)(\xi) \overline{\mathcal{F}(\varphi)(\xi)} \mathrm{d}\xi.$$

Using again Plancherel formula, one has

$$\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\mathcal{F}(u_k)(\xi)\overline{\mathcal{F}(\varphi)(\xi)}\mathrm{d}\xi=\int_{\mathbb{R}^d}u_k(x)\overline{\varphi(x)}\mathrm{d}x$$

so that

$$\int_{\mathbb{R}^d}f\overline{\frac{\partial\varphi}{\partial x_k}}\mathrm{d}x=-\int_{\mathbb{R}^d}u_k(x)\overline{\varphi(x)}\mathrm{d}x.$$

This shows that u_k is the weak-derivative of f in the k-th direction and, since $u_k\in L^2(\mathbb{R}^d)$ and $k\in\{1,\ldots,d\}$ is arbitrary, we get $f\in W_1^2(\mathbb{R}^d)$. ■

The above result admits the straightforward extension to higher-order derivatives:

Proposition 5.5.9 Given $k\in\mathbb{N}$, one has

$$W_k^2(\mathbb{R}^d)=\left\{f\in L^2(\mathbb{R}^d)\;;\;\int_{\mathbb{R}^d}\left(1+\|\xi\|^2\right)^k|\mathcal{F}(f)(\xi)|^2\mathrm{d}\xi<\infty\right\}.$$

5.5.2 Schwarz Class and tempered distributions

We introduce here the Schwarz class on $\mathbb{R}^d$ which is the space of rapidly decreasing functions:

$$\mathscr{S}=\left\{f\in\mathscr{C}^\infty(\mathbb{R}^d)\;:\;\|f\|_{\boldsymbol{\alpha},\beta}<\infty\quad\forall\boldsymbol{\alpha},\in\mathbb{N}^d,\beta\in\mathbb{N}\right\},$$

where

$$\|f\|_{\boldsymbol{\alpha},\beta}=\sup_{x\in\mathbb{R}^d}\left(1+\|x\|\right)^\beta\|\boldsymbol{\partial}^{\boldsymbol{\alpha}}f(x)\|.$$

■ **Example 5.9** If $f\in\mathscr{C}_c^\infty(\mathbb{R}^d)$, then $f\in\mathscr{S}$. ■

■ **Example 5.10** If P is a polynomial function over $\mathbb{R}^d$ and $a>0$, then

$$f(x)=P(x)\mathbb{G}_a(x)\in\mathscr{S}.$$

■

Remark 5.5.1 Notice that $\mathscr{S}$ is clearly a vector space. Moreover, if $f,g\in\mathscr{S}$, using Leibniz rule one has

$$fg\in\mathscr{S}.$$

Notice also that $\|\cdot\|_{\boldsymbol{\alpha},\beta}$ is a family of semi-norms on $\mathscr{S}$.

Definition 5.5.4 Let $f\in\mathscr{S}$ and $(f_n)_n\subset\mathscr{S}$. We say that (f_n) converges to f in $\mathscr{S}$ if

$$\lim_n\|f_n-f\|_{\boldsymbol{\alpha},\beta}=0$$

for any $\boldsymbol{\alpha}\in\mathbb{N}^d$ and any $\beta\in\mathbb{N}$.

Notice that $\mathscr{S}\subset L^1(\mathbb{R}^d)\cap L^2(\mathbb{R}^d)$ so that, for any $f\in\mathscr{S}$, the Fourier transform

$$\mathcal{F}(f)(\xi)=\int_{\mathbb{R}^d}\mathbf{e}^{-i\xi\cdot x}f(x)\mathrm{d}x$$

is well-defined for any $\xi\in\mathbb{R}^d$. One has then the following

Theorem 5.5.10 If $f \in \mathscr{S}$ then $\mathcal{F}(f) \in \mathscr{S}$.

Proof. Since, for any $\beta \in \mathbb{N}$, one has $x \mapsto (1+\|x\|)^{\beta} f(x)$ still belongs to $\mathscr{S}$ and thus to $L^1(\mathbb{R}^d)$, one can apply Corollary 5.1.4 to check that $\mathcal{F}(f) \in \mathscr{C}^{\infty}(\mathbb{R}^d)$ with

$$\boldsymbol{\partial}^{\boldsymbol{\alpha}} \mathcal{F}(f)(\xi) = (-i)^{\beta} \mathcal{F}(\boldsymbol{x}^{\boldsymbol{\alpha}} f)(\xi) \qquad \forall \xi \in \mathbb{R}^d, \quad \boldsymbol{\alpha} \in \mathbb{N}^d \text{ with } |\boldsymbol{\alpha}| = \beta.$$

and actually, using the second part of Corollary 5.1.4, one also gets that $\mathcal{F}(f) \in \mathscr{S}$ ***(Check the details)***. ∎

One has then easily the following

Theorem 5.5.11 If $f \in \mathscr{S}$ then

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathbf{e}^{i\xi \cdot x} \mathcal{F}(f)(\xi) \mathrm{d}\xi \qquad \forall \xi \in \mathbb{R}^d$$

with moreover

$$\int_{\mathbb{R}^d} \overline{g}(x) f(x) \mathrm{d}x = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathcal{F}(f)(\xi) \overline{\mathcal{F}(g)(\xi)} \mathrm{d}\xi \qquad \forall f, g \in \mathscr{S}.$$

Definition 5.5.5 We define then the "dual space" $\mathscr{S}'$ as the space of all linear applications $\Phi : \mathscr{S} \to \mathbb{R}$ such that

$$\lim_n \Phi(f_n) = \Phi(f)$$

for any sequence (f_n) which converges to f in $\mathscr{S}$.
The elements of $\mathscr{S}'$ are called ***tempered distributions*** over $\mathbb{R}^d$.

Definition 5.5.6 Let $(\Phi_n)_n \subset \mathscr{S}'$ and $\Phi \in \mathscr{S}'$. We say that (Φ_n) converges to Φ in $\mathscr{S}'$ and write $\Phi_n \rightharpoonup \Phi$ if

$$\Phi_n(f) \to \Phi(f) \qquad \forall f \in \mathscr{S}.$$

■ **Example 5.11** Given $\boldsymbol{x} \in \mathbb{R}^d$, the Dirac mass $\delta_{\boldsymbol{x}}$ induces a tempered distribution by

$$\delta_{\boldsymbol{x}}(f) = f(\boldsymbol{x}) \qquad \forall f \in \mathscr{S}.$$

Clearly, the mapping $f \in \mathscr{S} \mapsto f(x) \in \mathbb{R}$ is continuous in the sense of $\mathscr{S}$, so $\delta_{\boldsymbol{x}} \in \mathscr{S}'$. ■

■ **Example 5.12** Suppose μ is a positive Borel measure on $\mathbb{R}^d$ such that

$$I_k := \int_{\mathbb{R}^d} \left(1+\|x\|^2\right)^{-k} \mu(\mathrm{d}x) < \infty$$

for some positive integer $k \in \mathbb{N}$, then the mapping

$$\Phi(f) = \int_{\mathbb{R}^d} f(x) \mu(\mathrm{d}x) \qquad \forall f \in \mathscr{S}$$

defines a tempered distribution. Indeed, suppose $(f_n)_n \in \mathscr{S}$ converges to 0 in $\mathscr{S}$. Then, setting

$$\varepsilon_n := \sup_{x \in \mathbb{R}^d} \left(1+\|x\|^2\right)^k |f_n(x)|$$

one has

$$\lim_n \varepsilon_n = 0$$

(Explain why). Since

$$|\Phi(f_n)| \leqslant \varepsilon_n I_k$$

one gets that $\lim_n |\Phi(f_n)| = 0$ which proves the continuity.

One usually simply says that μ is a tempered distribution, using implicitly the identification between μ and $\Phi(f) = \int_{\mathbb{R}^d} f \mathrm{d}\mu$. ■

■ **Example 5.13** If g is a measurable mapping such that

$$I := \int_{\mathbb{R}^d} \left[\left(1 + \|x\|^2\right)^{-k} |g(x)|\right]^p \mathrm{d}x$$

for some integer $k \in \mathbb{N}$ and $1 \leqslant p < \infty$. Then

$$\Phi(f) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} f\, g \mathrm{d}x \qquad \forall f \in \mathscr{S}$$

is a tempered distribution. In particular, any L^p-function defines a tempered distribution and any polynomial function defines a tempered distribution. ■

The space $\mathscr{S}$ being invariant under the Fourier transform $\mathcal{F}$ (see Theorem 5.5.10), it is very easy to extend the Fourier transform to the dual space $\mathscr{S}'$ as follows:

Definition 5.5.7 Let $\Phi \in \mathscr{S}'$, define then

$$\widehat{\Phi}(f) = \Phi(\widehat{f}) \qquad \forall f \in \mathscr{S}.$$

Then, $\widehat{\Phi} \in \mathscr{S}'$ is called the Fourier transform of Φ.

■ **Example 5.14** If $g \in L^1(\mathbb{R}^d)$, we saw that

$$\Phi_g(f) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} f g \mathrm{d}x \qquad \forall f \in \mathscr{S}$$

defines a tempered distribution. One has

$$\widehat{\Phi_g}(f) = \Phi_g(\mathcal{F}(f)) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathcal{F}(f)(x) g(x) \mathrm{d}x \qquad \forall f \in \mathscr{S}$$

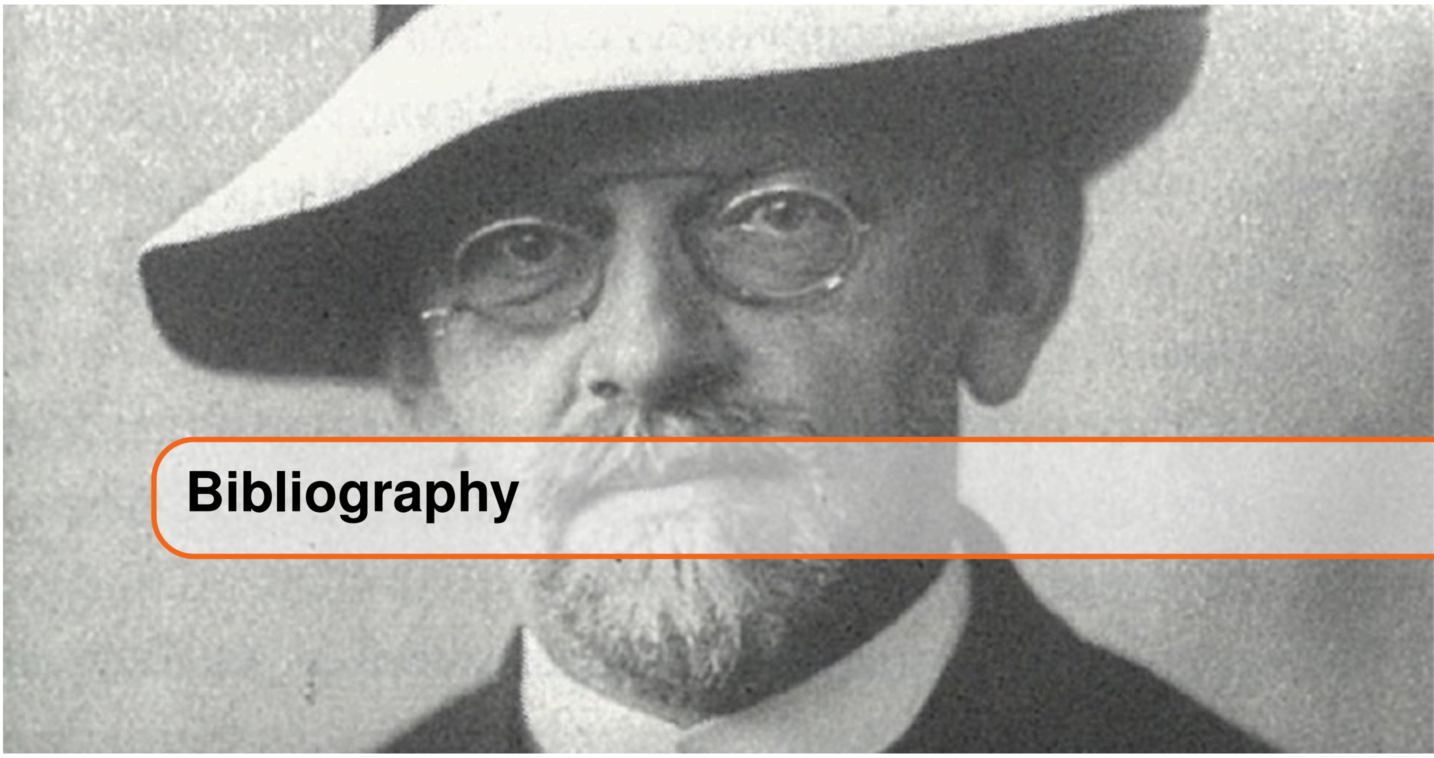
and one checks easily that

$$\frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathcal{F}(f)(x) g(x) \mathrm{d}x = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathcal{F}(g)(x) f(x) \mathrm{d}x$$

so that

$$\widehat{\Phi_g} = \Phi_{\mathcal{F}(g)}$$

i.e. $\widehat{\Phi_g}$ is the tempered distribution associated to $\widehat{g}$. ■



Bibliography

[1] Blanchard P., Devaney R. L., Hall G. R. — **Differential Equations,** Fourth Edition, Brooks/Cole, Cengage Learning, 2012.

[2] Brezis, H. — **Functional Analysis, Sobolev Spaces and Partial Differential Equations**, Springer, 2010.

[3] Dudley, R. M. — **Real Analysis and Probability,** Cambridge University Press, 2000.

[4] Durrett, R. — **Probability: Theory and Examples**, Second Edition, Duxbury Press, 1995.

[5] Heil, C. — **Five Courses on Analysis,** Lecture notes, 2015.

[6] Hunter, J. K. — **Measure Theory**, Lecture notes, 2011, https://www.math.ucdavis.edu/~hunter/measure_theory/measure_notes.pdf

[7] Malliavin, P. — **Integration and Probability**, Springer, 1995.

[8] Munkres J. R. — **Topology: A First Course,** Prentice-Hall, Englewood Cliffs, NJ, 1975.

[9] Rynne B. P. & Youngson M. A. — **Linear Functional Analysis,** Second Edition, Springer, 2008.

[10] Villani, C. — **Optimal transport, Old and New**, Springer, 2009.

[11] Williams, D. — **Probability with Martingales** Cambridge University Press, 1991.