Université Paris Dauphine
Master 1 de Mathématiques et Applications
2024-2025

Short reminder of mathematical analysis (and application to probability theory)

Pierre Cardaliaguet

Contents

It is obviously impossible to summarize in 12 hours and a few pages the 3 years of the License program. The following notes are only summary reminders of these courses. As they are reminders, the concepts are not necessarily approached in a natural order (in particular, the examples and exercises often call upon concepts discussed later).

In the following: The "basics", not detailed, are assumed to be very well known. If this is not the case, the reader is strongly advised to consult the courses cited in reference to review them as soon as possible. The exercises proposed are very classic and at the level of a student at the end of L3.

Notation : (Indicator function) If A is a subset of a set X, we denote $\mathbf{1}_A : X \to \{0,1\}$ the indicator function of A, i.e., the function defined by $\mathbf{1}_A(x) := 1$ if $x \in A$ and $\mathbf{1}_A(x) := 0$ otherwise, for all $x \in X$.

0. Linear algebra

In this part, we recall some basic results in linear algebra. *Vector spaces here are always finite dimensional.* This part is very widely borrowed from Guillaume Legendre's course "Algèbre linéaire 3".

0.1. Rank-nullity theorem.

Basic notions : dimension theory (vector space, vector subspace, linearly independent family, generating family, basis, dimension) ; linear maps, sum and composition of linear maps, endomorphisms.

Theorem 0.1. *Let E and F be two finite dimensional vector spaces (over $\mathbb{R}$ or aver $\mathbb{C}$) and $f : E \to F$ be a linear map. Then*

- *f is injective, if and only if, $Ker(f) := f^{-1}(\{0\}) = \{0\}$ (holds also in infinite dimension).*
- *(Rank-nullity theorem) $\dim(Im(f)) + \dim(Ker(f)) = \dim(E)$.*
- *(Consequence of the Rank-nullity theorem) If $\dim(E) = \dim(F)$, then f is injective, if and only if, f est onto (and then f is of course bijective).*

0.2. Diagonalization.

Basic notions : matrix relative to a linear map, link between matrix product and composition. Invertible matrix and bijective linear maps, determinant. Similar matrices. Eigenvalue and eigenvector of a linear map, link with the characteristic polynomial.

We denote by $M_{n,p}(\mathbb{R})$ (resp. $M_{n,p}(\mathbb{C})$) the set of real matrices (resp. complex matrices) with n lines and p columns. If $p = n$, we simply set $M_n(\mathbb{R}) := M_{n,n}(\mathbb{R})$ and $M_n(\mathbb{C}) := M_{n,n}(\mathbb{C})$. If $A \in M_{n,p}(K)$ (where $K = \mathbb{R}$ or $K = \mathbb{C}$), A^T is the transpose matrix of A (i.e., $A^T \in M_{p,n}(K)$ with, if $A = (a_{i,j})_{1\le i\le n,\ 1\le j\le p}$, $A^T = (a_{j,i})_{1\le i\le p,\ 1\le j\le n}$).

The **characteristic polynomial** of a matrix $A \in M_n(K)$ (where $K = \mathbb{R}$ or $K = \mathbb{C}$) is the polynomial $\chi_A(X) = \det(XI_n - A)$ (where I_n is the identity matrix of size $n \times n$). It is a monic polynomial of degree n. One has

$$\chi_{A^T} = \chi_A, \qquad \chi_{AB} = \chi_{BA} \qquad \text{if } A, B \in M_n(K).$$

In particular, two similar matrices have the same characteristic polynomial, and one can thus define the characteristic polynomial χ_f of an endomorphism $f : E \to E$ as the characteristic polynomial of any matrix representing f. Recall that λ is an eigenvalue of A, if and only if, λ is a root of χ_A. We then call **geometric multiplicity of** λ the dimension of $\mathrm{Ker}(\lambda I_d - A)$. The **algebraic multiplicity of** λ is the order of multiplicity of λ as root of χ_A (i.e., the integer $r \in \{1, \dots, n\}$ such that $(X - \lambda)^r$ divides χ_A and $(X - \lambda)^{r+1}$ does not divide by χ_A).

Theorem 0.2 (Diagonalization). *Let E be a vector space of dimension n over $K = \mathbb{R}$ or $K = \mathbb{C}$, and $f : E \to E$ be an endomorphism. Then f is diagonalizable if f satisfies one (and thus all) the equivalent statements:*

1. *There exists a basis of E in which the matrix of f is diagonal (this basis is necessarily made of eigenvectors of f).*

(2) The exists a basis of E made of eigenvectors of f.
(3) If $A \in M_n(K)$ is the matrix of f is a basis $\mathcal{B}$ of E, then A is diagonalizable, i.e., there exists a invertible matrix $P \in M_n(K)$ such that the matrix $P^{-1}AP$ is diagonal (then the columns of P are the coefficients in the basis $\mathcal{B}$ of a basis of eigenvectors of f).
(4) If $A \in M_n(K)$ is the matrix of f is a basis $\mathcal{B}$ of E, χ_A can be written as a product of polynomials of degree 1 and any eigenvalue of A has a geometric multiplicity equal to its algebraic multiplicity.

Trigonalization. An endomorphism $f : E \to E$ is said to be **trigonalizable** if there exists a basis in which the matrix of f is upper triangular. We show that f is trigonalizable if and only if its characteristic polynomial can be written as a product of polynomial of degree 1 (in particular, in $\mathbb{C}$, every endomorphism is trigonalizable).

Endomorphism polynomial. If $f : E \to E$ is an endomorphism and $P(X) = \sum_{k=0}^{m} a_k X^k$, we write $P(f)$ the endomorphism $P(f) = \sum_{k=0}^{m} a_k f^{(k)}$ (where $f^{(k)} = \underbrace{f \circ \cdots \circ f}_{(k \text{ times})}$ and, by convention, $f^{(0)} = id_E$). The **Cayley-Hamilton** theorem states that $\chi_f(f) = 0$. The **minimal polynomial** of f is the smallest (in the sense of Euclidean division) monic polynomial P such that $P(f) = 0$. The Cayley-Hamilton theorem therefore states that the minimal polynomial divides the characteristic polynomial. Moreover, one can show that f is diagonalizable, if and only if, the minimal polynomial of f is split and has only simple roots.

0.3. **Quadratic forms.**

Basics : Bilinear map, symmetric bilinear map, matrix representation and base change. Quadratic form, associated polar form. Signature of a quadratic form.

Theorem 0.3 (spectral)**.** *Any symmetric matrix $A \in M_n(\mathbb{R})$ admits an orthonormal basis of real eigenvectors: there exists $O \in M_n(\mathbb{R})$ such that $OO^T = O^T O = I_n$ (whose columns are eigenvectors of A) and $D := O^T AO$ is diagonal (the coefficients of D being the eigenvalues of A).*

Positivity. Recall that a quadratic form $q : E \to \mathbb{R}$ is positive if $q(x) \geq 0$ for all $x \in E$. It is positive definite if $q(x) > 0$ for all $x \in E^*$. A consequence of the spectral theorem is that q is positive (resp. positive definite), if and only if, its matrix (in any basis) has all its positive eigenvalues (resp. strictly positive).

1. Topology and functional analysis

This part is largely borrowed from the course of Paul Pegon "Mise à niveau en analyse".

1.1. **Metric spaces.**

1.1.1. *Basic notions.*
Distance, open sets, closed sets, intersections of open or closed sets, closure, convergence of sequences, continuous maps, Lipschitz maps, ...

Examples of metric spaces: $\mathbb{R}$ endowed with the absolute value, $\mathbb{R}^n$ or $\mathbb{C}^n$ (with $n \in \mathbb{N}\backslash\{0\}$) endowed with the euclidean norm, normed spaces, subsets of normed spaces endowed with the induced distance, etc...

1.1.2. *Density, separability.*

Definition 1.1. *Let (X, d) be a metric space. We say that a subset A of X is dense in X if every open subset of X contains a point of A.*

Equivalently, A is dense in X if, for all $x \in X$, there exists a sequence of elements of A which tends to x.

Classic examples are: $\mathbb{Q}$ is dense in $\mathbb{R}$ (and therefore $\mathbb{Q}^N$ is dense in $\mathbb{R}^N$) ; the density of polynomials in the set of continuous functions on an interval $[a,b]$ of $\mathbb{R}$ for the uniform norm, and the density of trigonometric polynomials in the space of continuous functions $2\pi-$periodic (see theorem 1.20, version of the Stone-Weierstrass theorem) ; step functions are dense in L^p (see theorem 3.10) ; similarly, the set $C_c^\infty(\mathbb{R})$ of functions C^∞ with compact support is dense in $L^p(\mathbb{R}, dx)$ (equipped with the Lebesgue measure, see the theorem 3.11).

Definition 1.2. *Let (X,d) be a metric space. We say that X is separable if there exists a dense countable set in X (in other words if there exists a sequence $(x_n)_{n\in\mathbb{N}}$ of elements of X which is dense in X).*

Classical examples are: $\mathbb{R}^N$ is separable (for $N \in \mathbb{N}^*$) since $\mathbb{Q}^N$, which is countable, is dense in $\mathbb{R}^N$; the set of continuous functions on an interval $[a,b]$ of $\mathbb{R}$ endowed with the uniform norm ; spaces L^p for $p \in [1,\infty[$ (on the other hand L^∞ is not separable).

Exercise 1. Show that if (X,d) is a separable metric space and if $Y \subset X$ is nonempty, then (Y,d) is also separable.

Exercise 2. Using the density of polynomials in $L^1([0,1])$, show that if $f \in Ł^1([0,1])$ is such that $\int_0^1 x^k f(x)dx = 0$ for all $k \in \mathbb{N}$, then $f = 0$ p.p.

Classic applications of density are:
- the fact that the Fourier coefficients $c_n(f)$ of a function of $f \in L^1([a,b])$ tends towards 0 when $|n| \to \infty$
- the extension to $L^2(\mathbb{R})$ of the Fourier transform (by density of $L^1 \cap L^2$ in L^2).
- in M1, in the "Discrete Processes" course, we will also see the extension of the conditional expectation from L^2 to L^1 (by density of L^2 in L^1).

Exercise 3. Recall that the space $C_c^1(\mathbb{R})$ of C^1 maps vanishing outside pf a compact set is dense in $L^1(\mathbb{R})$ (for the L^1 norm) show the Riemann-Lebesgue Lemma : if $f \in L^1(\mathbb{R})$,

$$\lim_{n\to+\infty} \int_{\mathbb{R}} \cos(nx) f(x)dx = 0.$$

1.1.3. *Complete metric spaces.*

Definition 1.3. *A sequence (x_n) of a metric space (X,d) is a Cauchy sequence if, for any $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $d(x_n, x_m) \le \varepsilon$ for any $n, m \ge N$.*

Exercise 4. Show that every convergent sequence is a Cauchy sequence.

Definition 1.4. *A metric space (X,d) is is said to be complete if any Cauchy sequence has a limit there.*

Exercise 5. Show that any closed subset of a complete metric space is itself complete (for the induced distance).

Theorem 1.5 (Extension theorem)**.** *Let A be a dense subset of a metric space (X, d_X) and let (Y, d_Y) be a complete metric space. If a map $f : A \to Y$ is uniformly continuous on A (for the metric d_X), then there exists a unique continuous map $\tilde{f} : X \to Y$ such that $\tilde{f}(x) = f(x)$ for all $x \in A$. Moreover $\tilde{f}$ is uniformly continuous on X.*

Theorem 1.6 (fixed point theorem)**.** *Let (X, d) be a complete metric space. If the map $f : X \to X$ is a contraction, i.e., there exists a real number $k \in (0, 1)$ such that*

$$d(f(x), f(y)) \leq kd(x, y) \qquad \forall x, y \in X,$$

then f has a unique fixed point. Moreover, if $x_0 \in X$ and one defines inductively the sequence (x_n) by setting $x_{n+1} = f(x_n)$ for all $n \in \mathbb{N}$, then (x_n) converges to the unique fixed point of f.

Application : The local inversion theorem 2.2; Cauchy-Lipschitz Theorem 2.4, on the existence and the uniqueness of the solution of an ordinary differential equation (used in the courses "Optimization" and "Numerical analysis: time dependent problems").

1.1.4. *Compact metric spaces.* Let (x_n) be a sequence in a metric space (X, d). We recall that a sequence extracted from (x_n) is any sequence of the form $(x_{\phi(n)})$ where $\phi : \mathbb{N} \to \mathbb{N}$ is strictly increasing. We often denote such an extracted sequence (x_{n_k}). It is easy to see that, if (x_n) converges, then any sequence extracted from (x_n) converges to the limit of (x_n).

Definition 1.7. *We say that a metric space (X, d) is compact if every sequence of elements of X has an extracted sequence which converges in X.*

Exercise 6. Show the following statements:

(1) any closed subset of a compact space is compact.
(2) any compact subset of a set is closed and bounded (the reciprocal is false in general).
(3) any continuous function on a compact space is bounded and has a maximum and a minimum.

Exercise 7 (a convergence criterium)**.** Let (X, d) be a compact metric space and (x_n) be a sequence in X. We assume that there exists $x \in X$ such that any **converging** extracted subsequence of (x_n) converges to x. Show that the whole sequence (x_n) converges to x.

1.2. **Normed vector spaces, Banach spaces.**

1.2.1. *Basic notions.* Norms, equivalent norms, convergence in terms of norm, continuity in terms of norm. Recall that a normed space is a metric space.

1.2.2. *The case of finite dimensional spaces.*

Theorem 1.8. *In a finite dimensional vector space, all norms are equivalent.*

This is not at all the case in infinite dimension.

Exercise 8. For example, show that on $X = C^0([0, 1])$ the space of continuous functions on $[0, 1]$, the norms $\|\cdot\|_\infty$ and $\|\cdot\|_1$ defined by

$$\|f\|_\infty = \sup_{x\in[0,1]} |f(x)| \qquad \text{et} \qquad \|f\|_1 := \int_0^1 |f(x)|dx \qquad \forall f \in X,$$

are not equivalent (show that there exists a sequence of continuous maps $f_n : [0, 1] \to [0, 1]$ such that $f_n(0) = 1$ for all $n \in \mathbb{N}$ but $\|f_n\|_1 \to 0$).

Another peculiarity of finite dimensional spaces is that compact subsets are quite "numerous" there.

Theorem 1.9. *In a finite-dimensional normed vector space, the compact subsets are the closed and bounded subsets.*

In infinite dimension, this characterization is false (in fact, the Riesz theorem even asserts that an normed space whose closed unit ball is compact is of finite dimension). In the "Functional Analysis" course of the M1 will be introduced a weaker topology, including more compact sets. The notion of convergence in law (in probability) is a similar means of circumventing this difficulty.

Exercise 9. Show that, in $\ell^1 := \{x = (x_k)_{k\in\mathbb{N}},\ \sum_{k=0}^{\infty} |x_k| < +\infty\}$ endowed with the norm $\|x\|_1 = \sum_{k=0}^{\infty} |x_k|$, the sequence (e^n) of elements of ℓ^1 defined by $e^n_k = \delta_{kn}$ (where $\delta_{kn} = 0$ if $k \neq n$ and $= 1$ if $k = n$) has no converging subsequence.

1.2.3. *Linear maps in a normed space.*

Let $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ be two normed spaces and $f : E \to F$ be a linear map.

Theorem 1.10. *The following statements are equivalent:*

(1) there exists $x_0 \in E$ such that f is continuous at x_0,
(2) f is continuous on E,
(3) $\sup_{\|x\|_E\leq 1} \|f(x)\|_F$ is finite.

If one of the assumptions above holds, then f is uniformly continuous on E. Moreover, $f \to \|f\|_{L(E,F)} := \sup_{\|x\|_E\leq 1} \|f(x)\|_F$ is a norm on the vector space $L(E, F)$ of continuous linear maps from E to F.

Exercise 10. Let $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ be two normed spaces, with F complete. Show that $L(E, F)$ is also complete.

The particular case of linear forms on E (i.e., continuous linear maps from E to $F := \mathbb{R}$) is often denoted by E^*. This space plays a central role in the course "Functional analysis" of the M1.

1.2.4. *Banach spaces.*

Definition 1.11. *A Banach space is a complete normed space.*

Most classical functional spaces are Banach spaces (see the part "Examples of functional spaces" below).

A central property of Banach spaces is the normal convergence of the series:

Theorem 1.12 (Normal convergence)**.** *Let $(X, \|\cdot\|)$ be a Banach space. If a sequence (x_n) of X is normally convergent, i.e., if the series of general term $(\|x_n\|)$ converges, then the series of general term (x_n) converges (in X) : in other words,*

$$\sum_{i=0}^{\infty} \|x_n\| < +\infty \qquad \Longrightarrow \qquad \lim_{m\to+\infty} \sum_{n=0}^{m} x_n \text{ exists (dans } X).$$

Exercise 11. Let E be a Banach space. We recall (cf. Exercise 10) that $L(E) := L(E, E)$ is also a Banach space. For $f \in L(E)$, we note $f^{(0)} = Id_E$ (the identity map of E), $f^{(1)} = f$, $f^{(2)} = f \circ f$ and $f^{(n+1)} = f \circ f^{(n)}$ for all $n \in \mathbb{N}$. We assume that $\|f\|_{L(E)} < 1$. Show that $Id_E - f$ is bijective and of continuous inverse with $(Id_E - f)^{-1} = \lim_{N\to+\infty} \sum_{n=0}^{N} f^{(n)}$.

1.3. **Hilbert spaces.**

1.3.1. *Basic notions.* Scalar product, norm associated to a scalar product, Cauchy-Schwarz inequality.

Definition 1.13. *A Hilbert space is a vector space endowed with a scalar product, which is complete for the norm associated with this scalar product. When this space is of finite dimension, we speak of Euclidean space.*

The Hilbert spaces most often encountered in practice are the Euclidean spaces $\mathbb{R}^N$ endowed with the Euclidean norm, the spaces L^2 in integration and the space ℓ^2 of summable square sequences.

Exercise 12 (Parallelogram law)**.** Let $(E, \|\cdot\|)$ be a Banach space. Show that E is a Hilbert space (in the sense that there exists an scalar product $\langle\cdot,\cdot\rangle$ such that $\|x\|^2 = \langle x, x\rangle$ for all $x \in H$), if and only if, E verifies the parallelogram law

$$\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2) \qquad \forall x, y \in E.$$

1.3.2. *The projection theorem.*

Theorem 1.14 (projection theorem (case of a closed convex set))**.** *Let C be a closed convex subset of a Hilbert space $(H, \langle\cdot,\cdot\rangle)$. Then every $x \in H$ has a unique projection $\Pi_C(x)$ on C, i.e. there exists a unique element $\Pi_C(X) \in C$ such that*

$$\|x-y\| \geq \|x - \Pi_C(x)\| \qquad \forall y \in C.$$

Moreover $\Pi_C(x)$ is the unique element of C satisfying the variational inequality

$$\langle x - \Pi_C(x), y - \Pi_C(x)\rangle \leq 0 \qquad \forall y \in C.$$

The remarkable aspect of this result is that C is not supposed to be compact. Moreover it provides a characterization of the projection. This result will be used frequently in M1 Math, especially in optimization. The following version, where C is a closed vector subspace, will be used in M1 in almost all courses (for the definition of conditional expectation in "Discrete Processes", in statistics (least squares), in "Time series", etc...).

Exercise 13. We assume that C is a closed convex subset of a Hilbert space $(H, \langle\cdot,\cdot\rangle)$. Show that Π_C is a continuous linear map from H to H, if and only if C is a vector subspace of H.

Theorem 1.15 (projection theorem (case of a closed linear space))**.** *If C a closed vector subspace of a Hilbert space H, then the map $x \to \Pi_C(x)$ (defined in the previous theorem) is a continuous linear map (of norm 1). Moreover, $\Pi_C(x)$ is the unique element of C satisfying the equality*

$$\langle x - \Pi_C(x), y\rangle = 0 \qquad \forall y \in C.$$

(in other words, $x - \Pi_C(x)$ belongs to the orthogonal $C^\perp$ of C).

.

Exercise 14. Show Theorem 1.15 from Theorem 1.14.

Theorem 1.16 (Riesz representation theorem)**.** *Let H be a Hilbert space and $f : H \to \mathbb{R}$ a continuous linear map. Then there is a unique element $x_0 \in H$ such that $f(x) = \langle x_0, x\rangle$ for all $x \in H$.*

1.3.3. *Orthogonal and Hilbert bases.*

Definition 1.17 (Orthogonal)**.** *Let H be a Hilbert space and A a subset of A. The orthogonal of A, denoted $A^\perp$, is the set defined by*

$$A^\perp = \{x \in H \text{ such that } \langle x, y\rangle = 0 \ \forall y \in A\}.$$

Exercise 15. Let A be a nonempty subset of a Hilbert space H. Show that

(1) the set $A^\perp$ is a closed vector subspace of H,
(2) $A^\perp = (Vect(A))^\perp$, where $Vect(A)$ is the vector subspace spanned by A,
(3) $H^\perp = \{0\}$

The last equality in the exercise actually characterizes H (this is where we use the fact that H is complete):

Theorem 1.18. *Let A be a* vector subspace *of H. Then*

$$(A^\perp)^\perp = \overline{A} \qquad \textit{(where } \overline{A} \textit{ is the closure of } A).$$

In particular, if $A^\perp = \{0\}$, then A is dense in H. If A is closed and $A^\perp = \{0\}$, then $A = H$.

Definition 1.19 (Hilbert basis). *A sequence $(x_n)_{n\in\mathbb{N}}$ is a Hilbert basis of a Hilbert space H if it is orthonormal*

$$\langle x_n, x_m\rangle = \begin{cases} 0 & si\, n \neq m \\ 1 & if\, n = m \end{cases}$$

and total:

$$(\{x_n,\ n \in \mathbb{N}\})^{\perp} = \{0\}.$$

If H is separable, then H has a Hilbert basis (this is shown by Schmidt's orthonormalization process).

The emblematic example of a Hilbertian basis is the family consisting of the functions $t \to \sqrt{2}\cos(2\pi nt)$ $(n \in \mathbb{N})$ and $t \to \sqrt{2}\sin(2\pi nt)$ $(n \in \mathbb{N}^*)$, which is a Hilbertian basis of $L^2([0,1])$.

Exercise 16 (Bessel inequality and equality). Check that, if $(e_n)_{n\in\mathbb{N}}$ is an orthonormal family of a Hilbert space H, then for all $x \in H$ we have $\sum_{n=0}^{\infty}\langle x, e_n\rangle^2 \leq \|x\|^2$. Show that if $(e_n)_{n\in\mathbb{N}}$ is a Hilbert basis of H, then $\sum_{n=0}^{\infty}\langle x, e_n\rangle^2 = \|x\|^2$.

1.4. **Some classic functional spaces.**

1.4.1. *Spaces of continuous functions.* If (X, d) is a compact metric space, then the set $C^0(X)$ of continuous maps from X to $\mathbb{R}$, endowed with the uniform norm:

$$\|f\|_\infty = \sup_{x\in X}|f(x)| \qquad \forall f \in C^0(X).$$

is a Banach space[1].

Theorem 1.20 (Density).

- **(density of polynomials)** *If X is a compact subset of $\mathbb{R}^n$ (where $n \in \mathbb{N}\backslash\{0\}$), the set of (restrictions to X of) polynomials[a] forms a dense subset of $C^0(X)$ for the $\|\cdot\|_\infty$. In particular, $(C^0(X), \|\cdot\|_\infty)$ is separable.*
- **(density of trigonometric polynomials)** *The set of trigonometric polynomials[b] is dense in the space of continuous and $2\pi-$periodic functions from $\mathbb{R}$ to $\mathbb{C}$.*

[a] A polynomial $P : \mathbb{R}^n \to \mathbb{R}$ is an expression of the form $P(x) = \sum_{i\in\mathbb{N}^n} a_i x_1^{i_1}\dots x_n^{i_n}$ where only a finite number of coefficients $a_i \in \mathbb{R}$ are nonzero and where $i = (i_1, \dots, i_n)$.

[b] A trigonometric polynomial $P : \mathbb{R} \to \mathbb{C}$ is an expression of the form $P(x) = \sum_{i=0}^{n} a_i e^{2i\pi x}$ where $n \in \mathbb{N}$, $a_i \in \mathbb{C}$.

Both of these results are a consequence of the Stone-Weierstrass theorem below (and are often known by this name). Recall that a subalgebra of $C^0(X)$ is a vector subspace of $C^0(X)$ which is stable under product. We say that this subalgebra separates the points if, for all $x_1, x_2 \in X$, there exists an element f of this subalgebra such that $f(x_1) \neq f(x_2)$.

Theorem 1.21 (Stone-Weierstrass). *Let (X, d) be a compact space and $C^0(X)$ the set of continuous functions from X to $\mathbb{R}$, endowed with the uniform norm. A subalgebra of $C^0(X)$ is dense in $C^0(X)$, if and only if, it separates the points and contains, for every point x of X, a function which does not vanish at x.*

[1] In this statement, $\mathbb{R}$ can be replaced by any Banach space $(E, \|\cdot\|_E)$. In this case, $\|f\|_\infty := \sup_{x\in X}\|f(x)\|_E$.

1.4.2. *The spaces ℓ^p.* These spaces will be used frequently, for instance in the course "Time Series".

Definition 1.22. *For $p \in [1, \infty)$, $\ell^p(\mathbb{R})$ (resp. $\ell^p(\mathbb{C})$) is the set of real (resp. complex) sequences $x = (x_n)_{n\in\mathbb{N}}$ such that $\|x\|_p := \left(\sum_{n=1}^{\infty} |x_n|^p\right)^{1/p} < \infty$. For $p = \infty$, $\ell^\infty(\mathbb{R})$ (resp. $\ell^\infty(\mathbb{C})$) is the set of real (resp. complex) bounded sequences.*

For $p \in [1, \infty[$, $(\ell^p(\mathbb{R}), \|\cdot\|_p)$ and $(\ell^p(\mathbb{C}), \|\cdot\ |_p)$ are normed vector spaces. The same is true for $\ell^\infty(\mathbb{R})$ and $\ell^\infty(\mathbb{C})$ equipped with the norm $\|x\|_\infty := \sup_n |x_n|$.

Theorem 1.23. *The ℓ^p are Banach spaces for $p \in [1, \infty]$ and, if $1 \leq p \leq q \leq \infty$, then*

$$\ell^p \subset \ell^q.$$

Finally, $\ell^2(\mathbb{R})$ is a Hilbert space for the inner product

$$\langle x, y\rangle = \sum_{n=0}^{\infty} x_n y_n \qquad \forall x = (x_n), y = (y_n) \in \ell^2(\mathbb{R}).$$

1.4.3. *The spaces L^p.* These spaces will be discussed later, in the section 3.4.

2. Differential calculus

This part is very largely borrowed from the course of Jacques Fejoz"Calcul différentiel et optimisation".

2.1. **Differentiability.**

2.1.1. *Definition and basic properties.* Let $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ be two normed vector spaces, $\mathcal{O}$ an open set of E and $x_0 \in\ mathcalO$. We say that $f : \mathcal{O} \to F$ is differentiable at x_0 if there exists a continuous linear map $df(x_0) : E \to F$ such that

$$\lim_{x\to x_0} \frac{\|f(x) - f(x_0) - df(x_0)(x - x_0)\|_F}{\|x - x_0\|_E} = 0 \tag{1}$$

The differential $df(x_0)$ is then unique.

(1) Equality (1) is also written
$$f(x) = f(x_0) + df(x_0)(x - x_0) + \|x - x_0\|_E\varepsilon(x - x_0),$$
where $\varepsilon : E \to F$ is a mapping that tends to 0 when x tends to x_0.

(2) The differential, denoted here $df(x_0)$, can also be simply called the derivative and be denoted $f'(x_0)$ or $Df(x_0)$, etc...

(3) If f is differentiable at x_0, then f is continuous at x_0.

(4) If $\mathcal{O} = E$ and f is a continuous linear map from E to F, then f is differentiable at every point $x_0 \in E$ and $df(x_0) = f$ (in particular, $df(x_0)$ does not depend on x_0).

(5) **Stability of differentiability by addition and by multiplication by a scalar :** if f_1, f_2 are differentiable at $x_0 \in E$ and $\alpha \in \mathbb{R}$, then $\alpha f_1 + f_2$ is differentiable at x_0 and
$$d(\alpha f_1 + f_2)(x_0) = \alpha df_1(x_0) + df_2(x_0).$$

(6) **Notion of gradient :** When $E = \mathbb{R}^n$ and $F = \mathbb{R}$, $df(x_0)$ is a linear application from $\mathbb{R}^n$ to $\mathbb{R}$ and can therefore be identified with a vector of $\mathbb{R}^n$, denoted $\nabla f(x_0)$ (called *gradient* from f to x_0). So we have
$$df(x_0)(v) = \langle \nabla f(x_0), v\rangle \qquad \forall v \in \mathbb{R}^n.$$
(this notation generalizes to the case where E is a Hilbert space, thanks to the Riesz representation theorem (Theorem 1.15)).

We say that f is C^1 on $\mathcal{O}$ if f is differentiable at every point of $\mathcal{O}$ and if the map $x \to df(x)$ is continuous from $\mathcal{O}$ in the space $L(E,F)$ of continuous linear maps from E to F.

Note:

(1) (Integral of the differential) If f is of class C^1 on $\mathcal{O}$ with values in $\mathbb{R}^m$ ($m \in \mathbb{N}^*$) and $[a,b] \subset \mathcal{O}$, then
$$f(b) - f(a) = \int_0^1 df((1-t)a + tb)(b-a)dt$$
(2) (mean value inequality) Under the same assumptions,
$$\|f(b) - f(a)\|_F \leq \sup_{z\in[a,b]} \|df(z)\|_{L(E,F)} \|b-a\|_E$$
(where the segment $[a,b]$ is the closed convex subset of E defined by $[a,b] = \{(1-t)a+tb,\ t \in [0,1]\}$).

2.1.2. *The finite dimensional case.* We suppose here that $E = \mathbb{R}^n$ and $G = \mathbb{R}^p$ (with $n, p \in \mathbb{N}^*$), $\mathcal{O}$ is an open of $\mathbb{R}^n$ and $f : \mathcal{O} \to \mathbb{R}^p$ with $f = (f_1, \ldots, f_p)$. We assume that f admits continuous partial derivatives on $\mathcal{O}$, i.e., that for all $x \in \mathcal{O}$, $i \in \{1, \ldots, p\}$ and $j \in \{1, \ldots, n\}$, the following limit exists
$$\frac{\partial f_i(x)}{\partial x_j} = \lim_{h\to 0} \frac{f_i(x_1, \ldots, x_{j-1}, x_j + h, x_{j+1}, \ldots, x_n) - f_i(x_1, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots, x_n)}{h},$$
and that the maps $\frac{\partial f_i(x)}{\partial x_j} : \mathcal{O} \to \mathbb{R}$ are continuous on $\mathcal{O}$. Then f is of class C^1 over $\mathcal{O}$ and we have, for all $v \in \mathbb{R}^n$ (in matrix notation for the right term of equality)
$$df(x)(v) = J_f(x)v \qquad \text{where } J_f(x) = \begin{pmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_1(x)}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_p(x)}{\partial x_1} & \cdots & \frac{\partial f_p(x)}{\partial x_n} \end{pmatrix} \quad v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}.$$
The matrix $J_f(x)$ is called **the Jacobian matrix from f to x.**

2.1.3. *Differentiability and composition.*

Theorem 2.1. *Let $(E, \|\cdot\|_E)$, $(F, \|\cdot\|_F)$ and $(G, \|\cdot\|_G)$ be three vector spaces normed, $\mathcal{O}$ an open set of E, $\mathcal{O}'$ an open set of F, and $x_0 \in \mathcal{O}$. We assume that $f : \mathcal{O} \to \mathcal{O}'$ and $g : \mathcal{O}' \to G$ are differentiable at x_0 and at $f(x_0)$ respectively. Then $g \circ f$ is differentiable at x_0 and*
$$d(g \circ f)(x_0) = dg(f(x_0)) \circ df(x_0).$$
In finite dimension, we obtain the matrix equality :
$$J_{g\circ f}(x_0) = J_g(f(x_0))J_f(x_0).$$

In particular, if f is bijective from $\mathcal{O}$ into $\mathcal{O}'$ and if $g = f^{-1}$ is differentiable at $y_0 := f(x_0)$, then $df(x_0)$ is invertible and
$$d(f^{-1})(y_0) = (df(x_0))^{-1}$$
(compare with the local inversion theorem below).

2.1.4. *Higher order differentials.* Let $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ be two normed vector spaces, $\mathcal{O}$ an open set of E and $f : \mathcal{O} \to F$ an application. We say that f is of class C^2 if f is of class C^1 and if $df : E \to L(E,F)$ is also of class C^1. The derivative of df at a point $x_0 \in E$, denoted $d^2 f(x_0)$, is an element of $L(E, L(E,F))$ and is therefore identified with a symmetric continuous bilinear map (the symmetry coming from the **Schwarz lemma**). If $[a,b] \subset \mathcal{O}$, we have the equality
$$f(b) = f(a) + df(a)(b-a) + \frac{1}{2}\int_0^1 d^2 f(a + t(b-a))(b-a, b-a)(1-t)dt.$$

Especially,

$$\|f(b)-f(a)-df(a)(b-a)\|_F \le \frac{1}{2}\sup_{z\in[a,b]} |||d^2f(z)||| \|b-a\|^2$$

(where $|||h||| = \sup_{\|x\|_E\le 1, \|y\|_E\le 1} \|h(x,y)\|_F$ if $h : E\times E \to F$ is bilinear).

If $E = \mathbb{R}^n$, $F = \mathbb{R}$ and $x_0 \in \mathbb{R}^n$, $d^2f(x_0)$ is a symmetric bilinear form on $\mathbb{R}^n$ and can be represented by a symmetric matrix, the Hessian matrix from f to x_0:

$$\mathrm{Hess}_f(x_0) = \left(\frac{\partial^2 f(x_0)}{\partial x_i \partial x_j}\right)_{1\le i,j\le n}.$$

By induction, we can also define the notion of class function C^k (where $k\in\mathbb{N}^*$), where $d^kf(x_0)$ is a symmetric $k-$linear map. Taylor's formula of order k is then written, for all $a, b \in \mathcal{O}$ such that $[a,b]\subset\mathcal{O}$,

$$\begin{aligned} f(b) =& f(a) + df(a)(b-a) + \frac{1}{2}d^2f(a)(b-a)^2 + \cdots + \frac{1}{(k-1)!}d^{(k-1)}f(a)(b-a)^{k-1} \\ &+ \int_0^1 d^kf(a+t(b-a))(b-a)^k \frac{(1-t)^{k-1}}{(k-1)!}dt \end{aligned}$$

2.1.5. *Application to convexity.* Let E be a real vector space and $\mathcal{O}\subset E$ a convex subset of E, i.e.,

$$\forall x,y\in\mathcal{O},\ \forall\lambda\in[0,1], \qquad (1-\lambda)x+\lambda y\in\mathcal{O}.$$

We say that $f : \mathcal{O}\to\mathbb{R}\cup\{+\infty\}$ is convex in $\mathcal{O}$ if

$$\forall x,y\in\mathcal{O},\ \forall\lambda\in[0,1], \qquad f((1-\lambda)x+\lambda y)\le(1-\lambda)f(x)+\lambda f(y).$$

Suppose that E is normed space, that $\mathcal{O}$ is a convex open set of E and that $f:\mathcal{O}\to\mathbb{R}$ is of class C^1. We then have the equivalence

$$f \text{ is convex in } \mathcal{O} \qquad \Longleftrightarrow \qquad [\forall x,y\in\mathcal{O}, \qquad (df(x)-df(y))(y-x)\ge 0].$$

If f is of class C^2, then

$$f \text{ is convex in } \mathcal{O} \qquad \Longleftrightarrow \qquad \left[\forall x\in\mathcal{O},\ \forall\xi\in E, \qquad d^2f(x)(\xi,\xi)\ge 0\right].$$

2.2. **Local inversion and implicit functions.** We assume that $\mathcal{O}$ is an open set of $\mathbb{R}^n$ (where $n\in\mathbb{N}^*$) and $f:\mathcal{O}\to\mathbb{R}^n$ is of class C^1.

Theorem 2.2 (local inversion)**.** *If $x_0\in\mathcal{O}$ and $df(x_0):\mathbb{R}^n\to\mathbb{R}^n$ is bijective, then there exists an open set $\mathcal{O}'\subset\mathcal{O}$ of $\mathbb{R}^n$ containing x_0 and an open set $\mathcal{O}''$ of $\mathbb{R}^n$ containing $f(x_0)$ such that f is a diffeomorphism of $\mathcal{O}'$ in $\mathcal{O}''$, i.e. f is bijective from $\mathcal{O}'$ in $\mathcal{O}''$ with f and f^{-1} of class C^1.*

The proof is based on the fixed point theorem. The result can therefore be easily extended to the case where f is defined on Banach spaces.

Now suppose that $f:\mathbb{R}^n\times\mathbb{R}^p\to\mathbb{R}^p$ (where $n,p\in\mathbb{N}$, $n,p\ge 1$) is of class C^1.

Theorem 2.3 (implicit functions)**.** *Let $(x_0,y_0)\in\mathbb{R}^n\times\mathbb{R}^p$. We assume that $f(x_0,y_0)=0$ and that the partial derivative $d_yf(x_0,y_0)$ is invertible. Then there exists an open set $\mathcal{O}$ of $\mathbb{R}^{n+p}$ containing (x_0,y_0), an open set $\mathcal{O}'$ of $\mathbb{R}^n$ containing x_0 and an application $g:\mathcal{O}'\to\mathbb{R}^p$ of class C^1 such that, for all $(x,y)\in\mathcal{O}$,*

$$\Big[\ f(x,y)=0\ \Big] \qquad \Leftrightarrow \qquad \Big[\ y=g(x)\ \Big]$$

In particular, $g(x_0)=y_0$ and $dg(x_0)=-(d_yf(x_0,y_0))^{-1}\circ d_xf(x_0,y_0)$.

The last relation is obtained by deriving the equality $f(x,g(x))=0$: by derivation of the composite functions, we have $d_xf(x,g(x))+d_yf(x,g(x))\circ dg(x)=0$, which gives the result (and also a mnemonic to remember the assumptions of the theorem).

2.3. **Ordinary differential equations.** This part will be useful in the M1 course of optimization, for the part concerning optimal control.

2.3.1. *The Cauchy-Lipschitz theorem.* Let $\mathcal{O}$ be an open set of $\mathbb{R} \times \mathbb{R}^n$, $f : \mathcal{O} \to \mathbb{R}^n$ *a vector field* and (t_0, x_0) *an initial condition.* We are looking for a couple (I, x) where I is an open interval containing t_0 and $x : I \to \mathbb{R}^n$ is of class C^1 and satisfies

$$x'(t) = f(t, x(t)) \; \forall t \in I, \qquad x(t_0) = x_0.$$

We say that the solution (I, x) is maximal if any other solution $(\tilde{I}, \tilde{x})$ such that $\tilde{x} = x$ on $I \cap \tilde{I}$ satisfies $\tilde{I} \subset I$. In other words, we cannot extend the solution x to a larger interval.

Theorem 2.4 (Cauchy-Lipschitz)**.** *We suppose that f is continuous on $\mathcal{O}$ and that f is $L-$Lipschitzian in x, i.e., there is a constant $L > 0$ such that*

$$\|f(t, x) - f(t, y)\| \leq L\|x - y\| \qquad \forall (t, x), (t, y) \in \mathcal{O}.$$

Then there is a unique maximal solution to the differential equation. If moreover $\mathcal{O} =]t_1, t_2[\times \mathbb{R}^n$ with $t_1 < t_0 < t_2$ and if f has linear growth, i.e., if there exists a constant $C > 0$ such that

$$\|f(t, x)\| \leq C(\|x\| + 1) \qquad \forall (t, x) \in (t_1, t_2) \times \mathbb{R}^n,$$

then the maximal solution (I, x) is defined on the entire interval $I = (t_1, t_2)$.

2.3.2. *Gronwall's lemma.* The second part of the 2.4 theorem is based on Gronwall's lemma, which is very useful in many contexts. Let $a, b : [0, +\infty[\to \mathbb{R}$ be two continuous functions with $a(t) \geq 0$ for all $t \in [0, +\infty[$.

Lemma 2.5 (from Gronwall)**.** *If a continued function $x : [0, +\infty[\to \mathbb{R}$ checks*

$$x(t) \leq b(t) + \int_0^t a(s)x(s)ds \qquad \forall t \geq 0,$$

SO

$$x(t) \leq b(t) + \int_0^t b(s)a(s) \exp\{\int_s^t a(\tau)d\tau\}ds \qquad \forall t \geq 0.$$

2.3.3. *Linear differential equations.* In this part, we identify vectors and matrices to simplify the notations. Let $A : \mathbb{R} \to M_n(\mathbb{R})$ be a continuous map (with $n \in \mathbb{N}^*$) and $b : \mathbb{R} \to \mathbb{R}^n$ continuous. We are interested in the solutions of the linear differential equation

$$x'(t) = A(t)x(t) + b(t) \qquad t \in \mathbb{R}. \tag{2}$$

A consequence of the Cauchy-Lipschitz theorem is the following result :

Theorem 2.6. *The set of solutions of the differential equation* (2) *is an affine space of dimension n.*

If A does not depend on t and $b \equiv 0$, then the solutions of (2) *are of the form $t \to e^{At}v$ where $v \in R^n$ and where the exponential of a square matrix B is defined by $e^B = \sum_{m=0}^{\infty} \frac{B^m}{m!}$.*

3. Integration

This part is largely borrowed from Jacques Fejoz's course "Intégrale de Lebesgue et Probabilités"

3.1. Measured spaces, measures.

3.1.1. *Basic notions:* $\sigma-$algebra, $\sigma-$algebra generated by a part, Borel $\sigma-$algebra, measurable space (= set equipped with a $\sigma-$algebra), measurable functions, reciprocal image of a $\sigma-$algebra by a measurable map, indicator function (of a set A, denoted $\mathbf{1}_A$).

3.1.2. *Measures.*

Definition 3.1. *If $(E,\mathcal{A})$ is a measure space, a map $\mu : \mathcal{A} \to [0,+\infty]$ is a measure if μ is $\sigma-$additive, i.e.,*

$$\mu(\bigcup_{n=0}^{\infty} A_n) = \sum_{n=0}^{\infty} \mu(A_n)$$

for any **enumerable disjoint family** *(A_n) of $\mathcal{A}$. In particular, $\mu(\emptyset) = 0$.*

Recall that, if μ is a measure, then

- μ is increasing for inclusion: if $A_1 \subset A_2$ with $A_1, A_2 \in \mathcal{A}$, then $\mu(A_1) \le \mu(A_2)$.
- if (A_n) is an increasing family of elements of $\mathcal{A}$ (i.e., $A_n \subset A_{n+1}$ for all n), then $\mu(\bigcup_n A_n) = \lim_n \mu(A_n)$.
- if (A_n) is a decreasing family of elements of $\mathcal{A}$ *and if one of $\mu(A_n)$ is finite*, then $\mu(\bigcap_n A_n) = \lim_n \mu(A_n)$.

The space $(E,\mathcal{A},\mu)$ is then called a **measure space**. When $\mu(E) = 1$, we say that μ is a **probability**.

Set of measure zero : A subset B of E is said to have measure zero (for μ) if there exists $A \in \mathcal{A}$ such that $B \subset A$ and $\mu(A) = 0$.
Property true $\mu-$almost everywhere. A property P_x (indexed by $x \in E$) is said to be true $\mu-$almost everywhere if there is a measure set null $\mathcal{N} \subset E$ such that P_x is true for all $x \in E\backslash\mathcal{N}$. For example, a map $f : E \to \mathbb{R}$ (not necessarily measurable) is zero almost everywhere if the set $\{x \in E,\ f(x) \neq 0\}$ is a zero measure set (here P_x is the property "$f(x) = 0$").

Completion: Given a measurable space $(E,\mathcal{A})$ and a measure μ on this space, it is often useful to consider the $\sigma-$algebra $\overline{\mathcal{A}}$ (the completion of the $\sigma-$algebra $\mathcal{A}$) whose elements are given by the union of an element of $\mathcal{A}$ and a set of zero measure for μ.

Standard examples of measures are:

(1) If $(E,\mathcal{A})$ is a measurable space and $a \in E$, the Dirac measure δ_a is defined by

$$\delta_a(A) = \begin{cases} 1 & \text{si } a \in A \\ 0 & \text{else} \end{cases}$$

It is a probability.

(2) On the measurable space $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$ (where $\mathcal{B}(\mathbb{R}^N)$ is the Borel $\sigma-$algebra on $\mathbb{R}^N$) , the Lebesgue measure λ is the unique measure on $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$ satisfying $\lambda(]a_1,b_1[\times\cdots\times]a_N,b_N[) = (b_1-a_1)\cdots(b_N-a_N)$ for all reals $a_1 < b_1, \ldots, a_N < b_N$. The existence and uniqueness of this measure are non-trivial results.

(3) If $(\Omega,\mathcal{A},\mathbb{P})$ is a measure space and $X : \Omega \to \mathbb{R}$ a random variable, we define the measure image $\mathbb{P}_X$ of $\mathbb{P}$ by X on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ by setting

$$\mathbb{P}_X(B) = \mathbb{P}\left[\{\omega \in \Omega,\ X(\omega) \in B\}\right] \qquad \forall B \in \mathcal{B}(\mathbb{R}) \qquad \text{(noted most often } \mathbb{P}[X \in B].)$$

(4) Of course all the classical laws in probability :
  (a) the discrete laws :
    - Bernoulli $Ber(p)$, where $p \in [0,1]$ (which is $(1-p)\delta_0 + p\delta_1$),
    - Binomial $Bin(n,p)$, where $n \in \mathbb{N}$, $p \in [0,1]$ (= sum of independent n Bernoulli),
    - Poisson $\mathcal{P}(\lambda)$, where $\lambda \in]0,+\infty[$ $(= e^{-\lambda}\sum_{n=0}^{\infty} \frac{lambda^n}{n!}\delta_n)$,
    - Geometric $\mathcal{G}(p)$ with $p \in]0,1[$ $(= \sum_{n=1}^{\infty} p(1-p)^{n-1}\delta_n)$,
  (b) continuous laws :
    - Uniform distribution $U(]a,b[)$, of density $(b-a)^{-1}\mathbf{1}_{[a,b]}$, where $a < b$.
    - Normal distribution $\mathcal{N}(m,\sigma^2)$, of density $(2\pi\sigma)^{-1/2}\exp\{-(x-m)^2/(2\sigma^2)\}$, where

$m \in \mathbb{R}$, $\sigma \in]0,+\infty[$,
- Exponential law $\mathcal{E}(\lambda)$, of density $\lambda e^{-\lambda x}\mathbf{1}_{]0,+\infty[}$, ...

Exercise 17 (Borel-Cantelli)**.** Let $(E,\mathcal{A},\mu)$ be a measured space. Let (A_n) be a sequence of elements of $\mathcal{A}$. We define $\limsup_n A_n$ and $\liminf_n A_n$ by

$$\limsup_n A_n = \bigcap_{n\in\mathbb{N}}\bigcup_{k\geq n} A_k, \qquad \liminf_n A_n = \bigcup_{n\in\mathbb{N}}\bigcap_{k\geq n} A_k.$$

Show that $\limsup_n A_n$ and $\liminf_n A_n$ are in $\mathcal{A}$, that $\liminf_n A_n \subset \limsup_n A_n$ and that, if $\sum_n \mu(A_n) < +\infty$, then $\mu(\limsup_n A_n) = 0$.

3.2. **Construction of the Lebesgue integral.** The following construction should be well known: it plays an important role in the construction of the conditional expectation (in the "Discrete Processes" course). In the sequel, $(E,\mathcal{A},\mu)$ is a measure space and $\mathbb{R}$ is endowed with the Borel $\sigma-$algebra.

3.2.1. *Integration of step functions.* A measurable map $f : E \to [0,+\infty]$ is said to be a **step function** if it takes only a finite number of values. There then exists an integer $n \in \mathbb{N}^*$, a finite number of sets $A_i \in \mathcal{A}$ and real numbers $a_i \in [0,+\infty]$ (for $i = 1,\dots,n$) such that

$$f(x) = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}(x)$$

where $\mathbf{1}_A$ is the indicator function of a set A.

If f is a nonnegative step function, we define the integral of f by

$$\int_E f(x)\mu(dx) = \sum_{i=1}^{n} a_i \mu(A_i), \tag{3}$$

with the convention that $0 \times \infty = 0$. The term on the right is well defined since it only involves sums and products of positive quantities (possibly equal to $+\infty$). The notation $\int_E f d\mu$ is also frequently used. One can show (and it is not immediate) that the expression on the right does not depend on the representation of f chosen.

3.2.2. *Integration of nonnegative functions.* Let $f : E \to \mathbb{R}$ be a **nonnegative** measurable map. We then define

$$\int_E f(x)\mu(dx) = \sup_{g \text{ step function, } 0\leq g\leq f} \int_E g(x)\mu(dx).$$

It is an element of $[0,+\infty]$.

One can be show that this definition yields formula (3) if f is a step function. In particular, if $A \in \mathcal{A}$, then $\int_E \mathbf{1}_A(x)\mu(dx) = \mu(A)$.

3.2.3. *Integration of integrable functions.* Let $f : E \to \mathbb{R}$ be a measurable map. Let $f^+(x) = \max\{0, f(x)\}$ and $f^-(x) = \max\{0, -f(x)\}$. The functions f^+ and f^- are nonnegative measurable functions with $f = f^+ - f^-$.

We say that f **is integrable** if the map $|f|$ (which is also measurable) satisfies

$$\int_E |f(x)|\mu(dx) < \infty.$$

Note that then $\displaystyle\int_E f^+(x)\mu(dx) < \infty \qquad \text{et} \qquad \int_E f^-(x)\mu(dx) < \infty.$

If f is integrable, we set

$$\int_E f(x)\mu(dx) = \int_E f^+(x)\mu(dx) - \int_E f^-(x)\mu(dx).$$

(we also use the notation $\int_E f d\mu$).

Remember that we can define $\int_E f d\mu$

(1) when $f : E \to \mathbb{R} \cup \{+\infty\}$ is **measurable nonnegative** and, in this case, $\int_E f d\mu \in [0, +\infty]$;
(2) or when $f : E \to \mathbb{R} \cup \{+\infty\}$ is **integrable** and, in this case, $\int_E f d\mu \in \mathbb{R}$.

Basic properties. The space $\mathcal{L}^1(E, \mathcal{A}, \mu)$ of integrable functions is a vector space and the integral is a linear map on this vector space. Moreover the integral is increasing in the sense that, if f_1 and f_2 are integrable and such that $f_1 \leq f_2$, then $\int_E f_1 d\mu \leq \int_E f_2 d\mu$.

Warning! $\mathcal{L}^1(E, \mathcal{A}, \mu)$ is not a normed vector space in general, because we can have $\int_E |f(x)|\mu(dx) = 0$ without f being necessarily zero (we will discuss this question later, in subsection 3.4).

Here are some usual inequalities in integration: for $p \geq 1$, let $\mathcal{L}^p(E)$ denote the set of measurable functions $f : E \to \mathbb{R}$ such that $\int_E |f|^p d\mu < \infty$. The set $\mathcal{L}^\infty(E)$ is the set of measurable functions $f : E \to \mathbb{R}$ for which there exists $M > 0$ such that $|f(x)| \leq M$ for $\mu-$almost all $x \in E$.

Theorem 3.2.

(1) *(Triangular inequality) Let $(E, \mathcal{A}, \mu)$ be a measure space. If f_1 and f_2 are integrable, then*
$$\Big|\int_E (f_1 + f_2)(x)\mu(dx)\Big| \leq \int_E |f_1(x)|\mu(dx) + \int_E |f_2(x)|\mu(dx).$$
(2) *(Hölder's inequality) Let $(E, \mathcal{A}, \mu)$ be a measure space, $p, q \in [1, +\infty]$ with $1/p+1/q = 1$ (with the convention $1/\infty = 0$). If $f \in \mathcal{L}^p(E)$ and $g \in \mathcal{L}^q(E)$, then $fg \in \mathcal{L}^1(E)$ and*
$$\left|\int_E f(x)g(x)\mu(dx)\right| \leq \left(\int_E |f(x)|^p \mu(dx)\right)^{1/p} \left(\int_E |g(x)|^q \mu(dx)\right)^{1/q}$$
(3) *(Minkowski inequality) Let $(E, \mathcal{A}, \mu)$ be a measure space and $p \geq 1$. then for all $f, g \in \mathcal{L}^p(E)$, we have $f + g \in \mathcal{L}^p(E)$ and*
$$\left(\int_E |f(x) + g(x)|^p \mu(dx)\right)^{1/p} \leq \left(\int_E |f(x)|^p d\mu(dx)\right)^{1/p} + \left(\int_E |g(x)|^p d\mu(dx)\right)^{1/p}.$$
(4) *(Jensen's inequality) If $(E, \mathcal{A}, \mu)$ is a probability space (= measure space with μ probability), then for any convex function $\phi : \mathbb{R} \to R$ and for any integrable function f on E we have*
$$\phi\left(\int_E f(x)\mu(dx)\right) \leq \int_E \phi(f(x))\mu(dx).$$

Exercise 18. (1) Prove Hölder's inequality: we can first assume that $\int_E |f|^p d\mu = \int_E |g|^q d\mu = 1$ and use Young's inequality $ab \leq \frac{a^p}{p} + \frac{b^q}{q}$ for $a, b \geq 0$.
(2) Prove Hölder's inequality from Jensen's inequality (hint: use the probability measure $\mu_f := |f|^p/(\int_E |f|^p d\mu)$).
(3) Prove Minkowski's inequality from Hölder's inequality.

Case of complex valued functions: When $f : E \to \mathbb{C}$ takes complex values, we say that f is integrable if f is measurable and if $|f|$ (the modulus of f) is integrable. In this case the real and complex parts $\mathcal{R}e(f)$ and $\mathcal{I}m(f)$ are integrable and we set
$$\int_E f(x)\mu(dx) = \int_E \mathcal{R}e(f)(x)\mu(dx) + i \int_E \mathcal{I}m(f)(x)\mu(dx).$$

3.3. **Central properties of the Lebesgue integral.** These are the monotone convergence theorem, Fatou's lemma, dominated convergence theorem and Fubini's theorem.

Theorem 3.3 (of monotone convergence)**.** *If (f_n) is an increasing sequence of nonnegative measurable functions on E (in the sense that $0 \le f_n \le f_{n+1}$ for all n), then*

$$\lim_{n\to+\infty} \int_E f_n d\mu = \int_E \lim_n f_n d\mu.$$

The advantage of this result is that it does not require any integrability on the f_n. The price to pay is the condition that the f_n are positive. Note carefully the dissymmetry of the result: it does not hold if we just suppose that (f_n) is a decreasing sequence of positive functions for example.

Lemma 3.4 (de Fatou[a])**.** *If (f_n) is a sequence of nonnegative measurable functions on E, then*

$$\int_E \liminf_n f_n d\mu \le \liminf_n \int_E f_n d\mu.$$

[a]We recall that, if (a_n) is a real sequence, then $\liminf_n a_n = \lim_{k\to+\infty} \inf_{n\ge k} a_n$ and $\limsup_n a_n = \lim_{k\to+\infty} \sup_{n\ge k} a_n$. Of course, if (a_n) has a limit (finite or infinite), then $\liminf_n a_n = \limsup_n a_n = \lim_n a_n$.

As for the monotone convergence theorem, the sign condition is essential and leads to an asymmetrical statement (with an inequality this time, but without assuming monotonicity in the sequence).

Exercise 19.
(1) Show that the monotone convergence theorem and Fatou's lemma remain true without sign condition on the (f_n) if there exists integrable map g such that $f_n \ge g$ for all n.
(2) More generally, show that, if there exist integrable functions g_1 and g_2 such that $g_1 \le f_n \le g_2$ for all n, then

$$\int_E \liminf_n f_n d\mu \le \liminf_n \int_E f_n d\mu \le \limsup_n \int_E f_n d\mu \le \int_E \limsup_n f_n d\mu.$$

Theorem 3.5 (Lebesgue dominated convergence theorem)**.** *If f_n is a sequence of measurable functions from E to $\mathbb{R}$ such that*

(1) there is an integrable function $g : E \to \mathbb{R}$ verifying:

$$|f_n(x)| \le g(x) \qquad \text{for } \mu\text{−almost all } x \in E,$$

(2) there is a (measurable) function $f : E \to \mathbb{R}$ such that

$$\lim_{n\to+\infty} f_n(x) = f(x) \qquad \text{for } \mu\text{−almost all } x \in E,$$

then f is integrable and

$$\lim_{n\to+\infty} \int_E f_n d\mu = \int_E f d\mu.$$

Unlike the two previous results, the domination condition is perfectly symmetric. This makes this result the main tool for manipulating the Lebesgue integral.

Exercise 20 (Sum and integral inversion)**.** Let (f_n) be a sequence of measurable functions from E into $\mathbb{R}$ for which there exists an integrable function g such that

$$\sum_n |f_n(x)| \le g(x) \qquad \text{for } \mu\text{−almost all } x \in E.$$

Show that the function $x \to \sum_n f_n(x)$ is defined for μ−almost all $x \in E$ and integrable and that

$$\int_E (\sum_n f_n(x))\mu(dx) = \sum_n \int_E f_n(x)\mu(dx).$$

Exercise 21 (A.e. convergence and convergence in measure)**.** Let $(E, \mathcal{A}, \mu)$ be a measure space with $\mu(E) < \infty$. A sequence $f_n : E \to \mathbb{R}$ of measurable functions is said to converge $\mu-$p.p. to a measurable function $f : E \to \mathbb{R}$ if the set of $x \in E$ such that $(f_n(x))$ does not converge to $f(x)$ has measure zero. We say that (f_n) converges in measure to f if

$$\forall \varepsilon > 0, \lim_{n \to +\infty} \mu(\{x \in E, \ |f_n(x) - f(x)| \geq \varepsilon\}) = 0.$$

(1) Show that, if (f_n) converges to f $\mu-$p.p., then (f_n) converges in measure to f.
(2) Give a counterexample to the previous assertion if we remove the assumption that $\mu(E) < \infty$.
(3) Prove that, if (f_n) converges in measure to f, then there exists a subsequence (f_{n_k}) which converges $\mu-$p.p. to f. (we can show that there exists an extracted sequence (n_k) such that, for all $k \geq 1$, $\mu\{x \in E, \ |f_{n_k}(x) - f(x)| \geq 1/k\} \leq 1/k^2$ and use the Borel-Cantelli lemma (Exercise 17)).

Fubini's theorem deals with integration over product spaces. Let $(E_1, \mathcal{A}_1, \mu_1)$ and $(E_2, \mathcal{A}_2, \mu_2)$ be two measurable spaces. The $\sigma-$algebra $\mathcal{A}_1 \otimes \mathcal{A}_2$ is the $\sigma-$algebra on $E_1 \times E_2$ generated by the set of products $A_1 \times A_2$ with $A_1 \in \mathcal{A}_1$ and $A_2 \in \ mathcal A_2$, while the measure $\mu_1 \otimes \mu_2$ is the only measure on $(E_1 \times E_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ such that $\mu_1 \otimes \mu_2(A_1 \ times A_2) = \mu_1(A_1)\mu_2(A_2)$ for all $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$.

Theorem 3.6 (Fubini)**.** *The theorem has two parts, the first for nonnegative functions, and the second for integrable functions :*

(1) If $f : E_1 \times E_2 \to \mathbb{R}$ is measurable and nonnegative, then

$$\int_{E_1 \times E_2} f(x_1, x_2)(\mu_1 \otimes \mu_2)(dx_1, dx_2) = \int_{E_1} \Big(\int_{E_2} f(x_1, x_2)\mu_2(dx_2) \Big) \mu_1(dx_1)$$
$$= \int_{E_2} \Big(\int_{E_1} f(x_1, x_2)\mu_1(dx_1) \Big) \mu_2(dx_2). \tag{4}$$

(2) If $f : E_1 \times E_2 \to \mathbb{R}$ is integrable for $\mu_1 \otimes \mu_2$, then
(a) for μ_1-almost all $x_1 \in E_1$, $x_2 \to f(x_1, x_2)$ is integrable for μ_2,
(b) for μ_2-almost all $x_2 \in E_2$, $x_1 \to f(x_1, x_2)$ is integrable for μ_1,
(c) and finally the two equalities in (4) *remain true.*

One often uses the first part of the theorem to show the integrability condition of the second part.

Exercise 22. Using Fubini's theorem, check that if $(E, \mathcal{A}, \mu)$ is a measure space and if $f \in \mathcal{L}^p(E)$ (for $p \geq 1$), then

$$\int_E |f(x)|^p \mu(dx) = \int_0^\infty p t^{p-1} \mu(\{x \in E, \ |f(x)| \geq t\}) dt$$

3.4. **The spaces $\mathcal{L}^p$ and L^p.**

Definition 3.7. *Let $(E, \mathcal{A}, \mu)$ be a measure space. For $p \in [1, \infty)$, we define the sets $\mathcal{L}^p(E)$ as the set of measurable functions $f : E \to \mathbb{R}$ (or in $\mathbb{C}$) such that $\int_E |f(x)|^p \mu(dx)$ is finite. For $p = 0$, $\mathcal{L}^0(E)$ is the set of measurable functions from E to $\mathbb{R}$. Note that $\mathcal{L}^p(E) \subset \mathcal{L}^0(E)$ for all $p \in [1, \infty)$.*

The spaces $\mathcal{L}^p(E)$ are vector spaces. However, for $p \in [1, +\infty)$ these spaces do not have a "natural" normed vector space structure because we can very well have $\int_E |f(x)|^p \mu(dx) = 0$ without f being zero. In fact, $\int_E |f(x)|^p \mu(dx) = 0$, if and only if $f(x) = 0$ for $\mu-$almost all $x \in E$.

To overcome this (small) difficulty, we introduce the following equivalence relation. Let $f, g \in \mathcal{L}^0(E)$. We say that $f \sim g$ if the function $f - g = 0$ $\mu-$almost-everywhere. It is easy to see that this relation is indeed an equivalence relation, compatible with the vector space structure of $\mathcal{L}^0(E)$. Moreover, for $p \in [1, +\infty)$, if $f \in \mathcal{L}^p(E)$, $g \in \mathcal{L}^0(E)$ and $f \sim g$, then $g \in \mathcal{L}^p(E)$. So $\sim$ is also an equivalence relation on $\mathcal{L}^p(E)$.

Definition 3.8. *The space $L^p(E)$ is the set of equivalence classes of the elements of $\mathcal{L}^p(E)$ for the relation $\sim$. For all $\dot{f} \in L^p(E)$, we set*

$$\|\dot{f}\|_p := \left(\int_E |f(x)|^p \mu(dx)\right)^{1/p}$$

where f is any element of the equivalence class $\dot{f}$ (denoted $f \in \dot{f}$).

Moreover, if $f \in \dot{f}$, $g \in \dot{g}$ and $\alpha \in \mathbb{R}$, then $\alpha\dot{f} + \dot{g} = \overset{\cdot}{\overbrace{\alpha f + g}}$. Therefore the space $L^p(E)$ has a vector space structure and we show that $\|\cdot\|_p$ is a norm on this space. **By abuse of notation, we simply denote by f, g etc... the elements of $L^p(E)$ instead of $\dot{f}$, $\dot{g}$, etc. ..., even though they are actually equivalence classes and not functions.**

The case $p = \infty$. We define $L^\infty(E)$ as being the set of equivalence classes of elements of $\mathcal{L}^0(E)$ for which there exists $M > 0$ with $|f(x)| \leq M$ for $\mu-$almost all $x \in E$. We pose

$$\|f\|_\infty = \inf\{M > 0,\ |f(x)| \leq M \text{ for } \mu\text{−almost all } x \in E.\}.$$

The space $(L^\infty(E), \|\cdot\|_\infty)$ is a normed vector space.

Theorem 3.9. *The spaces $(L^p(E), \|\cdot\|_p)$ are Banach spaces for $p \in [1, \infty]$. For $p = 2$, the space $(L^2(E), \|\cdot\|_2)$ is a Hilbert space, with associated scalar product*

$$\langle f, g\rangle = \int_E f(x)g(x)\mu(dx).$$

Finally, if $\mu(E) < \infty$, then

$$L^p(E) \subset L^q(E) \qquad \text{si } 1 \leq p \leq q \leq \infty.$$

Here are some density properties.

Theorem 3.10. *The set of step functions is dense in $L^p(E)$ for $p \in [1, \infty]$.*

Theorem 3.11. *We suppose that $E = (a, b)$ (with $-\infty \leq a < b \leq +\infty$) and that μ is the Lebesgue measure on (a, b). Then, for all $p \in [1, +\infty)$, the set $C_c^\infty(]a, b[)$ of functions of class C^∞ with compact support in (a, b) is dense in $L^p((a, b))$. In particular, $L^p((a, b))$ is separable.*

More generally, if $\mathcal{O}$ is an open set of $\mathbb{R}^n$ (where $n \in \mathbb{N}^*$), μ is the Lebesgue measure and $p \in [1, +\infty)$, then $C_c^\infty(\mathcal{O})$ of functions of class C^∞ with compact support in $\mathcal{O}$ is dense in $L^p(\mathcal{O})$. In particular, $L^p(\mathcal{O})$ is separable.

Warning! this last result is false in L^∞.

3.5. **Functions defined by integrals.** A classical consequence of the dominated convergence theorem is the analysis of function depending on a parameter. In this part we fix $(E, \mathcal{A}, \mu)$ a measure space, A an open set of $\mathbb{R}^N$ (where $N \in \mathbb{N}$) and $f : E \times A \to \mathbb{R}$ an application. We are interested in the application (when it is defined)

$$g(a) := \int_E f(x, a)\mu(dx).$$

Theorem 3.12 (Continuity)**.** *Let $a_0 \in A$. We suppose that*

1. *for $\mu-$almost all $x \in E$, $a \to f(x, a)$ is continuous at a_0,*
2. *for all $a \in A$, $x \to f(x, a)$ is measurable,*
3. *there is an integrable function $h : E \to [0, +\infty[$ such that, for all $a \in A$, for $\mu-$almost all $x \in E$,*

$$|f(x, a)| \leq h(x).$$

Then the map g is continuous at a_0.

Theorem 3.13 (Differentiability). *Suppose that*

1. *for $\mu-$almost all $x \in E$, the map $a \to f(x,a)$ is differentiable on A (with differential noted $d_a f(x,a)$) ,*
2. *for all $a \in A$, $x \to f(x,a)$ is integrable,*
3. *there is an integrable function $h : E \to [0, +\infty[$ such that, for all $a \in A$, for $\mu-$almost all $x \in E$,*

$$\|d_a f(x,a)\| \leq h(x).$$

Then the map g is differentiable on A and its differential at a point $a_0 \in A$ is given by

$$dg(a_0) = \int_E d_a f(x,a_0)\mu(dx).$$

Finally, if the map $a \to d_a f(x,a)$ is continuous on A for $\mu-$almost all $x \in E$, then g is of class C^1 on A.

The transposition to case where f takes its values ??in $\mathbb{C}$ is immediate (changing absolute value in modulus).

The most common examples of application are the **regularization by convolution**, the **Fourier transform** of a function or a measure (in probability language, we speak of a characteristic function- up to the complex conjugation).

Exercise 23. Let μ be a measure of total finite mass on $\mathbb{R}$. We set

$$\hat{\mu}(t) := \int_{\mathbb{R}} e^{-itx}\mu(dx) \qquad \forall t \in \mathbb{R}.$$

Show that the map $t \to \hat{\mu}(t)$ is continuous on $\mathbb{R}$. Check that it is of class C^k (where $k \in \mathbb{N}$, $k \geq 1$) if μ has a finite moment of order k, i.e., if $int_{\mathbb{R}}|x|^k\mu(dx) < +\infty$ and then

$$\hat{\mu}^{(k)}(t) = \int_{\mathbb{R}} (-ix)^k e^{-itx}\mu(dx) \qquad \forall t \in \mathbb{R}.$$

Exercise 24. Let $f : [0, +\infty[\to \mathbb{R}$ be the function defined by

$$f(x) = \left(\int_0^x e^{-t^2}dt\right)^2 + \int_0^1 \frac{e^{-x^2(1+t^2)}}{1+t^2}dt.$$

Show that $f(0) = \pi/4$, that $\lim_{x\to+\infty} f(x) = (\int_0^\infty e^{-t^2}dt)^2$ and that f is differentiable with zero derivative. Deduce that $\int_0^\infty e^{-t^2}dt = \sqrt{\pi}/2$.

3.6. **Change of variable formula.** We work here with the Lebesgue measure on $\mathbb{R}^n$, denoted dx. Let $\mathcal{O}$ and $\mathcal{O}'$ be two open subsets of $\mathbb{R}^n$ and $\phi : \mathcal{O} \to \mathcal{O}'$ a diffeomorphism (i.e. ϕ is a bijection of $\mathcal{O}$ in $\mathcal{O}'$ with ϕ and ϕ^{-1} of class C^1). Then, for any nonnegative or integrable function $f : \mathcal{O} \to \mathbb{R}$ (for the Lebesgue measure), the function $y \to f(\phi^{-1}(y))\left|\det(J_\phi(\phi^{-1})(y)\right|$ is nonnegative or integrable over $\mathcal{O}'$ and

$$\int_{\mathcal{O}} f(x)dx = \int_{\mathcal{O}'} f(\phi^{-1}(y))\left|\det(J_\phi(\ phi^{-1}(y))\right| dy.$$

We will see a classic application of this formula in the study of convolution.

For example, we have the formula for **"polar coordinates" in the plane** : if $f : \mathbb{R}^2 \to \mathbb{R}$ is integrable, then

$$\int_{\mathbb{R}^2} f(x_1,x_2)dx_1dx_2 = \int_0^\infty \int_0^{2\pi} rf(r\cos(\theta), r\sin(\theta))d\theta dr.$$

3.7. **Fourier transform of a function.** Let $f \in L^1(\mathbb{R}^n)$. We recall that the Fourier transform of f is defined by

$$\hat{f}(\xi) = \int_{\mathbb{R}^n} f(x) e^{-i\langle x,\xi\rangle} dx \qquad \forall \xi \in \mathbb{R}^n.$$

Proposition 3.14. *The function $\xi \to \hat{f}(\xi)$ is a continuous function, which tends to 0 when $\|\xi\| \to +\infty$ and verifies*

$$\sup_{\xi \in \mathbb{R}^n} |f(\xi)| \leq \int_{\mathbb{R}^n} |f(x)| dx.$$

Exercise 25. This exercise shows how to quantify the convergence of $\hat{f}$ to 0. We assume that $f : \mathbb{R} \to \mathbb{C}$ is of class C^1 and that f and f' are in $L^1(\mathbb{R})$. Show that $\widehat{f'}(\xi) = (i\xi)\hat{f}(\xi)$ for all $\xi \in \mathbb{R}$ and deduce that $\lim_{|\xi| \to \infty} |\xi||\hat{f}(\xi)| = 0$.

Exercise 26 (A classic application of Fubini's theorem)**.** Show that if $f, g \in L^1(\mathbb{R}^n)$ then $\int_{\mathbb{R}^n} \hat{f}(\xi) g(\xi) d\xi = \int_{\mathbb{R}^n} f(\xi) \hat{g}(\xi) d\xi$.

If $f \in L^1(\mathbb{R}^n)$, we set

$$\check{f}(x) = \int_{\mathbb{R}^n} e^{i\langle x,\xi\rangle} f(\xi) d\xi \qquad \forall x \in \mathbb{R}^n.$$

Theorem 3.15 (Inversion formula)**.** *If $f \in L^1(\mathbb{R}^n)$ and $\hat{f} \in L^1(\mathbb{R}^n)$, then*

$$f = \frac{1}{(2\pi)^n} \check{f} \qquad p.p.$$

More importantly for its applications in probability:

Theorem 3.16. *The map that associates its Fourier transform to a Borel measure μ on $\mathbb{R}^d$:*

$$\hat{\mu}(\xi) := \int_{\mathbb{R}^n} e^{-i\langle x,\xi\rangle} \mu(dx) \qquad \forall \xi \in \mathbb{R}^n$$

is injective.

In other words, to demonstrate the equality between two Borel measures μ_1 and μ_2 on $\mathbb{R}^n$, it suffices to show that $\hat{\mu}_1(\xi) = \hat{\mu}_2(\ xi)$ for all $\xi \in \mathbb{R}^n$.

3.8. **Convolution product.**

Theorem 3.17. *Let f and g be two integrable functions on $\mathbb{R}^n$ (for the Lebesgue measure). Then the map $x \to f * g(x) := \int_{\mathbb{R}^n} f(x-y)g(y)dy$ is finite for almost all $x \in \mathbb{R}^n$ and $f * g$ is integrable over $\mathbb{R}^n$.*

Notes:

(1) The convolution product is bilinear and symmetric: if $\alpha \in \mathbb{R}$ and if $f_1, f_2, g : \mathbb{R}^n \to \mathbb{R}$ are integrable, then $(\alpha f_1 + f_2) * g = \alpha(f_1 * g) + f_2 * g$ and $f * g = g * f$.
(2) The convolution product plays a central role in probability: if X and Y are two independent random variables on $\mathbb{R}^n$ with respective densities f_X and f_Y, then the sum $X + Y$ also has a density on $\mathbb{R}^n$ and this density is $f_X * f_Y$: in particular, for any continuous function $\phi : \mathbb{R}^n \to \mathbb{R}$ bounded on R^n, we have

$$\mathbb{E}[\phi(X+Y)] = \int_{\mathbb{R}^n} \phi(z)(f_X * f_Y)(z) dz.$$

Proof of theorem. The map $(x,y) \to f(y-x)g(y)$ is integrable on $\mathbb{R}^{2n}$ because

$$\int_{\mathbb{R}^{2n}} |f(x-y)g(y)|dxdy = \int_{\mathbb{R}^{2n}} |f(z)g(y)|dzdy = \left(\int_{\mathbb{R}^n} |f(z)|dz\right)\left(\int_{\mathbb{R}^n} |g(y)|dy\right) < +\infty,$$

where the first equality comes by change of variable $(x,y) \to (y-x,y)$, whose determinant of the Jacobian has absolute value equal to 1, and the second equality is given by Fubini's theorem (first part). Therefore, still by Fubini's theorem (but second part), the map $x \to f*g(x) := \int_{\mathbb{R}^n} f(x-y)g(y)dy$ is finite for almost all $x \in \mathbb{R}^n$ and $f*g$ is integrable on $\mathbb{R}^n$. $\square$

Exercise 27. Show that if $f \in L^1(\mathbb{R}^n)$ and $f \in L^p(\mathbb{R}^n)$ (with $p \geq 1$), then $f*g \in L^p(\mathbb{R}^n)$ with $\|f*g\|_p \leq \|f\|_1\|g\|_p$.

Exercise 28. Show that if $f, g \in L^1(\mathbb{R}^n)$, then

$$\widehat{f*g}(\xi) = \hat{f}(\xi)\hat{g}(\xi) \qquad \forall \xi \in \mathbb{R}^n.$$

A more analytical application of convolution is the approximation of integrable functions by regular functions. Let $\phi : \mathbb{R}^n \to \mathbb{R}$ be a map of class C^∞, with compact support, positive and integral 1 : $\int_{\mathbb{R}^n} \phi(x)dx = 1$ (such an application exists, but its existence is not simple). For $\varepsilon > 0$, let $\phi^\varepsilon(x) = \varepsilon^{-n}\phi(x/\varepsilon)$.

Theorem 3.18. *If $f : \mathbb{R}^n \to \mathbb{R}$ is integrable over $\mathbb{R}^n$, then, for all $\varepsilon > 0$, the function $f^\varepsilon := \phi^\varepsilon * f$ is integrable, of class C^∞ on $\mathbb{R}^n$ and*

$$\lim_{\varepsilon \to 0} \int_{\mathbb{R}^n} |f(x) - f^\varepsilon(x)|dx = 0.$$

4. Application to probability theory

4.1. Vocabulary.

- **A probability space** $(\Omega, \mathcal{A}, \mathbb{P})$ is the data of a set Ω equipped with a $\sigma-$algebra $\mathcal{A}$ and a probability measure $\mathbb{P}$ (i.e., $\mathbb{P}$ is a measure on $\mathcal{A}$ such that $\mathbb{P}(\Omega) = 1$). The elements of $\mathcal{A}$ are called events.
- **A random vector** X is a measurable map from Ω into $\mathbb{R}^n$, where $n \in \mathbb{N}^*$ and $\mathbb{R}^n$ is endowed with the Borel $\sigma-$algebra; if $n = 1$, we speak of **random variable**. If $f : \mathbb{R}^n \to \mathbb{R}$ (or in $\mathbb{C}$) is Borel and bounded, we set

$$\mathbb{E}[f(X)] := \int_\Omega f(X(\omega))\mathbb{P}(d\omega).$$

Note that, although X depends on ω, we almost never write $X(\omega)$.

If X is a random variable, **the cumulative distribution function of** X (CDF function), is the function $F_X : \mathbb{R} \to [0,1]$ defined by

$$F_X(t) = \mathbb{P}[X \leq t].$$

Finally, if X is a random variable, $\sigma(X)$ is the $\sigma-$algebra $\{X^{-1}(A),\ A \in \mathcal{A}\}$.
- **Law of a random vector.** In probability, a random vector X (over $\mathbb{R}^n$) is very often apprehended by its **law** $\mathbb{P}_X$, which is a probability on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, where $\mathcal{B}(\mathbb{R}^n)$ is the Borel $\sigma-$algebra; $\mathbb{P}_X$ is defined by

$$\mathbb{P}_X[B] = \mathbb{P}[X \in B] \qquad (= \mathbb{P}[\{\omega \in \Omega,\ X(\omega) \in B\}]) \qquad \forall B \in \mathcal{B}(\mathbb{R}^n)$$

or, equivalently, by equality

$$\mathbb{E}[\phi(X)] = \int_{\mathbb{R}^n} \phi(x)\mathbb{P}_X(dx) \qquad \text{for all } \phi : \mathbb{R}^n \to \mathbb{R} \text{ borelian bounded.}$$

(this last equality is sometimes known as the **transfer theorem**). For example, if X is a random variable with distribution function F_X, then

$$F_X(t) = \mathbb{P}_X(]-\infty, t]) \qquad \forall t \in \mathbb{R}.$$

- The random vector X on $\mathbb{R}^n$ **is said to have a density** if there exists an integrable function $f_X : \mathbb{R}^n \to \mathbb{R}$ such that

$$\mathbb{P}_X[B] = \int_B f_X(x)dx \qquad \forall B \in \mathcal{B}(\mathbb{R}^n)$$

or, equivalently,

$$\mathbb{E}[\phi(X)] = \int_{\mathbb{R}^n} \phi(x) f_X(x)dx \qquad \forall \phi : \mathbb{R}^n \to \mathbb{R} \text{ Borel and bounded.}$$

The function f_X (which is uniquely defined in $L^1(\mathbb{R}^n, dx)$ by one of the above relations) is nonnegative and has an integral equal to 1.

- **The characteristic function** Φ_X of a random vector X on $\mathbb{R}^n$ is (the complex conjugate of) the Fourier transform of $\mathbb{P}_X$:

$$\Phi_X(t) = \mathbb{E}\left[e^{i\langle X,t\rangle}\right] = \int_{\mathbb{R}^n} e^{i\langle x,t\rangle}\mathbb{P}_X(dx) \qquad \forall t \in \mathbb{R}^n,$$

where $\langle \cdot, \cdot \rangle$ is the scalar product in $\mathbb{R}^n$.

The characteristic function characterizes the distribution : if X and Y are two random vectors such that $\Phi_X = \Phi_Y$, then X and Y have the same distribution, i.e., $\mathbb{P}_X = \mathbb{P}_Y$.

4.2. **Variance.** If X, Y are two square-integrable random variables (i.e., $X, Y \in L^2(\Omega)$), the covariance of (X, Y) is

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}(Y))] = \mathbb{E}[XY] - \mathbb{E}[X]\ E[Y].$$

The variance of X is

$$\text{var}(X) = \text{cov}(X, X) = \mathbb{E}\left[|X - \mathbb{E}(X)|^2\right] = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2.$$

We have

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y).$$

Proposition 4.1 (Two fundamental inequalities)**.**

- *(Markov inequality) If $X \in L^1(\Omega)$, then, for all $a > 0$,*

$$\mathbb{P}[|X| \geq a] \leq \frac{\mathbb{E}[|X|]}{a}.$$

- *(Bienaymé-Tchebichev inequality) If $X \in L^2(\Omega)$, then for all $a > 0$,*

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq a] \leq \frac{\text{var}(X)}{a^2}.$$

More generally if $X = (X_1, \ldots, X_n) \in (L^2(\Omega))^n$ is a random vector, the covariance matrix of X is the symmetric matrix

$$K_X := (\text{cov}(X_i, X_j))_{1 \leq i,j \leq n}.$$

This equality can also be written (by identifying a vector with the associated column matrix)

$$K_X = \mathbb{E}\left[(X - \mathbb{E}[X])\ (X - \mathbb{E}[X])^T\right],$$

where the expectation of a matrix is ??simply the matrix of the expectation of its coefficients.

Exercise 29. Let $X = (X_1, \dots, X_n) \in (L^2(\Omega))^n$ be a random vector, M be a matrix $p \times n$ and $b \in \mathbb{R}^p$. Show that

$$K_{MX+b} = MK_X\ M^T.$$

Show that K_X is a positive semidefinite matrix.

Be careful not to confuse the fact that random variables are uncorrelated (i.e., in the case of two v.a., such as $\mathrm{cov}(X_1, X_2) = 0$) from the fact that they are independent. We will see right after that the notion of independence is much stronger : basically, X_1 and X_2 are independent if "any" function of X_1 is decorrelated from "any" function of X_2. These two notions, quite distinct in general, coincide when the vector (X_1, X_2) is Gaussian (see below).

4.3. **Independence.** We say that events A_1 and A_2 are independent if $\mathbb{P}[A_1 \cap A_2] = \mathbb{P}[A_1]\mathbb{P}[A_2]$. More generally, the events $\{A_i\}_{i\in I}$ (where I is any set) are independent if, for any finite subset $J \subset I$, we have

$$\mathbb{P}\left[\bigcap_{i\in J} A_i\right] = \prod_{i\in J} \mathbb{P}\left[A_i\right].$$

We say that $\sigma-$algebras $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent if, for all $(A_1, A_2) \in \mathcal{A}_1 \times \mathcal{A}_2$, the events A_1 and A_2 are independent. This notion generalizes as above to an arbitrary family of $\sigma-$algebras.

Definition 4.2. *Two random variables X_1 and X_2 are said to be independent if one of the equivalent assertions holds:*

(1) the $\sigma-$algebras $\sigma(X_1)$ and $\sigma(X_2)$ generated by X_1 and X_2 are independent.

(2) For any pair of bounded measurable functions $f_1, f_2 : \mathbb{R} \to \mathbb{R}$, we have

$$\mathbb{E}\left[f_1(X_1)f_2(X_2)\right] = \mathbb{E}\left[f_1(X_1)\right]\mathbb{E}\left[f_2(X_2)\right].$$

(3) (characterization in terms of law) $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$; in other words, for any bounded Borel function $f : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$, we have

$$\mathbb{E}\left[f(X,Y)\right] = \int_{\mathbb{R}} (\int_{\mathbb{R}} f(x,y)\mathbb{P}_Y(dy))\mathbb{P}_X(dx) = \int_R (\int_{\mathbb{R}} f(x,y)\mathbb{P}_X(dx))\mathbb{P}_Y(dy).$$

(4) (characterization in terms of characteristic functions)

$$\Phi_{(X_1,X_2)}(\xi_1,\xi_2) = \Phi_{X_1}(\xi_1)\Phi_{X_2}(\xi_2) \qquad \forall(\xi_1,\xi_2) \in \mathbb{R}^2.$$

The statements (2) and (4) allow *to show* that the v.a. are independent; one *most often uses* this independence via the assertion (3). These equivalences generalize to any number of random variables. Random variables $(X_n)_{n\in\mathbb{N}}$ are said to be **i.i.d.** if they are **independent and identically distributed**.

Exercise 30 (Box-Muller algorithm)**.** Let U_1 and U_2 be two independent variables with uniform distribution on $[0,1]$. Show that the random variables

$$Z_1 = \sqrt{-2\ln(U_1)}\cos(2\pi U_2), \qquad Z_2 = \sqrt{-2\ln(U_1)}\sin(2\pi U_2)$$

are independent and of normal distribution $\mathcal{N}(0,1)$ (Hint: switch to polar coordinates).

4.4. **Convergence of random variables.**

Definition 4.3. *Let (X_n) be a sequence of random variables defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and X a random variable. We say that (X_n) converges to X*

- **almost surely** *if*
$$\mathbb{P}\left[\lim_n X_n = X\right] = 1 \qquad (\text{noted } X_n \xrightarrow{a.s.} X)$$
- **uniformly** *if*
$$\lim_n \|X_n - X\|_\infty \longrightarrow 0.$$
(it is implied here that the X_n and X are bounded).
- **in average of order** p **(or in** L^p**)** *(where $p \in [1, +\infty)$) if*
$$\lim_n \mathbb{E}\left[|X_n - X|^p\right] = 0 \qquad (\text{denoted } X_n \xrightarrow{L^p} X)$$
(It is implied here that the X_n and X are in $L^p(\Omega)$). If $p = 1$, we speak of **mean convergence** *and, if $p = 2$, of* **mean quadratic convergence**.
- **in probability** *if, for all $\varepsilon > 0$,*
$$\lim_n \mathbb{P}\left[|X_n - X| \geq \varepsilon\right] = 0. \qquad (\text{denoted } X_n \xrightarrow{\mathbb{P}} X)$$
- **in law** *if, for any continuous bounded function $f : \mathbb{R} \to \mathbb{R}$,*
$$\lim_n \mathbb{E}\left[f(X_n)\right] = \mathbb{E}\left[f(X)\right]. \qquad (\text{denoted } X_n \xrightarrow{\mathcal{L}} X)$$
Note that, for this last form of convergence, it is not useful for the X_n and X to be defined on the same probability space.

The notions above generalize easily to the case of random vectors (if the vectors are in $\mathbb{R}^d$, replace the absolute value by any norm on $\mathbb{R}^d$, the function f in the convergence in law having to go from $\mathbb{R}^d$ in $\mathbb{R}$).

There are many relationships between these different forms of convergence, it is difficult to retain them all. It is better to know by heart the following relations (to save time) :

Proposition 4.4 (Link between convergences)**.**

(1) Convergence in probability implies convergence in law.
(2) Convergence a.s. as well as mean convergence of order p (where $p \in [1, +\infty[$) imply convergence in probability and convergence in law.
(3) Convergence in average implies the existence of a subsequence which converges a.s..
(4) If $1 \leq p \leq q$, convergence in L^q implies convergence in L^p.
(5) Uniform convergence implies all other convergences.

The following characterization of convergence in law in terms of distribution functions and characteristic functions is very useful:

Theorem 4.5. *Under the assumptions of the definition 4.3, the following assertions are equivalent:*

- *(X_n) converges in law to X,*
- *$F_{X_n}(t) \to F_X(t)$ at any point t of continuity of F_X,*
- *$\Phi_{X_n}(t) \to \Phi_X(t)$ for all $t \in \mathbb{R}$,*

where F_{X_n} (resp. Φ_{X_n}) is the cumulative distribution function (resp. characteristic function) of X_n.

The last equivalence is known as Lévy's theorem and holds true for random vectors. While it occurs very frequently in probability and statistics, convergence in law is tricky to handle because it is incompatible with usual operations. Remember though:

Theorem 4.6 (Slutsky)**.** *If (X_n) converges in law to a random variable X and (Y_n) converges in probability to a constant c, then (X_n, Y_n) converges in law to (X, c). In particular, $(X_n + Y_n)$ converges in distribution to $(X + c)$ and $(X_n Y_n)$ converges in distribution to cX.*

4.5. **Gaussian vectors.** Recall that the Gaussian (or normal) law $\mathcal{N}(m, \sigma^2)$ (with mean $m \in \mathbb{R}$ and standard deviation $\sigma \geq 0$) is

- if $\sigma > 0$, the probability on $\mathbb{R}$ of density $\dfrac{1}{\sigma\sqrt{2\pi}} \exp\{-(x-m)^2/(2\sigma^2)\}$.
- if $\sigma = 0$, the Dirac mass δ_m.

We say that the distribution is reduced if $\sigma = 0$, centered if $m = 0$ and degenerate if $\sigma = 0$.

Definition 4.7. *A random vector $(X_1, \ldots, X_n)$ over $(\Omega, \mathcal{A}, \mathbb{P})$ is said to be Gaussian if any linear combination of $(X_i)_{i=1,\ldots,n}$ follows a Gaussian distribution (in other words, for all $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$, the random variable $\alpha_1 X_1 + \cdots + \alpha_n X_n$ follows a Gaussian distribution).*

In particular, if $(X_1, \ldots, X_n)$ is a Gaussian vector, all random variables X_i $(i = 1, \ldots, n)$ are Gaussian.

Theorem 4.8 (Characterization by characteristic functions)**.** *A random vector $X = (X_1, \ldots, X_n)$ is Gaussian if and only if there exists a (deterministic) vector $m_X \in \mathbb{R}^n$ and a symmetric (deterministic) matrix $K_X \in \mathbb{R}^{n\times n}$ such that*

$$\Phi_X(\xi) = \exp\left\{ i\langle m_X, \xi\rangle - \frac{1}{2}\langle K_X\xi, \xi\rangle \right\} \qquad \forall \xi \in \mathbb{R}^n ot.$$

Moreover, m_X is the expectation of the vector X and K_X is the covariance matrix.

Proposition 4.9. *We assume that the random vector $X = (X_1, \ldots, X_n)$ is Gaussian. Then the random variables $\{X_1, \ldots, X_n\}$ are independent, if and only if, the covariance matrix K_X of X is diagonal.*

This result is a direct consequence of the characterization of the independence in terms of characteristic functions and of the specific form of the characteristic functions for Gaussian vectors.

Theorem 4.10 (Gaussian Density Vectors)**.** *We assume that the random vector $X = (X_1, \ldots, X_n)$ is Gaussian. Then this vector has a density, if and only if $\det(K_X) \neq 0$ (where K_X is the covariance matrix of X). In this case, the density of X is given by*

$$f_X(x) = \frac{1}{(2\pi)^{n/2}(\det(K_X))^{1/2}} \exp\{-\frac{1}{2}\langle K_X^{-1}(x - m_X), (x - m_X)\rangle\} \qquad \forall x = (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Appendix A. Additional exercises

A.1. **Topology and functional analysis.**

Exercise 31. If (X, d) is a metric space, prove the inverse triangular inequality
$\Big|d(x, y_1) - d(x, y_2)\Big| \leq d(y_1, y_2)$ for all $x, y_1, y_2 \in X$.

Exercise 32 (Point of closure)**.** Let (x_n) be a sequence in a metric space (X, d) and $\ell \in X$. Show the equivalence between the following assertions:

(1) For all $\varepsilon > 0$, the set $\{n \in \mathbb{N},\ d(x_n, \ell) < \varepsilon\}$ is infinite,
(2) $\ell \in \bigcap_n \overline{\{x_p : p \geq n\}}$ (where $\bar{A}$ is the closure of a set $A \subset X$),
(3) there is a subsequence $(x_{\phi(n)})$ such that $\lim x_{\phi(n)} = \ell$.

We then say that ℓ is a point of closure of (x_n).

Exercise 33. Assume that $(X_1, d_1), (X_2, d_2)$ are two compact metric spaces. Show that the product $X_1 \times X_2$, endowed with the distance $d((x_1, x_2), (y_1, y_2)) = \max(d_1(x_1, y_1), d_2(x_2, y_2))$, is compact.

Exercise 34. Show that a continuous bijection $f : X \to Y$ where X is compact metric and Y metric, is of continuous reciprocal. (hint: use Exercise 7).

Exercise 35. Let (X, d) be a metric space and $A \subset X$ a non-empty subset of X. Assume that $f : A \to \mathbb{R}$ is a L-Lipschitz function on A and define for all $x \in X$

$$\bar{f}(x) = \inf_{y \in A} f(y) + Ld(x, y).$$

Show that $\bar{f}$ is a $L-$Lipschitz function on X which is equal to f on A.

Exercise 36. Let (X, d) be a metric space.

(1) Assume that A is a closed subset of X. Show that there exists a continuous function $f : X \to [0, +\infty[$ which vanishes exactly on A.
(2) Assume that A and B are two closed subsets of X with $A \cap B = \emptyset$. Show that there exists a continuous function $g : X \to [0, 1]$ such that $A = \{x \in X,\ g(x) = 0\}$ and $B = \{x \in X,\ g(x) = 1\}$.

Exercise 37. Let $(E, \|\cdot\|)$ be a normed vector space and F a vector subspace of E. Prove that, if F has a non-empty interior, then $F = E$.

Exercise 38 (A version of the Cauchy-Lipschitz theorem)**.** Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a $L-$Lipschitz function (where $L \in \mathbb{R}^*$), $x_0 \in \mathbb{R}^n$ and, for $T > 0$, $E_T := C^0([-T, T], \mathbb{R}^n)$, endowed with the norm

$$\|x\|_{E_T} = \sup_{t \in [-T, T]} \|x(t)\|.$$

We define the function $\Phi : E_T \to E_T$ by

$$\Phi(x)(t) = x_0 + \int_0^t f(x(s))ds \qquad \forall t \in [-T, T],\ \forall x \in E_T.$$

(1) Show that, if $T > 0$ is sufficiently small, the map Φ_T is a contraction on E_T.
(2) Deduce that, if $T > 0$ is small enough, there is a unique solution to the differential equation

$$x't) = f(x(t)),\ \forall t \in [-T, T], \qquad x(0) = x_0.$$

A.2. **Differential calculus.**

Exercise 39. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x, y) = \frac{xy^2}{x^2 + y^4}$ if $(x, y) \neq (0, 0)$ and $f(0, 0) = 0$. Show that the restriction of f to any line passing through the origin is continuous, but that f is not continuous at $(0, 0)$.

Exercise 40. Let $\mathcal{O} := \{(x,y) \in \mathbb{R}^2,\ x \neq y\}$. Show that the map $f : \mathcal{O} \to \mathbb{R}$ defined by $f(x,y) = \frac{\sin(x)-\sin(y)}{x-y}$ can be extended by continuity to $\mathbb{R}^2$. Show that this extension is of class C^1 over $\mathbb{R}^2$.

Exercise 41. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x,y) = \frac{xy^2}{x^2+y^2}$ if $(x,y) \neq (0,0)$ and $f(0,0) = 0$. Show that f is continuous at $(0,0)$, that $\frac{\partial f}{\partial x}(0,0)$ and $\frac{\partial f}{\partial y}(0,0)$ exist, but f is not differentiable at $(0,0)$.

Exercise 42. Let $f : \mathbb{R}^n \to \mathbb{R}$ (where $n \in \mathbb{N}^*$) be defined by $f(x) = \sin(\|x\|^2 + 1)$ for all $x \in \mathbb{R}^n$ ($\|x\|$ being the Euclidean norm of $x \in \mathbb{R}^n$). Show that f is of class C^2 over $\mathbb{R}^n$, compute its gradient and its Hessian matrix at every point.

Exercise 43 (Euler relation)**.** Let $f : (\mathbb{R}^n \backslash \{0\}) \to \mathbb{R}$ be a function of class C^1 and $k \in \mathbb{R}^*$. Show that f is $k-$homogeneous (i.e., $f(\lambda x) = \lambda^k f(x)$ for all $x \in (\mathbb{R}^n)^*$ and for all $\lambda > 0$), if and only if, $\langle f(x), x\rangle = kf(x)$ for all $x \in \mathbb{R}^n \backslash \{0\}$.

Exercise 44. Let $M_n(\mathbb{R})$ be the set of matrices of format $n \times n$ and $GL_n(\mathbb{R})$ the set of invertible matrices of $M_n(\mathbb{R})$.

(1) Show that $GL_n(\mathbb{R})$ is an open subset of $M_n(\mathbb{R})$.
(2) Show that the following maps are of class C^1 on their domain of definition and compute their differential.

$$f(A) = A^2, \qquad \forall A \in M_n(\mathbb{R}); \qquad g(A) = A^{-1}, \qquad \forall A \in GL_n(\mathbb{R}).$$

A.3. **Integration.**

Exercise 45. Let $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, where λ is the Lebesgue measure.

(1) Let $\mathcal{O}$ be an open set of $\mathbb{R}$. Show that $\lambda(\mathcal{O}) > 0$.
(2) Let $\mathcal{O}$ be an open set of $\mathbb{R}$. We assume that $\lambda(\mathcal{O}) < \infty$. Is the set $\mathcal{O}$ necessarily bounded ?
(3) Construct a dense open set of $\mathbb{R}$ of Lebesgue measure equal to 3.

Exercise 46. Let μ and ν be two finite Borel measures on $\mathbb{R}^d$. We set

$$\sigma(A) := \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbf{1}_{x+y \in A} (\mu \otimes \nu)(dx, dy), \quad \text{for all Borel set } A \subseteq \mathbb{R}^d.$$

(1) Show that σ is a positive measure on $\mathbb{R}^d$, we write $\mu * \nu := \sigma$.
(2) Note that $\mu * \nu$ is a finite measure and that $\mu * \nu = \nu * \mu$.
(3) Show that if $\mu = f\lambda^d, \nu = g\lambda^d$, where λ^d is the Lebesgue measure on $\mathbb{R}^d$ and $f, g \geq 0$ are integrable, then $\mu \star \nu = (f \star g)\lambda^d$.

Exercise 47. In the following cases (where $f_n : \mathbb{R}^+ \to \mathbb{R}$) show that the sequence $(\int_{\mathbb{R}^+} f_n(x)dx)_{n\in\mathbb{N}}$ converges and compute its limit:

(1) $f_n(x) = \dfrac{ne^{-x}}{\sqrt{1+n^2x^2}}$,
(2) $f_n(x) = \dfrac{ne^{-nx}}{\sqrt{1+n^2x^2}}$,
(3) $f_n(x) = \sin(nx)\mathbf{1}_{[0,n]}(x)$,
(4) $f_n(x) = |\cos(x)|^{1/n} e^{-x}$,
(5) $f_n(x) = \dfrac{ne^{-x}}{nx+1}\mathbf{1}_{[0,1]}$,
(6) $f_n(x) = \dfrac{\sin(nx^n)}{nx^{n+\frac{1}{2}}}$.

Exercise 48. Compute the limit of the following sequences:

$$\int_{\mathbb{R}} e^{-|x|/n} dx, \qquad \int_{\mathbb{R}} \frac{e^{-x^2}}{2\cos(\frac{x}{n}) - 1}\ \mathbf{1}_{\{3|\cos(\frac{x}{n})| \geq 2\}}\ dx, \qquad \sum_{m \geq 1} \frac{n}{m} \sin(\frac{1}{nm}).$$

Exercise 49. (1) Show that the map φ defined by $\varphi(u,v) = (u^2+v^2, 2uv)$ is a C^1-diffeomorphism from $\Delta = \{(u,v) \in \mathbb{R}^2; u > v > 0\}$ onto $D = \{(x,y) \in \mathbb{R}^2; x > y > 0\}$.
(2) Deduce the value of $\int_{(\mathbb{R}_+)^2} |u^4 - v^4| e^{-(u+v)^2} dudv$.

Exercise 50. Compute the integral

$$I = \int_{y>x>0} e^{-y+x} \frac{\sqrt{y-x}}{y^2} dxdy.$$

[Hint : we can consider the change of variable $u = y - x, v = y/x$.]

A.4. **Probability.**

Exercise 51. Let X, Y be two real independent random variables. Compute the law of $X + Y$ in the following cases:

(1) X and Y have uniform law on $[-1, 1]$.
(2) X and Y have respectively a density law $\gamma_{a,\lambda}$ and $\gamma_{b,\lambda}$ where

$$\gamma_{a,\lambda}(x) = \frac{\lambda^a}{\Gamma(a)} e^{-\lambda x} x^{a-1} \mathbf{1}_{\mathbb{R}_+}(x), \quad \Gamma(a) := \int_0^\infty e^{-x}\, x^{a-1}\, dx.$$

We can check and use the fact that $\int_{\mathbb{R}_+} \gamma_{a,\lambda} = 1$.

Exercise 52. Let (X_1, X_2) be a pair of random variables of density

$$f(x_1, x_2) = e^{-(x_1+x_2)} \mathbf{1}_{\mathbb{R}_+}(x_1) \mathbf{1}_{\mathbb{R}_+}(x_2) \qquad \forall (x_1, x_2) \in \mathbb{R}^2.$$

(1) What is the density of $Y_1 = X_1 + X_2$?
(2) What is the density of $Y_2 = X_1/X_2$?
(3) Are the random variables Y_1 and Y_2 independent ?

Exercise 53. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. random variable such as $X_0 \in L^2(\Omega)$. Show that

$$\bar{X}_N := \frac{1}{N} \sum_{n=1}^{N} X_n \text{ tends to } \mathbb{E}[X_0] \text{ in } L^2(\Omega).$$

Exercise 54. Let X be a random vector in $\mathbb{R}^d$ (where $d \in \mathbb{N}^*$) and M be a real matrix of size $k \times d$.

(1) Show that

$$\mathbb{E}[MX] = M\mathbb{E}[X] \qquad \text{and} \qquad K_{MX} = M\ K_X\ M^T.$$

(2) Assume that X is a Gaussian vector with distribution $\mathcal{N}(0, I_d)$. What is the law of MX ?
(3) Suppose that X is a Gaussian vector on $\mathbb{R}^3$ with distribution $\mathcal{N}(0, D)$ with

$$D = \begin{pmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Show that there exists an orthogonal matrix $P \in \mathbb{R}^{3\times 3}$ such that $P^{-1}DP$ is diagonal. Determine the law of the vector $P^{-1}X$.

Exercise 55. Let $X_1, \ldots, X_n$ be independent random variables with the same distribution $\mathcal{N}(0, 1)$ and $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ reals. Let $Y = \sum_{i=1}^n a_i X_i$ and $Z = \sum_{i=1}^n b_i X_i$.

(1) Show that (Y, Z) is a Gaussian vector.
(2) Show that Y and Z are independent, if and only if, $\sum_{i=1}^n a_i b_i = 0$.
(3) Show that the random variables $X_1+X_2+X_3$, $2X_1-X_2-X_3$ and X_2-X_3 are independent.

Index