

Appendix A

Topics in measure theory, topology and analysis

In this series of appendices we recall several basic concepts and facts in measure theory, topology and functional analysis that are useful throughout the book. Our purpose is to provide the reader with a quick, accessible source of references to measure and integration, general and differential topology and spectral theory, to try and make this book as self-contained as possible. We have not attempted to make the material in these appendices completely sequential: it may happen that a notion mentioned in one section is defined or discussed in more depth in a later one (check the index).

As a general rule, we omit the proofs. For Appendices A.1, A.2 and A.5, the reader may find detailed information in the books of Castro [Cas04], Fernandez [Fer02], Halmos [Hal50], Royden [Roy63] and Rudin [Rud87]. The presentation in Appendix A.3 is a bit more complete, including the proofs of most results, but the reader may find additional relevant material in the books of Billingsley [Bil68, Bil71]. We recommend the books of Hirsch [Hir94] and do Carmo [dC79] to all those interested in going further into the topics in Appendix A.4. For more information on the subjects of Appendices A.6 and A.7, including proofs of the results quoted here, check the book of Halmos [Hal51] and the treatise of Dunford and Schwarz [DS57, DS63], especially Section IV.4 of the first volume and the initial sections of the second volume.

A.1 Measure spaces

Measure spaces are the natural environment for the definition of the Lebesgue integral, which is the main topic to be presented in Appendix A.2. We begin by introducing the notions of algebra and σ -algebra of subsets of a set, which lead to the concept of measurable space. Next, we present the notion of measure on a σ -algebra and we analyze some of its properties. In particular, we mention a few results on the construction of measures, including Lebesgue measures in

Euclidean spaces. The last part is dedicated to measurable maps, which are the maps that preserve the structure of measurable spaces.

A.1.1 Measurable spaces

Given a set X , we often denote by A^c the complement $X \setminus A$ of each subset A .

Definition A.1.1. An algebra of subsets of a set X is a family \mathcal{B} of subsets of X that contains the empty set and is closed under the elementary operations of set theory:

- (i) $\emptyset \in \mathcal{B}$;
- (ii) $A \in \mathcal{B}$ implies $A^c \in \mathcal{B}$;
- (iii) $A \in \mathcal{B}$ and $B \in \mathcal{B}$ implies $A \cup B \in \mathcal{B}$;
- (iv) $A \in \mathcal{B}$ and $B \in \mathcal{B}$ implies $A \cap B \in \mathcal{B}$;
- (v) $A \in \mathcal{B}$ and $B \in \mathcal{B}$ implies $A \setminus B \in \mathcal{B}$.

The two last properties are immediate consequences of the previous ones, since $A \cap B = (A^c \cup B^c)^c$ and $A \setminus B = A \cap B^c$. Moreover, by associativity, properties (iii) and (iv) imply that the union and the intersection of any finite family of elements of \mathcal{B} are also in \mathcal{B} .

Definition A.1.2. A σ -algebra of subsets of a set X is an algebra \mathcal{B} of subsets of X that is also closed under countable unions:

$$A_j \in \mathcal{B} \text{ for } j = 1, \dots, n, \dots \quad \text{implies} \quad \bigcup_{j=1}^{\infty} A_j \in \mathcal{B}.$$

Then \mathcal{B} is also closed under countable intersections:

$$A_j \in \mathcal{B} \text{ for } j = 1, \dots, n, \dots \quad \text{implies} \quad \bigcap_{j=1}^{\infty} A_j = \left(\bigcup_{j=1}^{\infty} A_j^c \right)^c \in \mathcal{B}.$$

Definition A.1.3. A measurable space is a pair (X, \mathcal{B}) where X is a set and \mathcal{B} is a σ -algebra of subsets of X . The elements of \mathcal{B} are called measurable sets.

Next, we describe a few examples of constructions of σ -algebras.

Example A.1.4. For any set X , the following families of subsets are σ -algebras:

$$\{\emptyset, X\} \quad \text{and} \quad 2^X = \{ \text{all subsets of } X \}.$$

Moreover, clearly, if \mathcal{B} is any algebra of subsets of X then $\{\emptyset, X\} \subset \mathcal{B} \subset 2^X$. So, $\{\emptyset, X\}$ is the smallest and 2^X is the largest of all algebras of subsets of X .

In the statement that follows, \mathcal{I} is an arbitrary set whose sole use is to index the elements of the family of σ -algebras.

Proposition A.1.5. Consider any non-empty family $\{\mathcal{B}_i : i \in \mathcal{I}\}$ of σ -algebras of subsets of the same set X . Then the intersection $\mathcal{B} = \bigcap_{i \in \mathcal{I}} \mathcal{B}_i$ is also a σ -algebra of subsets of X .

Given any family \mathcal{E} of subsets of X , we may apply Proposition A.1.5 to the family of all σ -algebras that contain \mathcal{E} . Note that this family is non-empty, since it contains the σ -algebra 2^X of all subsets of X . According to the previous proposition, the intersection of all these σ -algebras is also a σ -algebra. By construction, this σ -algebra contains \mathcal{E} and is contained in every σ -algebra that contains \mathcal{E} . In other words, it is the smallest σ -algebra that contains \mathcal{E} . This leads to the following definition:

Definition A.1.6. The σ -algebra generated by a family \mathcal{E} of subsets of X is the smallest σ -algebra $\sigma(\mathcal{E})$ that contains \mathcal{E} or, in other words, the intersection of all the σ -algebras that contain \mathcal{E} .

Recall that a *topological space* is a pair (X, τ) where X is a set and τ is a family of subsets of X that contains $\{\emptyset, X\}$ and is closed under finite intersections and arbitrary unions. Such a family τ is called a *topology* and its elements are called *open subsets* of X . In this book we take all topological spaces to be *Hausdorff*, that is, such that for any pair of distinct points there exists a pair of disjoint open subsets each of which contains one of the points.

Definition A.1.7. The *Borel σ -algebra* of a topological space is the σ -algebra $\sigma(\tau)$ generated by the topology τ , that is, the smallest σ -algebra that contains all the open subsets of X . The elements of $\sigma(\tau)$ are called *Borel subsets* of X . The *closed* subsets of X , being the complements of the open subsets, are also in the Borel σ -algebra.

Analogously to Proposition A.1.5, the intersection of any non-empty family $\{\tau_i : i \in \mathcal{I}\}$ of topologies of the same set X is also a topology of X . Then, by the same argument as we used before for σ -algebras, given any family \mathcal{E} of subsets of X there exists a smallest topology $\tau(\mathcal{E})$ that contains \mathcal{E} . We call it the *topology generated by \mathcal{E}* .

Example A.1.8. Let (X, \mathcal{B}) be a measurable space. The *limit superior* of a sequence of sets $E_n \in \mathcal{B}$ is the set $\limsup_n E_n$ formed by the points $x \in X$ such that $x \in E_n$ for infinitely many values of n . Analogously, the *limit inferior* of $(E_n)_n$ is the set $\liminf_n E_n$ of points $x \in X$ such that $x \in E_n$ for every value of n sufficiently large. In other words,

$$\liminf_n E_n = \bigcup_{n \geq 1} \bigcap_{m \geq n} E_m \quad \text{and} \quad \limsup_n E_n = \bigcap_{n \geq 1} \bigcup_{m \geq n} E_m.$$

Observe that $\liminf_n E_n \subset \limsup_n E_n$ and both sets are in \mathcal{B} .

Example A.1.9. The *extended line* $\bar{\mathbb{R}} = [-\infty, \infty]$ is the union of the real line $\mathbb{R} = (-\infty, +\infty)$ with the two points $\pm\infty$ at infinity. This space has a natural

topology, generated by the intervals $[-\infty, b)$ and $(a, +\infty]$, with $a, b \in \mathbb{R}$. It is easy to see that the extended line is homeomorphic to a compact interval in the real line: for example, the function $\arctan : \mathbb{R} \rightarrow (-\pi/2, \pi/2)$ extends straightforwardly to a *homeomorphism* (that is, a continuous bijection whose inverse is also continuous) between $\overline{\mathbb{R}}$ and $[-\pi/2, \pi/2]$. We always consider on the extended line the Borel σ -algebra associated with this topology.

Of course, the real line \mathbb{R} is a subspace (measurable as well as topological) of the extended line. The Borel subsets of the real line constitute a large family and one might even be led to think that every subset of \mathbb{R} is a Borel subset. However, this is not true: a counterexample is constructed in Exercise A.1.4.

A.1.2 Measure spaces

Let (X, \mathcal{B}) be a measurable space. The following notions have a central role in this book:

Definition A.1.10. A *measure* on (X, \mathcal{B}) is a function $\mu : \mathcal{B} \rightarrow [0, +\infty]$ such that $\mu(\emptyset) = 0$ and

$$\mu\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mu(A_j)$$

for any countable family of pairwise disjoint sets $A_j \in \mathcal{B}$. This last property is called *countable additivity* or σ -*additivity*. Then the triple (X, \mathcal{B}, μ) is called a *measure space*. If $\mu(X) < \infty$ then we say that μ is a *finite* measure and if $\mu(X) = 1$ then we call μ a *probability measure*. In this last case, (X, \mathcal{B}, μ) is called a *probability space*.

Example A.1.11. Let X be an arbitrary set, endowed with the σ -algebra $\mathcal{B} = 2^X$. Given any $p \in X$, consider the function $\delta_p : 2^X \rightarrow [0, +\infty]$ defined by:

$$\delta_p(A) = \begin{cases} 1 & \text{if } p \in A \\ 0 & \text{if } p \notin A. \end{cases}$$

It is easy to see that δ_p is a measure. It is usually called the *Dirac measure*, or *Dirac mass* at p .

Definition A.1.12. We say that a measure μ is σ -*finite* if there exists a sequence A_1, \dots, A_n, \dots of subsets of X such that $\mu(A_i) < \infty$ for every $i \in \mathbb{N}$ and

$$X = \bigcup_{i=1}^{\infty} A_i.$$

We say that a function $\mu : \mathcal{B} \rightarrow [0, +\infty]$ is *finitely additive* if

$$\mu\left(\bigcup_{j=1}^N A_j\right) = \sum_{j=1}^N \mu(A_j)$$

for any finite family $A_1, \dots, A_N \in \mathcal{B}$ of pairwise disjoint subsets. Note that if μ is σ -additive then it is also finitely additive. Moreover, if μ is finitely additive and is not constant equal to $+\infty$ then $\mu(\emptyset) = 0$.

The main tool for constructing measures is the following theorem:

Theorem A.1.13 (Extension). *Let \mathcal{A} be an algebra of subsets of X and let $\mu_0 : \mathcal{A} \rightarrow [0, +\infty]$ be a σ -additive function with $\mu_0(X) < \infty$. Then there exists a unique measure μ defined on the σ -algebra \mathcal{B} generated by \mathcal{A} that is an extension of μ_0 , meaning that it satisfies $\mu(A) = \mu_0(A)$ for every $A \in \mathcal{A}$.*

Theorem A.1.13 remains valid for σ -finite measures. Moreover, there is a version for finitely additive functions: if μ_0 is finitely additive then it admits a finitely additive extension to σ -algebra \mathcal{B} generated by \mathcal{A} . However, in this context the extension need not be unique.

The most useful criterion for proving that a given function is σ -additive is provided by the following theorem:

Theorem A.1.14 (Continuity at the empty set). *Let \mathcal{A} be an algebra of subsets of X and $\mu : \mathcal{A} \rightarrow [0, +\infty)$ be a finitely additive function with $\mu(X) < \infty$. Then μ is σ -additive if and only if*

$$\lim_n \mu(A_n) = 0 \tag{A.1.1}$$

for every sequence $A_1 \supset \dots \supset A_j \supset \dots$ of elements of \mathcal{A} with $\bigcap_{j=1}^\infty A_j = \emptyset$.

The proof of this theorem is proposed in Exercise A.1.7. Exercise A.1.9 deals with some variations of the statement.

Definition A.1.15. We say that an algebra \mathcal{A} is *compact* if any decreasing sequence $A_1 \supset \dots \supset A_n \supset \dots$ of non-empty elements of \mathcal{A} has non-empty intersection.

An *open cover* of a topological space is a family of open subsets whose union is the whole of K . A *subcover* is just a subfamily of elements of a cover whose union is still the whole space. A topological space is *compact* if every open cover admits some finite subcover. A subset K of a topological space X is *compact* if the topology of X restricted to K turns the latter into a compact topological space. Every closed subset of a compact space is compact. Conversely, (assuming X is a Hausdorff space) then every compact subset is closed. Another important fact is that the intersection $\bigcap_n K_n$ of any decreasing sequence $K_1 \supset \dots \supset K_n \supset \dots$ of compact subsets is non-empty.

Example A.1.16. It follows from what we have just said that if X is a (Hausdorff) topological space and every element of the algebra \mathcal{A} is compact then \mathcal{A} is a compact algebra.

It follows from Theorem A.1.14 that if \mathcal{A} is a compact algebra then every finitely additive function $\mu : \mathcal{A} \rightarrow [0, +\infty)$ with $\mu(X) < \infty$ is σ -additive.

Hence, by Theorem A.1.13, μ extends uniquely to a measure defined on the σ -algebra generated by \mathcal{A} .

Definition A.1.17. We say that a non-empty family \mathcal{C} of subsets of X is a *monotone class* if \mathcal{C} contains X and is closed under countable monotone unions and intersections:

- if $A_1 \subset A_2 \subset \dots$ are in \mathcal{C} then $\bigcup_{n \geq 1} A_n \in \mathcal{C}$, and
- if $A_1 \supset A_2 \supset \dots$ are in \mathcal{C} then $\bigcap_{n \geq 1} A_n \in \mathcal{C}$.

Clearly, the two families $\{\emptyset, X\}$ and 2^X are monotone classes. Moreover, if $\{\mathcal{C}_i : i \in \mathcal{I}\}$ is any family of monotone classes then the intersection $\bigcap_{i \in \mathcal{I}} \mathcal{C}_i$ is a monotone class. Thus, for every subset \mathcal{A} of 2^X there exists the smallest monotone class that contains \mathcal{A} .

Theorem A.1.18 (Monotone class). *The smallest monotone class that contains an algebra \mathcal{A} coincides with the σ -algebra $\sigma(\mathcal{A})$ generated by \mathcal{A} .*

Another important result about σ -algebras that will be useful later states that every element of a σ -algebra \mathcal{B} generated by an algebra \mathcal{A} is approximated by the elements of \mathcal{A} , in the sense that the measure of the symmetric difference

$$A \Delta B = (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B)$$

can be made arbitrarily small. More precisely:

Theorem A.1.19 (Approximation). *Let (X, \mathcal{B}, μ) be a probability space and \mathcal{A} be an algebra \mathcal{A} of subsets of X that generates the σ -algebra \mathcal{B} . Then, for every $\varepsilon > 0$ and every $B \in \mathcal{B}$ there exists $A \in \mathcal{A}$ such that $\mu(A \Delta B) < \varepsilon$.*

Definition A.1.20. A measure space is *complete* if every subset of a measurable set with zero measure is also measurable.

It is possible to transform any measure space (X, \mathcal{B}, μ) into a complete space, as follows. Let $\bar{\mathcal{B}}$ be the family of all subsets $A \subset X$ such that there exist $B_1, B_2 \in \mathcal{B}$ with $B_1 \subset A \subset B_2$ and $\mu(B_2 \setminus B_1) = 0$. Then $\bar{\mathcal{B}}$ is a σ -algebra and it contains \mathcal{B} . Consider the function $\bar{\mu} : \bar{\mathcal{B}} \rightarrow [0, +\infty]$ defined by $\bar{\mu}(A) = \mu(B_1) = \mu(B_2)$, for any $B_1, B_2 \in \mathcal{B}$ as before. The function $\bar{\mu}$ is well defined, it is a measure on $\bar{\mathcal{B}}$ and its restriction to \mathcal{B} coincides with μ . By construction, $(X, \bar{\mathcal{B}}, \bar{\mu})$ is a complete measure space. It is called the *completion* of (X, \mathcal{B}, μ) .

Given subsets \mathcal{U}_1 and \mathcal{U}_2 of the σ -algebra \mathcal{B} , we say that $\mathcal{U}_1 \subset \mathcal{U}_2$ up to measure zero if for every $B_1 \in \mathcal{U}_1$ there exists $B_2 \in \mathcal{U}_2$ such that $\mu(B_1 \Delta B_2) = 0$. By definition, $\mathcal{U}_1 = \mathcal{U}_2$ up to measure zero if $\mathcal{U}_1 \subset \mathcal{U}_2$ up to measure zero and $\mathcal{U}_2 \subset \mathcal{U}_1$ up to measure zero. We say that a set $\mathcal{U} \subset \mathcal{B}$ generates the σ -algebra \mathcal{B} up to measure zero if the σ -algebra generated by \mathcal{U} is equal to \mathcal{B} up to measure zero. Equivalently, \mathcal{U} generates \mathcal{B} up to measure zero if the completion of the σ -algebra generated by \mathcal{U} coincides with the completion of \mathcal{B} .

By definition, a measure takes values in $[0, \infty]$. Whenever it is convenient to stress that fact, we speak of *positive measure* instead. But it is possible to weaken that requirement and, indeed, such generalizations are useful for our purposes.

We call a *signed measure* on a measurable space (X, \mathcal{B}) any σ -additive function $\mu : \mathcal{B} \rightarrow [-\infty, \infty]$ such that $\mu(\emptyset) = 0$. More precisely, μ may take either the value $-\infty$ or the value $+\infty$, but not both; this is to avoid the “indetermination” $\infty - \infty$ in the additivity condition.

Theorem A.1.21 (Hahn decomposition). *If μ is a signed measure then there exist measurable sets $P, N \subset X$ such that $P \cup N = X$ and $P \cap N = \emptyset$, and*

$$\mu(E) \geq 0 \text{ for every } E \subset P \quad \text{and} \quad \mu(E) \leq 0 \text{ for every } E \subset N.$$

This means that we may write $\mu = \mu^+ - \mu^-$, where μ^+ and μ^- are the (positive) measures defined by

$$\mu^+(E) = \mu(E \cap P) \quad \text{and} \quad \mu^-(E) = -\mu(E \cap N).$$

In particular, the sum $|\mu| = \mu^+ + \mu^-$ is also a positive measure; it is called the *total variation* of the signed measure μ .

If μ takes values in $(-\infty, \infty)$ only, we call it a *finite signed measure*. In this case, the measures μ^+ and μ^- are finite. The set $\mathcal{M}(X)$ of finite signed measures is a real vector space and the function $\|\mu\| = |\mu|(X)$ is a complete norm in this space (see Exercise A.1.10). In other words, $(\mathcal{M}(X), \|\cdot\|)$ is a real Banach space. When X is a compact metric space, this Banach space is isomorphic to the dual of the space $C^0(X)$ of continuous real functions X (theorem of Riesz–Markov).

More generally, we call a *complex measure* on a measurable space (X, \mathcal{B}) any σ -additive function $\mu : \mathcal{B} \rightarrow \mathbb{C}$. Observe that $\mu(\emptyset)$ is necessarily zero. Clearly, we may write $\mu = \Re\mu + i\Im\mu$, where the real part $\Re\mu$ and the imaginary part $\Im\mu$ are finite signed measures. The *total variation* of μ is the finite measure defined by

$$|\mu|(E) = \sup_{\mathcal{P}} \sum_{P \in \mathcal{P}} |\mu(P)|,$$

where the supremum is taken over all countable partitions of the measurable set E into measurable subsets (this definition coincides with the one we gave previously in the special case when μ is real). The function $\|\mu\| = |\mu|(X)$ defines a norm in the vector space of complex measures on X , which we also denote as $\mathcal{M}(X)$. Moreover, this norm is complete. When X is a compact metric space, the complex Banach space $(\mathcal{M}(X), \|\cdot\|)$ is isomorphic to the dual of the space $C^0(X)$ of continuous complex functions on X (theorem of Riesz–Markov).

A.1.3 Lebesgue measure

The notion of Lebesgue measure corresponds to the notion of volume of subsets of the Euclidean space \mathbb{R}^d . It is defined as follows.

Let $X = [0, 1]$ and \mathcal{A} be the family of all subsets of the form $A = I_1 \cup \dots \cup I_N$ where I_1, \dots, I_N are pairwise disjoint intervals. It is easy to check that \mathcal{A} is an algebra of subsets of X . Let $m_0 : \mathcal{A} \rightarrow [0, 1]$ be the function defined on this algebra by

$$m_0(I_1 \cup \dots \cup I_N) = |I_1| + \dots + |I_N|,$$

where $|I_j|$ represents the length of each interval I_j . Note that $m_0(X) = 1$. In Exercise A.1.8 we ask the reader to show that m_0 is σ -additive.

Note that the σ -algebra \mathcal{B} generated by \mathcal{A} coincides with the Borel σ -algebra of X , since every open subset is a countable union of pairwise intervals. So, by Theorem A.1.13, there exists a unique probability measure m defined on \mathcal{B} that is an extension of m_0 . It is called the *Lebesgue measure* on $[0, 1]$.

More generally, one defines the Lebesgue measure m on the *cube* $X = [0, 1]^d$ of any dimension $d \geq 1$, in the following way. First, we call a *rectangle* in X any subset of the form $R = I_1 \times \dots \times I_d$ where the I_j are intervals. Then we define:

$$m_0(R) = |I_1| \times \dots \times |I_d|.$$

Next, we consider the algebra \mathcal{A} of subsets of X of the form $A = R_1 \cup \dots \cup R_N$, where R_1, \dots, R_N are pairwise disjoint rectangles, and we define

$$m_0(A) = m_0(R_1) + \dots + m_0(R_N)$$

for every A in that algebra. The σ -algebra generated by \mathcal{A} coincides with the Borel σ -algebra of X . The *Lebesgue measure* on the cube $X = [0, 1]^d$ is the extension of m_0 to that σ -algebra.

In order to define the Lebesgue measure on the whole Euclidean space \mathbb{R}^d , we decompose the space into cubes of unit size:

$$\mathbb{R}^d = \bigcup_{k_1 \in \mathbb{Z}} \dots \bigcup_{k_d \in \mathbb{Z}} [k_1, k_1 + 1) \times \dots \times [k_d, k_d + 1).$$

Each cube $[k_1, k_1 + 1) \times \dots \times [k_d, k_d + 1)$ may be identified with $[0, 1)^d$ through the translation $T_{k_1, \dots, k_d}(x) = x - (k_1, \dots, k_d)$ that maps (k_1, k_2, \dots, k_d) to the origin. That allows us to define a measure m_{k_1, k_2, \dots, k_d} on C , by setting

$$m_{k_1, k_2, \dots, k_d}(B) = m_0(T_{k_1, \dots, k_d}(B))$$

for every measurable set $B \subset C$. Finally, given any measurable set $B \subset \mathbb{R}^d$, define:

$$m(B) = \sum_{k_1 \in \mathbb{Z}} \dots \sum_{k_d \in \mathbb{Z}} m_{k_1, \dots, k_d}(B \cap [k_1, k_1 + 1) \times \dots \times [k_d, k_d + 1)).$$

Note that this measure m is σ -finite but not finite.

Example A.1.22. It is worthwhile outlining a classical alternative construction of the Lebesgue measure (see Chapter 2 of Royden [Roy63] for details). We call the *Lebesgue exterior measure* of an arbitrary set $E \subset \mathbb{R}^d$ the number

$$m^*(E) = \inf \sum_k m_0(R_k),$$

where the infimum is taken over all countable covers $(R_k)_k$ of E by open rectangles. The function $E \mapsto m(E)$ is defined for every $E \subset \mathbb{R}^d$, but is not finitely additive (although it is countably subadditive). We say that E is a *Lebesgue measurable set* if

$$m^*(A) = m^*(A \cap E) + m^*(A \cap E^c) \quad \text{for every } A \subset \mathbb{R}^d.$$

Every rectangle R is a Lebesgue measurable set and satisfies $m^*(R) = m_0(R)$. The family \mathcal{M} of all Lebesgue measurable sets is a σ -algebra. Moreover, the restriction of m^* to \mathcal{M} is σ -additive and, hence, a measure. By the previous observation, \mathcal{M} contains every Borel set of \mathbb{R}^d . The restriction of m^* to the Borel σ -algebra \mathcal{B} of \mathbb{R}^d coincides with the Lebesgue measure on \mathbb{R}^d .

Actually, \mathcal{M} coincides with the completion of the Borel σ -algebra of \mathbb{R}^d with respect to the Lebesgue measure. This and other related properties are part of Exercise A.1.13.

Example A.1.23. Let $\phi : [0, 1] \rightarrow \mathbb{R}$ be a positive continuous function. Given any interval I , with endpoints $0 \leq a < b \leq 1$, define

$$\mu_\phi(I) = \int_a^b \phi(x) dx \quad (\text{Riemann integral}).$$

Next, extend the definition of μ_ϕ to the algebra \mathcal{A} formed by the finite unions $A = I_1 \cup \dots \cup I_k$ of pairwise disjoint intervals, through the relation

$$\mu_\phi(A) = \sum_{j=1}^k \mu_\phi(I_j).$$

The basic properties of the Riemann integral ensure that μ_ϕ is finitely additive. We leave it to the reader to check that the measure μ_ϕ is σ -additive in the algebra \mathcal{A} (see Exercise A.1.7). Moreover, $\mu_\phi(\emptyset) = 0$ and $\mu_\phi([0, 1]) < \infty$, because ϕ is continuous and, hence, bounded. With the help of Theorem A.1.13, we may extend μ_ϕ to the whole Borel σ -algebra of $[0, 1]$.

The measure μ_ϕ that we have just constructed has the following special property: if a set $A \subset [0, 1]$ has Lebesgue measure zero then $\mu_\phi(A) = 0$. This property is called *absolute continuity* (with respect to the Lebesgue measure) and is studied in a lot more depth in Appendix A.2.4.

Here is an example of a measure that is positive on any open set but is not absolutely continuous with respect to Lebesgue measure:

Example A.1.24. Fix any enumeration $\{r_1, r_2, \dots\}$ of the set \mathbb{Q} of rational numbers. Consider the measure μ defined on \mathbb{R} by

$$\mu(A) = \sum_{r_i \in A} \frac{1}{2^i}.$$

On the one hand, the measure of any non-empty open subset of the real line is positive, for such a subset must contain some r_i . On the other hand, the measure of \mathbb{Q} is

$$\mu(\mathbb{Q}) = \sum_{r_i \in \mathbb{Q}} \frac{1}{2^i} = 1.$$

Since \mathbb{Q} has Lebesgue measure zero (because it is a countable set), this implies that μ is not absolutely continuous with respect to the Lebesgue measure.

This example also motivates the concept of the support of a measure on a topological space (X, τ) , which we introduce next. For that, we must recall a few basic ideas from topology.

A subset τ' of the topology τ is a *basis of the topology*, or a *basis of open sets*, if for every $x \in X$ and every open set U containing x there exists $U' \in \tau'$ such that $x \in U' \subset U$. We say that the topological space admits a *countable basis of open sets* if such a subset τ' may be chosen to be countable. A set $V \subset X$ is a *neighborhood* of a point $x \in X$ if there exists some open set U such that $x \in U \subset V$. Thus, a subset X is open if and only if it is a neighborhood of each one of its points. A family v' of subsets of X is a *basis of neighborhoods* of a point $x \in X$ if for every neighborhood V there exists some $V' \in v'$ such that $x \in V' \subset V$. We say that x admits a *countable basis of neighborhoods* if v' may be chosen to be countable. If the topological space admits a countable basis of open sets then every $x \in X$ admits a countable basis of neighborhoods, namely, the family of elements of the countable basis of open sets that contain x .

Definition A.1.25. Let (X, τ) be a topological space and μ be a measure on the Borel σ -algebra of X . The *support* of the measure μ is the set $\text{supp } \mu$ formed by the points $x \in X$ such that $\mu(V) > 0$ for any neighborhood V of x .

It follows immediately from the definition that the support of a measure is a closed set. In Example A.1.24 above, the support of μ is the whole real line, despite the fact that $\mu(\mathbb{Q}) = 1$.

Proposition A.1.26. *If X is a topological space with a countable basis of open sets and μ is a non-zero measure on X , then the support $\text{supp } \mu$ is non-empty.*

Proof. If $\text{supp } \mu$ is empty then for each point $x \in X$ we may find an open neighborhood V_x such that $\mu(V_x) = 0$. Let $\{A_j : j = 1, 2, \dots\}$ be a countable

basis of the topology of X . Then, for each $x \in X$ we may choose $i(x) \in \mathbb{N}$ such that $x \in A_{i(x)} \subset V_x$. Hence,

$$X = \bigcup_{x \in X} V_x = \bigcup_{x \in X} A_{i(x)}$$

and so

$$\mu(X) = \mu\left(\bigcup_{x \in X} A_{i(x)}\right) \leq \sum_{i=1}^{\infty} \mu(A_i) = 0.$$

This is a contradiction, and so $\text{supp } \mu$ cannot be empty.

A.1.4 Measurable maps

Measurable maps play a role in measure theory similar to the role of continuous maps in topology: measurability corresponds to the idea that the map preserves the family of measurable subsets, just as continuity means that the family of open subsets is preserved by the map.

Definition A.1.27. Given measurable spaces (X, \mathcal{B}) and (Y, \mathcal{C}) , we say that a map $f : X \rightarrow Y$ is *measurable* if $f^{-1}(C) \in \mathcal{B}$ for every $C \in \mathcal{C}$.

In general, the family of sets $C \in \mathcal{C}$ such that $f^{-1}(C) \in \mathcal{B}$ is a σ -algebra. So, to prove that f is measurable it suffices to show that $f^{-1}(C_0) \in \mathcal{B}$ for every set C_0 in some family $\mathcal{C}_0 \subset \mathcal{C}$ that generates the σ -algebra \mathcal{C} . See also Exercise A.1.1.

Example A.1.28. A function $f : X \rightarrow [-\infty, \infty]$ is measurable if and only if the set $f^{-1}((c, +\infty))$ belongs to \mathcal{B} for every $c \in \mathbb{R}$. This follows from the previous observation, since the family of intervals $(c, +\infty]$ generates the Borel σ -algebra of the extended line (recall Example A.1.9). In particular, if a function f takes values in $(-\infty, +\infty)$ then it is measurable if and only if $f^{-1}((c, +\infty))$ belongs to \mathcal{B} for every $c \in \mathbb{R}$.

Example A.1.29. If X is a topological space and \mathcal{B} is the corresponding Borel σ -algebra, then every continuous function $f : X \rightarrow \mathbb{R}$ is measurable. Indeed, continuity means that the pre-image of every open subset of \mathbb{R} is an open subset of X and, hence, is in \mathcal{B} . Since the family of open sets generates the Borel σ -algebra of \mathbb{R} , it follows that the pre-image of every Borel subset of the real line is also in \mathcal{B} .

Example A.1.30. The *characteristic function* $\mathcal{X}_B : X \rightarrow \mathbb{R}$ of a set $B \subset X$ is defined by:

$$\mathcal{X}_B(x) = \begin{cases} 1, & \text{if } x \in B; \\ 0, & \text{otherwise.} \end{cases}$$

Observe that the function \mathcal{X}_B is measurable if and only if B is a measurable subset: indeed, $\mathcal{X}_B^{-1}(A) \in \{\emptyset, B, X \setminus B, X\}$ for any $A \subset \mathbb{R}$.

Among the basic properties of measurable functions, let us highlight:

Proposition A.1.31. *Let $f, g : X \rightarrow [-\infty, +\infty]$ be measurable functions and let $a, b \in \mathbb{R}$. Then the following functions are also measurable:*

$$(af + bg)(x) = af(x) + bg(x) \quad \text{and} \quad (f \cdot g)(x) = f(x) \cdot g(x).$$

Moreover, if $f_n : X \rightarrow [-\infty, +\infty]$ is a sequence of measurable functions, then the following functions are also measurable:

$$s(x) = \sup\{f_n(x) : n \geq 1\} \quad \text{and} \quad i(x) = \inf\{f_n(x) : n \geq 1\},$$

$$f^*(x) = \limsup_n f_n(x) \quad \text{and} \quad f_*(x) = \liminf_n f_n(x).$$

In particular, if $f(x) = \lim f_n(x)$ exists then f is measurable.

The linear combinations of characteristic functions form an important class of measurable functions:

Definition A.1.32. We say that a function $s : X \rightarrow \mathbb{R}$ is *simple* if there exist constants $\alpha_1, \dots, \alpha_k \in \mathbb{R}$ and pairwise disjoint measurable sets $A_1, \dots, A_k \in \mathcal{B}$ such that

$$s = \sum_{j=1}^k \alpha_j \mathcal{X}_{A_j}, \tag{A.1.2}$$

where \mathcal{X}_A is the characteristic function of the set A .

Note that every simple function is measurable. In the converse direction, the result that follows asserts that every measurable function is the limit of a sequence of simple functions. This fact will be very useful in the next appendix, when defining the Lebesgue integral.

Proposition A.1.33. *Let $f : X \rightarrow [-\infty, +\infty]$ be a measurable function. Then there exists a sequence $(s_n)_n$ of simple functions such that $|s_n(x)| \leq |f(x)|$ for every n and*

$$\lim_n s_n(x) = f(x) \text{ for every } x \in X.$$

If f takes values in \mathbb{R} , we may take every s_n with values in \mathbb{R} . If f is bounded, the sequence $(s_n)_n$ may be chosen such that the convergence is uniform. If f is non-negative, we may take $0 \leq s_1 \leq s_2 \leq \dots \leq f$.

In Exercise A.1.16 the reader is invited to prove this proposition.

A.1.5 Exercises

- A.1.1. Let X be a set and (Y, \mathcal{C}) be a measurable space. Show that, for any transformation $f : X \rightarrow Y$ there exists some σ -algebra \mathcal{B} of subsets of X such that the transformation is measurable with respect to the σ -algebras \mathcal{B} and \mathcal{C} .
- A.1.2. Let X be a set and consider the family of subsets

$$\mathcal{B}_0 = \{A \subset X : A \text{ is finite or } A^c \text{ is finite}\}.$$

Show that \mathcal{B}_0 is an algebra. Moreover, \mathcal{B}_0 is a σ -algebra if and only if the set X is finite. Show also that, in general,

$$\mathcal{B}_1 = \{A \subset X : A \text{ is finite or countable or } A^c \text{ is finite or countable}\}$$

is the σ -algebra generated by the algebra \mathcal{B}_0 .

A.1.3. Prove Proposition A.1.5.

A.1.4. The purpose of this exercise is to exhibit a *non-Borel subset* of the real line. Let α be any irrational number. Consider the following relation on $\mathbb{R} : x \sim y \Leftrightarrow$ there are $m, n \in \mathbb{Z}$ such that $x - y = m + n\alpha$. Check that \sim is an equivalence relation and every equivalence class intersects $[0, 1)$. Let E_0 be a subset of $[0, 1)$ containing exactly one element of each equivalence class (the existence of such a set is a consequence of the Axiom of Choice). Show that E_0 is not a Borel set.

A.1.5. Let (X, \mathcal{B}, μ) be a measure space. Show that if A_1, A_2, \dots are in \mathcal{B} then

$$\mu\left(\bigcup_{j=1}^{\infty} A_j\right) \leq \sum_{j=1}^{\infty} \mu(A_j).$$

A.1.6. (Lemma of Borel–Cantelli). Let $(E_n)_n$ be a countable family of measurable sets. Let F be the set of points that belong to E_n for infinitely many values of n , that is, $F = \limsup_n E_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} E_n$. Show that if $\sum_n \mu(E_n) < \infty$ then $\mu(F) = 0$.

A.1.7. Prove Theorem A.1.14.

A.1.8. Let \mathcal{A} be the collection of subsets of $X = [0, 1]$ that may be written as finite unions of pairwise disjoint intervals. Check that \mathcal{A} is an algebra of subsets of X . Let $m_0 : \mathcal{A} \rightarrow [0, 1]$ be the function defined on this algebra by

$$m_0(I_1 \cup \dots \cup I_N) = |I_1| + \dots + |I_N|,$$

where $|I_j|$ represents the length of I_j . Show that m_0 is σ -additive.

A.1.9. Let \mathcal{B} be an algebra of subsets of X and $\mu : \mathcal{B} \rightarrow [0, +\infty)$ be a finitely additive function with $\mu(X) < \infty$. Show that μ is σ -additive if and only if any one of the following conditions holds:

- (a) $\lim_n \mu(A_n) = \mu(\bigcap_{j=1}^{\infty} A_j)$ for any decreasing sequence $A_1 \supset \dots \supset A_j \supset \dots$ of elements of \mathcal{B} ;
- (b) $\lim_n \mu(A_n) = \mu(\bigcup_{j=1}^{\infty} A_j)$ for any increasing sequence $A_1 \subset \dots \subset A_j \subset \dots$ of elements of \mathcal{B} .

A.1.10. Show that $\|\mu\| = |\mu|(X)$ defines a complete norm in the vector space of finite signed measures on a measurable space (X, \mathcal{B}) .

A.1.11. Let $X = \{1, \dots, d\}$ be a finite set, endowed with the discrete topology, and let $M = X^{\mathcal{I}}$ with $\mathcal{I} = \mathbb{N}$ or $\mathcal{I} = \mathbb{Z}$.

- (a) Check that (A.2.7) defines a distance on M and that the topology defined by this distance coincides with the product topology on M . Describe the open balls and the closed balls around any point $x \in X^{\mathcal{I}}$.
- (b) Without using the theorem of Tychonoff, show that (M, d) is a compact space.
- (c) Let \mathcal{A} be the algebra generated by the elementary cylinders of M . Show that every additive function $\mu : \mathcal{A} \rightarrow [0, 1]$ with $\mu(M) = 1$ extends to a probability measure on the Borel σ -algebra of M .

- A.1.12. Let $K \subset [0, 1]$ be the *Cantor set*, that is, $K = \bigcap_{n=0}^{\infty} K_n$ where $K_0 = [0, 1]$ and each K_n is the set obtained by removing from each connected component C of K_{n-1} the open interval whose center coincides with the center of C and whose length is one third of the length of C . Show that K has Lebesgue measure equal to zero.
- A.1.13. Given a set $E \subset \mathbb{R}^d$, prove that the following conditions are equivalent:
- E is a Lebesgue measurable set.
 - E belongs to the completion of the Borel σ -algebra relative to the Lebesgue measure, that is, there exist Borel sets $B_1, B_2 \subset \mathbb{R}^d$ such that $B_1 \subset E \subset B_2$ and $m(B_2 \setminus B_1) = 0$.
 - (Approximation from above by open sets) Given $\varepsilon > 0$ we can find an open set A such that $E \subset A$ and $m^*(A \setminus E) < \varepsilon$.
 - (Approximation from below by closed sets) Given $\varepsilon > 0$ we can find a closed set F such that $F \subset E$ and $m^*(E \setminus F) < \varepsilon$.
- A.1.14. Prove Proposition A.1.31.
- A.1.15. Let $g_n : M \rightarrow \mathbb{R}$, $n \geq 1$ be a sequence of measurable functions such that $f(x) = \sum_{n=1}^{\infty} g_n(x)$ converges at every point. Show that the sum f is a measurable function.
- A.1.16. Prove Proposition A.1.33.
- A.1.17. Let $f : X \rightarrow X$ be a measurable transformation and ν be a measure on X . Define $(f_*\nu)(A) = \nu(f^{-1}(A))$. Show that $f_*\nu$ is a measure and note that it is finite if and only if ν itself is finite.
- A.1.18. Let $\omega_5 : [0, 1] \rightarrow [0, 1]$ be the function assigning to each $x \in [0, 1]$ the *upper frequency of the digit 5* in the decimal expansion of x . In other words, writing $x = 0.a_0a_1a_2 \dots$ with $a_i \neq 9$ for infinitely many values of i ,

$$\omega_5(x) = \limsup_n \frac{1}{n} \#\{0 \leq j \leq n-1 : a_j = 5\}.$$

Prove that the function ω_5 is measurable.

A.2 Integration in measure spaces

In this appendix we define the Lebesgue integral of a measurable function with respect to a measure. This generalizes the notion of Riemann integral that is usually presented in calculus or introductory analysis courses to a much broader class of functions. Indeed, the Riemann integral is not defined for many useful functions, for example the characteristic functions of arbitrary measurable sets (see Example A.2.5 below). In contrast, the Lebesgue integral makes sense for the whole class of measurable functions, which, as we have seen in Proposition A.1.31, is closed under all the main operations in analysis.

Also in this appendix, we state some important results about the behavior of the (Lebesgue) integral under limits of sequences. Moreover, we describe the product of any finite family of finite measures; for probability measures we

even extend this construction to countable families. Near the end, we discuss the related notions of absolute continuity and Lebesgue derivation.

A.2.1 Lebesgue integral

Throughout this section, we always take (X, \mathcal{B}, μ) to be a measure space. We are going to introduce the notion of Lebesgue integral in a certain number of steps. The first one deals with the integral of a simple function:

Definition A.2.1. Let $s = \sum_{j=1}^k \alpha_j \chi_{A_j}$ be a simple function. The *integral* of s is given by:

$$\int s \, d\mu = \sum_{j=1}^k \alpha_j \mu(A_j).$$

It is easy to check (Exercise A.2.1) that this definition is consistent: if two different linear combinations of characteristic functions define the same function then the values of the integrals obtained from those two linear combinations are equal.

The next step is to define the integral of a non-negative measurable function. The idea is to approximate the function by a monotone sequence of simple functions, using Proposition A.1.33:

Definition A.2.2. Let $f : X \rightarrow [0, \infty]$ be a non-negative measurable function. Then

$$\int f \, d\mu = \lim_n \int s_n \, d\mu,$$

where $s_1 \leq s_2 \leq \dots$ is a non-decreasing sequence of simple functions such that $\lim_n s_n(x) = f(x)$ for every $x \in X$.

It is not difficult to check (Exercise A.2.2) that this definition is consistent: the value of the integral does not depend on the choice of the sequence $(s_n)_n$.

Next, to extend the definition of integral to an arbitrary measurable function, let us observe that given any function $f : X \rightarrow [-\infty, +\infty]$ we can always write $f = f^+ - f^-$ with

$$f^+(x) = \max\{f(x), 0\} \quad \text{and} \quad f^-(x) = \max\{-f(x), 0\}.$$

It is clear that the functions f^+ and f^- are non-negative. Moreover, by Proposition A.1.31, they are measurable whenever f is measurable.

Definition A.2.3. Let $f : X \rightarrow [-\infty, +\infty]$ be a measurable function. Then

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu,$$

as long as at least one of the integrals on the right-hand side is finite (with the usual conventions that $(+\infty) - a = +\infty$ and $a - (+\infty) = -\infty$ for every $a \in \mathbb{R}$).

Definition A.2.4. A function $f : X \rightarrow [-\infty, +\infty]$ is *integrable* if it is measurable and its integral is a real number. We denote the set of all integrable functions as $\mathcal{L}^1(X, \mathcal{B}, \mu)$ or, simply, as $\mathcal{L}^1(\mu)$.

Given a measurable function $f : X \rightarrow [-\infty, \infty]$ and a measurable set E , we define the *integral of f over E* to be

$$\int_E f d\mu = \int f \chi_E d\mu,$$

where χ_E is the characteristic function of the set E .

Example A.2.5. Consider $X = [0, 1]$ endowed with the Lebesgue measure m . Let $f = \chi_B$, where B is the subset of rational numbers. Then $m(B) = 0$ and so, using Definition A.2.2, the Lebesgue integral of f is equal to zero. On the other hand, a direct calculation shows that every lower Riemann sum of f is equal to 0, while every upper Riemann sum of f is equal to 1. So, the Riemann integral of f does not exist. Indeed, more generally, the Riemann integral of the characteristic function of a measurable set exists if and only if the boundary of the set has zero Lebesgue measure. Note that in the present case the boundary is the whole of $[0, 1]$, which has positive Lebesgue measure.

Example A.2.6. Let $x_1, \dots, x_m \in X$ and $p_1, \dots, p_m > 0$ with $p_1 + \dots + p_m = 1$. Let μ be the probability measure μ defined on 2^X by

$$\mu = \sum_{i=1}^m p_i \delta_{x_i} \quad \text{where } \delta_{x_i} \text{ is the Dirac mass at } x_i.$$

In other words, $\mu(A) = \sum_{x_i \in A} p_i$ for every subset A of X . Then, for any function $f : X \rightarrow [-\infty, +\infty]$,

$$\int f d\mu = \sum_{i=1}^m p_i f(x_i).$$

Proposition A.2.7. The set $\mathcal{L}^1(\mu)$ of all real integrable functions is a real vector space. Moreover, the map $I : \mathcal{L}^1(\mu) \rightarrow \mathbb{R}$ given by $I(f) = \int f d\mu$ is a positive linear functional:

- (1) $\int af + bg d\mu = a \int f d\mu + b \int g d\mu$, and
- (2) $\int f d\mu \geq \int g d\mu$ if $f(x) \geq g(x)$ for every x .

In particular, $|\int f d\mu| \leq \int |f| d\mu$ if $|f| \in \mathcal{L}^1(\mu)$. Moreover, $|f| \in \mathcal{L}^1(\mu)$ if and only if $f \in \mathcal{L}^1(\mu)$.

The notion of the Lebesgue integral may be extended to an even broader class of functions, in two different ways. On the one hand, we may consider complex functions $f : X \rightarrow \mathbb{C}$. In this case, we say that f is integrable if and only if the real part $\Re f$ and the imaginary part $\Im f$ are both integrable. Then, by definition,

$$\int f d\mu = \int \Re f d\mu + i \int \Im f d\mu.$$

On the other hand, we may consider functions that are not necessarily measurable but coincide with some measurable function on a subset of the domain with total measure. To explain this, we need the following notion, which is used frequently throughout the text:

Definition A.2.8. We say that a property holds *at μ -almost every point* (or *μ -almost everywhere*) if the subset of points of X for which it does not hold is contained in some zero measure set.

For example, we say that a sequence of functions $(f_n)_n$ converges to some function at μ -almost every point if there exists some measurable set $N \subset X$ with $\mu(N) = 0$ such that $f(x) = \lim_n f_n(x)$ for every $x \in X \setminus N$. Analogously, we say that two functions f and g are equal at μ -almost every point if there exists a measurable set $N \subset X$ with $\mu(N) = 0$ such that $f(x) = g(x)$ for every $x \in X \setminus N$. Clearly, this is an equivalence relation in the space of functions defined on X . Moreover, assuming that the two functions are integrable, it implies that the two integrals coincide:

$$\int f d\mu = \int g d\mu \quad \text{if } f = g \text{ at } \mu\text{-almost every point.}$$

This observation permits the definition of the integral for any function f , possibly non-measurable, that coincides at μ -almost every point with some measurable function g : it suffices to take $\int f d\mu = \int g d\mu$.

To close this section, let us observe that the notion of integral may also be extended to signed measures and even complex measures, as follows. Let μ be a signed measure and $\mu = \mu^+ - \mu^-$ be its Hahn decomposition. We say that a function ϕ is integrable with respect to μ if it is integrable with respect to both μ^+ and μ^- . Then we define:

$$\int \phi d\mu = \int \phi d\mu^+ - \int \phi d\mu^-.$$

Similarly, let μ be a complex measure. By definition, a function ϕ is integrable with respect to μ if it is integrable with respect to both the real part $\Re\mu$ and the imaginary part $\Im\mu$. Then we define:

$$\int \phi d\mu = \int \phi d\Re\mu - \int \phi d\Im\mu.$$

A.2.2 Convergence theorems

Next, we mention three important results concerning the convergence of functions under the integral sign. The first one deals with monotone sequences of functions:

Theorem A.2.9 (Monotone convergence). *Let $f_n : X \rightarrow [-\infty, +\infty]$ be a non-decreasing sequence of non-negative measurable functions. Consider the*

function $f : X \rightarrow [-\infty, +\infty]$ defined by $f(x) = \lim_n f_n(x)$. Then

$$\lim_n \int f_n d\mu = \int f(x) d\mu.$$

The next result applies to much more general sequences, not necessarily monotone:

Theorem A.2.10 (Lemma of Fatou). *Let $f_n : X \rightarrow [0, +\infty]$ be a sequence of non-negative measurable functions. Then the function $f : X \rightarrow [-\infty, +\infty]$ defined by $f(x) = \liminf_n f_n(x)$ is integrable and satisfies*

$$\int \liminf_n f_n(x) d\mu \leq \liminf_n \int f_n d\mu.$$

The most powerful of the results in this section is the dominated convergence theorem, which asserts that we may take the limit under the integral sign whenever the sequence of functions is bounded by some integrable function:

Theorem A.2.11 (Dominated convergence). *Let $f_n : X \rightarrow \mathbb{R}$ be a sequence of measurable functions and assume that there exists some integrable function $g : X \rightarrow \mathbb{R}$ such that $|f_n(x)| \leq |g(x)|$ for μ -almost every x in X . Assume moreover that the sequence $(f_n)_n$ converges at μ -almost every point to some function $f : X \rightarrow \mathbb{R}$. Then f is integrable and satisfies*

$$\lim_n \int f_n d\mu = \int f d\mu.$$

In Exercise A.2.7 we invite the reader to deduce the dominated convergence theorem from the Lemma of Fatou.

A.2.3 Product measures

Let $(X_j, \mathcal{A}_j, \mu_j), j = 1, \dots, n$ be finite measure spaces, that is, such that $\mu_j(X_j) < \infty$ for every j . One can endow the Cartesian product $X_1 \times \dots \times X_n$ with the structure of a finite measure space in the following way. Consider on $X_1 \times \dots \times X_n$ the σ -algebra generated by the family of all subsets of the form $A_1 \times \dots \times A_n$ with $A_j \in \mathcal{A}_j$. This is called the *product σ -algebra* and is denoted by $\mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n$.

Theorem A.2.12. *There exists a unique measure μ on the measurable space $(X_1 \times \dots \times X_n, \mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n)$ such that $\mu(A_1 \times \dots \times A_n) = \mu_1(A_1) \dots \mu_n(A_n)$ for every $A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n$. In particular, μ is a finite measure.*

The proof of this result (see Theorem 35.B in Halmos [Hal50]) combines the extension theorem (Theorem A.1.13) with the monotone convergence theorem (Theorem A.2.9). The measure μ in the statement is the *product* of the measures μ_1, \dots, μ_n and is denoted by $\mu_1 \times \dots \times \mu_n$. In this way one defines the *product measure space*

$$(X_1 \times \dots \times X_n, \mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n, \mu_1 \times \dots \times \mu_n).$$

Theorem A.2.12 remains valid when the measures μ_j are just σ -finite, except that in this case the product measure μ is also only σ -finite.

Next, we describe the product of a *countable* family of measure spaces. Actually, for now we restrict ourselves to the case of probability spaces. Let $(X_j, \mathcal{B}_j, \mu_j), j \in \mathcal{I}$ be probability measure spaces with $\mu_j(X_j) = 1$ for every $j \in \mathcal{I}$. What follows holds for both $\mathcal{I} = \mathbb{N}$ and $\mathcal{I} = \mathbb{Z}$. Consider the Cartesian product

$$\Sigma = \prod_{j \in \mathcal{I}} X_j = \{(x_j)_{j \in \mathcal{I}} : x_j \in X_j\}. \tag{A.2.1}$$

We call *cylinders* of Σ all subsets of the form

$$[m; A_m, \dots, A_n] = \{(x_j)_{j \in \mathcal{I}} : x_j \in A_j \text{ for } m \leq j \leq n\}, \tag{A.2.2}$$

where $m \in \mathcal{I}$ and $n \geq m$ and $A_j \in \mathcal{B}_j$ for each $m \leq j \leq n$. Note that X itself is a cylinder: we may write $X = [1; X_1]$, for example. By definition, the *product σ -algebra* on Σ is the σ -algebra \mathcal{B} generated by the family of all cylinders. The family \mathcal{A} of all finite unions of pairwise disjoint cylinders is an algebra and it generates the product σ -algebra \mathcal{B} .

Theorem A.2.13. *There exists a unique measure μ on (Σ, \mathcal{B}) such that*

$$\mu([m; A_m, \dots, A_n]) = \mu_m(A_m) \cdots \mu_n(A_n) \tag{A.2.3}$$

for every cylinder $[m; A_m, \dots, A_n]$. In particular, μ is a probability measure.

The proof of this theorem (see Theorem 38.B in Halmos [Hal50]) uses the extension theorem (Theorem A.1.13) together with the theorem of continuity at the empty set (Theorem A.1.14). The probability measure μ is called the *product* of the measures μ_j and is denoted as $\prod_{j \in \mathcal{I}} \mu_j$. The probability space $(\Sigma, \mathcal{B}, \mu)$ is called the *product* of the spaces $(X_j, \mathcal{B}_j, \mu_j), j \in \mathcal{I}$.

An important special case is when the spaces $(X_i, \mathcal{B}_i, \mu_i)$ are all equal to a given (X, \mathcal{C}, ν) . The corresponding product space may be used to model a sequence of identical random experiments such that the outcome of each experiment is independent of all the others. To explain this, take X to be the set of possible outcomes of each experiment and let ν be the probability distribution of those outcomes. In this context, the measure $\mu = \nu^{\mathcal{I}} = \prod_{j \in \mathcal{I}} \nu$ is usually called the *Bernoulli measure* defined by ν . Property (A.2.3) corresponds to the identity

$$\mu([m; A_m, \dots, A_n]) = \prod_{j=m}^n \nu(A_j), \tag{A.2.4}$$

which may be read in the following way: the probability of any composite event $\{x_m \in A_m, \dots, x_n \in A_n\}$ is equal to the product of the probabilities of the individual events $x_i \in A_i$. So, (A.2.4) does reflect the assumption that the successive experiments are mutually independent.

We have a special interest in the case when X is a finite set, endowed with the σ -algebra $\mathcal{C} = 2^X$ of all its subsets. In this case, it is useful to consider the *elementary cylinders*

$$[m; a_m, \dots, a_n] = \{(x_j)_{j \in \mathcal{I}} \in X : x_m = a_m, \dots, x_n = a_n\}, \tag{A.2.5}$$

corresponding to subsets A_j consisting of a single point a_j . Observe that every cylinder is a finite union of pairwise disjoint elementary cylinders. In particular, the σ -algebra generated by the elementary cylinders coincides with the σ -algebra generated by all the cylinders, and the same is true for the generated algebra. Moreover, the relation (A.2.4) may be written as

$$\mu([m; a_m, \dots, a_n]) = p_{a_m} \cdots p_{a_n} \quad \text{where } p_a = \nu(\{a\}) \text{ for } a \in X. \tag{A.2.6}$$

Consider the finite set X endowed with the discrete topology. The product topology on $\Sigma = X^{\mathcal{I}}$ coincides with the topology generated by the elementary cylinders. Moreover (see Exercise A.1.11), it coincides with the topology associated with the distance defined by

$$d((x_i)_{i \in \mathcal{I}}, (y_i)_{i \in \mathcal{I}}) = \theta^N, \tag{A.2.7}$$

where $\theta \in (0, 1)$ is fixed and $N = N((x_i)_{i \in \mathcal{I}}, (y_i)_{i \in \mathcal{I}}) \geq 0$ is the largest integer such that $x_i = y_i$ for every $i \in \mathcal{I}$ with $|i| < N$.

A.2.4 Derivation of measures

Let m be the Lebesgue measure on \mathbb{R}^d . Given a measurable subset A of \mathbb{R}^d , we say that $a \in \mathbb{R}^d$ is a *density point* of A if the subset A occupies most of every small neighborhood of a , in the following sense:

$$\lim_{\delta \rightarrow 0} \frac{m(B(a, \delta) \cap A)}{m(B(a, \delta))} = 1. \tag{A.2.8}$$

Theorem A.2.14. *Let A be a measurable subset of \mathbb{R}^d with Lebesgue measure $m(A)$ positive. Then m -almost every $a \in A$ is a density point of A .*

In Exercise A.2.11 we propose a proof of this result. It is also a direct consequence of the theorem that we state next. We say that a function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is *locally integrable* if the product $f \chi_K$ is integrable for every compact set $K \subset \mathbb{R}^d$.

Theorem A.2.15 (Lebesgue derivation). *Let $X = \mathbb{R}^d$ and \mathcal{B} be a Borel σ -algebra and m be the Lebesgue measure on \mathbb{R}^d . Let $f : X \rightarrow \mathbb{R}$ be a locally integrable function. Then*

$$\lim_{r \rightarrow 0} \frac{1}{m(B(x, r))} \int_{B(x, r)} |f(y) - f(x)| dm = 0 \quad \text{at } m\text{-almost every point } x.$$

In particular,

$$\lim_{r \rightarrow 0} \frac{1}{m(B(x, r))} \int_{B(x, r)} f(y) dm = f(x) \quad \text{at } m\text{-almost every point } x.$$

The crucial ingredient in the proof of these results is the following geometric fact:

Theorem A.2.16 (Lemma of Vitali). *Let m be the Lebesgue measure on \mathbb{R}^d and suppose that for every $x \in \mathbb{R}$ one is given a sequence $(B_n(x))_n$ of balls centered at x with radii converging to zero. Let $A \subset \mathbb{R}^d$ be a measurable set with $m(A) > 0$. Then, for every $\varepsilon > 0$ there exist sequences $(x_j)_j$ in \mathbb{R} and $(n_j)_j$ in \mathbb{N} such that*

1. *the balls $B_{n_j}(x_j)$ are pairwise disjoint;*
2. *$m\left(\bigcup_j B_{n_j}(x_j) \setminus A\right) < \varepsilon$ and $m\left(A \setminus \bigcup_j B_{n_j}(x_j)\right) = 0$.*

This theorem remains valid if, instead of balls, we take for $(B_n(x))_n$ any sequence of sets such that $\bigcap_n B_n(x) = \{x\}$ and

$$\sup_{x, n} \frac{\sup\{d(x, y) : y \in B_n(x)\}}{\inf\{d(x, z) : z \notin B_n(x)\}} < \infty.$$

The set of measures defined on the same measurable space possesses a natural partial order relation:

Definition A.2.17. Let μ and ν be two measures in the same measurable space (X, \mathcal{B}) . We say that ν is *absolutely continuous* with respect to μ if every measurable set E that satisfies $\mu(E) = 0$ also satisfies $\nu(E) = 0$; then we write $\nu \ll \mu$. We say that μ and ν are *equivalent* if each one of them is absolutely continuous with respect to the other; then we write $\mu \sim \nu$. In other words, two measures are equivalent if they have exactly the same zero measure sets.

Another very important result, known as the theorem of Radon–Nikodym, asserts that if $\nu \ll \mu$ then the measure ν may be seen as the product of μ by some measurable function ρ :

Theorem A.2.18 (Radon–Nikodym). *If μ and ν are finite measures such that $\nu \ll \mu$ then there exists a measurable function $\rho : X \rightarrow [0, +\infty]$ such that $\nu = \rho \mu$, meaning that*

$$\int \phi d\nu = \int \phi \rho d\mu \quad \text{for any bounded measurable function } \phi : X \rightarrow \mathbb{R}. \tag{A.2.9}$$

In particular, $\nu(E) = \int_E \rho d\mu$ for every measurable set $E \subset X$. Moreover, ρ is essentially unique: any two functions satisfying (A.2.9) coincide at μ -almost every point.

We call ρ the *density*, or *Radon–Nikodym derivative*, of ν relative to μ and we write

$$\rho = \frac{d\nu}{d\mu}.$$

Definition A.2.19. Let μ and ν be two measures in the same measurable space (X, \mathcal{B}) . We say that μ and ν are *mutually singular* if there exist disjoint measurable subsets A and B such that $A \cup B = X$ and $\mu(A) = 0$ and $\nu(B) = 0$. Then we write $\mu \perp \nu$.

The Lebesgue decomposition theorem states that, given any two finite measures μ and ν in the same measurable space, we may write $\nu = \nu_a + \nu_s$ where ν_a and ν_s are finite measures such that $\nu_a \ll \mu$ and $\nu_s \perp \mu$. Combining this with the theorem of Radon–Nikodym, we get:

Theorem A.2.20 (Lebesgue decomposition). *Given any finite measures μ and ν , there exist a measurable function $\rho : X \rightarrow [0, +\infty]$ and a finite measure η such that $\nu = \rho\mu + \eta$ and $\eta \perp \mu$.*

A.2.5 Exercises

- A.2.1. Prove that the integral of a simple function is well defined: if two linear combinations of characteristic functions define the same function, then the values of the integrals obtained from the two combinations coincide.
- A.2.2. Show that if $(r_n)_n$ and $(s_n)_n$ are non-decreasing sequences of non-negative functions converging at μ -almost every point to the same function $f : M \rightarrow [0, +\infty)$, then $\lim_n \int r_n d\mu = \lim_n \int s_n d\mu$.
- A.2.3. Prove Proposition A.2.7.
- A.2.4. (Tchebysheff–Markov inequality) Let $f : M \rightarrow \mathbb{R}$ be a non-negative function integrable with respect to a finite measure μ . Then, given any real number $a > 0$,

$$\mu(\{x \in M : f(x) \geq a\}) \leq \frac{1}{a} \int_X f d\mu.$$

In particular, if $\int |f| d\mu = 0$ then $\mu(\{x \in X : f(x) \neq 0\}) = 0$.

- A.2.5. Let f be an integrable function. Show that for every $\varepsilon > 0$ there exists $\delta > 0$ such that $|\int_E f d\mu| < \varepsilon$ for every measurable set E with $\mu(E) < \delta$.
- A.2.6. Let $\psi_1, \dots, \psi_N : M \rightarrow \mathbb{R}$ be bounded measurable functions defined on a probability space (M, \mathcal{B}, μ) . Show that for any $\varepsilon > 0$ there exist $x_1, \dots, x_s \in M$ and positive numbers $\alpha_1, \dots, \alpha_s$ such that $\sum_{j=1}^s \alpha_j = 1$ and

$$\left| \int \psi_i d\mu - \sum_{j=1}^s \alpha_j \psi_i(x_j) \right| < \varepsilon \quad \text{for every } i = 1, \dots, N.$$

- A.2.7. Deduce the dominated convergence theorem (Theorem A.2.11) from the Lemma of Fatou (Theorem A.2.10).

- A.2.8. A set \mathcal{F} of measurable functions $f : M \rightarrow \mathbb{R}$ is said to be *uniformly integrable* with respect to a probability measure μ if for every $\alpha > 0$ there exists $C > 0$ such that $\int_{\{|f|>C\}} |f| d\mu < \alpha$ for every $f \in \mathcal{F}$. Show that
- (a) \mathcal{F} is uniformly integrable with respect to μ if and only if there exists $L > 0$ and for every $\varepsilon > 0$ there exists $\delta > 0$ such that $\int |f| d\mu < L$ and $\int_A |f| d\mu < \varepsilon$ for every $f \in \mathcal{F}$ and every measurable set A with $\mu(A) < \delta$.
 - (b) If there exists a function $g : M \rightarrow \mathbb{R}$ integrable with respect to μ such that $|f| \leq |g|$ for every $f \in \mathcal{F}$ (we say that \mathcal{F} is *dominated* by g) then the set \mathcal{F} is uniformly integrable with respect to μ .
 - (c) If the set \mathcal{F} is uniformly integrable with respect to μ then $\lim_n \int f_n d\mu = \int \lim_n f_n d\mu$ for any sequence $(f_n)_n$ in \mathcal{F} such that $\lim_n f_n$ exists at μ -almost every point.
- A.2.9. Show that a is a density point of a set $A \subset \mathbb{R}^d$ if and only if

$$\liminf_{\delta \rightarrow 0} \left\{ \frac{m(B \cap A)}{m(B)} : B \text{ a ball with } a \in B \subset B(a, \delta) \right\} = 1. \tag{A.2.10}$$

- A.2.10. Let $\mathcal{P}_n, n \geq 1$ be a sequence of countable partitions of \mathbb{R}^d into measurable subsets. Assume that the diameter $\text{diam } \mathcal{P}_n = \sup\{\text{diam } P : P \in \mathcal{P}_n\}$ converges to zero when $n \rightarrow \infty$. Show that, given any measurable set $A \subset \mathbb{R}^d$ with positive Lebesgue measure, it is possible to choose sets $P_n \in \mathcal{P}_n, n \geq 1$ in such a way that $m(A \cap P_n)/m(P_n) \rightarrow 1$ when $n \rightarrow \infty$.
- A.2.11. Prove Theorem A.2.14.
- A.2.12. Consider $x_1, x_2 \in M$ and $p_1, p_2, q_1, q_2 > 0$ with $p_1 + p_2 = q_1 + q_2 = 1$. Let μ and ν be the probability measures given by

$$\mu(A) = \sum_{x_i \in A} p_i, \quad \nu(A) = \sum_{x_i \in A} q_i,$$

that is, $\mu = p_1\delta_{x_1} + p_2\delta_{x_2}$ and $\nu = q_1\delta_{x_1} + q_2\delta_{x_2}$. Check that $\nu \ll \mu$ and $\mu \ll \nu$ and calculate the corresponding Radon–Nikodym derivatives.

- A.2.13. Construct a probability measure μ on $[0, 1]$ absolutely continuous with respect to the Lebesgue measure m and such that there exists a measurable set $K \subset [0, 1]$ with $\mu(K) = 0$ and $m(K) = 1/2$. In particular, m is not absolutely continuous with respect to μ . Could we require that $m(K) = 1$?
- A.2.14. Assume that $f : X \rightarrow X$ is such that there exists a countable cover of M by measurable sets $B_n, n \geq 1$, such that the restriction of f to each B_n is a bijection onto its image, with measurable inverse. Let η be a probability measure on M such that $A \subset B_n$ and $\eta(A) = 0$ implies $\eta(f(A)) = 0$. Show that there exists a function $J_\eta : X \rightarrow [0, +\infty]$ such that

$$\int_{f(B_n)} \psi d\eta = \int_{B_n} (\psi \circ f) J_\eta d\eta$$

for every bounded measurable function $\psi : X \rightarrow \mathbb{R}$ and every n . Moreover, J_η is essentially unique.

- A.2.15. Let $\mu = \mu^+ - \mu^-$ be the Hahn decomposition of a finite signed measure μ . Show that there exist functions ρ^\pm and τ^\pm such that $\mu^+ = \rho^+|\mu| = \tau^+\mu$ and $\mu^- = \rho^-|\mu| = \tau^-\mu$. Which functions are these?

A.2.16. Let $(\mu_n)_n$ and $(\nu_n)_n$ be two sequences of measures such that $\mu = \sum_n \mu_n$ and $\nu = \sum_n \nu_n$ are finite measures. Let $\hat{\mu}_n = \sum_{i=1}^n \mu_i$ and $\hat{\nu}_n = \sum_{i=1}^n \nu_i$. Show that if $\hat{\mu}_n \ll \hat{\nu}_n$ for every n then $\mu \ll \nu$ and

$$\frac{d\mu}{d\nu} = \lim_n \frac{d\hat{\mu}_n}{d\hat{\nu}_n} \quad \text{at } \nu\text{-almost every point.}$$

A.3 Measures in metric spaces

In this appendix, unless stated otherwise, μ is a Borel probability measure on a metric space M , that is, a probability measure defined on the Borel σ -algebra of M . Most of the results extend immediately to finite Borel measures, in fact.

Recall that a *metric space* is a pair (M, d) where M is a set and d is a *distance* in M , that is, a function $d : M \times M \rightarrow \mathbb{R}$ satisfying:

1. $d(x, y) \geq 0$ for any x, y and the equality holds if and only if $x = y$;
2. $d(x, y) = d(y, x)$ for any x, y ;
3. $d(x, y) \leq d(x, z) + d(z, y)$ for any x, y, z .

We denote $B(x, r) = \{y \in M : d(x, y) < r\}$ and call it the *ball* of center $x \in M$ and radius $r > 0$.

Every metric space has a natural structure of a topological space where the family of balls centered at each point is a basis of neighborhoods for that point. Equivalently, a subset of M is open if and only if it contains some ball centered at each one of its points. In the converse direction, one says that a topological space is *metrizable* if its topology can be defined in this way, from some distance function.

A.3.1 Regular measures

A first interesting fact is that any probability measure on a metric space is completely determined by the values it takes on the open subsets (or the closed subsets) of the space.

Definition A.3.1. A (Borel) measure μ on a topological space is *regular* if for every measurable subset B and every $\varepsilon > 0$ there exists a closed set F and an open set A such that $F \subset B \subset A$ and $\mu(A \setminus F) < \varepsilon$.

Proposition A.3.2. Any probability measure on a metric space is regular.

Proof. Let \mathcal{B}_0 be the family of all Borel subsets B for which the condition in the definition holds, that is, such that for every $\varepsilon > 0$ there exist a closed set F and an open set A satisfying $F \subset B \subset A$ and $\mu(A \setminus F) < \varepsilon$. Begin by noting that \mathcal{B}_0 contains all the closed subsets of M . Indeed, let B be any closed set and let B^δ denote the (open) set of points whose distance to B is less than δ .

By Theorem A.1.14, we have that $\mu(B^\delta \setminus B) \rightarrow 0$ when $\delta \rightarrow 0$. Hence, we may take $F = B$ and $A = B^\delta$ for $\delta > 0$ sufficiently small.

It is immediate that the family \mathcal{B}_0 is closed under taking the complement, that is, $B^c \in \mathcal{B}_0$ whenever $B \in \mathcal{B}_0$. Furthermore, consider any countable family $B_n, n = 1, 2, \dots$ of elements of \mathcal{B}_0 and let $B = \bigcup_{n=1}^\infty B_n$. By hypothesis, for every $n \in \mathbb{N}$ and $\varepsilon > 0$ there exist a closed set F_n and an open set A_n satisfying $F_n \subset B_n \subset A_n$ and $\mu(A_n \setminus F_n) < \varepsilon/2^{n+1}$. The union $A = \bigcup_{n=1}^\infty A_n$ is an open set and any finite union $F = \bigcup_{n=1}^m F_n$ is a closed set. Fix m large enough that

$$\mu\left(\bigcup_{n=1}^\infty F_n \setminus F\right) < \varepsilon/2$$

(recall Theorem A.1.14). Then $F \subset B \subset A$ and

$$\mu(A \setminus F) \leq \sum_{n=1}^\infty \mu(A_n \setminus F_n) + \mu\left(\bigcup_{n=1}^\infty F_n \setminus F\right) < \sum_{n=1}^\infty \frac{\varepsilon}{2^{n+1}} + \frac{\varepsilon}{2} = \varepsilon.$$

This shows that $B \in \mathcal{B}_0$. In this way, we have shown that \mathcal{B}_0 is a σ -algebra. Hence, \mathcal{B}_0 contains all the Borel subsets of M .

It follows that, as stated above, the values that the probability measure μ takes on the closed subsets of M determine μ completely: if ν is another probability measure such that $\mu(F) = \nu(F)$ for every closed set F then, taking the complement, $\mu(A) = \nu(A)$ for every open set A and, using the theorem, $\mu(B) = \nu(B)$ for every Borel set B . In other words, $\mu = \nu$. The same argument shows that the values of μ on the open sets also determine the measure completely.

The proposition that we state and prove next implies that the values of the integrals of the bounded continuous functions also determine the probability measure completely. Indeed, the same is true for the (smaller) set of bounded Lipschitz functions.

Recall that a map $h : M \rightarrow N$ is *Lipschitz* if there exists some constant $C > 0$ such that $d(h(x), h(y)) \leq Cd(x, y)$ for every $x, y \in M$. When it is necessary to specify the constant, we say that the function h is C -Lipschitz. More generally, we say that h is *Hölder* if there exist $C, \theta > 0$ such that $d(h(x), h(y)) \leq Cd(x, y)^\theta$ for every $x, y \in M$. Then we also say that h is θ -Hölder or even (C, θ) -Hölder.

Proposition A.3.3. *If μ and ν are probability measures on a metric space M with $\int \varphi d\mu = \int \varphi d\nu$ for every bounded Lipschitz function $\varphi : M \rightarrow \mathbb{R}$ then $\mu = \nu$.*

Proof. We are going to use the following simple topological fact:

Lemma A.3.4. *Given any closed subset F of M and any $\delta > 0$, there exists a Lipschitz function $g_\delta : M \rightarrow [0, 1]$ such that $g_\delta(x) = 1$ for every $x \in F$ and $g_\delta(x) = 0$ for every $x \in M$ such that $d(x, F) \geq \delta$.*

Proof. Consider the function $h : \mathbb{R} \rightarrow [0, 1]$ given by $h(s) = 1$ if $s \leq 0$ and $h(s) = 0$ if $s \geq 1$ and $h(s) = 1 - s$ if $0 \leq s \leq 1$. Define

$$g : M \rightarrow [0, 1], \quad g(x) = h\left(\frac{1}{\delta}d(x, F)\right).$$

Note that g is Lipschitz, since it is a composition of Lipschitz functions. The other properties in the lemma follow immediately from the definition.

Now we may finish the proof of Proposition A.3.3. Let F be any closed subset of M and, for every $\delta > 0$, let $g_\delta : M \rightarrow [0, 1]$ be a function as in the lemma above. By assumption,

$$\int g_\delta d\mu = \int g_\delta dv \quad \text{for every } \delta > 0.$$

Moreover, by the dominated convergence theorem (Theorem A.2.11),

$$\lim_{\delta \rightarrow 0} \int g_\delta d\mu = \mu(F) \quad \text{and} \quad \lim_{\delta \rightarrow 0} \int g_\delta dv = v(F).$$

This shows that $\mu(F) = v(F)$ for every closed subset F . As pointed out before, the latter implies $\mu = v$.

As observed in Example A.1.29, continuous maps are automatically measurable relative to the Borel σ -algebra. The result that we prove next asserts that, under a simple condition on the metric space, there is a kind of converse: measurable maps are continuous, restricted to certain subsets with almost full measure.

A subset of a topological space M is *dense* if it intersects every open subset of M . We say that the space M is *separable* if it admits some countable dense subset. In the special case of metric spaces this is equivalent to saying that the topology admits a countable basis of open sets (Exercise A.3.1).

Theorem A.3.5 (Lusin). *Let $\varphi : M \rightarrow N$ be a measurable map with values in some separable metric space N . Given any $\varepsilon > 0$, there exists a closed set $F \subset M$ such that $\mu(M \setminus F) < \varepsilon$ and the restriction of φ to F is continuous.*

Proof. Let $\{x_n : n \in \mathbb{N}\}$ be a countable dense subset of N and, for every $k \geq 1$, let $B_{n,k}$ be the ball of center x_n and radius $1/k$. Fix $\varepsilon > 0$. By Proposition A.3.2, for every (n, k) we may find an open set $A_{n,k} \subset M$ containing $\varphi^{-1}(B_{n,k})$ and satisfying $\mu(A_{n,k} \setminus \varphi^{-1}(B_{n,k})) < \varepsilon/2^{n+k+1}$. Define

$$E = \bigcap_{n,k=1}^{\infty} (\varphi^{-1}(B_{n,k}) \cup A_{n,k}^c).$$

On the one hand,

$$\mu(M \setminus E) \leq \sum_{n,k=1}^{\infty} \mu(A_{n,k} \setminus \varphi^{-1}(B_{n,k})) < \sum_{n,k=1}^{\infty} \frac{\varepsilon}{2^{n+k+1}} = \frac{\varepsilon}{2}.$$

On the other hand, every $\varphi^{-1}(B_{n,k})$ is an open subset of $\varphi^{-1}(B_{n,k}) \cup A_{n,k}^c$, since the complement is the closed set $A_{n,k}^c$. Consequently, $\varphi^{-1}(B_{n,k})$ is open in E for every (n, k) . This shows that the restriction of φ to the set E is continuous. To conclude the proof it suffices to use Proposition A.3.2 once more to find a closed set $F \subset E$ such that $\mu(E \setminus F) < \varepsilon/2$.

A.3.2 Separable complete metric spaces

Next, we discuss another important property of measures on metric spaces that are both separable and complete. Recall that the latter means that every Cauchy sequence converges.

Definition A.3.6. A (Borel) measure μ on a topological space is *tight* if for every $\varepsilon > 0$ there exists a compact subset K such that $\mu(K^c) < \varepsilon$.

Since every closed subset of a compact metric space is also compact, it follows immediately from Proposition A.3.2 that every probability measure on a compact metric space is tight. However, this conclusion is a lot more general:

Proposition A.3.7. Every probability measure on a separable complete metric space is tight.

Proof. Let $\{p_k : k \in \mathbb{N}\}$ be a countable dense subset of M . Then, for every $n \geq 1$, the closed balls $\bar{B}(p_k, 1/n)$, $k \in \mathbb{N}$ form a countable cover of M . Given $\varepsilon > 0$ and $n \geq 1$, fix $k(n) \geq 1$ in such a way that the (closed) set

$$L_n = \bigcup_{k=1}^{k(n)} \bar{B}(p_k, 1/n)$$

satisfies $\mu(L_n) > 1 - \varepsilon/2^n$. Take $K = \bigcap_{n=1}^{\infty} L_n$. Note that K is closed and

$$\mu(K^c) \leq \mu\left(\bigcup_{n=1}^{\infty} L_n^c\right) < \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = \varepsilon.$$

It remains to check that K is compact. For that, it is enough to show that every sequence $(x_i)_i$ in K admits some Cauchy subsequence (since M is complete, this subsequence converges). Such a subsequence may be found as follows. Since $x_i \in L_1$ for every i , there exists $l(1) \leq k(1)$ such that the set of indices

$$\mathcal{I}_1 = \{i \in \mathbb{N} : x_i \in B(p_{l(1)}, 1)\}$$

is infinite. Let $i(1)$ be the smallest element of \mathcal{I}_1 . Next, since $x_i \in L_2$ for every i , there exists $l(2) \leq k(2)$ such that

$$\mathcal{I}_2 = \{i \in \mathcal{I}_1 : x_i \in B(p_{l(2)}, 1/2)\}$$

is infinite. Let $i(2)$ be the smallest element of $\mathcal{I}_2 \setminus \{i(1)\}$. Repeating this procedure, we construct a decreasing sequence \mathcal{I}_n of infinite subsets of \mathbb{N} , and an increasing sequence $i(1) < i(2) < \dots < i(n) < \dots$ of integers such that

$i(n) \in \mathcal{I}_n$ and all the $x_i, i \in \mathcal{I}_n$ are contained in the same closed ball of radius $1/n$. In particular,

$$d(x_{i(a)}, x_{i(b)}) \leq 2/n \quad \text{for every } a, b \geq n.$$

This shows that the subsequence $(x_{i(n)})_n$ is indeed Cauchy.

Corollary A.3.8. *Assume that M is a separable complete metric space and μ is a probability measure on M . For every $\varepsilon > 0$ and every Borel set $B \subset M$ there exists a compact set $L \subset B$ such that $\mu(B \setminus L) < \varepsilon$.*

Proof. By Proposition A.3.2, we may find some closed set $F \subset B$ such that $\mu(B \setminus F) < \varepsilon/2$. By Theorem A.3.5, there exists a compact subset $K \subset M$ such that $\mu(M \setminus K) < \varepsilon/2$. Take $L = F \cap K$. Then L is compact and $\mu(B \setminus L) < \varepsilon$.

Analogously, when the metric space M is separable and complete we can improve the statement of Lusin's theorem, replacing "closed" with "compact" in the conclusion:

Theorem A.3.9 (Lusin). *Assume that M is a separable complete metric space and μ is a probability measure on M . Let $\varphi : M \rightarrow N$ be a measurable map with values in a separable metric space N . Then, given any $\varepsilon > 0$ there exists a compact set $K \subset M$ such that $\mu(M \setminus K) < \varepsilon$ and the restriction of φ to K is continuous.*

We close this section with another important fact about measures on separable complete metric spaces. A measure μ is called *atomic* if there exists some point x such that $\mu(\{x\}) > 0$; any such point is called an *atom*. Otherwise, the measure μ is said to be non-atomic.

The next theorem states that every non-atomic probability measure on a separable complete metric space is equivalent to the Lebesgue measure in the interval. The proof is given in Section 8.5.

Theorem A.3.10. *Let M be a separable complete metric space and μ be a non-atomic probability measure on M . Then there exists a measurable map $\psi : M \rightarrow [0, 1]$ such that ψ is a bijection with measurable inverse, restricted to a subset with full measure, and $\psi_*\mu$ is the Lebesgue measure on $[0, 1]$.*

A.3.3 Space of continuous functions

Let M be a compact metric space. We are going to describe some important properties of the vector space $C^0(M)$ of continuous functions, real or complex, defined on M . We consider on this space the norm of uniform convergence, given by:

$$\|\phi\| = \sup\{|\phi(x)| : x \in M\}.$$

This norm is complete and, hence, endows $C^0(M)$ with the structure of a Banach space.

The conclusions of the previous sections hold in this setting, since every compact metric space is separable and complete. Another useful fact about compact metric spaces is that every open cover admits some *Lebesgue number*, that is, some number $\rho > 0$ such that for every $x \in M$ there exists some element of the cover that contains the ball $B(x, \rho)$.

A linear functional $\Phi : C^0(M) \rightarrow \mathbb{C}$ is said to be *positive* if $\Phi(\varphi) \geq 0$ for every function $\varphi \in C^0(M)$ with $\varphi(x) \geq 0$ for every $x \in M$. The theorem of Riesz–Markov (see Theorem 6.19 in Rudin [Rud87]) shows that the only positive linear functionals on $C^0(M)$ are the integrals:

Theorem A.3.11 (Riesz–Markov). *Let M be a compact metric space. Consider any positive linear functional $\Phi : C^0(M) \rightarrow \mathbb{C}$. Then there exists a unique finite Borel measure μ on M such that*

$$\Phi(\varphi) = \int \varphi \, d\mu \quad \text{for every } \varphi \in C^0(M).$$

Moreover, μ is a probability measure if and only if $\Phi(1) = 1$.

The next result, which is also known as the theorem of Riesz–Markov, gives an analogous representation for *continuous* linear functionals in $C^0(M)$, not necessarily positive. Recall that the *norm* of a linear functional $\Phi : C^0(M) \rightarrow \mathbb{C}$ is defined by

$$\|\Phi\| = \sup \left\{ \frac{|\Phi(\varphi)|}{\|\varphi\|} : \varphi \neq 0 \right\} \tag{A.3.1}$$

and that Φ is continuous if and only if the norm is finite.

Theorem A.3.12 (Riesz–Markov). *Let M be a compact metric space. Consider any continuous linear functional $\Phi : C^0(M) \rightarrow \mathbb{C}$. Then there exists some complex Borel measure μ on M such that*

$$\Phi(\varphi) = \int \varphi \, d\mu \quad \text{for every } \varphi \in C^0(M).$$

The norm $\|\mu\| = |\mu|(X)$ of the measure μ coincides with the norm $\|\Phi\|$ of the functional Φ . Moreover, μ takes values in $[0, \infty)$ if and only if Φ is positive and μ takes values in \mathbb{R} if and only if $\Phi(\varphi) \in \mathbb{R}$ for every real function φ .

In other words, this last theorem asserts that *the dual space of $C^0(M)$ is isometrically isomorphic to $\mathcal{M}(M)$* . Theorems A.3.11 and A.3.12 extend to locally compact topological spaces, with suitable assumptions on the behavior of the functions at infinity. In this context the measure μ is still regular, but not necessarily finite.

We also use the fact that the space $C^0(M)$ has countable dense subsets (Exercise A.3.6 is a particular instance):

Theorem A.3.13. *If M is a compact metric space then $C^0(M)$ is separable.*

Proof. We treat the case of real functions; the complex case is entirely analogous. Every compact metric space is separable. Let $\{x_k : k \in \mathbb{N}\}$ be a countable dense subset of M . For each $k \in \mathbb{N}$, consider the function $f_k : M \rightarrow \mathbb{R}$ defined by $f_k(x) = d(x, x_k)$. Represent by \mathcal{A} the set of all functions $f : M \rightarrow \mathbb{R}$ of the form

$$f = c + \sum_{k_1, \dots, k_s} c_{k_1, \dots, k_s} f_{k_1} \cdots f_{k_s} \quad (\text{A.3.2})$$

with $c \in \mathbb{R}$ and $c_{k_1, \dots, k_s} \in \mathbb{R}$ for every $k_1, \dots, k_s \in \mathbb{N}$. It is clear that \mathcal{A} contains all constant functions. Observe also that \mathcal{A} is an *algebra of functions*, that is, it is closed under the operations of addition and multiplication (including multiplication by any constant). Moreover, \mathcal{A} *separates* the points of M , in the sense that for any $x \neq y$ there exists some $f \in \mathcal{A}$ such that $f(x) \neq f(y)$. To see that, fix $\varepsilon > 0$ such that $d(x, y) > 2\varepsilon$, consider $k \in \mathbb{N}$ such that $d(x, x_k) < \varepsilon$ and then take $f = f_k$. Note that $f(x) = d(x, x_k) < \varepsilon$ while, by the triangle inequality, $f(y) = d(y, x_k) \geq d(x, y) - d(x, x_k) > \varepsilon$. So, the algebra of functions \mathcal{A} is separating, as we claimed. Now, the theorem of Stone–Weierstrass (see [DS57, Theorem 4.6.16]) asserts that *every separating subalgebra of the space of continuous functions that contains the constant function 1 is dense in $C^0(M)$* . The previous observations show that this applies to \mathcal{A} . It follows that the (countable) set of functions of the form (A.3.2) with $c \in \mathbb{Q}$ and $c_{k_1, \dots, k_s} \in \mathbb{Q}$ is also dense in $C^0(M)$.

A.3.4 Exercises

- A.3.1. Let M be a metrizable topological space. Justify that every point of M admits a countable basis of neighborhoods. Check that M is separable if and only if it admits a countable basis of open sets. Give examples of separable metric spaces and non-separable metric spaces.
- A.3.2. Let μ be a finite measure on a metric space M . Show that for every closed set $F \subset M$ there exists some finite or countable set $E \subset (0, \infty)$ such that
- $$\mu(\{x \in M : d(x, F) = r\}) = 0 \quad \text{for every } r \in (0, \infty) \setminus E.$$
- A.3.3. Let μ be a finite measure on a separable metric space M . Show that for every $\varepsilon > 0$ there exists a countable partition of M into measurable subsets with diameter less than ε and whose boundaries have measure zero.
- A.3.4. Let μ be a probability measure on $[0, 1]$ and $\phi : [0, 1] \rightarrow [0, 1]$ be the function given by $\phi(x) = \mu([0, x])$. Check that ϕ is continuous if and only if μ is non-atomic. Check that ϕ is absolutely continuous if and only if μ is absolutely continuous with respect to the Lebesgue measure.
- A.3.5. Let μ be a probability measure on some metric space M . Show that for every integrable function $\psi : M \rightarrow \mathbb{R}$ there exists a sequence $\psi_n : M \rightarrow \mathbb{R}$, $n \geq 1$ of uniformly continuous functions converging to ψ at μ -almost every point. Moreover, if ψ is bounded then we may choose the sequence in such a way

that $\sup |\psi_n| \leq \sup |\psi|$ for every n . Do these claims remain true if we require convergence at *every* point?

- A.3.6. Without using Theorem A.3.13, show that the space $C^0([0, 1]^d)$ of continuous functions, real or complex, on the compact unit cube is separable, for every $d \geq 1$.

A.4 Differentiable manifolds

In this appendix we review some fundamental notions and facts from differential topology and Riemannian geometry.

A.4.1 Differentiable manifolds and maps

A *differentiable manifold* of dimension d is a (Hausdorff) topological space M endowed with a *differentiable atlas* of dimension d , that is, a family of homeomorphisms $\varphi_\alpha : U_\alpha \rightarrow X_\alpha$ such that

1. each U_α is an open subset of M and each X_α is an open subset of \mathbb{R}^d and $M = \bigcup_\alpha U_\alpha$;
2. the map $\varphi_\beta \circ \varphi_\alpha^{-1} : \varphi_\alpha(U_\alpha \cap U_\beta) \rightarrow \varphi_\beta(U_\alpha \cap U_\beta)$ is differentiable, for any α and β such that $U_\alpha \cap U_\beta \neq \emptyset$.

More generally, instead of \mathbb{R}^d we may consider any Banach space E . Then we say that M is a differentiable manifold *modelled* on the space E .

The homeomorphisms φ_α are called *local charts*, or *local coordinates*, and the transformations $\varphi_\beta \circ \varphi_\alpha^{-1}$ are called *coordinate changes*. Exchanging the roles of α and β , we see that the inverse $(\varphi_\beta \circ \varphi_\alpha^{-1})^{-1} = \varphi_\alpha \circ \varphi_\beta^{-1}$ is also differentiable. So, the definition of a differentiable manifold requires the coordinate changes to be diffeomorphisms between open subsets of Euclidean space.

Unless explicitly stated otherwise, we only consider manifolds such that M admits a countable basis of open sets and is *connected*. The latter means that no subset of M is both open and closed, except for M and \emptyset .

Let $r \in \mathbb{N} \cup \{\infty\}$. If every coordinate change is of class C^r (that is, all its partial derivatives up to order r exist and are continuous), we say that the manifold M (and the atlas $\{\varphi_\alpha : U_\alpha \rightarrow X_\alpha\}$) are of class C^r . Clearly, every manifold of class C^r is also of class C^s for every $s \leq r$.

Example A.4.1. The following are manifolds of class C^∞ and dimension d :

Euclidean space \mathbb{R}^d : consider the atlas consisting of a unique map, namely, the identity map $\mathbb{R}^d \rightarrow \mathbb{R}^d$.

Sphere $S^d = \{(x_0, x_1, \dots, x_d) \in \mathbb{R}^{d+1} : x_0^2 + x_1^2 + \dots + x_d^2 = 1\}$: consider the atlas formed by the two *stereographic projections*:

$$S^d \setminus \{(1, 0, \dots, 0)\} \rightarrow \mathbb{R}^d, \quad (x_0, x_1, \dots, x_d) \mapsto (x_1, \dots, x_d)/(1 - x_0)$$

$$S^d \setminus \{(-1, 0, \dots, 0)\} \rightarrow \mathbb{R}^d, \quad (x_0, x_1, \dots, x_d) \mapsto (x_1, \dots, x_d)/(1 + x_0).$$

Torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$: consider the atlas formed by the inverses of the maps $g_z : (0, 1)^d \rightarrow \mathbb{T}^d$, defined by $g_z(x) = z + x \pmod{\mathbb{Z}^d}$ for every $z \in \mathbb{R}^d$.

Example A.4.2 (Grassmannian manifolds). Given $0 \leq k \leq d$, denote by $\text{Gr}(k, d)$ the set of all vector subspaces of dimension k of the Euclidean space \mathbb{R}^d . For each $j_1 < \dots < j_k$, denote by $\text{Gr}(k, d, j_1, \dots, j_k)$ the subset of elements of $\text{Gr}(k, d)$ that are transverse to $\{(x_j)_j \in \mathbb{R}^d : x_{j_1} = \dots = x_{j_k} = 0\}$. For every $V \in \text{Gr}(k, d, j_1, \dots, j_k)$ there exists a unique matrix $(u_{i,j})_{i,j}$ with $(d - k)$ rows and k columns such that

$$V = \{(x_j)_j \in \mathbb{R}^d : x_i = u_{i,j_1}x_{j_1} + \dots + u_{i,j_k}x_{j_k} \text{ for every } i \notin \{j_1, \dots, j_k\}\}.$$

The maps $\text{Gr}(k, d, j_1, \dots, j_k) \rightarrow \mathbb{R}^{(d-k)k}$ associating with each V the corresponding matrix $(u_{i,j})_{i,j}$ constitute an atlas of class C^∞ for $\text{Gr}(k, d)$. So, every $\text{Gr}(k, d)$ is a manifold of class C^∞ and dimension $(d - k)k$.

Let M be a manifold of dimension d and $\mathcal{A} = \{\varphi_\alpha : U_\alpha \rightarrow X_\alpha\}$ be the corresponding atlas. Let S be a subset of M . We say that S is a *submanifold* of dimension $k < d$ if there exists some atlas $\mathcal{B} = \{\psi_\beta : V_\beta \rightarrow Y_\beta\}$ of M such that

- (i) \mathcal{A} and \mathcal{B} are *compatible*: the coordinate changes $\psi_\beta \circ \varphi_\alpha^{-1}$ and $\varphi_\alpha \circ \psi_\beta^{-1}$ are differentiable in their domains, for every α and every β ;
- (ii) for every β , the local chart ψ_β maps $V'_\beta = V_\beta \cap S$ onto an open subset Y'_β of $\mathbb{R}^k \times \{0^{d-k}\}$.

Identifying $\mathbb{R}^k \times \{0^{d-k}\} \simeq \mathbb{R}^k$, we get that the family formed by the restrictions $\psi_\beta : V'_\beta \rightarrow Y'_\beta$ constitutes an atlas for S . Hence, S is a manifold of dimension k . If M is a manifold of class C^r and the atlases \mathcal{A} and \mathcal{B} are C^r -compatible, that is, if all the coordinate changes in (i) are of class C^r , then S is a (sub)manifold of class C^r .

We say that a map $f : M \rightarrow N$ between two manifolds is *differentiable* if

$$\psi_\beta \circ f \circ \varphi_\alpha^{-1} : \varphi_\alpha(U_\alpha \cap f^{-1}(V_\beta)) \rightarrow \psi_\beta(V_\beta \cap f(U_\alpha)) \tag{A.4.1}$$

is a differentiable map for every local chart $\varphi_\alpha : U_\alpha \rightarrow X_\alpha$ of M and every local chart $\psi_\beta : V_\beta \rightarrow Y_\beta$ of N with $f(U_\alpha) \cap V_\beta \neq \emptyset$. Moreover, we say that f is of *class* C^r if M and N are manifolds of class C^r and every map $\psi_\beta \circ f \circ \varphi_\alpha^{-1}$ in (A.4.1) is of class C^r . A *diffeomorphism* $f : M \rightarrow N$ is a bijection between two manifolds such that both f and f^{-1} are differentiable. If both maps are of class C^r then we say that the diffeomorphism is of *class* C^r .

Let $C^r(M, N)$ be the space of maps of class C^r between two manifolds M and N . We are going to introduce in this space a certain topology, called the C^r topology, for which two maps are close if and only if they are uniformly close and the same is true for their derivatives up to order r . The definition may be given in a very broad context (see Section 2.1 of Hirsch [Hir94]), but we restrict ourselves to the case when M and N are compact. In this case, the C^r topology may be defined in the following way.

Fix finite families of local charts $\varphi_i : U_i \rightarrow X_i$ of M and $\psi_j : V_j \rightarrow Y_j$ of N , such that $\bigcup_i U_i = M$ and $\bigcup_j V_j = N$. Let $\delta > 0$ be a Lebesgue number for the open cover $\{U_i \cap f^{-1}(V_j)\}$ of M . For each pair (i, j) such that $U_i \cap f^{-1}(V_j) \neq \emptyset$, let $K_{i,j}$ be the set of points whose distance to the complement of $U_i \cap f^{-1}(V_j)$ is greater than or equal to δ . Then $K_{i,j}$ is a compact set contained in $U_i \cap f^{-1}(V_j)$ and the union $\bigcup_{i,j} K_{i,j}$ is the whole M . Consider

$$\mathcal{U}(f) = \{g \in C^r(M, N) : g(K_{i,j}) \subset V_j \text{ for any } (i, j)\}.$$

It is clear that $f \in \mathcal{U}(f)$. For each $g \in \mathcal{U}(f)$ and each pair (i, j) such that $K_{i,j}$ is non-empty, denote by $g_{i,j}$ the restriction of $\psi_j \circ g \circ \varphi_i^{-1}$ to the set $\varphi_i(K_{i,j})$. For each $r \in \mathbb{N}$ and $\varepsilon > 0$, define

$$\mathcal{U}^r(f, \varepsilon) = \{g \in \mathcal{U}(f) : \sup_{s,x,i,j} \|D^s f_{i,j}(x) - D^s g_{i,j}(x)\| < \varepsilon\}, \tag{A.4.2}$$

where the supremum is over every $s \in \{1, \dots, r\}$, every $x \in \varphi_i(K_{i,j})$ and every pair (i, j) such that $K_{i,j} \neq \emptyset$. By definition, the family $\{\mathcal{U}^r(f, \varepsilon) : \varepsilon > 0\}$ is a basis of neighborhoods of each $f \in C^r(M, N)$ relative to the C^r topology. Also by definition, the family $\{\mathcal{U}^r(f, \varepsilon) : \varepsilon > 0 \text{ and } r \in \mathbb{N}\}$ is a basis of neighborhoods of $f \in C^\infty(M, N)$ relative to the C^∞ topology.

The C^r topology has very nice properties: in particular, it admits a countable basis of open sets and is *completely metrizable*, that is, it is generated by some complete distance. An interesting consequence is that $C^r(M, N)$ is a *Baire space*: every intersection of a countable family of open dense subsets is dense in the space. The set $\text{Diffeo}^r(M)$ of diffeomorphisms of class C^r is an open subset of $C^r(M, M)$ relative to the C^r topology.

A.4.2 Tangent space and derivative

Let M be a manifold. For each $p \in M$, consider the set $\mathcal{C}(p)$ of all the curves $c : I \rightarrow M$ whose domain is some open interval I containing $0 \in \mathbb{R}$, such that $c(0) = p$ and c is differentiable at the point 0. The latter means that the map $\varphi_\alpha \circ c$ is differentiable at the point 0 for every local chart $\varphi_\alpha : U_\alpha \rightarrow X_\alpha$ with $p \in U_\alpha$. We say that two curves $c_1, c_2 \in \mathcal{C}(p)$ are *equivalent* if $(\varphi_\alpha \circ c_1)'(0) = (\varphi_\alpha \circ c_2)'(0)$ for every local chart $\varphi_\alpha : U_\alpha \rightarrow X_\alpha$ with $p \in U_\alpha$. Actually, if the equality holds for some local chart then it holds for all the other charts as well. We denote by $[c]$ the equivalence class of any curve $c \in \mathcal{C}(p)$.

The *tangent space* to the manifold M at the point p is the set of such equivalence classes. We denote this set by T_pM . For any fixed local chart $\varphi_\alpha : U_\alpha \rightarrow X_\alpha$ with $p \in U_\alpha$, the map

$$D\varphi_\alpha(p) : T_pM \rightarrow \mathbb{R}^d, \quad [c] \mapsto (\varphi_\alpha \circ c)'(0)$$

is well defined and is a bijection. We may use this bijection to identify T_pM with \mathbb{R}^d . In this way, the tangent space acquires the structure of a vector space, transported from \mathbb{R}^d via $D\varphi_\alpha(p)$. Although this identification $D\varphi_\alpha(p)$ depends on the choice of the local chart, the vector space structure on T_pM does not. That is because, for any other local chart $\varphi_\beta : U_\beta \rightarrow X_\beta$ with $p \in U_\beta$, the corresponding map $D\varphi_\beta(p)$ is given by

$$D\varphi_\beta(p) = D(\varphi_\beta \circ \varphi_\alpha^{-1})(\varphi_\alpha(p)) \circ D\varphi_\alpha(p).$$

Since $D(\varphi_\beta \circ \varphi_\alpha^{-1})(\varphi_\alpha(p))$ is a linear isomorphism, it follows that the vector space structures transported from Euclidean space to T_pM by $D\varphi_\alpha(p)$ and $D\varphi_\beta(p)$ coincide, as we stated.

If $f : M \rightarrow N$ is a differentiable map, its *derivative* at a point $p \in M$ is the linear map $Df(p) : T_pM \rightarrow T_{f(p)}N$ defined by

$$Df(p) = D\psi_\beta(f(p))^{-1} \circ D(\psi_\beta \circ f \circ \varphi_\alpha^{-1})(\varphi_\alpha(p)) \circ D\varphi_\alpha(p),$$

where $\varphi_\alpha : U_\alpha \rightarrow X_\alpha$ is a local chart of M with $p \in U_\alpha$ and $\psi_\beta : V_\beta \rightarrow Y_\beta$ is a local chart of N with $f(p) \in V_\beta$. The definition does not depend on the choice of these local charts.

The *tangent bundle* to M is the (disjoint) union $TM = \bigcup_{p \in M} T_pM$ of all the tangent spaces to M . For each local chart $\varphi_\alpha : U_\alpha \rightarrow X_\alpha$, consider the union $T_{U_\alpha}M = \bigcup_{p \in U_\alpha} T_pM$ and the map

$$D\varphi_\alpha : T_{U_\alpha}M \rightarrow X_\alpha \times \mathbb{R}^d$$

that associates with each $[c] \in T_{U_\alpha}M$ the pair

$$((\varphi_\alpha \circ c)(0), (\varphi_\alpha \circ c)'(0)) \in X_\alpha \times \mathbb{R}^d.$$

We consider on TM the (unique) topology that turns every $D\varphi_\alpha$ into a homeomorphism. Assuming that the atlas $\{\varphi_\alpha : U_\alpha \rightarrow X_\alpha\}$ of the manifold M is of class C^r , the coordinate change

$$D\varphi_\beta \circ D\varphi_\alpha^{-1} : \varphi_\alpha(U_\alpha \cap U_\beta) \times \mathbb{R}^d \rightarrow \varphi_\beta(U_\alpha \cap U_\beta) \times \mathbb{R}^d$$

is a map of class C^{r-1} for any α and β such that $U_\alpha \cap U_\beta \neq \emptyset$. So, the tangent bundle TM is endowed with the structure of a manifold of class C^{r-1} and dimension $2d$.

The *derivative* $Df : TM \rightarrow TN$ of a differentiable map $f : M \rightarrow N$ is the map whose restriction to each tangent space T_pM is given by $Df(p)$. If f is of class C^r then Df is of class C^{r-1} , relative to the manifold structure on the tangent bundles TM and TN that we introduced in the previous paragraph. For example,

the canonical projection $\pi : TM \rightarrow M$, associating with each $v \in TM$ the unique point $p \in M$ such that $v \in T_pM$, is a map of class C^{r-1} (Exercise A.4.9).

A *vector field* on a manifold M is a map that associates with each point $p \in M$ an element $X(p)$ of the tangent space T_pM , that is, a map $X : M \rightarrow TM$ such that $\pi \circ X = \text{id}$. We say that the vector field is of class C^k , with $k \leq r - 1$, if this map is of class C^k .

Assuming that $k \geq 1$, we may apply the theorem of existence and uniqueness of solutions of ordinary differential equations to conclude that for every point $p \in M$ there exists a unique curve $c_p : I_p \rightarrow M$ such that

- $c_p(0) = p$ and $c'_p(t) = X(c(t))$ for every $t \in I_p$, and
- I_p is the largest open interval where such a curve can be defined.

If M is compact then $I_p = \mathbb{R}$ for any $p \in M$. Moreover, the maps $f^t : M \rightarrow M$ defined by $f^t(p) = c_p(t)$ are diffeomorphisms of class C^k , with $f^0 = \text{id}$ and $f^s \circ f^t = f^{s+t}$ for any $s, t \in \mathbb{R}$. The family $\{f^t : t \in \mathbb{R}\}$ is called the *flow* of the vector field X .

A.4.3 Cotangent space and differential forms

The *cotangent space* T^*_pM to a manifold M at a point p is the dual of the tangent space T_pM , that is, the space of linear functionals $\zeta : T_pM \rightarrow \mathbb{R}$. For any local chart $\varphi_\alpha : U_\alpha \rightarrow X_\alpha$ with $p \in U_\alpha$, the isomorphism $D\varphi_\alpha(p) : T_pM \rightarrow \mathbb{R}^d$ induces an isomorphism

$$D\varphi^*_\alpha(p) : T^*_pM \rightarrow \mathbb{R}^d$$

as follows. For each $i = 1, \dots, d$, let $dx_i = \pi_i \circ D\varphi_\alpha(p)$, where $\pi_i : \mathbb{R}^d \rightarrow \mathbb{R}$ is the projection to the i -th coordinate. Then $dx_i \in T^*_pM$ and, in fact, the family $\{dx_1, \dots, dx_d\}$ is a basis of T^*_pM . For each $\zeta \in T^*_pM$, define

$$D\varphi^*_\alpha(p)\zeta = (\zeta_1, \dots, \zeta_d) \iff \zeta = \sum_{i=1}^d \zeta_i dx_i.$$

The *cotangent bundle* of M is the (disjoint) union $T^*M = \bigcup_{p \in M} T^*_pM$ of all the cotangent spaces to M . For each local chart $\varphi_\alpha : U_\alpha \rightarrow X_\alpha$, consider the union $T^*_{U_\alpha}M = \bigcup_{p \in U_\alpha} T^*_pM$ and the map

$$D\varphi^*_\alpha : T^*_{U_\alpha}M \rightarrow X_\alpha \times \mathbb{R}^d$$

defined by $D\varphi^*_\alpha \zeta = (\varphi_\alpha(p), D\varphi^*_\alpha(p)\zeta)$ if $\zeta \in T^*_pM$. It is clear that this is a bijection. We consider on T^*M the unique topology that turns every $D\varphi^*_\alpha$ into a homeomorphism. If $\{\varphi_\alpha : U_\alpha \rightarrow X_\alpha\}$ is an atlas of class C^r for the manifold M then

$$\{D\varphi^*_\alpha : T^*_{U_\alpha}M \rightarrow X_\alpha \times \mathbb{R}^d\}$$

is an atlas of class C^{r-1} for T^*M . So, the cotangent bundle T^*M is also endowed with the structure of a manifold of class C^{r-1} and dimension $2d$.

Moreover, the canonical map $\pi^* : T^*M \rightarrow M$ defined by $\pi^* \mid T_p^*M = p$ is of class C^{r-1} .

A differential 1-form in M is a differentiable map $\theta : M \rightarrow T^*M$ such that $\pi^* \circ \theta = \text{id}$. In other words, θ assigns to each point $p \in M$ a linear functional (or linear form) $\theta_p : T_pM \rightarrow \mathbb{R}$ that depends differentiably on the point.

More generally, for any $0 \leq k \leq d$, an alternate k -linear form in T_pM is a map¹

$$\theta_p : (T_pM)^k \rightarrow \mathbb{R}, \quad (v_1, \dots, v_k) \mapsto \theta_p(v_1, \dots, v_k)$$

such that θ_p is linear on each variable v_i and

$$\theta_p(v_1, \dots, v_i, v_{i+1}, \dots, v_k) = -\theta_p(v_1, \dots, v_{i+1}, v_i, \dots, v_k)$$

for any $1 \leq i < k$ and any $(v_1, \dots, v_k) \in (T_pM)^k$.

Let $\{dx_1, \dots, dx_d\}$ be the basis of the cotangent space associated with a local chart $\varphi_\alpha : U_\alpha \rightarrow X_\alpha$ and let $\{\partial/\partial x_1, \dots, \partial/\partial x_d\}$ be the dual basis of T_pM , defined by

$$dx_i(\partial/\partial x_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

If $i_1, \dots, i_k \in \{1, \dots, d\}$ are all distinct, there exists a unique alternate k -linear form $dx_{i_1} \wedge \dots \wedge dx_{i_k}$ such that

- $dx_{i_1} \wedge \dots \wedge dx_{i_k}(\partial/\partial x_{i_1}, \dots, \partial/\partial x_{i_k}) = 1$, and
- $dx_{i_1} \wedge \dots \wedge dx_{i_k}(\partial/\partial x_{j_1}, \dots, \partial/\partial x_{j_k}) = 0$ when $\{i_1, \dots, i_k\} \neq \{j_1, \dots, j_k\}$.

The family $\{dx_{i_1} \wedge \dots \wedge dx_{i_k} : 1 \leq i_1 < \dots < i_k \leq d\}$ is a basis of the vector space of alternate k -linear forms in T_pM .

A differential k -form in M is a map θ assigning to each point $p \in M$ an alternate k -linear form in the tangent space T_pM that depends differentiably on the point. In local coordinates, this may be written as

$$\theta_p = \sum_{1 \leq i_1 < \dots < i_k \leq d} a_{i_1, \dots, i_k}(p) dx_{i_1} \wedge \dots \wedge dx_{i_k}.$$

The differentiability condition means that the coefficients $a_{i_1, \dots, i_k}(p)$ depend differentiably on the point p .

Assuming that $k < d$, the exterior derivative of θ is the differential $(k + 1)$ -form $d\theta$ determined by

$$d\theta_p = \sum_{1 \leq i_1 < \dots < i_k \leq d} \sum_j \frac{\partial a_{i_1, \dots, i_k}}{\partial x_j}(p) dx_j \wedge dx_{i_1} \wedge \dots \wedge dx_{i_k},$$

where the second sum is over all $j \notin \{i_1, \dots, i_k\}$; one can check that the expression on the right-hand side does not depend on the choice of the local chart. A differential k -form θ is closed if $d\theta = 0$ (or else $k = d$) and it is exact if

¹ An alternate 0-linear form is just a real number.

there exists some $(k - 1)$ -form η such that $d\eta = \theta$ (or else $k = 0$). Every exact differential form is closed.

For much more information on the subject of differential forms, see the book of Henri Cartan [Car70].

A.4.4 Transversality

The result that we state next is an important tool for constructing new manifolds. We say that $y \in N$ is a *regular value* of a differentiable map $f : M \rightarrow N$ if the derivative $Df(x) : T_xM \rightarrow T_yN$ is surjective for every $x \in f^{-1}(y)$. Note that this holds, automatically, if y is not in the image of f , that is, if $f^{-1}(y)$ is the empty set. On the other hand, in order that some point $y \in f(M)$ is a regular value of f it is necessary that $\dim M \geq \dim N$.

Theorem A.4.3. *Let $f : M \rightarrow N$ be a map of class C^r and $y \in f(M)$ be a regular value of f . Then $f^{-1}(y)$ is a submanifold (not necessarily connected) of class C^r of M , with dimension equal to $\dim M - \dim N$.*

Example A.4.4. For any $d \geq 1$, the space of square matrices of dimension d with real coefficients is isomorphic to the Euclidean space $\mathbb{R}^{(d^2)}$ and, hence, it is a manifold of dimension d^2 and class C^∞ . The *linear group* $GL(d, \mathbb{R})$ of invertible matrices is an open subset of that space and, hence, it is also a manifold of dimension d^2 and class C^∞ . The function $\det : GL(d, \mathbb{R}) \rightarrow \mathbb{R}$ that maps each matrix to its determinant is of class C^∞ and $y = 1$ is a regular value (see Exercise A.4.5). Using Theorem A.4.3, it follows that the *special linear group* $SL(d, \mathbb{R})$ formed by the matrices with determinant equal to 1 is a submanifold of class C^∞ of $GL(d, \mathbb{R})$, with dimension equal to $d^2 - 1$.

It is possible to generalize Theorem A.4.3, using the notion of transversality. We say that a submanifold S of N is *transverse* to f if

$$Df(x)(T_xM) + T_{f(x)}S = T_{f(x)}N \quad \text{for every } x \in f^{-1}(S). \tag{A.4.3}$$

For example, if S is a submanifold of dimension zero, that is, if it consists of a unique point, then S is transverse to f if and only if that point is a regular value of f . Therefore, the following statement generalizes Theorem A.4.3:

Theorem A.4.5. *Let $f : M \rightarrow N$ be a map of class C^r and let S be a submanifold of class C^r of N transverse to f . Then $f^{-1}(S)$ is a submanifold (not necessarily connected) of class C^r of M , with dimension equal to $\dim M - \dim N + \dim S$.*

The next theorem asserts that, for every map $f : M \rightarrow N$ of class C^r with r sufficiently high, “almost all” points $y \in N$ are regular values. We say that a set $X \subset N$ is *residual* if it contains some countable intersection of open and dense subsets. Every residual set is dense in the manifold, because manifolds are Baire spaces. We say that a set $Z \subset N$ has *volume zero* if for every local

chart $\psi_\beta : V_\beta \rightarrow Y_\beta$ the image $\psi_\beta(Z \cap V_\beta)$ is a subset of the Euclidean space with volume zero, that is, it may be covered by balls in such a way that the sum of the volumes of those balls is arbitrarily small.

Theorem A.4.6 (Sard). *Assume that $f : M \rightarrow N$ is a map of class C^r with $r > \max\{0, \dim M - \dim N\}$. Then the set of regular points of f is a residual subset of N and its complement has volume zero.*

A.4.5 Riemannian manifolds

A *Riemannian metric* on a manifold M is a map that associates with each point $p \in M$ an inner product in the tangent space T_pM , that is, a symmetric bilinear map

$$\cdot_p : T_pM \times T_pM \rightarrow \mathbb{R}$$

such that $v \cdot_p v > 0$ for every non-zero vector $v \in T_pM$. As part of the definition, this inner product is required to vary in a differentiable way with the point p , in the following sense. Consider any local chart $\varphi_\alpha : U_\alpha \rightarrow X_\alpha$ of M . As explained previously, for every $p \in U_\alpha$ we may identify T_pM with \mathbb{R}^d , through the map $D\varphi_\alpha(p)$. Thus, we may view \cdot_p as an inner product in the Euclidean space. Let e_1, \dots, e_d be a basis of \mathbb{R}^d . Then the functions $g_{\alpha,ij}(p) = e_i \cdot_p e_j$ are required to be differentiable, for every pair (i,j) and any choice of the local chart φ_α and the basis e_1, \dots, e_d .

We call a *Riemannian manifold* any manifold endowed with a Riemannian metric. Every submanifold S of a Riemannian manifold M inherits the structure of a Riemannian manifold, given by the restriction of the inner product \cdot_p of M to the tangent subspace T_pS of each point $p \in S$. Every compact manifold admits (infinitely many) Riemannian metrics. That follows from the theorem of Whitney (see Section 1.3 of Hirsch [Hir94]), according to which every compact manifold may be realized as a submanifold of some Euclidean space. Actually, this remains true in the much larger class of *paracompact* manifolds (which we do not define here): every paracompact manifold of dimension d is diffeomorphic to some submanifold of \mathbb{R}^{2d} . In particular, paracompact manifolds are always metrizable.

Starting from the Riemannian metric, we may define the *length* of a differentiable curve $\gamma : [a, b] \rightarrow M$, by

$$\text{length}(\gamma) = \int_a^b \|\gamma'(t)\|_{\gamma(t)} dt, \quad \text{where } \|v\|_p = (v \cdot_p v)^{1/2}.$$

This also allows us to define on the manifold M the following *distance associated with the Riemannian metric*: the distance $d(p, q)$ between two points $p, q \in M$ is the infimum of the lengths of all the differentiable curves connecting the two points. We say that a differentiable curve $\gamma : [a, b] \rightarrow M$ is *minimizing*

if it realizes the distance between its endpoints, that is, if

$$\text{length}(\gamma) = d(\gamma(a), \gamma(b)).$$

Any two points $p, q \in M$ are connected by some minimizing curve; in other words, the infimum in the definition of $d(p, q)$ is always realized.

A differentiable curve $\gamma : I \rightarrow M$ defined on an open interval I is called a *geodesic* if it is locally minimizing, in the following sense: for every $c \in I$ there exists $\delta > 0$ such that the restriction of γ to the interval $[c - \delta, c + \delta]$ is minimizing. Every minimizing curve is a geodesic, but the converse is not true: for example, the great circles are geodesics on the sphere S^2 , but closed curves cannot be minimizing. An important fact is that if γ is a geodesic then the norm $\|\gamma'(t)\|_{\gamma(t)}$ is constant on the domain I . The theory of ordinary differentiable equations may be used to show that for every $p \in M$ and every $v \in T_pM$ there exists a unique geodesic $\gamma_{p,v} : I_{p,v} \rightarrow M$ such that $\gamma_{p,v}(0) = p$, $\gamma'_{p,v}(0) = v$ and $I_{p,v}$ is a maximal interval such that $\gamma_{p,v}$ is locally minimizing.

If the manifold M is compact then $I_{p,v} = \mathbb{R}$ for every $p \in M$ and every $v \in T_pM$. Then we define the *exponential map* at each point $p \in M$:

$$\exp_p : T_pM \rightarrow M, \quad v \mapsto \gamma_{p,v}(1).$$

This is a differentiable map and its derivative at $v = 0$ is the identity transformation on the tangent space T_pM . We also define the *geodesic flow* on the tangent bundle:

$$f^t : TM \rightarrow TM, \quad (p, v) \mapsto (\gamma_{p,v}(t), \gamma'_{p,v}(t)).$$

Most of the time, one considers the restriction of the geodesic flow to the *unit tangent bundle* $T^1M = \{(p, v) \in TM : \|v\|_p = 1\}$. This is well defined since, as we mentioned before, the norm of the velocity vector of any geodesic is constant.

A.4.6 Exercises

- A.4.1. Check that every set X with the cardinality of \mathbb{R} may be endowed with the structure of a differentiable manifold of class C^∞ and dimension d , for any $d \geq 1$.
- A.4.2. Consider the differentiable manifolds $M = (\mathbb{R}, \mathcal{A})$ and $N = (\mathbb{R}, \mathcal{B})$, where \mathcal{A} is the atlas consisting of the map $\phi(x) = x$ and \mathcal{B} is the atlas consisting of the map $\psi(x) = x^3$. Is the map $f : M \rightarrow N$ defined by $f(x) = x$ a diffeomorphism between these manifolds?
- A.4.3. A topological space is *path connected* if any two points are connected by some continuous curve. Show that every (connected) manifold is path connected.
- A.4.4. For each $d \geq 2$, the *projective space of dimension d* is the set \mathbb{P}^d of all subspaces of \mathbb{R}^{d+1} with dimension 1. Equivalently, \mathbb{P}^d is the quotient space of $\mathbb{R}^{d+1} \setminus \{0\}$ for the equivalence relation defined by:

$$(x_0, \dots, x_d) \sim (y_0, \dots, y_d) \Leftrightarrow \text{there exists } c \neq 0 \text{ such that } x_i = cy_i \text{ for every } i.$$

Show that the family of maps $\varphi_i : U_i \rightarrow \mathbb{R}^d, i = 0, \dots, d$ defined by

$$U_i = \{[x_0 : \dots : x_d] \in \mathbb{P}^d : x_i \neq 0\}$$

(where $[x_0 : \dots : x_d]$ denotes the equivalence class of (x_0, \dots, x_d)) and

$$\varphi_i([x_0 : \dots : x_d]) = \left(\frac{x_0}{x_i}, \dots, \frac{x_{i-1}}{x_i}, \frac{x_{i+1}}{x_i}, \dots, \frac{x_d}{x_i} \right),$$

constitutes an atlas of class C^∞ and dimension d for \mathbb{P}^d .

- A.4.5. Check the claims in Example A.4.4.
- A.4.6. Let M and N be two compact (connected) manifolds with the same dimension. A map $f : M \rightarrow N$ of class C^1 is a *local diffeomorphism* if the derivative $Df(x) : T_x M \rightarrow T_{f(x)} N$ is an isomorphism for every $x \in M$. Show that in that case there exists an integer $k \geq 1$ such that every $y \in N$ has exactly k pre-images:

$$\#f^{-1}(y) = k \quad \text{for every } y \in N.$$

[Observation: The number k is called the *degree* of f and is denoted $\text{degree}(f)$.]

- A.4.7. Consider on $\mathbb{R}_+ = \{x \in \mathbb{R} : x > 0\}$ the Riemannian metric defined by $u \cdot_x v = uv/x^2$. Calculate the distance $d(a, b)$ between any two points $a, b \in \mathbb{R}_+$.
- A.4.8. Let M and N be submanifolds of \mathbb{R}^{m+n} with $\dim M = m$ and $\dim N = n$. Show that there exists a set $Z \subset \mathbb{R}^{m+n}$ with volume zero such that, for every v in the complement of Z , the translate $M + v$ is *transverse* to N :

$$T_x(M + v) + T_x N = \mathbb{R}^d \quad \text{for every } x \in (M + v) \cap N.$$

- A.4.9. Show that if M is a manifold of class C^r then the canonical projection $\pi : TM \rightarrow M$ is a map of class C^{r-1} .

A.5 $L^p(\mu)$ spaces

In this appendix we review certain Banach spaces formed by functions with special integrability properties. Throughout, (X, \mathcal{B}, μ) is a measure space. Recall that a Banach space is a vector space endowed with a norm relative to which the space is complete. We also state some properties of the norms in these spaces.

A.5.1 $L^p(\mu)$ spaces with $1 \leq p < \infty$

Given any $p \in [1, \infty)$, we say that a function $f : X \rightarrow \mathbb{C}$ is *p-integrable* with respect to μ if the function $|f|^p$ is integrable with respect to μ . For $p = 1$ this is the same as saying that the function f is integrable (Definition A.2.4 and Proposition A.2.7).

Definition A.5.1. We denote by $L^p(\mu)$ the set of all complex functions *p-integrable* with respect to μ , modulo the equivalence relation that identifies any two functions that are equal at μ -almost every point.

Note that if the measure μ is finite, which is the case in most of our examples, then all bounded measurable functions are in $L^p(\mu)$:

$$\int |f|^p d\mu \leq (\sup |f|)^p m(X) < \infty.$$

In particular, if X is a compact topological space then every continuous function is in $L^p(\mu)$. In other words, the space $C^0(X)$ of all continuous functions is contained in $L^p(\mu)$ for every $p \geq 1$.

For every function $f \in L^p(\mu)$, define the L^p -norm of f by:

$$\|f\|_p = \left(\int |f|^p d\mu \right)^{\frac{1}{p}}.$$

The next theorem asserts that $\|\cdot\|_p$ turns $L^p(\mu)$ into a Banach space:

Theorem A.5.2. *The set $L^p(\mu)$ is a complex vector space. Moreover, $\|\cdot\|_p$ is a norm in $L^p(\mu)$ and this norm is complete.*

The most interesting part of the proof of this theorem is to establish the triangle inequality, which in this context is known as the *Minkowski inequality*:

Theorem A.5.3 (Minkowski inequality). *Let $f, g \in L^p(\mu)$. Then:*

$$\left(\int |f + g|^p d\mu \right)^{\frac{1}{p}} \leq \left(\int |f|^p d\mu \right)^{\frac{1}{p}} + \left(\int |g|^p d\mu \right)^{\frac{1}{p}}.$$

In Exercises A.5.2 and A.5.5 we invite the reader to prove the Minkowski inequality and to complete the proof of Theorem A.5.2.

A.5.2 Inner product in $L^2(\mu)$

The case $p = 2$ deserves special attention. The reason is that the norm $\|\cdot\|_2$ introduced in the previous section arises from an (Hermitian) inner product. Indeed, consider:

$$f \cdot g = \int f\bar{g} d\mu. \tag{A.5.1}$$

It follows from the properties of the Lebesgue integral that this expression does define an inner product on $L^2(\mu)$. Moreover, this product gives rise to the norm $\|\cdot\|_2$ through:

$$\|f\|_2 = (f \cdot f)^{1/2}.$$

In particular, we have the Cauchy–Schwarz inequality:

Theorem A.5.4 (Cauchy–Schwarz Inequality). *For every $f, g \in L^2(\mu)$ we have that $f\bar{g} \in L^1(\mu)$ and*

$$\left| \int f\bar{g} d\mu \right| \leq \int |f\bar{g}| d\mu \leq \left(\int |f|^2 d\mu \right)^{1/2} \left(\int |g|^2 d\mu \right)^{1/2}.$$

This inequality has the following interesting consequence. Assume that the measure μ is finite and consider any $f \in L^2(\mu)$. Then, taking $g \equiv 1$,

$$\int |f| d\mu = \int |f\bar{g}| d\mu \leq \left(\int |f|^2 d\mu \right)^{1/2} \left(\int 1 d\mu \right)^{1/2} < \infty. \tag{A.5.2}$$

This proves that every function in $L^2(\mu)$ is also in $L^1(\mu)$. In fact, when the measure μ is finite one has $L^p(\mu) \subset L^q(\mu)$ whenever $p \geq q$ (Exercise A.5.3).

The next result is a generalization of the Cauchy–Schwarz inequality for all values of $p > 1$:

Theorem A.5.5 (Hölder inequality). *Given $1 < p < \infty$, consider $q > 1$ defined by the relation $\frac{1}{p} + \frac{1}{q} = 1$. Then, for every $f \in L^p(\mu)$ and every $g \in L^q(\mu)$, we have that $f\bar{g} \in L^1(\mu)$ and*

$$\int |f\bar{g}| d\mu \leq \left(\int |f|^p d\mu \right)^{\frac{1}{p}} \left(\int |g|^q d\mu \right)^{\frac{1}{q}}.$$

A.5.3 Space of essentially bounded functions

Next, we extend the definition of $L^p(\mu)$ to the case $p = \infty$. For that we need the following notion. We say that a function $f : X \rightarrow \mathbb{C}$ is *essentially bounded* with respect to μ if there exists some constant $K > 0$ such that $|f(x)| \leq K$ at μ -almost every point. Then the infimum of all such constants K is called the *essential supremum* of f and is denoted by $\text{supess}_\mu(f)$.

Definition A.5.6. We denote by $L^\infty(\mu)$ the set of all complex functions essentially bounded with respect to μ , identifying any two functions that coincide at μ -almost every point.

We endow $L^\infty(\mu)$ with the following norm:

$$\|f\|_\infty = \text{supess}_\mu(f).$$

The conclusion of Proposition A.5.2 remains valid for $p = \infty$ (Exercise A.5.5): the space $L^\infty(\mu)$ is a Banach space for the norm $\|\cdot\|_\infty$. Clearly, if μ is a finite measure then $L^\infty(\mu) \subset L^p(\mu)$ for any $p \geq 1$.

The *dual* of a complex Banach space E is the space E^* of all continuous linear functionals $\phi : E \rightarrow \mathbb{C}$, endowed with the norm

$$\|\phi\| = \sup \left\{ \frac{|\phi(v)|}{\|v\|} : v \in E \setminus \{0\} \right\}. \tag{A.5.3}$$

The Hölder inequality (Theorem A.5.5) leads to the following explicit characterization of the dual space of $L^p(\mu)$ for every $p < \infty$:

Theorem A.5.7. *For each $p \in [1, \infty)$ consider $q \in (1, \infty]$ defined by the relation $\frac{1}{p} + \frac{1}{q} = 1$. The map $L^q(\mu) \rightarrow L^p(\mu)^*$ defined by $g \mapsto [f \mapsto \int fg d\mu]$ is an isomorphism and an isometry between $L^q(\mu)$ and the dual space of $L^p(\mu)$.*

This statement is false for $p = \infty$: in general, the dual space of $L^\infty(\mu)$ is not isomorphic to $L^1(\mu)$.

A.5.4 Convexity

We say that a function $\phi : I \rightarrow \mathbb{R}$ defined on an interval I of the real line is *convex* if

$$\phi(tx + (1 - t)y) \leq t\phi(x) + (1 - t)\phi(y)$$

for every $x, y \in I$ and $t \in [0, 1]$. Moreover, we say that ϕ is *concave* if $-\phi$ is convex. For functions that are twice differentiable we have the following practical criterion (Exercise A.5.1): ϕ is convex if $\phi''(x) \geq 0$ for every $x \in I$ and it is concave if $\phi''(x) \leq 0$ for every $x \in I$.

Theorem A.5.8 (Jensen inequality). *Let $\phi : I \rightarrow \mathbb{R}$ be a convex function. If μ is a probability measure on X and $f \in L^1(\mu)$ is such that $\int f d\mu \in I$, then:*

$$\phi\left(\int f d\mu\right) \leq \int \phi \circ f d\mu.$$

Example A.5.9. For any probability measure μ and any integrable positive function f , we have

$$\log \int f d\mu \geq \int \log f d\mu.$$

Indeed, this corresponds to the Jensen inequality for the function $\phi : (0, \infty) \rightarrow \mathbb{R}$ given by $\phi(x) = -\log x$. Note that ϕ is convex: $\phi''(x) = 1/x^2 > 0$ for every x .

Example A.5.10. Let $\phi : \mathbb{R} \rightarrow \mathbb{R}$ be a convex function, $(\lambda_i)_i$ be a sequence of non-negative real numbers satisfying $\sum_{i=1}^\infty \lambda_i \leq 1$ and $(a_i)_i$ be a bounded sequence of real numbers. Then

$$\phi\left(\sum_{i=1}^\infty \lambda_i a_i\right) \leq \sum_{i=1}^\infty \lambda_i \phi(a_i). \tag{A.5.4}$$

This may be seen as follows. Consider $X = [0, 1]$ endowed with the Lebesgue measure μ . Let $f : [0, 1] \rightarrow \mathbb{R}$ be a function of the form $f = \sum_{i=1}^\infty a_i \chi_{E_i}$, where the E_i are pairwise disjoint measurable sets such that $\mu(E_i) = \lambda_i$. The Jensen inequality applied to this function f gives precisely the relation (A.5.4).

A.5.5 Exercises

- A.5.1. Consider any function $\phi : (a, b) \rightarrow \mathbb{R}$. Show that if ϕ is twice differentiable and $\phi'' \geq 0$ then ϕ is convex. Show that if ϕ is convex then it is continuous.
- A.5.2. Consider $p, q > 1$ such that $1/p + 1/q = 1$. Prove:
 - (a) The Young inequality: $ab \leq a^p/p + a^q/q$ for every $a, b > 0$.
 - (b) The Hölder inequality (Theorem A.5.5).
 - (c) The Minkowski inequality (Theorem A.5.3).

- A.5.3. Show that if μ is a finite measure then we have $L^q(\mu) \subset L^p(\mu)$ for every $1 \leq p < q \leq \infty$.
- A.5.4. Let μ be a finite measure and $f \in L^\infty(\mu)$ be different from zero. Show that
- $$\|f\|_\infty = \lim_n \frac{\int |f|^{n+1} d\mu}{\int |f|^n d\mu}.$$
- A.5.5. Show that a normed vector space $(V, \|\cdot\|)$ is complete if and only if every series $\sum_k v_k$ that is absolutely summable (meaning that $\sum_k \|v_k\|$ converges) is convergent. Use this fact to show that if μ is a probability measure then $\|\cdot\|_p$ is a complete norm on $L^p(\mu)$ for every $1 \leq p \leq \infty$.
- A.5.6. Show that if μ is a finite measure and $1/p + 1/q = 1$ with $1 \leq p < \infty$ then the map $\Phi : L^q(\mu) \rightarrow L^p(\mu)^*$, $\Phi(g)f = \int fg d\mu$ is an isomorphism and an isometry.
- A.5.7. Show that if X is a metric space then, given any Borel probability measure μ , the set $C^0(X)$ of all continuous functions is dense in $L^p(\mu)$ for every $1 \leq p \leq \infty$. Indeed, the same holds for the subset of all uniformly continuous bounded functions.
- A.5.8. Let $f, g : X \rightarrow \mathbb{R}$ be two positive measurable functions such that $f(x)g(x) \geq 1$ for every x . Show that $\int f d\mu \int g d\mu \geq 1$ for every probability measure μ .

A.6 Hilbert spaces

Let H be a vector space, real or complex. An (Hermitian) *inner product* on H is a map $(u, v) \mapsto u \cdot v$ from $H \times H$ to the scalar field (\mathbb{R} or \mathbb{C} , respectively) satisfying: for any $u, v, w \in H$ and any scalar λ ,

1. $(u + w) \cdot v = u \cdot v + w \cdot v$ and $u \cdot (v + w) = u \cdot v + u \cdot w$;
2. $(\lambda u) \cdot v = \lambda(u \cdot v)$ and $u \cdot (\lambda v) = \bar{\lambda}(u \cdot v)$;
3. $u \cdot v = \overline{v \cdot u}$;
4. $u \cdot u \geq 0$ and $u \cdot u = 0$ if and only if $u = 0$.

Then we can define the *norm* of a vector $u \in H$ to be $\|u\| = (u \cdot u)^{1/2}$.

A *Hilbert space* is a vector space endowed with an inner product whose norm $\|\cdot\|$ is complete: relative to $\|\cdot\|$ every Cauchy sequence is convergent. Thus, in particular, $(H, \|\cdot\|)$ is a Banach space. A standard example of a Hilbert space is the space $L^2(\mu)$ of square-integrable functions that we introduced in Appendix A.5.2.

Given $v \in H$ and any family $(v_\alpha)_\alpha$ of vectors of H , we say that $v = \sum_\alpha v_\alpha$ if for every $\varepsilon > 0$ there exists a finite set I such that

$$\left\| v - \sum_{\beta \in I} v_\beta \right\| \leq \varepsilon \quad \text{for every finite set } J \supset I.$$

Given any family $(H_\alpha)_\alpha$ of subspaces of H , the set of all vectors of the form $v = \sum_\alpha v_\alpha$ with $v_\alpha \in H_\alpha$ for every α is a subspace of H (see Exercise A.6.2). It is called the *sum* of the family $(H_\alpha)_\alpha$ and it is denoted by $\sum_\alpha H_\alpha$.

A.6.1 Orthogonality

Let H be a Hilbert space. Two vectors $u, v \in H$ are said to be *orthogonal* if $u \cdot v = 0$. We call a subset of H *orthonormal* if its elements have norm 1 and are pairwise orthogonal.

A *Hilbert basis* of H is an orthonormal subset $B = \{v_\beta\}$ such that the set of all (finite) linear combinations of elements of B is dense in H . For example, the *Fourier basis*

$$\{x \mapsto e^{2\pi ikx} : k \in \mathbb{Z}\} \tag{A.6.1}$$

is a Hilbert basis of the space $L^2(m)$ of all measurable functions on the unit circle whose square is integrable with respect to the Lebesgue measure.

A Hilbert basis $B = \{v_\beta\}$ is usually not a basis of the vector space in the usual sense (Hammel basis): it is usually not true that every vector of H is a finite linear combination of the elements of B . However, every $v \in H$ may be written as an *infinite* linear combination of the elements of the Hilbert basis:

$$v = \sum_{\beta} (v \cdot v_{\beta})v_{\beta} \quad \text{and, moreover,} \quad \|v\|^2 = \sum_{\beta} |v \cdot v_{\beta}|^2.$$

In particular, $v \cdot v_{\beta} = 0$ except, possibly, for a countable subset of values of β .

Every orthonormal subset of H may be extended to a Hilbert basis. In particular, Hilbert bases always exist. Moreover, any two Hilbert bases have the same cardinal, which is called the *Hilbert dimension* of H . The Hilbert dimension depends monotonically on the space: if H_1 is a subspace of H_2 then $\dim H_1 \leq \dim H_2$. We say that two Hilbert spaces are *isometrically isomorphic* if there exists some isomorphism between the two that also preserves the inner product. A necessary and sufficient condition is that the two spaces have the same Hilbert dimension.

A Hilbert space is said to be *separable* if it admits some countable subset that is dense for the topology defined by the norm. This happens if and only if the Hilbert dimension is either finite or countable. In particular, all separable Hilbert spaces with infinite Hilbert dimension are isometrically isomorphic. For this reason, one often finds in the literature (especially in the area of mathematical physics) mentions of *the* Hilbert space, as if there were only one.

Given any family $(H_{\alpha})_{\alpha}$ of Hilbert spaces, we denote by $\bigoplus_{\alpha} H_{\alpha}$ their *orthogonal direct sum*, that is, the vector space of all $(v_{\alpha})_{\alpha} \in \prod_{\alpha} H_{\alpha}$ such that $\sum_{\alpha} \|v_{\alpha}\|_{\alpha}^2 < \infty$ (this implies that $v_{\alpha} = 0$ except, possibly, for a countable set of values of α), endowed with the inner product

$$(v_{\alpha})_{\alpha} \cdot (w_{\alpha})_{\alpha} = \sum_{\alpha} v_{\alpha} \bar{w}_{\alpha}.$$

The *orthogonal complement* of a subset S of a Hilbert space H is the set S^{\perp} of all the vectors of H that are orthogonal to every vector of S . It is easy to see that S^{\perp} is a closed subspace of H (Exercise A.6.7). If S itself is a closed subspace of H then $S = (S^{\perp})^{\perp}$ and every vector $v \in H$ may be decomposed as a

sum $v = s + s^\perp$ of some $s \in S$ and some $s^\perp \in S^\perp$. Moreover, this decomposition is unique and the vectors s and s^\perp are the elements of S and S^\perp , respectively, that are closest to v .

A.6.2 Duality

A *linear functional* on a Hilbert space H (or, more generally, on a Banach space) is a linear map from H to the scalar field (\mathbb{R} or \mathbb{C}). It is said to be *bounded* if

$$\|\phi\| = \sup \left\{ \frac{|\phi(v)|}{\|v\|} : v \neq 0 \right\} < \infty.$$

This is equivalent to saying that the linear functional is continuous, relative to the topology defined by the norm of H (see Exercise A.6.3). The *dual space* of a Hilbert space H is the vector space H^* formed by all the bounded linear functionals. The function $\phi \mapsto \|\phi\|$ is a complete norm on H^* and, hence, it endows the dual with the structure of a Banach space. The map

$$h : H \rightarrow H^*, \quad w \mapsto [v \mapsto v \cdot w] \tag{A.6.2}$$

is a bijection between the two spaces and it preserves the norms. In particular, h is a homeomorphism. Moreover, it satisfies $h(w_1 + w_2) = h(w_1) + h(w_2)$ and $h(\lambda w) = \bar{\lambda}h(w)$.

The *weak* topology in H is the smallest topology relative to which all the linear functionals $v \mapsto v \cdot w$ are continuous. In terms of sequences, it can be characterized as follows:

$$(w_n)_n \rightarrow w \text{ weakly} \iff (v \cdot w_n)_n \rightarrow v \cdot w \text{ for every } v \in H.$$

The *weak** topology in the dual space H^* is the smallest topology relative to which $\phi \mapsto \phi(v)$ is continuous for every $v \in H$.

It is known from the theory of Banach spaces (theorem of Banach–Alaoglu) that every bounded closed subset of the dual space is compact for the weak* topology. In the special case of Hilbert spaces, the weak topology in the space H is homeomorphic to the weak* topology in the dual space H^* : the map h in (A.6.2) is also a homeomorphism for these topologies. Since h preserves the class of bounded sets, it follows that the weak topology in the space H itself enjoys the property in the theorem of Banach–Alaoglu:

Theorem A.6.1 (Banach–Alaoglu). *Every bounded closed subset of a Hilbert space H is compact for the weak topology in H .*

A linear operator $L : H_1 \rightarrow H_2$ between two Hilbert spaces is *continuous* (or *bounded*) if

$$\|L\| = \sup \left\{ \frac{|L(v)|}{\|v\|} : v \neq 0 \right\}$$

is finite. The *adjoint* of a continuous linear operator is the linear operator $L^* : H_2 \rightarrow H_1$ defined by

$$v \cdot Lw = L^*v \cdot w \quad \text{for every } v, w \in H.$$

The adjoint operator is continuous, with $\|L^*\| = \|L\|$ and $\|L^*L\| = \|LL^*\| = \|L\|^2$. Moreover, $(L^*)^* = L$ and $(L_1 + L_2)^* = L_1^* + L_2^*$ and $(\lambda L)^* = \bar{\lambda}L^*$ (in Exercise A.6.5 we invite the reader to prove these facts).

A continuous linear operator $L : H \rightarrow H$ is *self-adjoint* if $L = L^*$. More generally, L is *normal* if it satisfies $L^*L = LL^*$. We are especially interested in the case when L is *unitary*, that is, $L^*L = \text{id} = LL^*$. We call *linear isometry* to every linear operator $L : H \rightarrow H$ such that $L^*L = \text{id}$. Hence, the unitary operators are the linear isometries that are also normal operators.

A.6.3 Exercises

A.6.1. Let H be a Hilbert space. Prove:

- (a) That every ball (either open or closed) is a convex subset of H .
- (b) The *parallelogram identity*: $\|v + w\|^2 + \|v - w\|^2 = \|v\|^2 + \|w\|^2$ for any $v, w \in H$.
- (c) The *polarization identity*: $4(v \cdot w) = \|v + w\|^2 - \|v - w\|^2$ (real case) or $4(v \cdot w) = (\|v + w\|^2 - \|v - w\|^2) + i(\|v + iw\|^2 - \|v - iw\|^2)$ (complex case).

A.6.2. Show that, given any family $(H_\alpha)_\alpha$ of subspaces of a Hilbert space H , the set of all the vectors of the form $v = \sum_\alpha v_\alpha$ with $v_\alpha \in H_\alpha$ for every α is a vector subspace of H .

A.6.3. Show that a linear operator $L : E_1 \rightarrow E_2$ between two Banach spaces is continuous if and only if there exists $C > 0$ such that $\|L(v)\|_2 \leq C\|v\|_1$ for every $v \in E_1$, where $\|\cdot\|_i$ denotes the norm in the space E_i (we say that L is a *bounded operator*).

A.6.4. Consider the Hilbert space $L^2(\mu)$. Let V be the subspace formed by the constant functions. What is the orthogonal complement of V ? Determine the (orthogonal) projection to V of an arbitrary function $g \in L^2(\mu)$.

A.6.5. Prove that if $L : H \rightarrow H$ is a bounded operator on a Hilbert space H then the adjoint operator L^* is also bounded and $\|L^*\| = \|L\|$ and $\|L^*L\| = \|LL^*\| = \|L\|^2$ and $(L^*)^* = L$.

A.6.6. Show that if K is a closed convex subset of a Hilbert space then for every $z \in H$ there exists a unique $v \in K$ such that $\|z - v\| = d(z, K)$.

A.6.7. Let S be a subspace of a Hilbert space H . Prove that:

- (a) The orthogonal complement S^\perp of S is a closed subspace of H and it coincides with the orthogonal complement of the closure \bar{S} . Moreover, $(S^\perp)^\perp = \bar{S}$.
- (b) Every $v \in H$ may be written, in a unique fashion, as a sum $v = s + s^\perp$ of some $s \in \bar{S}$ and some $s^\perp \in S^\perp$. The two vectors s and s^\perp are the elements of S and S^\perp that are closest to v .

- A.6.8. Let E be a closed subspace of a Hilbert space H . Show that E is also closed in the weak topology. Moreover, $U(E)$ is a closed subspace of H , for every isometry $U : H \rightarrow H$.
- A.6.9. Show that a linear operator $L : H \rightarrow H$ on a Hilbert space H is an isometry if and only if $\|L(v)\| = \|v\|$ for every $v \in H$. Moreover, L is a unitary operator if and only if L is an isometry and is invertible.

A.7 Spectral theorems

Let H be a complex Hilbert space. The *spectrum* of a continuous linear operator $L : H \rightarrow H$ is the set $\text{spec}(L)$ of all numbers $\lambda \in \mathbb{C}$ such that $L - \lambda \text{id}$ is not an isomorphism. The spectrum is closed and it is contained in the closed disk of radius $\|L\|$ around $0 \in \mathbb{C}$. In particular, $\text{spec}(L)$ is a compact subset of the complex plane. When H has finite dimension, $\text{spec}(L)$ consists of the eigenvalues of L , that is, the complex numbers λ such that $L - \lambda \text{id}$ is not injective. In general, the spectrum is strictly larger than the set of eigenvalues (see Exercise A.7.2).

A.7.1 Spectral measures

By definition, a *projection* in H is a continuous linear operator $P : H \rightarrow H$ that is idempotent ($P^2 = P$) and self-adjoint ($P^* = P$). Then the image and the kernel of P are closed subspaces of H and they are orthogonal complements to each other. In fact, the image coincides with the set of all fixed points of P .

Consider any map E associating with each measurable subset of the plane \mathbb{C} a projection in H . Such a map is called a *spectral measure* if it satisfies $E(\mathbb{C}) = \text{id}$ and

$$E\left(\bigcup_{n \in \mathbb{N}} B_n\right) = \sum_{n \in \mathbb{N}} E(B_n)$$

whenever the B_n are pairwise disjoint (σ -additivity). Then, given any $v, w \in H$, the function

$$Ev \cdot w : B \mapsto E(B)v \cdot w \tag{A.7.1}$$

is a complex measure in \mathbb{C} . Clearly, it depends on the pair (v, w) in a bilinear fashion.

We call the *support* of a spectral measure E the set $\text{supp} E$ of all the points $z \in \mathbb{C}$ such that $E(V) \neq 0$ for every neighborhood V of z . Note that the support is always a closed set. Moreover, the support of the complex measure $Ev \cdot w$ is contained in $\text{supp} E$ for every $v, w \in H$.

Example A.7.1. Consider $\{\lambda_1, \dots, \lambda_s\} \subset \mathbb{C}$ and let V_1, \dots, V_s be a finite family of subspaces of \mathbb{C}^d , pairwise orthogonal and such that $\mathbb{C}^d = V_1 \oplus \dots \oplus V_s$. For each set $J \subset \{1, \dots, s\}$, denote by P_J the projection in \mathbb{C}^d whose image is

$\bigoplus_{j \in J} V_j$. For each measurable set $B \subset \mathbb{C}$ define

$$E(B) : \mathbb{C}^d \rightarrow \mathbb{C}^d, \quad E(B) = P_{J(B)},$$

where $J(B)$ is the set of all $j \in \{1, \dots, s\}$ such that $\lambda_j \in B$. The function E is a spectral measure.

Example A.7.2. Let μ be a probability measure in \mathbb{C} and $H = L^2(\mu)$ be the space of all complex functions whose square is integrable with respect to μ . For each measurable set $B \subset \mathbb{C}$, let

$$E(B) : L^2(\mu) \rightarrow L^2(\mu), \quad \varphi \mapsto \chi_B \varphi.$$

Each $E(B)$ is a projection and the function E is a spectral measure.

The next lemma collects a few simple properties of the spectral measures:

Lemma A.7.3. *Let E be a spectral measure and A, B be measurable subsets of \mathbb{C} . Then:*

1. $E(\emptyset) = 0$ and $E(\text{supp } E) = \text{id}$;
2. if $A \subset B$ then $E(A) \leq E(B)$ and $E(B \setminus A) = E(B) - E(A)$;
3. $E(A \cup B) + E(A \cap B) = E(A) + E(B)$;
4. $E(A)E(B) = E(A \cap B) = E(B)E(A)$.

In what follows we always assume that E is a spectral measure with compact support. Then the support of every complex measure $E\nu \cdot w$ is also compact. Consequently, the integral $\int z d(E(z)\nu \cdot w)$ is well defined and it is a bilinear function of (ν, w) . Hence, there exists a bounded linear operator $L : H \rightarrow H$ such that

$$L\nu \cdot w = \int z d(E(z)\nu \cdot w) \quad \text{for every } \nu, w \in H. \tag{A.7.2}$$

We write, in shorter form:

$$L = \int z dE(z). \tag{A.7.3}$$

More generally, given any bounded measurable function ψ in the support of the spectral measure E , there exists a bounded linear operator $\psi(L) : H \rightarrow H$ that is characterized by

$$\psi(L)\nu \cdot w = \int \psi(z) d(E(z)\nu \cdot w) \quad \text{for every } \nu, w \in H. \tag{A.7.4}$$

We write

$$\psi(L) = \int \psi(z) dE(z). \tag{A.7.5}$$

Lemma A.7.4. *Let E be a spectral measure with compact support. Given bounded measurable functions φ, ψ and numbers $\alpha, \beta \in \mathbb{C}$,*

- (1) $\int (\alpha\varphi + \beta\psi)(z) dE(z) = \alpha \int \varphi dE(z) + \beta \int \psi dE(z)$;
- (2) $\int \bar{\varphi}(z) dE(z) = \left(\int \varphi(z) dE(z) \right)^*$;

$$(3) \int (\varphi \psi)(z) dE(z) = \left(\int \varphi(z) dE(z) \right) \circ \left(\int \psi(z) dE(z) \right).$$

In particular, by part (3) of this lemma,

$$L^j = \left(\int z dE(z) \right)^j = \int z^j dE(z) \quad \text{for every } j \in \mathbb{N}. \tag{A.7.6}$$

Analogously, using also part (2) of the lemma,

$$\begin{aligned} LL^* &= \left(\int z dE(z) \right) \left(\int \bar{z} dE(z) \right) = \int |z|^2 dE(z) \\ &= \left(\int \bar{z} dE(z) \right) \left(\int z dE(z) \right) = L^*L. \end{aligned} \tag{A.7.7}$$

Consequently, the linear operator defined by (A.7.3) is normal. Conversely, the spectral theorem asserts that every normal operator may be written in this way:

Theorem A.7.5 (Spectral). *For every normal operator $L : H \rightarrow H$ there exists a spectral measure E such that $L = \int z dE(z)$. This measure is unique and its support coincides with the spectrum of L . In particular, L is unitary if and only if $\text{supp } E$ is contained in the unit circle $\{z \in \mathbb{C} : |z| = 1\}$.*

Example A.7.6 (Spectral theorem in finite dimension). Let H be a complex Hilbert space with finite dimension. Then for every normal operator $L : H \rightarrow H$ there exists a basis of H formed by eigenvectors of L . Let $\lambda_1, \dots, \lambda_s$ be the eigenvalues of L . The eigenspaces $V_j = \ker(L - \lambda_j \text{id})$ are pairwise orthogonal, because L is normal. Moreover, by Theorem A.7.5, the direct sum $\bigoplus_{j=1}^s V_j$ is the whole of H . So

$$L = \sum_{j=1}^s \lambda_j \pi_j$$

where π_j denotes the orthogonal projection to V_j . In other words, the spectral measure E of the operator L is given by $E(\{\lambda_j\}) = \pi_j$ for every $j = 1, \dots, s$ and $E(B) = 0$ if B contains no eigenvalue of L .

Example A.7.7. Let $(\sigma_\alpha)_{\alpha \in \mathcal{A}}$ be any family of finite measures in the unit circle $\{z \in \mathbb{C} : |z| = 1\}$. Consider $H = \bigoplus_{\alpha \in \mathcal{A}} L^2(\sigma_\alpha)$ and the linear operator

$$L : H \rightarrow H, \quad (\varphi_\alpha)_\alpha \mapsto (z \mapsto z\varphi_\alpha(z))_\alpha.$$

Consider the spectral measure E given by

$$E(B) : H \rightarrow H, \quad (\varphi_\alpha)_\alpha \mapsto (\mathcal{X}_B \varphi_\alpha)_\alpha$$

(compare with Example A.7.2). Then, $L = \int z dE(z)$. Indeed, the definition of E gives that $E\varphi \cdot \psi = \sum_\alpha \varphi_\alpha \bar{\psi}_\alpha \sigma_\alpha$ for every $\varphi = (\varphi_\alpha)_\alpha$ and $\psi = (\psi_\alpha)_\alpha$ in the space H . Then,

$$L\varphi \cdot \psi = \sum_\alpha \int z\varphi_\alpha(z) \bar{\psi}_\alpha(z) d\sigma_\alpha(z) = \int z d(E(z)\varphi \cdot \psi) \tag{A.7.8}$$

for every φ, ψ .

We say that $\lambda \in \mathbb{C}$ is an *atom* of the spectral measure if $E(\{\lambda\}) \neq 0$ or, equivalently, if there exists some non-zero vector $\omega \in H$ such that $E(\{\lambda\})\omega \neq 0$. The proof of the next proposition is outlined in Exercise A.7.4.

Proposition A.7.8. *Every eigenvalue of L is an atom of the spectral measure E . Conversely, if λ is an atom of E then λ is an eigenvalue of the operator L and every non-zero vector of the form $v = E(\{\lambda\})\omega$ is an eigenvector.*

A.7.2 Spectral representation

Theorem A.7.5 shows that normal linear operators on a Hilbert space are essentially the same thing as spectral measures in that space. Theorems of this type, establishing a kind of dictionary between two classes of objects that a priori do not seem to be related, are among the most fascinating results in mathematics. Of course, just how useful such a dictionary is to study one of those classes (normal linear operators, say) depends on to what extent we are capable of understanding the other one (spectral measures, in this case). In the present situation this is handled, in a most satisfactory way, by the next result, which exhibits a canonical form (inspired by Example A.7.2) in which every normal linear operator may be written.

As before, we use \oplus to denote the orthogonal direct sum of Hilbert spaces. Given any cardinal χ , finite or infinite, and a Hilbert space V , we denote by V^χ the orthogonal direct sum of χ copies of V .

Theorem A.7.9 (Spectral representation). *Let $L : H \rightarrow H$ be a normal linear operator. Then there exist mutually singular finite measures $(\sigma_j)_j$ with support in the spectrum of L , there exist cardinals $(\chi_j)_j$ and there exists a unitary operator $U : H \rightarrow \bigoplus_j L^2(\sigma_j)^{\chi_j}$, such that the conjugate $ULU^{-1} = T$ is given by:*

$$T : \bigoplus_j L^2(\sigma_j)^{\chi_j} \rightarrow \bigoplus_j L^2(\sigma_j)^{\chi_j}, \quad (\varphi_{j,l})_{j,l} \mapsto (z \mapsto z\varphi_{j,l}(z))_{j,l}. \tag{A.7.9}$$

We call (A.7.9) the *spectral representation* of the normal operator L . Let us point out that the measures σ_j in Theorem A.7.9 are not uniquely determined. However, the spectral representation is unique, in the following sense.

Call the *multiplicity function* of the operator L the function associating with each finite measure θ in \mathbb{C} the smallest cardinal χ_θ such that the measures θ and σ_j are not mutually singular. One can prove that this function is uniquely determined by the operator L , that is, it does not depend on the choice of the measures σ_j in the statement. Moreover, two normal operators are conjugate by some unitary operator if and only if they have the same multiplicity function.

Example A.7.10 (Spectral representation in finite dimension). Let us go back to the setting of Example A.7.6. For each $j = 1, \dots, s$, let σ_j be the Dirac mass at the eigenvalue λ_j and χ_j be the dimension of the eigenspace V_j . Note that the space $L^2(\sigma_j)$ has dimension 1. Hence, we may choose a unitary operator

$U_j : V_j \rightarrow L^2(\sigma_j)$, for each $j = 1, \dots, s$. Since $L = \lambda_j \text{id}$ restricted to V_j , we have that $T_j = U_j L U_j^{-1} = \lambda_j \text{id}$, that is,

$$T_j : ((\varphi_\alpha)_\alpha) \mapsto (z \mapsto \lambda_j \varphi_\alpha(z))_\alpha = (z \mapsto z \varphi_\alpha(z))_\alpha.$$

In this way, we have found a unitary operator

$$U : \mathbb{C}^d \rightarrow \bigoplus_{j=1}^s L^2(\sigma_j)^{\otimes j}$$

such that $T = U L U^{-1}$ is a spectral representation of L .

A.7.3 Exercises

- A.7.1. Let $T : E \rightarrow E$ be a Banach space isomorphism, that is, a continuous linear bijection whose inverse is also continuous. Show that $T + H$ is a Banach space isomorphism for every continuous linear map $H : E \rightarrow E$ such that $\|H\| \|T^{-1}\| < 1$. Use this fact to prove that the spectrum of every continuous linear operator $L : E \rightarrow E$ is a closed set and is contained in the closed disk of radius $\|L\|$ around the origin.
- A.7.2. Show that if $L : H \rightarrow H$ is a linear operator in a Hilbert space H with finite dimension then $\text{spec}(L)$ consists of the eigenvalues of L , that is, the complex numbers λ for which $L - \lambda \text{id}$ is not injective. Give an example, in infinite dimension, such that the spectrum is strictly larger than the set of eigenvalues.
- A.7.3. Prove Lemma A.7.3.
- A.7.4. Prove Proposition A.7.8, along the following lines:

(a) Assume that $Lv = \lambda v$ for some $v \neq 0$. Consider the functions

$$\varphi_n(z) = \begin{cases} (z - \lambda)^{-1} & \text{if } |z - \lambda| > 1/n \\ 0 & \text{otherwise.} \end{cases}$$

Show that $\varphi_n(L)(L - \lambda \text{id}) = E(\{|z - \lambda| > 1/n\})$ for every n . Conclude that $E(\{\lambda\})v = v$ and, consequently, λ is an atom of E .

(b) Assume that there exists $w \in H$ such that $v = E(\{\lambda\})w$ is non-zero. Show that, given any measurable set $B \subset \mathbb{C}$,

$$E(B)v = \begin{cases} v & \text{if } \lambda \in B \\ 0 & \text{if } \lambda \notin B. \end{cases}$$

Conclude that $Lv = \lambda v$ and, consequently, λ is an eigenvalue of L .

- A.7.5. Let $(\sigma_j)_j$ be the family of measures given in Theorem A.7.9. Given any measurable set $B \subset \mathbb{C}$, check that $E(B) = 0$ if and only if $\sigma_j(B) = 0$ for every j . Therefore, given any measure η in \mathbb{C} , we have that $E \ll \eta$ if and only if $\sigma_j \ll \eta$ for every j .