Real analysis

Emilio Corso

Department of Mathematics, Pennsylvania State University, 203 McAllister Building, State College, PA 16801, U.S.
Email address: ebc5504@psu.edu

Contents

Notation and terminology 4

Chapter 1. Measures 7
1.1. A motivating discussion: the limitations of Riemann integration 7
1.2. Pitfalls of a naive approach: the paradoxes of Vitali and Banach-Tarski 9
1.3. Measurable spaces and maps 11
1.4. Measures 16
1.5. Outer measures and Carathéodory's construction 22
1.6. The Lebesgue measure on $\mathbb{R}^n$ 32
1.7. Lebesgue-Stieltjes measures on the real line 37
1.8. Metric outer measures and Hausdorff measures 38

Chapter 2. Integration 43
2.1. Measurable functions with values in number systems 43
2.2. Integral of simple functions 44
2.3. Integral of positive functions 47
2.4. Integral of real, extended-real, and complex-valued functions 50
2.5. The integrals of Riemann and Lebesgue on the real line 61
2.6. Measure-theoretic notions of convergence 62
2.7. Some further constructions of measures 67
2.8. Product measures and Fubini's theorem 68
2.9. Change of variable formula for Lebesgue integrals on Euclidean spaces 74

Chapter 3. L^p spaces 77
3.1. Functional-analytic preliminaries 77
3.2. The inequalities of Hölder and Minkowski 84
3.3. Relations between L^p spaces and some useful inequalities 89
3.4. Convolutions and regularization 92

Appendix A. Topology 99
A.1. Topological and metric spaces 99
A.2. Functions between metric spaces 102
A.3. Various topological properties 102
A.4. Convergence of functions 103

Appendix B. Infinite sums 105
B.1. Sums of families of real numbers 105
B.2. Sums of extended real numbers 106

Appendix. Bibliography 107

Notation and terminology

If X is a set, $\mathcal{P}(X)$ denotes the power set of X, namely the set of all subsets of X. If $E \in \mathcal{P}(X)$, then E^c denotes the complement of E in X, that is, the set-theoretic difference $X \setminus E$. The identity map $X \to X$, $x \mapsto x$ is denoted by id_X.

If X is a set and E is a subset of X, we denote by χ_E the characteristic function, also known as indicator function, of E, defined by

$$\chi_E = \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{if } x \notin E \end{cases} .$$

A family $(X_\alpha)_{\alpha\in\mathsf{A}}$ of sets is said to consist of pairwise disjoint sets if $X_\alpha \cap X_\beta = \emptyset$ for all $\alpha \neq \beta \in J$. The union of such a collection $(X_\alpha)_{\alpha\in\mathsf{A}}$ is usually denoted with the symbol

$$\bigsqcup_{\alpha\in\mathsf{A}} X_\alpha .$$

If X is a set and Y is a subset of X, a covering of Y is a family $(Y_\alpha)_{\alpha\in\mathsf{A}}$ of subsets $Y_\alpha \subset X$ such that $Y \subset \bigcup_{\alpha\in\mathsf{A}} Y_\alpha$. A partition of X is a covering $(X_\alpha)_{\alpha\in\mathsf{A}}$ of X consisting of pairwise disjoint sets.

If E is a finite set, we let $|E|$ denote its cardinality; we also declare $|E| = \infty$ if E is an infinite set, making to distinction between the various infinite cardinals.

A set X is countable if there is a bijection $f\colon X \to I$ where I is a subset of $\mathbb{N}$.

The following are the most important sets of numbers we will consider in these notes: $\mathbb{N} = \{0, 1, 2, \dots\}$ is the set of natural numbers, $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$, $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots,\}$ set of integers, $\mathbb{Q}$ the set of rational numbers, $\mathbb{R}$ the set of real numbers and $\mathbb{C}$ is the set of complex numbers. We indicate with $\mathbb{R}_{\geq 0}$ (resp. $\mathbb{R}_{>0}$) the set of positive (resp. strictly positive) real numbers.

If $(X, \leq_X)$ and $(Y, \leq_Y)$ are ordered sets, a function $f\colon X \to Y$ is increasing if, for all $x_1, x_2 \in X$, $x_1 \leq_X x_2$ implies $f(x_1) \leq_Y f(x_2)$. It is strictly increasing if, for all $x_1 \neq x_2 \in X$, $x_1 \leq_X x_2$ implies $f(x_1) \leq_Y f(x_2)$ and $f_{x_1} \neq f(x_2)$.

Unless explicit mention to the contrary, we adhere to the convention

$$\inf \emptyset = +\infty , \quad \sup \emptyset = 0 .$$

If $(G, +)$ is an abelian group and A, B are subsets of G, the sumset of A and B is

$$A + B = \{a + b : a \in A,\ b \in B\} ,$$

if $A = \{a\}$, we employ the notation $a + B$ in place of $\{a\} + B$. If V is a real (resp. complex) vector space, $A \subset V$ and $\lambda \in \mathbb{R}$ (resp. $\lambda \in \mathbb{C}$), we write λA for the set

$$\{\lambda x : x \in A\} .$$

The origin of $\mathbb{R}^n$ is always denoted by 0, without explicit mention of the dimension, as no confusion shall arise.

The extended real number system. We define the extended real number system as the set

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$$

obtained by adding to the set $\mathbb{R}$ of real numbers two new elements, denoted $-\infty$ and $+\infty$. We extend the total order on $\mathbb{R}$ in the natural way by declaring that

$$x \leq +\infty , \quad -\infty \leq x$$

for all $x \in \mathbb{R}$. We extend further the operations of addition and multiplication of real numbers in the following manner:

$$\pm\infty + x = \pm\infty \quad \text{for all } x \in \overline{\mathbb{R}},\ x \neq \mp\infty,$$

$$\pm\infty \cdot x = \pm\infty \quad \text{for all } x \in \overline{\mathbb{R}},\ x > 0,$$

$$\pm\infty \cdot x = \mp\infty \quad \text{for all } x \in \overline{\mathbb{R}},\ x < 0,$$
$$\pm\infty \cdot 0 = 0\ .$$

The sums $+\infty + (-\infty)$, $-\infty + (+\infty)$ are not defined.

We will often consider only the extended positive half-line $[0, +\infty]$, also indicated with $[0, \infty]$ as no confusion arises, where the interval notation $[a, b]$ is naturally extended to all $a, b\overline{\mathbb{R}}$ with the meaning

$$[a, b] = \{x \in \overline{\mathbb{R}} : a \leq x \leq b\}\ ;$$

similarly for the notation $[a, b)$ and $(a, b]$.

CHAPTER 1

Measures

The purpose of these notes is to provide a sufficiently extensive treatment of *measure and integration theory*, which play an indispensable role in modern real analysis, while simultaneously serving as the theoretical foundations of modern probability theory.

The theory, as is nowadays commonly presented, has been fully developed during the early decades of the twentieth century, the original driving impetus for it being the endeavour to devise a theory of integration which overcomes the limitations of Riemann's theory of integration, established in the second half of the nineteenth century, and which is still the one routinely introduced in the first Calculus courses. While still retaining the geometric flavour which informs Riemann's theory, this more novel theory of integration goes a step further in abstracting the axiomatic properties a satisfactory notion of integral should possess, thereby extending its scope much beyond the integration of functions defined on Euclidean spaces, which is the traditional purview of Riemann's theory.

1.1. A motivating discussion: the limitations of Riemann integration

Let us start by discussing some of the serious drawbacks of the Riemann integral, of which for convenience we recall the definition[1] on the real line. If $[a, b]$ is a closed interval in $\mathbb{R}$, a function $f\colon [a, b] \to \mathbb{R}$ is said to be Riemann-integrable if the infimum of the upper Riemann sums

$$\sum_{i=1}^{n} \Big(\sup_{a_{i-1} \le x \le a_i} f(x) \Big) (a_i - a_{i-1}) \tag{1.1.1}$$

equals the supremum of the lower Riemann sums

$$\sum_{i=1}^{n} \Big(\inf_{a_{i-1} \le x \le a_i} f(x) \Big) (a_i - a_{i-1}) \ , \tag{1.1.2}$$

both of which vary over all possible subdivisions

$$a = a_0 < a_1 < \cdots < a_{n-1} < a_n = b$$

of the interval $[a, b]$. In this case, the common value is called the Riemann integral of f over $[a, b]$, denoted

$$\int_a^b f(x)\, \mathrm{d}x \ .$$

In case $f \ge 0$, then the integral is a natural measure of the area of the plane region comprised between the x-axis and the graph of the function f.

Here are a number of natural features a good notion of integral should have, which the Riemann integral does not afford.

(1) Whereas it's a classical calculus fact that every continuous function $f\colon [a, b] \to \mathbb{R}$ is Riemann-integrable, there are many elementary instances of discontinuous functions

[1] We shall come back to it in §**??**, when we shall explicitly draw comparisons between the Riemann and the *Lebesgue* integral for functions of one real variable.

which are not Riemann-integrable. Perhaps the most standard example is the characteristic function of $\mathbb{Q} \cap [0,1]$, namely the function $f\colon [0,1] \to \mathbb{R}$ defined as

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q} \end{cases} .$$

Since both $\mathbb{Q} \cap [0,1]$ and $(\mathbb{R} \setminus \mathbb{Q}) \cap [0,1]$ are dense in $[0,1]$ (see §**??** for a recap of basic point-set topology), it is clear that the value of every upper Riemann sum for f is 1, while the value of every lower Riemann sum is 0.

The region enclosed by the x-axis and the graph of f is a countable union of segments of length 1; it is natural to consider every such set as having vanishing area in the plane; yet Riemann's theory of integration does not enable us to assign a well defined meaning for such an area.

(2) It is highly desirable, in a multitude of circumstances, to be able to exchange the fundamental analytical operations of limit and integral. More precisely, suppose for instance $(f_n)_{n\geq 0}$ is a sequence of continuous (thus Riemann-integrable) functions $f_n\colon [0,1] \to \mathbb{R}$ converging pointwise to a function $f\colon [0,1] \to \mathbb{R}$, that is, satisfying

$$\lim_{n\to\infty} f_n(x) = f(x)$$

for all $x \in [0,1]$. The equality

$$\int_a^b \lim_{n\to\infty} f_n(x)\, \mathrm{d}x = \lim_{n\to\infty} \int_a^b f_n(x)\, \mathrm{d}x \tag{1.1.3}$$

is not justified, for at least two fundamental reasons. To begin with, a pointwise limit of continuous functions is not necessarily continuous, and can even fail to be Riemann-integrable; thus the integral of the left-hand side of the last displayed equation may not make sense at all, in the Riemann theory. Secondly, even if f is continuous, its integral does not necessarily coincide with the limit of the integrals of the f_n. As an example, take f_n to be the positive function whose graph is the linear interpolation between the points $(0,0), (1/n, n)(2/n, 0)$ and $(1,0)$ on the plane. Then a moment's thought reveals that $(f_n)_{n\geq 0}$ converges pointwise to $f = 0$, and yet the area under the graph of f_n equals 1 for all n.

In order to ensure the validity of exchanging limits and integrals, as in (1.1.3), in the realm of Riemann integration, a very strong notion of functional convergence is required, namely uniform convergence:

$$\sup_{o\leq x\leq 1} |f(x) - f_n(x)| \stackrel{n\to\infty}{\longrightarrow} 0 .$$

It would be convenient to have at our disposal a more flexible notion of integral, which does not necessitate such a strong form of convergence for functions in order for (1.1.3) to be justified.

(3) Riemann integration does not offer a unified treatment of integrals over bounded and unbounded integrals. It starts by defining the integral over compact interval, and then relies fundamentally on such a notion to extend the definition to unbounded domains, via a limiting procedure. Lebesgue's theory, on the other hand, when specified to the real line, dispenses completely with the need to consider bounded intervals as a first building block, and treates every domain of integration[2] on equal footing.

(4) The limitations of Riemann's theory become even more apparent in higher dimensions, namely for integrals of functions defined over subsets of $\mathbb{R}^n$. By way of example, let $f(x, y)$ be a function defined over a rectangle $[a, b] \times [c, d] \subset \mathbb{R}^2$. Assuming all subsequent integrals are defined, it is natural to expect, at least when f satisfies some

[2] As we shall see, every *not too pathological* domain of integration.

minimal regularity assumptions, to be able to "exchange order of integration" and affirm that

$$\int_a^b \left(\int_c^d f(x,y)\, \mathrm{d}y \right) \mathrm{d}x = \int_c^d \left(\int_a^b f(x,y)\, \mathrm{d}x \right) \mathrm{d}y \, .$$

It turns out that the assumptions guaranteeing the validity of such an equality are so cumbersome to state, withing the realm of Riemann's theory, as to be virtually impractical.

The issues just presented call for a novel, more flexible (and potentially more widely applicable) theory of integration. Lebesgue's insight, for functions of a single real variable, was to stick to the geometric foundations of Riemann's definition of integrability (after all, for a sufficiently nice function $f\colon [a,b] \to \mathbb{R}_{\geq 0}$, the integral $\int_a^b f(x)\, \mathrm{d}x$ should always represent the area of the region lying below the graph of f), but to allow for more general "building blocks" of integration than closed intervals; in other words, to replace Riemann sums as in (1.1.1) and (1.1.2) by more general finite sums

$$\sum_{k=1}^{n} f(c_k)\, m(I_k)$$

where $(I_k)_{1\leq k\leq n}$ is a partition of the interval $[a,b]$, $c_k \in I_k$ for all k, and $m(I_k)$ is an appropriate measure of the size of I_k which extends the length $m([c,d]) = d-c$ for intervals. The function m should be subject to natural minimal properties, such as translation-invariance and additivity over finite collections of pairwise disjoint sets. Now, the latter property is not only satisfied by any reasonable notion of length for subsets of $\mathbb{R}$, but is shared by many other quantities of disparate origins. In physics, any distribution of matter in Euclidean spaces $\mathbb{R}^n$ should verify a similar requirement; so should any function measuring the probability of occurence of a certain event in a given random phenomenon. This motivates the abstract, axiomatic approach we shall develop starting with §1.3.

1.2. Pitfalls of a naive approach: the paradoxes of Vitali and Banach-Tarski

Let us go back to Lebesgue's quest of extending the interval-length function $m([c,d]) = d-c$ to a size function for arbitrary subsets of $\mathbb{R}$. The purpose of this section is to show that a lot of care is needed in such an endeavour.

We start with a classic construction, showing that the aforementioned, desirable properties of additivity over pairwise disjoint collections and translation invariance are mutually inconsistent.

Proposition 1.2.1 (Vitali's counterexample). *There is no function $\mu\colon \mathcal{P}(\mathbb{R}) \to [0,\infty]$ satisfying the following properties simultaneously:*

(1) for every countable collection $(E_i)_{i\in I}$ of pairwise disjoint subset of $\mathbb{R}$,

$$\mu\Big(\bigcup_{i\in I} E_i\Big) = \sum_{i\in I} \mu(E_i) \, ;$$

(2) for every subset $E \subset \mathbb{R}$ and every $x \in \mathbb{R}$,

$$\mu(E+x) = \mu(E) \, ;$$

(3) $\mu([0,1)) = 1$.

The argument presented here is well known in the subject, and originally due to the Italian mathematician Giuseppe Vitali. It relies on the following fundamental (and controversial) principle of set theory.

Axiom 1.2.2 (The Axiom of Choice). *Let $(X_i)_{i\in I}$ be a family of non-empty sets indexed by a non-empty set I. There exists a function $f\colon I \to \bigcup_{i\in I} X_i$ with the property that $f(i) \in X_i$ for all $i \in I$.*

Formulated differently, the Axiom of Choice asserts that the Cartesian product of a non-empty family of non-empty sets is non-empty. Indeed, the Cartesian product $\prod_{i\in I} X_i$ of a family $(X_i)_{i\in I}$ is precisely defined as the set

$$\left\{f\colon I \to \bigcup_{i\in I} X_i : f(i) \in X_i \text{ for all } i \in I\right\} .$$

Despite its intuitive appearance, the Axiom of Choice is not a consequence of all other axioms of the Zermelo-Fraenkel set theory[3]. In fact, it is equivalent to other two fundamental principles of set theory, Zorn's Lemma and the Well Ordering Principle; a classic reference for the axiomatic approach to naive set theory is Halmos' textbook [10], a particularly useful condensed version of which, useful for the entirety of our purposes, can be found in [9, Chapter 0].

PROOF. We introduce an equivalence relation $\sim$ on $\mathbb{R}$ declaring that $x \sim y$ if and only if $x-y \in \mathbb{Q}$. Since $\mathbb{Q}$ is an additive subgroup of $\mathbb{R}$, it is straightforward to verify that $\sim$ is reflexive, symmetric and transitive, and thus indeed an equivalence relation on $\mathbb{R}$[4]. The Axiom of Choice enables us to define a function $f\colon \mathbb{R}/\sim \to \mathbb{R}$ such that the image $N = f(\mathbb{R}/\sim)$ contains exactly one representative of each equivalence class $[x]_\sim \in \mathbb{R}/\sim$. Since each such equivalence class contains a representative in $[0,1)$, as is easily seen by considering the fractional part of any given $x \in \mathbb{R}$, we may harmlessly assume that $N \subset [0,1)$.

By definition of N, the real line can be written as the following countable disjoint union of translates of N:

$$\mathbb{R} = \bigsqcup_{q\in\mathbb{Q}} N + q \; . \tag{1.2.1}$$

For the purpose of a contradiction, suppose there is a function $\mu\colon \mathcal{P}(\mathbb{R}) \to [0,\infty]$ with all the properties claimed in the statement of the proposition. Call c the value $\mu(N)$; by the second property of μ,

$$\mu(N+q) = c$$

for all $q \in \mathbb{Q}$. There are now two possibilities.

(1) Suppose $c = 0$, then the first property of μ and (1.2.1) give

$$\mu(\mathbb{R}) = \sum_{q\in\mathbb{Q}} \mu(N+q) = \sum_{q\in\mathbb{Q}} 0 = 0 \; ;$$

on the other hand, the third property of μ, in conjunction with the third one, yields

$$\mu(\mathbb{R}) = \mu([0,1)) + \mu(\mathbb{R} \setminus [0,1)) \geq \mu([0,1)) = 1 \; .$$

The last two displayed inequalities are clearly in contradiction with each other.

(2) Suppose $c > 0$. Then clearly

$$[0,2) = \bigsqcup_{q\in\mathbb{Q}\cap[0,1]} N + q \; .$$

On the one hand we have

$$\mu([0,2)) = \sum_{q\in\mathbb{Q}\cap[0,1]} \mu(N+1) = \sum_{q\in\mathbb{Q}\cap[0,1]} c = \infty \; ,$$

by the first property of μ. On the other, the very same property together with the third one gives

$$\mu([0,2)) = \mu([0,1)) + \mu([1,2)) = 1 + 1 = 2 \; ,$$

[3] We hasten to observe here that the finitary version of the axiom, namely the statement for a finite index set I, *can* be deduced from the other axioms, as opposed to the version for arbitrary I.

[4] With the induced operation of addition, the quotient set is nothing but the quotient abelian group $\mathbb{R}/\mathbb{Q}$.

where the second equality follows from the second property of μ and the fact that $[1,2)$ is a translate of $[0,1)$. Once again, the last two displayed inequalities are mutually inconsistent.

We conclude that a function $f\colon \mathcal{P}(\mathbb{R}) \to [0,\infty]$ satisfying the three illustrated properties simultaneously does not exist. □

At this point, it is natural to wonder whether we can salvage the situation by requiring the additivity property to hold only for finite collections of pairwise disjoint sets. However, as shall clearly emerge in §2, countable additivity is absolutely indispensable for any limiting process to be valid when computing integrals. What is more, it turns out that, in higher dimensions, not even finite additivity can be guaranteed. This is due to the celebrated *Banach-Tarski paradox*.

PROPOSITION 1.2.3 (Banach-Tarski's paradox). *Let $n \geq 3$ be an integer, E, F arbitrary bounded open subsets of $\mathbb{R}^n$. Then there exists partitions $(E_i)_{1\leq i\leq m}$ and $(F_i)_{1\leq i\leq m}$ of E and F, respectively, such that, for all $1 \leq i \leq k$, $F_i = \phi_i(E_i)$ for some Euclidean isometry $\phi_i\colon \mathbb{R}^n \to \mathbb{R}^n$.*

The notions of bounded set and Euclidean isometry are recalled in §A.1 and §A.2, respectively. Mind the absolutely disconcerting import of the paradox, which amply justifies the adopted nomenclature: it is possible to cut up a tennis ball into finitely many pieces and rearrange them in space, via translations or rotations, to form the Sun! Needless to say, the sets E_i and F_i of Proposition 1.2.3 are far from being geometrically visualizable, and their construction relies once more on the Axiom of Choice; this fact notwithstanding, the proposition clearly prevents the existence, when $n \geq 3$, of any non-trivial function $\mu\colon \mathcal{P}(\mathbb{R}^n) \to [0,\infty]$ which is finitely additive and invariant under Euclidean isometries: any such μ would necessarily assign the same value to *any* bounded open set. For more on the Banach-Tarski paradox and its relation to the notion of *amenability* of topological groups, the reader is strongly invited to consult Terry Tao's blog post [18].

The only way out is to forgo the pretension to measure the size of every subset of $\mathbb{R}^n$, and to content ourselves with being able to measure a sufficiently ample class of sets which includes all those one is likely to ever encounter in practice.

1.3. Measurable spaces and maps

1.3.1. Measurable spaces. We start by axiomatizing the properties of those collections of sets we want to be able to assign a "measure" to.

DEFINITION 1.3.1 (σ-algebra, Borel space). Let X be a set. A **σ-algebra** on X is a family $\mathfrak{M} \subset \mathcal{P}(X)$ satisfying the following properties:

(1) $\emptyset \in \mathfrak{M}$ and $X \in \mathfrak{M}$;
(2) if $E \in \mathfrak{M}$, then $E^{c} \in \mathfrak{M}$;
(3) if $(E_i)_{i\in I}$ is a countable collection of elements of $\mathfrak{M}$, then $\bigcup_{i\in I} E_i \in \mathfrak{M}$.

The pair $(X,\mathfrak{M})$ is called a **measurable space**, or a **Borel space**.

Often in the literature, especially in older texts, the terminology *σ-field* is adopted in place of *σ-algebra*. As we shall see in §1.4, σ-algebras will serve as the domains of *measures*.

If $(X,\mathfrak{M})$ is a measurable space, any $E \in \mathfrak{M}$ is called a **measurable set**.

LEMMA 1.3.2. *Let $\mathfrak{M}$ be a σ-algebra on a set X.*

(1) If $E, F \in \mathfrak{M}$, then $E \setminus F \in \mathfrak{M}$.
(2) If $(E_i)_{i\in I}$ is a countable collection of elements of $\mathfrak{M}$, then $\bigcap_{i\in I} E_i \in \mathfrak{M}$.

PROOF. We start with the second assertion. For all $i \in I$, the assumption $E_i \in \mathfrak{M}$ implies that $E_i^{c} \in \mathfrak{M}$; it follows that $\bigcup_{i\in I} E_i^{c} \in \mathfrak{M}$. Since the complement in X of the latter set is $\bigcap_{i\in I} E_i$, the claim follows.

As for the first assertion, observe that $E \setminus F = E \cap F^{c}$, which is in $\mathfrak{M}$ since $F^{c} \in \mathfrak{M}$ and $\mathfrak{M}$ is closed under countable, hence in particular finite, intersections. □

Remark 1.3.3. Observe that, in the presence of (2) and (3), condition (1) in Definition 1.3.1 is equivalent to the requirement that $\mathfrak{M}$ is non empty. Indeed, clearly (1) implies non-emptiness of $\mathfrak{M}$; on the other hand, if $\mathfrak{M}$ is non-empty, then there is $E \in \mathfrak{M}$, whence by (2) also $E^c \in \mathfrak{M}$, so that $X = E \cup E^c$ and $\emptyset = E \cap E^c \in \mathfrak{M}$.

Example 1.3.4. If X is any set, the smallest[5] σ-algebra on X is $\{\emptyset, X\}$, which is sometimes called the trivial σ-algebra on X. The largest σ-algebra on X is $\mathcal{P}(X)$, known at times as the discrete σ-algebra on X. The pair $(X, \mathcal{P}(X))$ is then also known as a **discrete measurable space**.

Example 1.3.5. If X is any set, then the σ-algebra of countable or co-countable sets in X is defined as

$$\mathfrak{M} = \{E \subset X : E \text{ is countable or } E^c \text{ is countable}\} .$$

It is a σ-algebra: $\emptyset$ is countable, and if E is countable or E^c is countable, then the same automatically holds for E^c. Finally, if $(E_i)_{i\in I}$ is a countable family of elements of $\mathfrak{M}$, then either all of them are countable, and then so is $\bigcup_{i\in I} E_i$, or there is $i_0 \in I$ such that $E^c_{i_0}$ is countable, in which case

$$\left(\bigcup_{i\in I} E_i\right)^c = \bigcap_{i\in I} E_i^c$$

is also countable, being contained in $E^c_{i_0}$. In both cases, $\mathfrak{M}$ contains $\bigcup_{i\in I} E_i$.

Observe that, if X itself is countable, then trivially $\mathfrak{M} = \mathcal{P}(X)$. On the other hand, if X is uncountable, then $\mathfrak{M} \neq \mathcal{P}(X)$ (and $\mathfrak{M} \neq \{\emptyset, X\}$), since it is always possible to find an uncountable $E \subset X$ for which E^c is also uncountable[6].

Example 1.3.6. If $(\mathfrak{M}_\alpha)_{\alpha\in A}$ is a family of σ-algebras on a set X, then

$$\mathfrak{M} = \bigcap_{\alpha\in A} \mathfrak{M}_\alpha$$

is a σ-algebra on X. This follows directly from Definition 1.3.1.

From Example 1.3.6, it follows that, given a set X and any family $\mathfrak{E} \subset \mathcal{P}(X)$, there always exists a smallest[7] σ-algebra on X containing $\mathfrak{E}$. It is known as the **σ-algebra generated by** $\mathfrak{E}$, denoted as $\sigma(\mathfrak{E})$, and obtained as the intersection of all σ-algebras on X containing $\mathfrak{E}$. Notice that there is always at least one such σ-algebra, namely $\mathcal{P}(X)$. The collection $\mathfrak{E}$ is then called a **generating set** for the σ-algebra $\sigma(\mathfrak{E})$.

If (X, τ) is a topological space (see §A.1), then the **Borel σ-algebra** on X associated to the topology τ is the σ-algebra generated by τ, that is, by the collection of open subsets of X for the topology τ. Equivalently, it is the σ-algebra generated by all closed sets. It is denoted $\mathfrak{B}_X$, since every time the topology on X will be clear from the context, and thus omitted from notation; the elements of $\mathfrak{B}_X$ are called **Borel sets**.

Clearly, not only every open or closed set is a Borel set. Countable unions of closed sets are Borel sets (though not necessarily closed); these are called **F_σ-sets**. Similarly, countable intersections of open sets are Borel sets (though not necessarily open); they are called **G_δ-sets**[8].

Exercise 1.3.7. Let X be a set. An **algebra** on X is a collection $\mathfrak{A}$ of subsets of X satisfying the following properties:

(1) $\emptyset \in \mathfrak{A}$, $X \in \mathfrak{A}$;
(2) if $E \in \mathfrak{A}$, then $E^c \in \mathfrak{A}$;
(3) if $(E_i)_{i\in I}$ is a finite family of elements of $\mathfrak{A}$, then $\bigcup_{i\in I} E_i \in \mathfrak{A}$.

[5]With respect to the order relation of inclusion between subsets of $\mathcal{P}(X)$.

[6]This is an exercise in infinite cardinals: we gloss over the details as they are irrelevant for our purposes.

[7]Again, in the sense of inclusion.

[8]The subscripts σ and δ come from the German words "Summe" and "Durchschnitt" for "union" and "intersection", respectively.

Thus, any σ-algebra is an algebra, and any algebra which is closed under countable unions is a σ-algebra.

Show that an algebra $\mathfrak{A}$ is a σ-algebra provided that it is closed under infinite countable disjoint unions, namely provided that $\bigcup_{n\geq 0} E_n \in \mathfrak{A}$ whenever $(E_n)_{n\geq 0}$ is a sequence of pairwise disjoint element of $\mathfrak{A}$.

EXERCISE 1.3.8. Let $\mathfrak{M}$ be an infinite σ-algebra on a set X. Show that $\mathfrak{M}$ contains an infinite sequence $(E_n)_{n\geq 0}$ of pairwise disjoint sets, and deduce that $\mathfrak{M}$ has at least the cardinality of the *continuum*, namely that there is an injective map $\mathbb{R} \to \mathfrak{M}$.

1.3.2. Measurable maps. We now turn to the discussion of "structure-preserving" maps between measurable spaces.

DEFINITION 1.3.9 (Measurable map). Let $(X, \mathfrak{M}_X)$ and $(Y, \mathfrak{M}_Y)$ be measurable spaces. A function $f\colon X \to Y$ is **measurable** with respect to $\mathfrak{M}_X$ and $\mathfrak{M}_Y$ if, for all $E \in \mathfrak{M}_Y$, $f^{-1}(E) \in \mathfrak{M}_X$. We also say that f is $(\mathfrak{M}_X, \mathfrak{M}_Y)$-measurable.

Observe the formal analogy between the notion of measurable map and the topological notion of continuous map, which is recalled in §A.1. For both, several natural and desirable properties are a consequence of the fact that the definition involves inverse images which, as opposed to direct images, preserve set-theoretic operations.

EXAMPLE 1.3.10. If $(X, \mathfrak{M}_X)$ is a measurable space, then the identity map id_X is obviously measurable.

EXAMPLE 1.3.11. If $\mathfrak{M}_X = \mathcal{P}(X)$, then any map $f\colon X \to Y$ is $(\mathfrak{M}_X, \mathfrak{M}_Y)$-measurable. Similarly, if $\mathfrak{M}_Y = \{\emptyset, Y\}$, then any map $f\colon X \to Y$ is $(\mathfrak{M}_X, \mathfrak{M}_Y)$-measurable.

EXAMPLE 1.3.12. Every constant function $f\colon X \to Y$ is $(\mathfrak{M}_X, \mathfrak{M}_Y)$-measurable, since $f^{-1}(E) \in \{\emptyset, X\}$ for all $E \in \mathfrak{M}_Y$.

A partial converse of the previous statement is the following: if $\mathfrak{M}_X = \{\emptyset, X\}$ and $\mathfrak{M}_Y$ *separates points*[9], namely for all $y_1 \neq y_2 \in Y$ there is $E \in \mathfrak{M}_Y$ such that $y_1 \in E$ and $y_2 \notin E$, then any $(\mathfrak{M}_X, \mathfrak{M}_Y)$-measurable map $f\colon X \to Y$ is constant. To see this, argue by contradiction and assume there are $y_1 \neq y_2 \in f(X)$; if $E \in \mathfrak{M}_Y$ contains y_1 but not y_2, then $f^{-1}(E) \in \mathfrak{M}_X$ by measurability, and yet $f^{-1}(E)$ is a proper non-empty subset of X, which yields the desired contradiction.

LEMMA 1.3.13. *If $(X, \mathfrak{M}_X)$, $(Y, \mathfrak{M}_Y)$ and $(Z, \mathfrak{M}_Z)$ are measurable spaces, $f\colon X \to Y$ and $g\colon Y \to Z$ are measurable, then $g \circ f\colon X \to Z$ is measurable.*

The proof is obvious, and thus omitted.

Measurability can be checked on generating sets, as illustrated by the following statement.

LEMMA 1.3.14. *Let $(X, \mathfrak{M}_X)$ and $(Y, \mathfrak{M}_Y)$ be measurable spaces, $f\colon X \to Y$ a map. Let $\mathfrak{E} \subset \mathcal{P}(Y)$ be a generating set for the σ-algebra $\mathfrak{M}_Y$. Then f is measurable if and only if $f^{-1}(E) \in \mathfrak{M}_X$ whenever $E \in \mathfrak{E}$.*

PROOF. One direction is obvious from the definition of measurability. For the converse, suppose $f^{-1}(E) \in \mathfrak{M}_X$ for all $E \in \mathfrak{E}$, and consider the family

$$\mathfrak{M} = \{E \in \mathfrak{M}_Y : f^{-1}(E) \in \mathfrak{M}_X\} ;$$

since taking the inverse image is a set-theoretic operation which preserves complements and unions, it is clear that $\mathfrak{M}$ is a σ-algebra, and by definition $\mathfrak{M} \subset \mathfrak{M}_Y$. By assumption, $\mathfrak{E} \subset \mathfrak{M}$, whence $\sigma(\mathfrak{E}) \subset \mathfrak{M}$; but $\sigma(\mathfrak{E}) = \mathfrak{M}_Y$ by hypothesis, so that we conclude $\mathfrak{M} = \mathfrak{M}_Y$. This shows that f is measurable. □

COROLLARY 1.3.15. *Let X and Y be topological spaces, $f\colon X \to Y$ a continuous map. Then f is measurable with respect to the Borel σ-algebras on X and Y.*

[9] This is the case, for instance, for the Borel σ-algebra generated by a Hausdorff topology on Y.

PROOF. The Borel σ-algebra $\mathfrak{B}_Y$ is generated by the topology τ_Y on Y. By Lemma 1.3.14, it suffices to show that $f^{-1}(O)$ is in the Borel σ-algebra $\mathcal{B}_X$ for every open set $O \subset Y$; this is obvious since $f^{-1}(O)$ is open in X, by continuity of f. $\square$

If $(X, \mathfrak{M}_X)$ and (Y, τ) is a topological space, we shall call a map $f\colon X \to Y$ measurable, without further specification, if it is $(\mathfrak{M}_X, \mathfrak{B}_Y)$-measurable; a similar terminological convention shall apply to maps $X \to Y$ where (X, τ) is a topological space and $(Y, \mathfrak{M}_Y)$ is a Borel space.

1.3.3. Product and trace σ-algebras. Consider a family $(X_\alpha, \mathfrak{M}_\alpha)_{\alpha\in\mathsf{A}}$ of Borel spaces. On the product set

$$X = \prod_{\alpha\in\mathsf{A}} X_\alpha = \left\{ f\colon \mathsf{A} \to \bigcup_{\alpha\in\mathsf{A}} X_\alpha : f(\alpha) \in X_\alpha \ \forall \alpha \in \mathsf{A} \right\}$$

we define the **product σ-algebra**

$$\mathfrak{M} = \bigotimes_{\alpha\in\mathsf{A}} \mathfrak{M}_\alpha$$

as follows: $\mathfrak{M}$ is the smallest σ-algebra on X for which all canonical projection maps $\pi_\alpha\colon X \to X_\alpha$, $f \mapsto f(\alpha)$ are $(\mathfrak{M}, \mathfrak{M}_\alpha)$-measurable. Explicitly, $\mathfrak{M}$ is the intersection of all σ-algebras $\mathfrak{N}$ on X with the property that $\pi_\alpha^{-1}(E_\alpha) \in \mathfrak{N}$ for all $E_\alpha \in \mathfrak{M}_\alpha$ and all $\alpha \in \mathsf{A}$. Equivalently, a generating set for $\mathfrak{M}$ is

$$\{\pi_\alpha^{-1}(E_\alpha) : E_\alpha \in \mathfrak{M}_\alpha,\ \alpha \in \mathsf{A}\}\ . \tag{1.3.1}$$

The Borel space $(X, \mathfrak{M})$ is called the **product Borel space** of the $(X_\alpha, \mathfrak{M}_\alpha)$. It enjoys the following *universal property*[10]. Let $(Y, \mathfrak{M}_Y)$ be a Borel space; if $f\colon Y \to X$ is a $(\mathfrak{M}_Y, \mathfrak{M})$-measurable map, then the compositions $\pi_\alpha \circ f\colon Y \to X_\alpha$ are $(\mathfrak{M}_Y, \mathfrak{M}_\alpha)$-measurable for all $\alpha \in \mathsf{A}$. Conversely, if $f_\alpha\colon Y \to X_\alpha$ is a family of $(\mathfrak{M}_Y, \mathfrak{M}_\alpha)$-measurable maps, then there exists a unique $(\mathfrak{M}_Y, \mathfrak{M})$-measurable map $f\colon Y \to X$ such that $f_\alpha = \pi_\alpha \circ f$.

EXERCISE 1.3.16. Show the claimed universal property for the product Borel space.

REMARK 1.3.17. If the index set A is countable, then a generating set for the product σ-algebra is also given by the collection of "boxes"

$$\left\{ \prod_{\alpha\in\mathsf{A}} E_\alpha : E_\alpha \in \mathfrak{M}_\alpha,\ \alpha \in \mathsf{A} \right\}\ .$$

Indeed, it is clear that the σ-algebra generated by the latter collection contains $\mathfrak{M}$ (regardless of countability of A). When A is countable, then every box is a countable intersections of sets from (1.3.1), whence the reverse inclusion holds as well.

When A is uncountable, the σ-algebra generated by boxes is, in general, strictly finer[11] than the product σ-algebra.

Suppose we are in the particular case where the X_α's are topological spaces, and $\mathfrak{M}_\alpha = \mathfrak{B}_{X_\alpha}$ for all $\alpha \in \mathsf{A}$. On the product set X we can defined two *a priori* distinct σ-algebras. The first one is the product σ-algebra

$$\mathfrak{M} = \bigotimes_{\alpha\in\mathsf{A}} \mathfrak{B}_{X_\alpha}\ ;$$

the second one is the Borel σ-algebra $\mathfrak{B}_X$ determined by the product topology on X (see §A.1).

[10] We adopt here a standard terminology coming from the *theory of categories*, an abstract formalism to deal with mathematical structures and structure-preserving maps; the interested reader is invited to consult MacLane's classic textbook [14].

[11] That is, it contains more sets.

PROPOSITION 1.3.18. *Let $\mathfrak{M}$ and $\mathfrak{B}_X$ be defined as above. Then*

$$\mathfrak{M} \subset \mathfrak{B}_X \ .$$

If A is countable and X_α is second countable (cf. §A.1) for all $\alpha \in \mathsf{A}$, then

$$\mathfrak{M} = \mathfrak{B}_X \ .$$

PROOF. For every $\alpha \in \mathsf{A}$, the projection map $\pi_\alpha \colon X \to X_\alpha$ is continuous when X is endowed with the product topology, hence by Corollary 1.3.15 it is $(\mathfrak{B}_X, \mathfrak{B}_{X_\alpha})$-measurable. By definition of $\mathfrak{M}$, it follows that $\mathfrak{M} \subset \mathfrak{B}_X$.

Suppose now that A is countable and every X_α is second countable. For each $\alpha \in \mathsf{A}$, choose a countable basis $\mathcal{B}_\alpha$ of the topology on X_α. Then $\mathfrak{B}_X$ is generated by the collection

$$\mathcal{B} = \left\{ \bigcap_{\alpha \in J} O_\alpha : O_\alpha \in \mathcal{B}_\alpha \text{ for all } \alpha \in J,\ J \subset A \text{ finite} \right\} ;$$

to see this, observe that every open set in X is a union of elements of $\mathcal{B}$ (as follows rather directly from the definition of the product topology), and thus in particular a countable union thereof, since $\mathcal{B}$ itself is countable. It follows that $\sigma(\mathcal{B})$ contains $\mathfrak{B}_X$, and is therefore equal to it. It is on the other hand clear that $\mathcal{B} \subset \mathfrak{M}$, from which it follows that $\mathfrak{B}_X \subset \mathfrak{M}$, as desired. □

We let $\mathfrak{B}_{\mathbb{R}^n}$ denote, for every integer $n \geq 1$, the Borel σ-algebra generated by the Euclidean topology on $\mathbb{R}^n$. Since the the Euclidean topology on $\mathbb{R}$ is second countable (cf. Lemma A.1.7), Proposition 1.3.18 delivers at once:

COROLLARY 1.3.19. *For every integer $n \geq 1$,*

$$\mathfrak{B}_{\mathbb{R}^n} = \bigotimes_{i=1}^{n} \mathfrak{B}_{\mathbb{R}} \ .$$

Let $(X, \mathfrak{M})$ be a Borel space, Y a subset of X. We define the **trace of the σ-algebra $\mathfrak{M}$ on Y** to be the σ-algebra

$$\mathfrak{M}|_Y = \{Y \cap E : E \in \mathfrak{M}\} \ .$$

The validity of all axioms of a σ-algebra for $\mathfrak{M}|_Y$ is easily determined. Furthermore, it is straightforward to check that $\mathfrak{M}|_Y$ is the smallest σ-algebra on Y making the canonical incusion $\iota \colon Y \to X$ measurable. The Borel space $(Y, \mathfrak{M}|_Y)$, called a **measurable subspace** of $(X, \mathfrak{M})$, satisfies the following universal property. For every Borel space $(Z, \mathfrak{M}_Z)$, a map $f \colon Z \to Y$ is $(\mathfrak{M}_Z, \mathfrak{M}|_Y)$-measurable if and only if $\iota \circ f \colon Z \to X$ is $(\mathfrak{M}_Z, \mathfrak{M})$-measurable.

EXERCISE 1.3.20. Show the claimed universal property for the measurable subspace $(Y, \mathfrak{M}|_Y)$.

1.3.4. Measurability of real-valued functions. We shall be particularly concerned with measurable functions with values in $\mathbb{R}$ (or $\mathbb{C}$), since those are ultimately the functions we would like to integrate. For real-valued functions, it suffices to check measurability on subclasses of intervals.

PROPOSITION 1.3.21. *The Borel σ-algebra $\mathfrak{B}_{\mathbb{R}}$ is generated by each of the following collections:*

$$\{(a,b) : a < b \in \mathbb{R}\} , \quad \{[a,b) : a < b \in \mathbb{R}\} , \quad \{(a,b] : a < b \in \mathbb{R}\} , \quad \{[a,b] : a < b \in \mathbb{R}\} ,$$
$$\{(-\infty,a) : a \in \mathbb{R}\} , \quad \{(a,+\infty) : a \in \mathbb{R}\} , \quad \{(-\infty,a] : a \in \mathbb{R}\} , \quad \{[a,+\infty) : a \in \mathbb{R}\} .$$

PROOF. All listed sets are open, closed or the intersection of an open and a closed set; the σ-algebra generated by each collection is thus trivially contained in $\mathfrak{B}_{\mathbb{R}}$. We need to show the converse inclusion.

By definition of the Euclidean metric and topology, every open set in $\mathbb{R}$ is a union of open intervals (a,b), $a < b \in \mathbb{R}$. In fact, as the countable subcollection

$$\{(a,b) : a < b,\ a,b \in \mathbb{Q}\}$$

is a basis of the topology (see the proof of Lemma A.1.7), every open set is a countable union of bounded open intervals. As a consequence, the σ-algebra generated by bounded open intervals contains $\mathfrak{B}_{\mathbb{R}}$, and thus must be equal to it.

Every open interval (a,b), $a < b \in \mathbb{R}$ is the countable union of the half-open intervals $[a+1/n, b)$, $n \in \mathbb{N}^*$, whence

$$\mathfrak{B}_{\mathbb{R}} \subset \sigma(\{[a,b) : a < b \in \mathbb{R}\})$$

and must therefore be equal to it.

Similar arguments apply to all other cases; the (instructive) verification is left to the reader. □

Corollary 1.3.22. *Let $(X, \mathfrak{M})$ be a Borel space, $f\colon X \to \mathbb{R}$ a function. Let $\mathfrak{E}$ be one of the families listed in Proposition 1.3.21. The following are equivalent:*

(1) f is $(\mathfrak{M}, \mathfrak{B}_{\mathbb{R}})$-measurable;
(2) $f^{-1}(E) \in \mathfrak{M}$ for all $E \in \mathfrak{E}$.

Proof. The equivalence follows combining Lemma 1.3.14 and Proposition 1.3.21. □

We shall need the extension of the previous corollary to functions with values in the extended real number system $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$. We endow $\overline{\mathbb{R}}$ with the σ-algebra

$$\mathfrak{B}_{\overline{\mathbb{R}}} = \{E \subset \overline{\mathbb{R}} : E \cap \mathbb{R} \in \mathfrak{B}_{\mathbb{R}}\} \ .$$

The verification that $\mathfrak{B}_{\overline{\mathbb{R}}}$ is a σ-algebra is routine; the notation is justified by the fact that $\mathfrak{B}_{\overline{\mathbb{R}}}$ is the Borel σ-algebra generated by the natural topology on $\overline{\mathbb{R}}$ defined in §A.1.1.

Corollary 1.3.23. *Let $(X, \mathfrak{M})$ be a Borel space, $f\colon X \to \overline{\mathbb{R}}$ a function. Let*

$$\mathfrak{E} = \{(a, +\infty] : a \in \mathbb{R}\} \ .$$

The following are equivalent:

(1) f is $(\mathfrak{M}, \mathfrak{B}_{\mathbb{R}})$-measurable;
(2) $f^{-1}(E) \in \mathfrak{M}$ for all $E \in \mathfrak{E}$.

Proof. The argument relies once more on Lemma 1.3.14, in conjunction with the easily verified fact that $\mathfrak{E}$ generates $\mathfrak{B}_{\overline{\mathbb{R}}}$. □

Remark 1.3.24. The statement of Corollary 1.3.23 is unaltered when replacing $\mathfrak{E}$ by any of the following families:

$$\{[-\infty, a) : a \in \mathbb{R}\} \ , \quad \{[a, +\infty) : a \in \mathbb{R}\} \ , \quad \{[-\infty, a] : a \in \mathbb{R}\} \ .$$

This follows again from Lemma 1.3.14 and the fact that each of the above families generates $\mathfrak{B}_{\overline{\mathbb{R}}}$.

1.4. Measures

We are now in a position to introduce the abstract notion of a *measure*.

Definition 1.4.1 (Measure). Let $(X, \mathfrak{M})$ be a measurable space. A **measure** on $(X, \mathfrak{M})$ is a function $\mu\colon \mathfrak{M} \to [0, \infty]$ satisfying the following properties:

(1) $\mu(\emptyset) = 0$;
(2) if $(E_n)_{n \geq 0}$ is a sequence of pairwise disjoint elements of $\mathfrak{M}$, then

$$\mu\Big(\bigcup_{n=0}^{\infty} E_n\Big) = \sum_{n=0}^{\infty} \mu(E_n) \ .$$

The triple $(X, \mathfrak{M}, \mu)$ is called a **measure space**.

If $\mathfrak{M}$ is understood from the context, we also speak of μ being a measure on X.

Observe that the condition $\mu(\emptyset) = 0$ is a consequence of the second one in conjunction with the requirement that μ is not constantly equal to ∞: if $E_n = \emptyset$ for all $n \geq 0$, then we deduce

$$\mu(\emptyset) = \sum_{n=0}^{\infty} \mu(\emptyset) ,$$

where the right-hand side equals ∞ unless $\mu(\emptyset) = 0$. The possibility $\mu(\emptyset) = \infty$ would entail $\mu(E) = \infty$ for all $E \in \mathfrak{M}$ (owing to the monotonicity property of measures in Proposition 1.4.9), which is precluded.

Property (2) in the definition is called **σ-additivity** of a measure. It readily implies **finite additivity**, namely the fact that

$$\mu\Big(\bigcup_{n=0}^{N} E_n\Big) = \sum_{n=0}^{N} \mu(E_n)$$

for every finite sequence $(E_n)_{0\leq n\leq N}$ of pairwise disjoint measurable sets; it suffice to set $E_n = \emptyset$ for all $n > N$, and apply the σ-additivity property to the sequence $(E_n)_{n\geq 0}$ thus obtained, coupled with the fact that $\mu(\emptyset) = 0$.

We give a first list of elementary examples of measures, postponing the task of constructing much more interesting ones, of geometric nature, to §1.5.

Example 1.4.2 (Dirac measure). Let $(X, \mathfrak{M})$ be a Borel space, $x \in X$ a point. The **Dirac measure**, or **Dirac mass** at x is the measure δ_x given by

$$\delta_x(E) = \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{if } x \notin E \end{cases} .$$

The axioms of a measure are readily verified.

Example 1.4.3 (Counting measure). Let $(X, \mathfrak{M})$ be a Borel space. The **counting measure** on $(X, \mathfrak{M})$ is defined as

$$\mathsf{c}(E) = |E| = \sum_{x\in E} 1$$

for all $E \in \mathfrak{M}$. Once again, the axioms of a measure are trivially satisfied.

Example 1.4.4. The two previous examples can be generalized as follows. Let $(X, \mathfrak{M})$ be a Borel space, $f\colon X \to [0, \infty]$ a function. Then f induces a measure μ_f on $(X, \mathfrak{M})$ defined as

$$\mu_f(E) = \sum_{x\in E} f(x)$$

for all $E \in \mathfrak{M}$. That the axioms of a measure are verified can be established via the theory of infinite sums over arbitrary indexing sets, which is for convenience summarized in §B. If f is constantly equal to 1, we obtain $\mu_f = \mathsf{c}$, the counting measure seen in Example 1.4.3. The Dirac mass δ_x, for every base point $x \in X$, is obtained as μ_f for the function f defined by $f(x) = 1$, $f(y) = 0$ for all $y \neq x$.

Example 1.4.4 will be vastly generalized in Proposition 2.7.1 on measures with *densities* with respect to a given starting measure.

Example 1.4.5. Let X be a set, $\mathfrak{M}$ the σ-algebra of countable or co-countable subsets of X, contemplated in Example 1.3.5. Define a function $\mu\colon \mathfrak{M} \to [0, \infty]$ by

$$\mu(E) = \begin{cases} 0 & \text{if } E \text{ is countable} \\ 1 & \text{if } E \text{ is co-countable and uncountable} \end{cases} .$$

It is readily ascertained that μ is a measure on $(X, \mathfrak{M})$, since a countable sequence of pairwise disjoint sets can feature at most one co-countable element. If X is countable, then μ is the **zero measure**, namely the measure assigning mass 0 to every measurable set.

EXAMPLE 1.4.6 (Sum of measures). Let $(\mu_\alpha)_{\alpha\in\mathsf{A}}$ be a family of measures on a given Borel space $(X, \mathfrak{M})$. We define the sum $\mu = \sum_{\alpha\in\mathsf{A}} \mu_\alpha$ by the assignment

$$\mu(E) = \sum_{\alpha\in\mathsf{A}} \mu_\alpha(E)$$

for all $E \in \mathfrak{M}$. It is then clear, using the results of §B on infinite sums, that μ is a measure on $(X, \mathfrak{M})$.

We now introduce some standard terminology regarding measures. We say that a measure μ on a Borel space $(X, \mathfrak{M})$ is **finite** if $\mu(X) < \infty$[12]. If $\mu(X) = 1$, then μ is called a **probability measure**, and $(X, \mathfrak{M}, \mu)$ a **probability measure space**. A measure μ is **σ-finite** if there is a countable family $(X_i)_{i\in I}$ of sets $X_i \in \mathfrak{M}$ such that $X = \bigcup_{i\in I} X_i$ and $\mu(X_i) < \infty$ for all $i \in I$.

If (X, τ) is a topological space, a **Borel measure** on X is a measure on $(X, \mathfrak{B}_X)$.

REMARK 1.4.7. A finite measure is obviously σ-finite, the converse being not true: the counting measure on $(\mathbb{N}.\mathcal{P}(\mathbb{N}))$ (cf. Example 1.4.3) is σ-finite, since $\mathbb{N}$ is a countable union of singletons, but is not finite, since $\mathbb{N}$ is an infinite set.

1.4.1. A small aside in probability theory. In the realm of probability theory, the notion of a probability measure space $(X, \mathfrak{M}, \mu)$ formalizes the data of a set X of possible *outcomes* of a certain random experiment, a family of random *events* $\mathfrak{M}$, namely of collections of outcomes E to which a probability $\mu(E) \in [0, 1]$ of occurring is assigned. If $(X, \mathfrak{M}_X, \mu)$ is a probability measure space and $(Y, \mathfrak{M}_Y)$ is a measurable space, then a measurable function $f : X \to Y$ is routinely called, in the language of probability theory, a **random variable**.

If X is a set, $(Y, \mathfrak{M}_Y)$ is a measurable space and $f : X \to Y$ is a function, then the collection

$$\mathfrak{M}_f = \{f^{-1}(E) : E \in \mathfrak{M}_Y\}$$

is a σ-algebra on X, as is easily deduced from the fact that taking inverse images of sets under maps preserves all set-theoretic operations, in particular unions and complements. In the context of probability theory, it frequently occurs that $(X, \mathfrak{M}_X, \mu)$ is a probability measure space, so that f is a random variable; in this case the σ-algebra $\mathfrak{M}_f$ is said to be *generated by* f and plays a pivotal role in the study of *stochastic processes*, for instance in definining the notion of independence of random variables. We refer the interested reader to [**12**, **1**].

1.4.2. Properties of measures. We introduce now some standard terminology which shall be widely adopted in the sequel. Let $(X, \mathfrak{M}, \mu)$ be a measure space. A set $E \in \mathfrak{M}$ such that $\mu(E) = 0$ is called a **μ-null set**, or simply a **null set** if no confusion arises. A given assertion P concerning points of X is said to hold **μ-almost everywhere**, or **almost everywhere** if μ is clear from the context, if there is a μ-null set $N \subset X$ such that P holds for all $x \in X \setminus N$.

An observation that is routinely employed in measure-theoretic arguments is that countable unions of μ-null sets are μ-null, as follows readily from σ-subadditivity of μ.

REMARK 1.4.8. If $(X, \mathfrak{M}, \mu)$ is a probability measure space, then it is customary to say that a property P holds **μ-almost surely** instead of **μ-almost everywhere**, emphasizing the statistical aspect of the assertion.

If X is a set, we say that a sequence $(E_n)_{n\geq 0}$ of subsets of X is increasing if $E_n \subset E_{n+1}$ for all $n \geq 0$, and decreasing if $E_n \supset E_{n+1}$ for all $n \geq 0$.

PROPOSITION 1.4.9. *Let $(X, \mathfrak{M}, \mu)$ be a measure space.*

(1) (Monotonicity) If $E \subset F$ are measurable sets, then $\mu(E) \leq \mu(F)$.

[12]As a consequence of the forthcoming Proposition 1.4.9, this implies that μ takes values in the closed interval $[0, \mu(X)]$.

(2) *(σ-subadditivity) If $(E_n)_{n\geq 0}$ is a sequence of measurable sets, then*

$$\mu\Big(\bigcup_{n\geq 0} E_n\Big) \leq \sum_{n=0}^{\infty} \mu(E_n) \ .$$

(3) *(Continuity from below) If $(E_n)_{n\geq 0}$ is an increasing sequence of measurable sets, then*

$$\mu\Big(\bigcup_{n\geq 0} E_n\Big) = \lim_{n\to\infty} \mu(E_n) \ .$$

(4) *(Continuity from above) If $(E_n)_{n\geq 0}$ is a decreasing sequence of measurable sets, and $\mu(E_0) < \infty$, then*

$$\mu\Big(\bigcap_{n\geq 0} E_n\Big) = \lim_{n\to\infty} \mu(E_n) \ .$$

In the last assertion, the assumption that $\mu(E_0) < \infty$ is indispensable, as the following counterexample shows. Let μ be the counting measure on $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$, and let

$$E_n = \{m \in \mathbb{N} : m \geq n\}$$

for all $n \geq 0$. Then $(E_n)_{n\geq 0}$ is decreasing and $\bigcap_{n\geq 0} E_n = \emptyset$; now $\mu(\emptyset) = 0$, however

$$\lim_{n\to\infty} \mu(E_n) = \lim_{n\to\infty} \infty = \infty \ .$$

Remark 1.4.10. (1) For continuity from above to hold, it is sufficient that $\mu(E_{n_0}) < \infty$ for some $n_0 > 0$, as the result phrased in the proposition can then be applied to $(E_{n_0+n})_{n\geq 0}$, without altering the intersection nor the limit.

(2) Observe that, if $(E_n)_{n\geq 0}$, is increasing, the limit

$$\lim_{n\to\infty} \mu(E_n)$$

exists in $[0, \infty]$, since the sequence of real numbers $(\mu(E_n))_{n\geq 0}$ is increasing. Similarly, if $(E_n)_{n\geq 0}$ is decreasing and $\mu(E_0) < \infty$, then the same limit as above exists in $[0, \infty)$, as $(\mu(E_n))_{n\geq 0}$ is decreasing.

Proof. We start with the first assertion, and let $E \subset F$ be measurable sets. Then F is the disjoint union of E and $F \setminus E$, whence by finite additivity of μ

$$\mu(F) = \mu(E) + \mu(F \setminus E) \geq \mu(E) \ ,$$

the last inequality following from the fact that μ takes positive values.

As to the claimed σ-additivity property, let $(E_n)_{n\geq 0}$ be a sequence of measurable sets. We replace the union $\bigcup_{n\geq 0} E_n$ with a disjoint union, by setting $F_0 = E_0$ and, inductively on $n \geq 0$,

$$F_{n+1} = E_{n+1} \setminus \Big(\bigcup_{i=0}^{n} E_i\Big) \ .$$

Then it is clear that $(F_n)_{n\geq 0}$ is a sequence of pairwise disjoint measurable subsets satisfying

$$\bigcup_{n\geq 0} F_n = \bigcup_{n\geq 0} E_n \ .$$

By σ-additivity of μ, we deduce that

$$\mu\Big(\bigcup_{n\geq 0} E_n\Big) = \mu\Big(\bigcup_{n\geq 0} F_n\Big) = \sum_{n=0}^{\infty} \mu(F_n) \leq \sum_{n\geq 0} \mu(E_n) \ ,$$

where the last inequality follows from monotonicity of μ and the fact that $F_n \subset E_n$ for all $n \geq 0$.

We proceed with the third assertion. Let $(E_n)_{n\geq 0}$ be an increasing sequence of measurable sets. Define a new sequence $(F_n)_{n\geq 0}$ of measurable sets as follows: $F_0 = E_0$ and

$$F_{n+1} = E_{n+1} \setminus E_n$$

for all $n \geq 0$. Then it is clear that the F_n's are pairwise disjoint; moreover,

$$\bigcup_{n\geq 0} F_n = \bigcup_{n\geq 0} E_n \ ,$$

whence, by σ-additivity of μ,

$$\mu\Big(\bigcup_{n\geq 0} E_n\Big) = \sum_{n=0}^{\infty} \mu(F_n) = \lim_{N\to\infty} \sum_{n=0}^{N} \mu(F_n) \ .$$

By finite additivity of μ, we have

$$\sum_{n=0}^{N} \mu(F_n) = \mu\Big(\bigcup_{n=0}^{N} F_n\Big) = \mu(E_0 \cup (E_1 \setminus E_0) \cup \cdots \cup (E_N \setminus E_{N-1})) = \mu(E_N) \ ,$$

from which the claim follows.

Let us now turn to the final assertion, and let $(E_n)_{n\geq 0}$ be a decreasing sequence of measurable sets with $\mu(E_0) < \infty$. Define a new sequence $(F_n)_{n\geq 0}$ by

$$F_n = E_0 \setminus E_n$$

for all $n \geq 0$. Then clearly $(F_n)_{n\geq 0}$ is increasing, whence by the previous item of the proposition

$$\mu\Big(\bigcup_{n\geq 0} F_n\Big) = \lim_{n\to\infty} \mu(F_n) \ ;$$

now

$$\mu(F_n) = \mu(E_0) - \mu(E_n)$$

since $E_n \subset E_0$ for all n and $\mu(E_0) < \infty$, and by the same token

$$\mu\Big(\bigcup_{n\geq 0} F_n\Big) = \mu\Big(E \setminus \bigcap_{n\geq 0} E_n\Big) = \mu(E) - \mu\Big(\bigcap_{n\geq 0} E_n\Big) \ .$$

Combining the three last displayed equalities yields the desired result. □

Exercise 1.4.11. Let $(E_n)_{n\geq 0}$ be a sequence of subsets of a given set X. The **limit inferior** and **limit superior** of the sequence are defined, respectively, as

$$\liminf_{n\to\infty} E_n = \bigcup_{n\geq 0} \bigcap_{k\geq n} E_k \ , \quad \limsup_{n\to\infty} E_n = \bigcap_{n\geq 0} \bigcup_{k\geq n} E_k \ .$$

(1) Show that

$$\liminf_{n\to\infty} E_n = \{x \in X : x \in E_n \text{ for all but finitely many } n \geq 0\}$$

and

$$\limsup_{n\to\infty} E_n = \{x \in X : x \in E_n \text{ for infinitely many } n \geq 0\}; .$$

Deduce that the limit inferior is contained in the limit superior, and exhibit an example showing that the converse inclusion does not hold in general.

(2) Suppose $(X, \mathfrak{M}, \mu)$ is a measure space, and $E_n \in \mathfrak{M}$ for all n. Show that

$$\mu\Big(\liminf_{n\to\infty} E_n\Big) \leq \liminf_{n\to\infty} \mu(E_n)$$

and, if $\mu\big(\bigcup_{n\geq 0} E_n\big) < \infty$,

$$\mu\Big(\limsup_{n\to\infty} E_n\Big) \geq \limsup_{n\to\infty} \mu(E_n) \ .$$

EXERCISE 1.4.12. Let $(X, \mathfrak{M}, \mu)$ be a finite measure space.

(1) Suppose E, F are measurable sets with $\mu(E \triangle F) = 0$. Show that $\mu(E) = \mu(F)$. Exhibit an example of two measurable sets E, F with equal measure but satisfying $\mu(E \triangle F) > 0$.
(2) Define a relation $\sim$ on $\mathfrak{M}$ by declaring that $E \sim F$ if and only if $\mu(E \triangle F) = 0$. Show that $\sim$ is an equivalence relation.
(3) Let $\mathfrak{M}/\sim$ be the quotient set of $\mathfrak{M}$ with respect to the equivalence relation $\sim$. Show that the function

$$\rho\colon (\mathfrak{M}/\sim) \times (\mathfrak{M}/\sim) \to \mathbb{R}_{\geq 0}\ , \quad ([E]_\sim, [F]_\sim) \mapsto \mu(E \triangle F)$$

is well defined, and defines a metric on $\mathfrak{M}/\sim$.

1.4.3. Completion of a measure space. Let $(X, \mathfrak{M}, \mu)$ be a measure space, $E \in \mathfrak{M}$ a set with $\mu(E) = 0$. If F is a subset of X which is contained in F, then monotonicity of μ yields directly $\mu(F) = 0$, *provided that* $F \in \mathfrak{M}$ (which is not necessarily the case). At times, in measure-theoretic arguments, annoying technical points are avoided if the measure space under examination has the property that subsets of null sets are measurable (and thus null sets); in due course we shall see examples of arising potential complications when such a property is not assumed.

DEFINITION 1.4.13 (Complete measure space). A measure space $(X, \mathfrak{M}, \mu)$ is **complete** if, whenever $E \subset F$ are subsets of X such that $F \in \mathfrak{M}$ and $\mu(F) = 0$, then $E \in \mathfrak{M}$.

It is possible to "complete" an arbitrary measure space by simply adjoining to the σ-algebra all subsets of null sets, and extending the measure in the obvious way.

PROPOSITION 1.4.14. *Let* $(X, \mathfrak{M}, \mu)$ *be a measure space. Set*

$$\mathfrak{M}_{\mu,\mathrm{comp}} = \{E \cup F : E \in \mathfrak{M},\ F \subset A \textit{ for some } A \in \mathfrak{M} \textit{ with } \mu(A) = 0\}\ .$$

Define also a function $\mu_{\mathrm{comp}}\colon \mathfrak{M}_{\mu,\mathrm{comp}} \to [0, \infty]$ *via*

$$\mu_{\mathrm{comp}}(E \cup F) = \mu(E)$$

for all $E \cup F \in \mathfrak{M}_{\mu,\mathrm{comp}}$ *as described above. Then the following assertions hold:*

(1) $\mathfrak{M}_{\mu,\mathrm{comp}}$ *is a* σ*-algebra containing* $\mathfrak{M}$*;*
(2) μ_{comp} *is a measure on* $\mathfrak{M}_{\mu,\mathrm{comp}}$ *extending* μ*, namely satisfying* $\mu_{\mathrm{comp}}|_{\mathfrak{M}} = \mu$*;*
(3) the measure space $(X, \mathfrak{M}_{\mu,\mathrm{comp}}, \mu_{\mathrm{comp}})$ *is complete.*

The measure space $(X, \mathfrak{M}_{\mu,\mathrm{comp}}, \mu_{\mathrm{comp}})$ constructed in the foregoing proposition is called the **completion** of the measure space $(X, \mathfrak{M}, \mu)$.

PROOF. We begin by showing the first assertion. It is obvious that $\mathfrak{M}_{\mu,\mathrm{comp}}$ contains $\mathfrak{M}$ since one may take $F = \emptyset$ in the description of the elements of $\mathfrak{M}_{\mu,\mathrm{comp}}$. It is also clear that $\mathfrak{M}_{\mu,\mathrm{comp}}$ is closed under countable intersections, since a countable union of μ-null sets in $\mathfrak{M}$ is μ-null. To conclude that it is a σ-algebra, we must show that it is closed under complements. Let thus $E \cup F$ be an element of $\mathfrak{M}_{\mu,\mathrm{comp}}$, that is, $E \in \mathfrak{M}$ and $F \subset A$ for some $A \in \mathfrak{M}$ with $\mu(A) = 0$. Since $A^{\mathrm{c}} \subset F^{\mathrm{c}}$, the complement $(E \cup F)^{\mathrm{c}} = E^{\mathrm{c}} \cap F^{\mathrm{c}}$ can be written as the union

$$\left(E^{\mathrm{c}} \cap A^{\mathrm{c}}\right) \cup \left(F^{\mathrm{c}} \setminus A^{\mathrm{c}}\right)\ , \tag{1.4.1}$$

where we observe that $E^{\mathrm{c}} \cap A^{\mathrm{c}}$ is an element of $\mathfrak{M}$, and

$$F^{\mathrm{c}} \setminus A^{\mathrm{c}} = F^{\mathrm{c}} \cap (A^{\mathrm{c}})^{\mathrm{c}} = F^{\mathrm{c}} \cap A \subset A\ ,$$

so that (1.4.1) expresses $(E \cup F)^{\mathrm{c}}$ as an element of $\mathfrak{M}_{\mu,\mathrm{comp}}$.

Let us proceed with the second assertion. To begin with, the function μ_{comp} is well defined: if $E, E' \in \mathfrak{M}$ and F, F' are such that $E \cup F = E' \cup F'$ and there are $A, A' \in \mathfrak{M}$ with $\mu(A) = \mu(A') = 0$ and $F \subset A$, $F' \subset A'$, then

$$\mu(E) \leq \mu(E' \cup A') = \mu(E') + \mu(A' \setminus E') = \mu(E')\ ,$$

since $\mu(A' \setminus E') \leq \mu(A') = 0$. By symmetry, $\mu(E') \leq \mu(E)$, and thus $\mu(E) = \mu(E')$, whence μ_{comp} is well defined. The fact that μ_{comp} is σ-additive is rather obvious: if $(E_n \cup F_n)_{n\geq 0}$ is a collection of pairwise disjoint sets in $\mathfrak{M}_{\mu,\text{comp}}$, then *a fortiori* the E_n's are pairwise disjoint, so that

$$\mu_{\text{comp}}\Big(\bigcup_{n\geq 0} E_n \cup F_n\Big) = \mu_{\text{comp}}\Big(\Big(\bigcup_{n\geq 0} E_n\Big) \cup \Big(\bigcup_{n\geq 0} F_n\Big)\Big) = \mu\Big(\bigcup_{n\geq 0} E_n\Big)$$
$$= \sum_{n=0}^{\infty} \mu(E_n) = \sum_{n=0}^{\infty} \mu_{\text{comp}}(E_n \cup F_n) ,$$

using σ-additivity of μ in the process.

Finally, it is straightforward to check that $(X, \mathfrak{M}_{\mu,\text{comp}}, \mu_{\text{comp}})$ is complete. Suppose indeed that a subset N of X is contained in an element $E \cup F \in \mathfrak{M}_{\mu,\text{comp}}$ with $\mu_{\text{comp}}(E \cup F) = \mu(E) = 0$; then there is $A \in \mathfrak{M}$ with $\mu(A) = 0$ such that $F \subset A$, and thus $N \subset E \cup A$, the latter being a μ-null element of $\mathfrak{M}$. We infer that N is an element of $\mathfrak{M}_{\mu,\text{comp}}$ as desired. □

Remark 1.4.15. The completion $(X, \mathfrak{M}_{\mu,\text{comp}}, \mu_{\text{comp}})$ enjoys the following minimality property. If $(X, \mathfrak{M}', \mu')$ is a complete measure space extending $(X, \mathfrak{M}, \mu)$, namely verifying $\mathfrak{M} \subset \mathfrak{M}'$ and $\mu'|_{\mathfrak{M}} = \mu$, then $\mathfrak{M}_{\mu,\text{comp}}$ is contained in $\mathfrak{M}'$ and the restriction of μ' to $\mathfrak{M}_{\mu,\text{comp}}$ coincides with μ_{comp}. The (easy) verification of the statement is left to the reader.

1.5. Outer measures and Carathéodory's construction

In this section we present an abstract procedure, geometric in spirit, to define a measure on a class of sets, starting with an original "size function" which is preemptively assigned to a smaller class of sets. More precisely, suppose X is a set, and $\mathcal{E} \subset \mathcal{P}(X)$ is a collection of "elementary subsets" of X to which some notion of size is assigned, by means of a function $\rho\colon \mathcal{E} \to [0, \infty]$. We would like to extend ρ to a measure μ on a σ-algebra $\mathfrak{M}$ containing the original class $\mathcal{E}$, if possible in some reasonably unique way.

Example 1.5.1. Let $X = \mathbb{R}^n$. We would like to define a measure μ on the Borel σ-algebra $\mathfrak{B}_{\mathbb{R}^n}$ which, when restricted to sufficiently nice subsets, gives the ordinary geometric notion of volume. For instance, it is desirable to ask that

$$\mu\big([a_1, b_1) \times \cdots \times [a_n, b_n)\big) = (b_1 - a_1) \cdots (b_n - a_n)$$

for every *rectangle*[13] $\prod_{i=1}^{n} [a_i, b_i)$, where $a_i < b_i$ are real numbers for all $1 \leq i \leq n$.

The abstract construction presented in this section shall be applied in subsequent ones to the construction of the *Lebesgue measure* on $\mathbb{R}^n$ and of the *Hausdorff measure* on an arbitrary metric space.

1.5.1. Outer measures. We begin with the notion of *outer measure*, which furnishes the appropriate replacement of the notion of measure when we would like to assign a size to every possible subset of a given set, by relaxing the σ-additivity requirement (too strict for the purpose, as shown in §1.1), while preserving some of the features every reasonable notion of length, area, volume, mass... should possess.

Definition 1.5.2 (Outer measure). Let X be a set. An **outer measure** on X is a function $\mu^*\colon \mathcal{P}(X) \to [0, \infty]$ satisfying the following properties:

(1) $\mu^*(\emptyset) = 0$;
(2) if $E \subset F$, then $\mu^*(E) \leq \mu^*(F)$;

[13] The reason for the choice of *half-open rectangles* shall emerge when we will actually construct the Lebesgue measure, which is a measure with the desired properties.

(3) if $(E_n)_{n\geq 0}$ is a sequence of subsets of X, then

$$\mu^*\Big(\bigcup_{n\geq 0} E_n\Big) \leq \sum_{n=0}^{\infty} \mu^*(E_n) \ .$$

Thus, an outer measure is a function on the subsets of X which is monotone and σ-subadditive. Arguing as in the case of a measure, the condition $\mu^*(\emptyset) = 0$ is redundant in the presence of σ-subadditivity combined with the non-triviality requirement that μ^* is not constantly equal to ∞. Also, σ-subadditivity implies finite subadditivity: if $(E_n)_{0\leq n\leq N}$ is a finite sequence of subsets of X, then

$$\mu^*\Big(\bigcup_{n=0}^{N} E_n\Big) \leq \sum_{n=0}^{N} \mu^*(E_n) \ .$$

We are now ready to address our task, made explicit at the beginning of this section. The point is that any starting notion of size $\rho\colon \mathcal{E} \to [0,\infty]$ on an original family $\mathcal{E} \subset \mathcal{P}(X)$ produces an outer measure on X, by approximating every set from outside with countable covers by elements of $\mathcal{E}$.

Proposition 1.5.3. *Let X be a set, $\mathcal{E} \subset \mathcal{P}(X)$ a collection containing $\emptyset$, $\rho\colon \mathcal{E} \to [0,\infty]$ a function such that $\rho(\emptyset) = 0$. Define a function $\mu^*\colon \mathcal{P}(X) \to [0,\infty]$ by the assignment*

$$\mu^*(A) = \inf\Big\{\sum_{n=0}^{\infty} \rho(E_n) : A \subset \bigcup_{n\geq 0} E_n,\ E_n \in \mathcal{E} \text{ for all } n \geq 0\Big\} \ . \tag{1.5.1}$$

Then μ^* is an outer measure on X.

Here and throughout, the infimum over the empty set of positive real numbers is intended to be ∞. Observe that the conditions $\emptyset \in \mathcal{E}$ and $\rho(\emptyset) = 0$ is not restrictive at all, since it can always be met by artificially enlarging the domain of any function $\rho\colon \mathcal{E} \to [0,\infty]$, with $\mathcal{E}$ an arbitrary collection of subsets of X.

Proof. First of all, the function is well defined, which means that it clearly takes positive values, potentially ∞.

The empty set is covered by the family $E_n = \emptyset$ for all $n \geq 0$. Hence

$$0 \leq \mu^*(\emptyset) \leq \sum_{n=0}^{\infty} \rho(\emptyset) = 0 \ ,$$

and thus $\mu^*(\emptyset) = 0$.

If $A \subset B$, then every countably infinite cover of B by elements of $\mathcal{E}$ is also a cover of A, whence the infimum defining $\mu^*(A)$ is taken over a larger collection of real numbers, compared to the infimum defining $\mu^*(B)$. This establishes monotonicity.

We now show σ-subadditivity, the most delicate part of the argument. Let $(A_n)_{n\geq 0}$ be a sequence of subsets of X; fix $\varepsilon > 0$, and for every n, choose a countably infinite cover $(E_{j,n})_{j\geq 0}$ of A_n, consisting of elements of $\mathcal{E}$, such that

$$\sum_{j=0}^{\infty} \rho(E_{j,n}) \leq \mu^*(A) + \frac{\varepsilon}{2^{n+1}} \ .$$

Then $(E_{j,n})_{j,n\geq 0}$ is a countable cover of $\bigcup_{n\geq 0} A_n$, consisting of elements of $\mathcal{E}$, and satisfying

$$\sum_{j,n\geq 0} \rho(E_{j,n}) = \sum_{n\geq 0}\sum_{j\geq 0} \rho(E_{j,n}) \leq \sum_{n\geq 0} \mu^*(A_n) + \frac{\varepsilon}{2^{n+1}} = \Big(\sum_{n\geq 0} \mu^*(A_n)\Big) + \varepsilon \ .$$

It follows by definition of $\mu^*(A)$ that

$$\mu^*(A) \leq \varepsilon + \sum_{n\geq 0} \mu^*(A_n)$$

for all $\varepsilon > 0$; taking the infimum over all $\varepsilon > 0$ achieves the conclusion. □

1.5.2. Carathéodory's construction of a measure from an outer measure. Once an outer measure on X has been constructed, one would like to derive an actual measure from it. Carathéodory found the most successful construction for it, which we now turn to present. It consists of singling out those subsets of X which, together with their complement, "split" nicely the outer measure of every set, and then of restricting the outer measure to them.

Definition 1.5.4 (Measurable sets with respect to an outer measure). Let μ^* be an outer measure on a set X. A set $A \subset X$ is called **measurable with respect to** μ^*, or μ^*-**measurable**, if

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c)$$

for all $E \subset X$.

Observe that, by finite subadditivity of μ^*, it always holds that

$$\mu^*(E) \leq \mu^*(E \cap A) + \mu^*(E \cap A^c) ;$$

the actual requirement is thus that the opposite inequality holds as well.

Theorem 1.5.5. *Let μ^* be an outer measure on a set X.*

(1) The collection $\mathfrak{M}_{\mu^*}$ of μ^*-measurable subsets of X is a σ-algebra.

(2) The restriction of μ^* to $\mathfrak{M}_{\mu^*}$ is a measure on $(X, \mathfrak{M}_{\mu^*})$, and the measure space $(X, \mathfrak{M}_{\mu^*}, \mu^*|_{\mathfrak{M}_{\mu^*}})$ is complete.

Proof. We start by showing that $\mathfrak{M}_{\mu^*}$ is a σ-algebra. It is plain that $\emptyset \in \mathfrak{M}_{\mu^*}$ since, for every $E \subset X$,

$$\mu^*(E \cap \emptyset) + \mu^*(E \cap \emptyset^c) = \mu^*(\emptyset) + \mu^*(E) = \mu^*(E) .$$

Furthermore, the definition of μ^*-measurability is obviously stable under taking complements.

It remains to show that $\mathfrak{M}_{\mu^*}$ is closed under taking countable unions. Let $(A_n)_{n\geq 0}$ be a sequence of μ^*-measurable sets, and let $E \subset X$ be arbitrary. Since A_0 is μ^*-measurable,

$$\mu^*(E) = \mu^*(E \cap A_0) + \mu^*(E \cap A_0^c) .$$

Now, since A_1 is μ^*-measurable, we develop further

$$\begin{aligned} \mu^*(E) &= \mu^*(E \cap A_0) + \mu^*(E \cap A_0^c \cap A_1) + \mu^*(E \cap A_0^c \cap A_1^c) \\ &= \mu^*(E \cap A_0) + \mu^*(E \cap (A_1 \setminus A_0)) + \mu^*(E \cap (A_0 \cup A_1)^c) . \end{aligned}$$

It is straightforward to check, by induction on $n \geq 1$, that

$$\mu^*(E) = \mu^*\Big(E \cap \Big(\bigcup_{i=0}^{n} A_i\Big)^c\Big) + \sum_{m=0}^{n} \mu^*\Big(E \cap \Big(A_m \setminus \bigcup_{i=0}^{m-1} A_i\Big)\Big) ,$$

with the convention that $\bigcup_{i=0}^{-1} A_i = \emptyset$. Since

$$\Big(\bigcup_{i=0}^{n} A_i\Big)^c \supset \Big(\bigcup_{i\geq 0} A_i\Big)^c ,$$

it follows by monotonicity of μ^* that

$$\mu^*(E) \geq \mu^*\Big(E \cap \Big(\bigcup_{n\geq 0} A_n\Big)^c\Big) + \sum_{m=0}^{n} \mu^*\Big(E \cap \Big(A_m \setminus \bigcup_{i=0}^{m-1} A_i\Big)\Big) ;$$

taking the supremum over all $n \geq 0$, and applying σ-subadditivity of μ^*, we deduce that

$$\begin{aligned} \mu^*(E) &\geq \mu^*\Big(E \cap \Big(\bigcup_{n\geq 0} A_n\Big)^c\Big) + \sum_{n=0}^{\infty} \mu^*\Big(E \cap \Big(A_n \setminus \bigcup_{i=0}^{n-1} A_i\Big)\Big) \\ &\geq \mu^*\Big(E \cap \Big(\bigcup_{n\geq 0} A_n\Big)^c\Big) + \mu^*\Big(E \cap \bigcup_{n\geq 0} A_n\Big) , \end{aligned} \tag{1.5.2}$$

using for the last inequality that

$$\bigcup_{n\geq 0} E\cap\Big(A_n\setminus\bigcup_{i=0}^{n-1}A_i\Big)\big) = E\cap\bigcup_{n\geq 0}A_n\ .$$

Since E is arbitrary, we have thus shown that $\bigcup_{n\geq 0}A_n\in\mathfrak{M}_{\mu^*}$.

Suppose now, in addition, that the A_n's are pairwise disjoint. Then

$$A_n\setminus\bigcup_{i=0}^{n-1}A_i = A_n$$

for all $n\geq 0$. Apply (1.5.2) with $E=\bigcup_{n\geq 0}A_n$, so as to obtain

$$\mu^*\Big(\bigcup_{n\geq 0}A_n\Big)\geq \mu^*(\emptyset)+\sum_{n=0}^{\infty}\mu^*(A_n)=\sum_{n=0}^{\infty}\mu^*(A_n)\ ;$$

as the converse inequality holds by σ-subadditivity of μ^*, we have shown that μ^* is σ-additive on pairwise disjoint sequences of μ^*-measurable sets, that is, that $\mu^*|_{\mathfrak{M}_{\mu^*}}$ is a measure.

We are only left with the completeness claim. Suppose thus $A\subset B$ are subsets of X, B is μ^*-measurable and $\mu^*(B)=0$. We need to show that A is μ^*-measurable. Let thus $E\subset X$ be arbitrary; then

$$\mu^*(E\cap A)+\mu^*(E\cap A^{\mathrm{c}})\leq\mu^*(E\cap B)+\mu^*(E)\leq\mu^*(B)+\mu^*(E)=\mu^*(E)$$

using monotonicity of μ^* in the first two inequalities. Since E is arbitrary, we conclude that A is μ^*-measurable.

The theorem is fully established. □

1.5.3. The measure extension theorem. Let us recapitulate the discussion up to this point of the section. We start with an arbitrary "size function" $\rho\colon\mathcal{E}\to[0,\infty]$ on a collection $\mathcal{E}\subset\mathcal{P}(X)$ of "elementary" subsets of X. The function ρ induces an outer measure μ^* on X via the formula (1.5.1). In turn, μ^* can be restricted to the σ-algebra of μ^*-measurable sets, to form a measure μ on the latter. We are left to address three points, in the spirit of the discussion at the beginning of the section:

(1) is every elementary set $E\in\mathcal{E}$ μ^*-measurable?
(2) In the affirmative case, does μ give back the original size function ρ on $\mathcal{E}$?
(3) If the two previous points are met, is μ the unique extension of ρ to the σ-algebra generated by $\mathcal{E}$[14]?

If no further requirements are imposed on $\mathcal{E}$ and ρ, there is no reason for any of the previous three to hold. We shall thus impose some additional constraints on $\mathcal{E}$ and ρ, which are good enough for our purposes and yet still flexible enough to be satisfied in a wide variety of circumstances.

We begin by defining the notion of a *semiring* of subsets of a given set X.

Definition 1.5.6 (Semiring). Let X be a set. A **semiring** on X is a collection $\mathcal{E}\subset\mathcal{P}(X)$ with the following properties:

(1) $\emptyset\in\mathcal{E}$;
(2) if $A\in\mathcal{E}$ and $B\in\mathcal{E}$, then $A\cap B\in\mathcal{E}$;
(3) if $A\in\mathcal{E}$ and $B\in\mathcal{E}$, then the difference $B\setminus A$ is the disjoint union of finitely many elements of $\mathcal{E}$.

Before proceeding with the sought after measure extension theorem, we give three important examples of semirings.

[14]Strictly speaking, we should talk about the restriction of μ to $\sigma(\mathcal{E})$, for $\mathfrak{M}_{\mu^*}$ might *a priori* be bigger.

EXAMPLE 1.5.7 (Rectangles in Euclidean spaces). Let $\mathcal{E}$ be the collection of half-open **rectangles** in $\mathbb{R}^n$, namely of sets of the form

$$R = [a_1, b_1) \times \cdots \times [a_n, b_n)$$

where $a_i \leq b_i$ are real numbers for $1 \leq i \leq n$. Then $\mathcal{E}$ is a semiring. The empty set is obtained by picking $a_1 = b_1$; if

$$R_1 = \prod_{i=1}^{n} [a_i, b_i) \ , \quad R_2 = \prod_{i=1}^{n} [c_i, d_i) \ ,$$

then

$$R_1 \cap R_2 = \prod_{1 \leq i \leq n} [e_i, f_i)$$

where $e_i = \sup\{a_i, c_i\}$ and $f_i = \inf\{b_i, d_i\}$ for all $1 \leq i \leq n$. Finally, we can write

$$\begin{aligned} R_1 \setminus R_2 &= \bigcup_{1 \leq j \leq n} \Big(\prod_{i \neq j} [a_i, b_i) \times \big([a_j, b_j) \setminus [c_j, d_j) \big) \Big) \\ &= \Big(\big([a_1, b_1) \setminus [c_1, d_1) \big) \times \prod_{1 < i \leq n} [a_i, b_i) \Big) \bigsqcup \\ &\quad \bigsqcup \Big(\big([a_1, b_1) \cap [c_1, d_1) \big) \times \big([a_2, b_2) \setminus [c_2, d_2) \big) \times \prod_{2 < i \leq n} [a_i, b_i) \Big) \bigsqcup \ldots \ , \end{aligned}$$

the latter being a disjoint union of rectanges since intersections and differences of half-open intervals $[\alpha, \beta) \subset \mathbb{R}$ are of the same form.

EXAMPLE 1.5.8 (Cylinders in sequence spaces). Let A be a finite set, and consider the *sequence space* $X = A^{\mathbb{N}}$, consisting of all sequences $(x_n)_{n \geq 0}$ of elements of A. A **cylinder** in X is a set of the form

$$C^{a_1, \ldots, a_k}_{i_1, \ldots, i_k} = \{x = (x_n)_{n \geq 0} \in A^{\mathbb{N}} : x_{i_j} = a_j \text{ for all } 1 \leq j \leq k\}$$

where $k \in \mathbb{N}$, $i_1, \cdots, i_k$ are distinct natural numbers and $a_1, \ldots, a_k \in A$, with the convention that the cylinder is empty if $k = 0$. Then the collection of all cylinders in X is a semiring[15]. Verification that the intersection of two cyclinders is a cylinder is straightforward, and left to the reader; we focus on the complement property. Let thus

$$C_1 = C^{a_1, \ldots, a_k}_{i_1, \ldots, i_k} \ , \quad C_2 = C^{b_1, \ldots, b_\ell}_{j_1, \ldots, j_\ell}$$

be two cylinders. Assume that the sets of indices $\{i_1, \ldots, i_k\}$, $\{j_1, \ldots, j_\ell\}$ are disjoint; the other case is similar, and details are omitted. We can write

$$C_1 \setminus C_2 = \bigsqcup_{(c_1, \ldots, c_\ell) \in A^\ell} C^{a_1, \ldots, a_k, c_1, \ldots, c_\ell}_{i_1, \ldots, i_k, j_1, \ldots, j_\ell} \ ,$$

whence the symmetric difference is the disjoint union of finitely many cylinders, as desired.

EXAMPLE 1.5.9 (Products of measurable sets). Let $(X, \mathfrak{M}_X)$, $(Y, \mathfrak{M}_Y)$ be Borel spaces. Consider the collection of products

$$\mathcal{E} = \{E \times F : E \in \mathfrak{M}_X \ , F \in \mathfrak{M}_Y\} \subset \mathcal{P}(X \times Y) \ .$$

Then $\mathcal{E}$ is a semiring on $X \times Y$. We have $\emptyset = \emptyset \times \emptyset \in \mathcal{E}$; also, if $E_1, E_2 \in \mathfrak{M}_X$ and $F_1, F_2 \in \mathfrak{M}_Y$, then

$$(E_1 \times F_1) \cap (E_2 \times F_2) = (E_1 \cap E_2) \times (F_1 \cap F_2) \ ,$$

[15] Note, *en passant*, that the collection of cylinders is a basis for the product topology on X determined by the discrete topology $\mathcal{P}(A)$ on A.

which lies in $\mathcal{E}$ since $\mathfrak{M}_X$ and $\mathfrak{M}_Y$ are closed under finite intersections. Finally,

$$\begin{aligned}(E_1 \times F_1) \setminus (E_2 \times F_2) &= (E_1 \setminus E_2) \times F_1 \,\cup\, E_1 \times (F_1 \setminus F_2) \\ &= (E_1 \setminus E_2) \times F_1 \,\sqcup\, (E_1 \cap E_2) \times (F_1 \setminus F_2) ,\end{aligned}$$

as a direct set-theoretic verification allows to establish. Since $\mathfrak{M}_X$ and $\mathfrak{M}_Y$ are closed under complements and finite intersections, the third property of a semiring is thus proved.

REMARK 1.5.10. As regards the previous three examples, observe that half-open rectangles in $\mathbb{R}^n$ generate the Borel σ-algebra $\mathfrak{B}_{\mathbb{R}^n}$, cylinders in $X = A^{\mathbb{N}}$ generate the product σ-algebra $\mathcal{P}(A)^{\otimes\mathbb{N}} := \bigotimes_{n\geq 0} \mathcal{P}(A)$ (which, by virtue of Proposition 1.3.18, coincides with the Borel σ-algebra induced by the product topology on X), and product sets in $X \times Y$ generate the product σ-algebra $\mathfrak{M}_X \otimes \mathfrak{M}_Y$. These facts shall be relevant when applying the forthcoming Theorem 1.5.11 to manufacture important examples of measures on the three given spaces.

We are now ready to state and prove the measure extension theorem, which is the fundamental tool we will avail ourselves of in successive sections and chapters to construct the Lebesgue measure on $\mathbb{R}^n$, the *Bernoulli measure* on sequence spaces, and *product measures* on product sets. In order to better understand the role of the assumptions we shall place on the starting size function $\rho\colon \mathcal{E} \to [0, \infty]$ defined on a semiring $\mathcal{E}$, observe that, for ρ to extend to a measure μ on the σ-algebra $\sigma(\mathcal{E})$, it is necessary that at least the two following conditions are fulfilled:

(1) ρ is finitely additive: if $(E_i)_{i\in I}$ is a finite family of pairwise disjoint elements of $\mathcal{E}$ with $\bigcup_{i\in I} E_i \in \mathcal{E}$, then

$$\rho\Big(\bigcup_{i\in I} E_i\Big) = \sum_{i\in I} \rho(E_i) \;;$$

(2) ρ is σ-subadditive: if $E \in \mathcal{E}$ and $E \subset \bigcup_{n\geq 0} E_n$ with $E_n \in \mathcal{E}$ for all $n \geq 0$, then

$$\rho(E) \leq \sum_{n=0}^{\infty} \rho(E_n) \;.$$

The reason for them to hold is that any measure μ restricting to ρ on $\mathcal{E}$ satisfies the previous two properties. It turns out that the latter are sufficient for our purposes.

THEOREM 1.5.11 (Measure extension theorem). *Let X be a set, $\mathcal{E}$ a semiring of subsets of X, $\rho\colon \mathcal{E} \to [0, \infty]$ a function which is finitely additive and σ-subadditive as defined above, and satisfies $\rho(\emptyset) = 0$. Let μ^* be the outer measure on X determined by ρ as in Proposition 1.5.3. Then the following assertions hold.*

(i) Every set $A \in \mathcal{E}$ is μ^*-measurable.
(ii) The restriction of μ^* to $\mathcal{E}$ coincides with ρ.

The theorem thus allows to construct a measure μ on a set X starting from any finitely additive, σ-subadditive size function ρ defined on a given semiring $\mathcal{E}$ on X. Such a measure is defined on a σ-algebra containing $\mathcal{E}$, namely the σ-algebra of μ^*-measurable sets, where μ^* is the outer measure determined by ρ as in the theorem; furthermore, μ equals the size function ρ on $\mathcal{E}$. Actually, μ comes extended to a complete measure on the (potentially larger) σ-algebra of μ^*-measurable subsets.

PROOF. We start by proving that every $A \in \mathcal{E}$ is μ^*-measurable. Let E be an arbitrary subset of X; we need to show that

$$\mu^*(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^{\mathrm{c}}) \;. \tag{1.5.3}$$

Fix a covering $(A_n)_{n\geq 0}$ of E consisting of elements of $\mathcal{E}$. Then $(A_n \cap A)_{n\geq 0}$ is a covering of $E \cap A$ and $(A_n \cap A^{\mathrm{c}})_{n\geq 0}$ is a covering of $E \cap A^{\mathrm{c}}$. Whereas $A_n \cap A \in \mathcal{E}$ for all $n \geq 0$, directly from the fact that $\mathcal{E}$ is closed under intersections, this is not necessarily the case for $A_n \cap A^{\mathrm{c}}$; however, we know that, since $\mathcal{E}$ is a semiring $A_n \cap A^{\mathrm{c}} = A_n \setminus A$ is a union of pairwise disjoint sets $B_{n,1}, \dots, B_{n,k_n} \in \mathcal{E}$, whence $(B_{n,j})_{n\geq 0, 1\leq j\leq k_n}$ is a countable cover of $E \cap A^c$ consisting of elements of $\mathcal{E}$. It thus follows from the definition of μ^* that

$$\mu^*(E\cap A) \leq \sum_{n\geq 0} \rho(A_n \cap A) , \quad \mu^*(E\cap A^c) \leq \sum_{n\geq 0} \sum_{1\leq j\leq k_n} \rho(B_{n,j}) .$$

Now additivity of ρ yields, for all $n \geq 0$,

$$\rho(A_n \cap A) + \sum_{1\leq j\leq k_n} \rho(B_{n_j}) = \rho(A_n) .$$

Combining the last two displayed inequalities, we deduce

$$\mu^*(E\cap A) + \mu^*(E\cap A^c) \leq \sum_{n\geq 0} \rho(A_n) .$$

The covering $(A_n)_{n\geq 0}$ being arbitrary, this enables us to infer (1.5.3) by definition of $\mu^*(E)$.

We now show that $\mu^*|_{\mathcal{E}} = \rho$. The definition of μ^* delivers directly $\mu^*(E) \leq \rho(E)$ for all $E \in \mathcal{E}$, taking the trivial covering $(E_n)_{n\geq 0}$ of E consisting of $E_0 = E$, $E_n = \emptyset$ for all $n > 0$. The converse inequality is also a straightforward consequence of the assumptions; fix a covering $(E_n)_{n\geq 0}$ of E with $E_n \in \mathcal{E}$ for all $n \geq 0$, and apply σ-subadditivity of ρ, so as to get

$$\rho(E) \leq \sum_{n\geq 0} \rho(E_n) .$$

Since $(E_n)_{n\geq 0}$ is arbitrary, we deduce that $\rho(E) \leq \mu^*(E)$.

The proof is complete. □

1.5.4. Monotone class lemma and uniqueness in the measure extension theorem. In terms of our discussion at the beginning of the present section, it remains to address the uniqueness problem, namely whether the restriction $\mu^*|_{\sigma(\mathcal{E})}$ is the unique measure extending ρ to $\sigma(\mathcal{E})$. Under the assumptions of Theorem 1.5.11, this is not necessarily the case.

Example 1.5.12. Let X be an uncountable set,

$$\mathcal{E} = \{\{x\} : x \in X\} \cup \{\emptyset\} ,$$

and define $\rho(E) = 0$ for all $E \in \mathcal{E}$. It is clear that $\mathcal{E}$ is a semiring on X and that ρ satisfies the assumptions of Theorem 1.5.11. A moment's thought allows to realize that the σ-algebra $\sigma(\mathcal{E})$ generated by $\mathcal{E}$ is the σ-algebra of countable and co-countable sets, introduced in Example 1.3.5. For every $c \in \mathbb{R}_{\geq 0}$, define a set function $\mu_c \colon \sigma(\mathcal{E}) \to [0, \infty]$ via

$$\mu_c(A) = \begin{cases} 0 & \text{if } A \text{ is countable} \\ c & \text{if } A \text{ is co-countable} \end{cases} .$$

In Example 1.4.5, we showed that μ_c is a measure for $c = 1$; the same argument applies to any c. Now every measure μ_c extends ρ on $\mathcal{E}$, whence ρ admits an uncountable family of distinct extensions to a measure on $\sigma(\mathcal{E})$.

It suffices, however, to add an appropriate σ-finiteness assumption to the hypotheses of Theorem 1.5.11 to get uniqueness.

Theorem 1.5.13 (Uniqueness in the measure extension theorem). *Let assumptions and notation be as in Theorem 1.5.11. Suppose, in addition, that ρ is σ-finite, in the following sense: X admits a countable covering $(X_i)_{i\in\mathbb{N}}$ consisting of elements $X_i \in \mathcal{E}$ with $\rho(X_i) < \infty$ for all $i \in \mathbb{N}$. Let ν be a measure on the Borel space $(X, \sigma(\mathcal{E}))$ such that $\nu|_{\mathcal{E}} = \rho$. Then*

$$\nu = \mu^*|_{\sigma(\mathcal{E})} .$$

Remark 1.5.14. Under the assumptions of the theorem, the measure $\mu^*|_{\sigma(\mathcal{E})}$ is obviously σ-finite.

The proof of Theorem 1.5.13 hinges crucially upon a very important measure-theoretic fact[16], known as the *monotone class lemma*. Before stating and proving it, let us ask ourselves a natural question: suppose μ, ν are measures on a Borel space $(X, \mathfrak{M})$, what can we say about the properties of the collection

$$\mathscr{M} = \{E \in \mathfrak{M} : \mu(E) = \nu(E)\}$$

of measurable sets on which the two measures coincide? This leads us naturally to make the following definition.

DEFINITION 1.5.15 (Monotone class). Let X be a set. A **monotone class** on X is a collection $\mathscr{M} \subset \mathcal{P}(X)$ satisfying the following properties:

(1) $X \in \mathscr{M}$;
(2) if $A \subset B$ are elements of $\mathscr{M}$, then $B \setminus A \in \mathscr{M}$;
(3) if $(A_n)_{n\geq 0}$ is an increasing sequence of elements of $\mathscr{M}$, then $\bigcup_{n\geq 0} A_n \in \mathscr{M}$.

Thus, for instance, any σ-algebra is a monotone class, but the converse does not hold. Monotone classes arise naturally as sets of coincidence for two measures.

LEMMA 1.5.16. *Let $(X, \mathfrak{M})$ be a measurable space, μ and ν finite measures on $(X, \mathfrak{M})$ with $\mu(X) = \nu(X)$. Then the collection*

$$\mathscr{M} = \{E \in \mathfrak{M} : \mu(E) = \nu(E)\}$$

is a monotone class.

PROOF. The proof is elementary. First, $X \in \mathscr{M}$ directly from the assumption. Secondly, if $A \subset B$ satisfy $\mu(A) = \nu(A)$ and $\mu(B) = \nu(B)$, then finite additivity of measures gives

$$\mu(B \setminus A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B \setminus A) ,$$

using crucially in the first and last equality that μ and ν are finite measures. We have thus shown that $B \setminus A \in \mathscr{M}$. Finally, if $(A_n)_{n\geq 0}$ is an increasing sequence of measurable sets with $\mu(A_n) = \nu(A_n)$ for all $n \geq 0$, continuity from below of measures yields

$$\mu\Big(\bigcup_{n\geq 0} A_n\Big) = \lim_{n\to\infty} \mu(A_n) = \lim_{n\to\infty} \nu(A_n) = \nu\Big(\bigcup_{n\geq 0} A_n\Big) ,$$

whence $\bigcup_{n\geq 0} A_n \in \mathscr{M}$. □

Just as for σ-algebras, it is plain that the intersection of an arbitrary family of monotone classes is a monotone class. Thus, we can speak of the monotone class generated by an arbitrary collection $\mathcal{E} \subset \mathcal{P}(X)$, indicated with $\mathscr{M}(\mathcal{E})$, defined as the intersection of all monotone classes containing $\mathcal{E}$, $\mathcal{P}(X)$ being always among those. It is the coarsest monotone class containing $\mathcal{E}$.

Since every monotone class is a σ-algebra, it always holds that $\mathscr{M}(\mathcal{E}) \subset \sigma(\mathcal{E})$; equality fails in general, but holds in a very important case.

THEOREM 1.5.17 (Monotone class lemma). *Let X be a set, $\mathcal{E} \subset \mathcal{P}(X)$ a collection of subsets which is closed under finite intersections: for all $E, F \in \mathcal{E}$, $E \cap F \in \mathcal{E}$. Then*

$$\mathscr{M}(\mathcal{E}) = \sigma(\mathcal{E}) .$$

PROOF. Assigned in exercises. □

We are now in a position to prove uniqueness in the measure extension theorem.

PROOF OF THEOREM 1.5.13. Let ν be a measure on $\sigma(\mathcal{E})$ satisfying $\nu|_{\mathcal{E}} = \rho$, and set $\mu = \mu^*|_{\sigma(\mathcal{E})}$. We aim to show that $\mu = \nu$.

[16] It is, for instance, exceedingly relevant in probability theory.

Let $(X_i)_{i\in\mathbb{N}}$ be a countable cover of X with $X_i \in \mathcal{E}$ and $\rho(X_i) < \infty$ for all $i \in \mathbb{N}$. We claim that, without loss of generality, we can assume that the X_i's are disjoint. To see this, observe that it is possible to first replace the family $(X_i)_{i\in\mathbb{N}}$ with the family $(X_i')_{i\in\mathbb{N}}$ defined as

$$X_0' = X_0 \ , \quad X_n' = X_n \setminus \bigcup_{0\le m\le n-1} X_m = \bigcap_{0\le m\le n-1} X_n \setminus X_m \quad \text{for all } n \ge 1;$$

now, since $\mathcal{E}$ is a semiring, we can write, for all $0 \le m \le n-1$,

$$X_n \setminus X_m = \bigsqcup_{\lambda\in\Lambda_m} E_\lambda$$

with Λ_m a finite index set and $E_\lambda \in \mathcal{E}$ for all $\lambda \in \Lambda_m$. Distributivity of the intersection with respect to the union gives

$$\bigcap_{0\le m\le n-1} \bigsqcup_{\lambda\in\Lambda_m} E_\lambda = \bigsqcup_{(\lambda_0,\dots,\lambda_{n-1})\in\Lambda_0\times\cdots\times\Lambda_{n-1}} E_{\lambda_0} \cap \cdots \cap E_{\lambda_{n-1}} \ ,$$

where each element of the last disjoint union is in $\mathcal{E}$. We have thus expressed each X_n' as a finite disjoint union of elements of $\mathcal{E}$, all of which together (as n ranges in $\mathbb{N}$) can be reordered and taken as the new σ-finite cover of X, consisting now of pairwise disjoint sets.

For all $i \in \mathbb{N}$, define two measures μ_i and ν_i on $\sigma(\mathcal{E})$ by

$$\mu_i(A) = \mu(A \cap X_i) \ , \quad \nu_i(A) = \nu(A \cap X_i) \ .$$

Verification that the axioms of a measure are satisfied is straightforward, and thus left to the reader. We first claim that, if we show $\mu_i = \nu_i$ for all $i \in \mathbb{N}$, then we can conclude the proof. Indeed, we would then have, for all $A \in \sigma(\mathcal{E})$,

$$\mu(A) = \sum_{i\in\mathbb{N}} \mu(A \cap X_i) = \sum_{i\in\mathbb{N}} \mu_i(A) = \sum_{i\in I} \nu_i(A) = \sum_{i\in\mathbb{N}} \nu(A \cap X_i) = \nu(A) \ .$$

Let us thus fix an $i \in \mathbb{N}$. The family

$$\mathscr{M}_i = \{B \in \sigma(\mathcal{E}) : \mu_i(B) = \nu_i(B)\}$$

is a monotone class in light of Lemma 1.5.16; observe, for this purpose, that μ_i and ν_i are finite and

$$\mu_i(X) = \mu(X_i) = \rho(X_i) = \nu(X_i) = \nu_i(X)$$

since both μ and ν extend ρ. Furthermore, $\mathscr{M}_i$ contains $\mathcal{E}$: if $E \in \mathcal{E}$, then $E \cap X_i \in \mathcal{E}$ and thus

$$\mu_i(E) = \mu(E \cap X_i) = \rho(E \cap X_i) = \nu(E \cap X_i) = \nu_i(E) \ .$$

Since $\mathcal{E}$ is a semiring, it is closed under finite intersections; by the monotone class lemma (Theorem 1.5.17), we conclude that $\mathscr{M}_i$ contains $\mathscr{M}(\mathcal{E}) = \sigma(\mathcal{E})$, and is thus equal to $\sigma(\mathcal{E})$. This finishes the proof. □

1.5.5. A first example: Bernoulli measures on sequence spaces. Our first application of the measure extension theorem is of paramount relevance in probability theory: it concerns the construction of an appropriate probabilistic model for a random experiment, with finitely many outcomes, performed an infinite number of times, each independently from the previous trials.

Consider a sequence space $X = A^{\mathbb{N}}$, where A is a finite set; we have already introduced them in Example 1.5.8. Think of A as the set of possible outcomes of one single experiment. Each element of X is thus an infinite sequence of independent experiments. Let $p = (p_a)_{a\in A} \in \mathbb{R}^A$ be a *probability vector*, namely $p_a \ge 0$ for all $a \in A$ and $\sum_{a\in A} p_a = 1$. For all $a \in A$, p_a represents the probability we assign to the outcome a of a single experiment. If we want different trials to be independent, it is natural to assign a probability

$$\rho(C_{i_1,\dots,i_k}^{a_1,\dots,a_k}) = p_{a_1} \cdots p_{a_k}$$

to every cylinder $C_{i_1,\dots,i_k}^{a_1,\dots,a_k} \subset X$. If $\mathcal{E}$ is the family of all cylinders, then the previous assignment defines our starting size function $\rho \colon \mathcal{E} \to [0, 1]$.

Notice first that $\mathcal{E}$ generates the Borel σ-algebra $\mathfrak{B}_X$ determined by the topology τ defined as the product topology of the discrete topologies $\mathcal{P}(A)$ on each factor. Equivalently, in view of Proposition 1.3.18, $\mathfrak{B}_X = \mathcal{P}(A)^{\otimes\mathbb{N}}$ is the product of the discrete σ-algebras $\mathcal{P}(A)$ on each factor. To see that $\mathcal{E}$ generates $\mathfrak{B}_X$, observe first that each cylinder C is open for the product topology, being a finite intersections of preimages $\pi_n^{-1}(a)$ for $a \in A$ and $\pi_n \colon X \to A$ the projection onto the n-th coordinate. More is true, namely cylinders form a basis for the topology τ; the elementary verification is left to the reader. It follows that every open set for τ is a (necessarily countable, as there are countably many distinct cylinders) union of cylinders, and thus contained in $\sigma(\mathcal{E})$, the latter being hence equal to $\mathfrak{B}_X$.

In order to apply Theorem 1.5.11, we need to verify that ρ is finitely additive and σ-subadditive on cylinders. The argument for finite additivity is much akin in spirit to the one for σ-subadditivity, to which we thus turn. Let C be a cylinder covered by a sequence of cylinders $(C_n)_{n\geq 0}$. Observe that each cylinder is not only open, as already pointed out, but also closed, as the complement of a cylinder can easily be expressed as a finite union of cylinders. The topological space (X, τ) being compact, as product of finite (and thus compact) topological spaces (cf. Proposition A.3.2), every cylinder is compact as well. As a consequence, there is a finite subcollection $(C_j)_{j\in J}$ of $\{C_n : n \in \mathbb{N}\}$ which covers C, and it will clearly suffice to show that

$$\rho(C) \leq \sum_{j\in J} \rho(C_j)$$

in order to establish σ-subadditivity. Let thus

$$C = C^{a_1,\dots,a_k}_{i_1,\dots,i_k} \ , \quad C_j = C^{b_{1,j},\dots,b_{k_j,j}}_{i_{1,j},\dots,i_{k_j,j}}$$

for all $j \in J$. Let N denote the supremum of all indices

$$i_1, \dots, i_k \text{ and } i_{1,j}, \dots, i_{k_j,j} \text{ for all } j \in J;$$

Then we can consider the finite set $X' = A^{\{0,\dots,N\}}$, equipped with the discrete topology, and endow it with the probability measure μ_N defined by

$$\mu_N(a_0, \dots, a_N) = p_0 \cdots p_N$$

for all $(a_0, \dots, a_N) \in X'$[17]. Now the cylinders C, C_j correspond naturally to *cylinders* C', C'_j in X', obtained simply by discarding the tail $(a_n)_{n>N}$ of each word $(a_n)_{n\geq 0} \in X$. Obviously, C' is contained in the union of the C'_j's, and finite subadditivity for a measure yields

$$\mu_N(C') \leq \sum_{j\in J} \mu_N(C'_j) \ .$$

Finally, observe that an elementary calculation with finite sums, using the fact that

$$\sum_{(a_1,\dots,a_k)\in A^k} p_{a_1} \cdots p_{a_k} = \Big(\sum_{a\in A} p_a\Big)^k = 1$$

for all integers $k \geq 1$, yields

$$\mu_N(C') = \rho(C) \quad \text{and} \quad \mu_N(C'_j) = \rho(C_j) \quad \text{for all } j \in J,$$

from which our claim follows.

Combining Theorems 1.5.11 and 1.5.13, and observing as a last point that the size function ρ is obviously σ-finite (in fact, it is finite), the outcome of the discussion in this subsection if the following statement.

[17] It is an instance of the measures determined by functions, seen in Example 1.4.4.

PROPOSITION 1.5.18. *Let A be a finite set, $X = A^{\mathbb{N}}$, equipped with the σ-algebra $\mathfrak{B}_X = \mathcal{P}(X)^{\otimes\mathbb{N}}$. Let $p = (p_a)_{a\in A}$ be a probability vector. There exists a unique measure μ on $(X, \mathfrak{B}_X)$ such that*

$$\mu(\{(x_n) \in X : x_{i_1} = a_1, \ldots, x_{i_k} = a_k\}) = p_{a_1} \cdots p_{a_k}$$

for every $\{i_1, \ldots, i_k\} \subset \mathbb{N}$ and $a_1, \ldots, a_k \in A$. Moreover, μ is a probability measure.

The measure μ is called the **Bernoulli measure** on X determined by the probability vector p; it is routinely denoted by $p^{\mathbb{N}}$, a notation motivated by the setting of product measures which is the subject of Chapter **??**.

1.6. The Lebesgue measure on $\mathbb{R}^n$

Our prime application of Theorem 1.5.11 is the construction of the Lebesgue measure on Euclidean spaces $\mathbb{R}^n$. We let $\mathcal{E} \subset \mathcal{P}(\mathbb{R}^n)$ be the family of half-open rectangles of the form

$$R = [a_1, b_1) \times \cdots \times [a_n, b_n)$$

where $a_1 \le b_1, \ldots, a_n \le b_n$ are real numbers and the rectangle is meant to be empty as soon as $a_i = b_i$ for some $1 \le i \le n$. For every $R = \prod_{i=1}^n [a_i, b_i) \in \mathcal{E}$, we set

$$\rho(R) = \prod_{i=1}^{n} (b_i - a_i) .$$

We have already shown in Example 1.5.7 that $\mathcal{E}$ is a semiring; also, trivially $\rho(\emptyset) = 0$. Before pursuing the verification that ρ is finitely additive and σ-subadditive, let us immediately settle the following point.

LEMMA 1.6.1. *The collection $\mathcal{E}$ generates the Borel σ-algebra on $\mathbb{R}^n$.*

PROOF. First, every element

$$R = \prod_{i=1}^{n} [a_i, b_i) \in \mathcal{E}$$

is Borel-measurable, as it is the countable intersection of the open sets

$$\prod_{i=1}^{n} (a_i - 1/k, b_i) , \quad k \ge 1 \text{ integer.}$$

On the other hand, every open rectangle

$$\prod_{i=1}^{n} (a_i, b_i)$$

lies in $\sigma(\mathcal{E})$, being the countable union of the half-open rectangles

$$\prod_{i=1}^{n} [a_i + 1/k, b_i) , \quad k \ge 1 \text{ integer ;}$$

The collection of open rectangles is a basis for the Euclidean topology on $\mathbb{R}^n$. Since $\mathbb{R}^n$ is second countable, every open cover of a subset $Y \subset \mathbb{R}^n$ admits a countable subcover, and thus every open set is a countable union of open rectangles. As such, it is contained $\sigma(\mathcal{E})$, from which we concludee that $\mathfrak{B}_{\mathbb{R}^n} \subset \sigma(\mathcal{E})$, as desired. ☐

We now address the task of showing that ρ is finitely additive and σ-subadditive. As we shall see, σ-subadditivity shall follow from finite subadditivity via a compactness argument as in §1.5.5; in turn, finite subadditivity will be a fairly straightforward consequence of finite additivity. We thus begin with the latter property. Let $R = \prod_{1\le i\le n} [a_i, b_i) \in \mathcal{E}$ be a rectangle which is the finite disjoint union of rectangles $R_j \in \mathcal{E}$, $j \in J$. The crucial geometric observation, which considerably simplifies computations, is that we can further partition each R_j as a finite disjoint union of rectangles $R_\lambda \in \mathcal{E}$, $\lambda \in \Lambda_j$ (with the Λ_j's being pairwise disjoint sets of indices) in such a way that the resulting collection $(R_\lambda)_{\lambda\in\bigcup_{j\in J}\Lambda_j}$, whose union is R, consists of rectangles arranged in a "grid", in the following sense: for all $1 \le i \le n$, there is a partition of $[a_i, b_i)$ into finitely many half-open intervals I_α, $\alpha \in A_i$, such that for all $\lambda \in \bigcup_{j\in J}\Lambda_j$ we can write

$$R_\lambda = \prod_{1\le i\le n} I_{\alpha_i}$$

for some $\alpha_i \in A_i$, $1 \le i \le n$. The same "grid-property" holds for all collections $(R_\lambda)_{\lambda\in\Lambda_j}$, $j \in J$, as well. Figure **??** illustrates, in two dimensions, how to obtain such a grid-like configuration starting from an arbitrary cover by rectangles.

Now showing finite additivity of ρ when the disjoint union consists of rectangles arranged in a grid is an elementary matter, simply involving distributivity of the product with respect to the sum of real numbers. Verification of this fact is left to the reader. Using it, we get

$$\rho(R) = \sum_{\lambda\in\bigcup_{j\in J}\Lambda_j} \rho(R_\lambda) = \sum_{j\in J}\sum_{\lambda\in\Lambda_j} \rho(R_\lambda) = \sum_{j\in J} \rho(R_j)\ ,$$

which is the desired finite additivity.

Let us now show how finite additivity implies finite subadditivity[18], namely that if $R \in \mathcal{E}$ is contained in a finite union of sets $R_j \in \mathcal{E}$, $0 \le j \le J$, then

$$\rho(R) \le \sum_{0\le j\le J} \rho(R_j)\ .$$

Upon replacing R_j with $R_j \cap R$, which is an element of $\mathcal{E}$, and using monotonicity of ρ on rectangles (which follows readily from finite additivity and the semiring properties), we may assume without loss of generality that $R = \bigcup_{0\le j\le J} R_j$. Resorting to the same argument as in the beginning of the proof of Theorem 1.5.13, we define

$$R'_0 = R_0\ , \quad R'_j = R_j \setminus \bigcup_{0\le i\le j-1} R_i \quad \text{for all } 1 \le j \le J\ ,$$

noting that the union of the R'_j's is R, and write every R'_j[19] as

$$R'_j = \bigsqcup_{\lambda\in\Lambda_j} R_\lambda$$

for some finite index set Λ_j and some $R_\lambda \in \mathcal{E}$ for all $\lambda \in \Lambda_j$. Similarly, for all $0 \le j \le J$, we have

$$R_j \setminus R'_j = R_j \cap \Big(\bigcup_{0\le i\le j-1} R_i\Big) = \bigcup_{0\le i\le j-1} R_j \cap R_i\ ,$$

whence $R_j \setminus R'_j$ is a finite union of elements of $\mathcal{E}$, and we can thus write

$$R_j \setminus R'_j = \bigsqcup_{\sigma\in\Sigma_j} R_\sigma$$

for some finite set Σ_j and some $R_\sigma \in \mathcal{E}$ for all $\sigma \in \Sigma_j$. Note that both families $(\Lambda_j)_{0\le j\le J}$ and $(\Sigma_j)_{0\le j\le J}$ consists of pairwise disjoint sets. Now finite additivity of ρ gives us

$$\rho(R) = \sum_{0\le j\le J}\sum_{\lambda\in\Lambda_j} \rho(R_\lambda) \le \sum_{0\le j\le J}\ \sum_{\lambda\in\Lambda_j,\sigma\in\Sigma_j} \rho(R_\lambda) + \rho(R_\sigma) = \sum_{0\le j\le J} \rho(R_j)\ ,$$

which is the sought after inequality.

[18]What follows is a general argument, holding for any finitely additive function ρ defined on a semiring of subsets of a given set X.

[19]Notice that, despite the notation, each R'_j is not necessarily a rectangle.

Finally, let us demonstrate how a compactness argument enables us to extend finite subadditivity to σ-subadditivity. Let

$$R = \prod_{1\le i\le n} [a_i, b_i)$$

be contained in the union of a sequence $(R_m)_{m\in\mathbb{N}}$ with

$$R_m = \prod_{1\le i\le n} [a_{i,m}, b_{i,m}) \ ,$$

and fix $\varepsilon > 0$. Pick $b_i(\varepsilon) \in (a_i, b_i)$[20], for all $1 \le i \le n$, in such a way that

$$\rho\Big(\prod_{1\le i\le n} [a_i, b_i(\varepsilon))\Big) \ge \rho(R) - \frac{\varepsilon}{2} \ ; \tag{1.6.1}$$

similarly, for all $m \in \mathbb{N}$ and $1 \le i \le n$, pick $a_{i,m}(\varepsilon) < a_{i,m}$ in such a way that

$$\rho\Big(\prod_{1\le i\le n} [a_{i,m}(\varepsilon), b_{i,m})\Big) \le \rho(R_m) + \frac{\varepsilon}{2^{m+2}} \ . \tag{1.6.2}$$

The compact rectangle

$$\prod_{1\le i\le n} [a_i, b_i(\varepsilon)]$$

is contained in the union, as m ranges in $\mathbb{N}$, of the open rectangles

$$\prod_{1\le i\le n} (a_{i,m}(\varepsilon), b_{i,m}) \ ;$$

therefore, there is some $M \in \mathbb{N}$ such that

$$\prod_{1\le i\le n} [a_i, b_i(\varepsilon)) \subset \prod_{1\le i\le n} [a_i, b_i(\varepsilon)] \subset \bigcup_{0\le m\le M} \prod_{1\le i\le n} (a_{i,m}(\varepsilon), b_i) \subset \bigcup_{0\le m\le M} \prod_{1\le i\le n} [a_{i,m}(\varepsilon), b_{i,m}) \ .$$

Finite subadditivity of ρ, in conjunction with (1.6.1) and (1.6.2), gives

$$\begin{aligned} \rho(R) &\le \frac{\varepsilon}{2} + \rho\Big(\prod_{1\le i\le n} [a_i, b_i(\varepsilon))\Big) \le \frac{\varepsilon}{2} + \sum_{0\le m\le M} \rho\Big(\prod_{1\le i\le n} [a_{i,m}(\varepsilon), b_{i,m})\Big) \\ &\le \frac{\varepsilon}{2} + \varepsilon \sum_{0\le m\le M} \frac{1}{2^{m+2}} + \sum_{0\le m\le M} \rho(R_m) \le \varepsilon + \sum_{m\ge 0} \rho(R_m) \ . \end{aligned}$$

Taking the infimum over all $\varepsilon > 0$ achieves the conclusion.

The function ρ is obviously σ-finite: $\mathbb{R}^n$ is the union of the half-open rectangles $[-N, N)^n$, $N \in \mathbb{N}$. The combination of Theorems 1.5.11 and 1.5.13 allows us to conclude:

THEOREM 1.6.2. *For every integer $n \ge 1$, there exists a unique measure $\mathscr{L}^n$ on the Borel space $(\mathbb{R}^n, \mathfrak{B}_{\mathbb{R}^n})$ such that*

$$\mathscr{L}^n\big([a_1, b_1) \times \cdots \times [a_n, b_n)\big) = (b_1 - a_1)\cdots(b_n - a_n)$$

for all real numbers $a_1 < b_1, \dots, a_n < b_n$.

DEFINITION 1.6.3 (Lebesgue measure on $\mathbb{R}^n$). The unique measure $\mathscr{L}^n$ whose existence is guaranteed by Theorem 1.6.2 is called the **Lebesgue measure** on $(\mathbb{R}^n, \mathfrak{B}_{\mathbb{R}^n})$.

We shall also indicate with $\mathscr{L}^n$ the **Lebesgue outer measure**, that is, the outer measure determined (as in (1.5.1), as usual) by the volume function ρ defined on half-open rectangles. Thus, $\mathscr{L}^n$ is a function defined on $\mathcal{P}(\mathbb{R}^n)$, extending ρ and restricting to a measure on $\mathfrak{B}_{\mathbb{R}^n}$. Observe that, actually, and directly from Theorem 1.5.11, $\mathscr{L}^n$ restricts to a complete measure (which, again, we call $\mathscr{L}^n$, specifying every time the domain on which we are considering it) on a larger Borel space $(X, \mathfrak{M}_{\mathscr{L}^n})$, where $\mathfrak{M}_{\mathscr{L}^n}$ is the σ-algebra of $\mathscr{L}^n$-measurable subsets,

[20]Without loss of generality $a_i < b_i$ for all $1 \le i \le n$, else there is nothing to prove.

which we shall henceforth refer to as **Lebesgue measurable sets**. We shall prove in §**??** that the are Lebesgue measurable sets which are not Borel measurable, that is, that $\mathfrak{B}_{\mathbb{R}^n}$ is strictly contained in $\mathfrak{M}_{\mathscr{L}^n}$, and that $(\mathbb{R}^n, \mathfrak{M}_{\mathscr{L}^n}, \mathscr{L}^n)$ is the completion of $(\mathbb{R}^n, \mathfrak{B}_{\mathbb{R}^n}, \mathscr{L}^n)$.

If $f\colon \mathbb{R}^n \to \mathbb{C}$ is a function, we say that f is **Lebesgue-measurable** if it is measurable with respect to $\mathfrak{M}_{\mathscr{L}^n}$ on the domain (and $\mathfrak{B}_{\mathbb{C}}$, as usual, on the target set).

Remark 1.6.4. We will show this more constructively, but the fact that $\mathfrak{B}_{\mathbb{R}^n}$ is strictly contained in $\mathfrak{M}_{\mathscr{L}^n}$ (and, as a matter of fact, much smaller than it) could also be shown via a pure cardinality argument. It is a fact that the σ-algebra generated by an infinite collection of sets having at most the cardinality of the *continuum*, has the cardinality of the *continuum* (we refer the interested reader to [9, §1.6]). This applies in particular to the Borel σ-algebra $\mathfrak{B}_{\mathbb{R}^n}$, which is generated by open rectangles. On the other hand, we shall shortly see that there are Borel subsets of $\mathbb{R}^n$ with zero Lebesgue measure and the cardinality of the *continuum*. If $C \subset \mathbb{R}^n$ is any such set, then $\mathcal{P}(C) \subset \mathfrak{M}_{\mathscr{L}^n}$ by completeness. Now it is a well known set-theoretic fact that the cardinality of $\mathcal{P}(C)$ is strictly larger than the one of C, and thus $\mathfrak{M}_{\mathscr{L}^n}$ has cardinality strictly larger than the *continuum*.

Remark 1.6.5. Since every open rectangle $R = \prod_{1\le i\le n}(a_i, b_i)$ can be written as

$$\bigcup_{k\in\mathbb{N}^*} \prod_{1\le i\le n} [a_i + 1/k, b_i) \ ,$$

continuity from below of $\mathscr{L}^n$ gives

$$\mathscr{L}^n(R) = \lim_{k\to\infty} \prod_{1\le i\le n} (b_i - a_i - 1/k) = \prod_{1\le i\le n} (b_i - a_i) \ .$$

Similarly, any closed rectangle $R' = \prod_{1\le i\le n}[a_i, b_i]$ can be written as

$$\bigcap_{k\in\mathbb{N}^*} \prod_{1\le i\le n} [a_i, b_i + 1/k) \ ,$$

and continuity from below of $\mathscr{L}^n$ yields

$$\mathscr{L}^n(R') = \prod_{1\le i\le n} (b_i - a_i) \ .$$

Thus, as naturally expected, the Lebesgue measure of rectangles is not affected by whether we include endpoints of the intervals or not.

We collect the fundamental properties of the Lebesgue measure in the following proposition, phrasing them in their utmost generality, namely considering the Lebesgue outer measure.

Proposition 1.6.6. *Let $\mathscr{L}^n$ be the Lebesgue outer measure on $\mathbb{R}^n$. The following assertions hold.*

(1) $\mathscr{L}^n|_{\mathfrak{M}_{\mathscr{L}^n}}$ is an infinite, σ-finite measure.

(2) $\mathscr{L}^n$ is translation-invariant: for every set $E \subset \mathbb{R}^n$ and every $x \in \mathbb{R}^n$,

$$\mathscr{L}^n(E + x) = \mathscr{L}^n(E) \ .$$

(3) $\mathscr{L}^n$ scales as follows under homotheties: for every set $E \subset \mathbb{R}^n$ and every $\lambda \in \mathbb{R}$,

$$\mathscr{L}^n(\lambda E) = |\lambda|^n \mathscr{L}^n(E) \ .$$

(4) $\mathscr{L}^n(A) = 0$ for every countable set[21] *$A \subset \mathbb{R}^n$.*

(5) $\mathscr{L}^n(O) > 0$ for every non-empty open set $O \subset \mathbb{R}^n$.

(6) $\mathscr{L}^n(K) < \infty$ for every compact set $K \subset \mathbb{R}^n$.

Proof. It is assigned in the exercises. □

[21] Notice that countable sets are countable unions of singletons, and as such F_σ sets: in particular, they are Borel measurable.

Remark 1.6.7. (1) In subsequent chapters, we shall see how $\mathscr{L}^n$ behaves, more generally with respect to (2) and (3) in the proposition, with respect to *diffeomorphisms* between open subsets of $\mathbb{R}^n$. In particular, we shall prove that $\mathscr{L}^n$ is invariant under all Euclidean isometries.

(2) When examining regularity properties of Borel measures on Euclidean spaces, we shall prove that $\mathscr{L}^n$, as an outer measure, satisfies the following two approximation properties, often of tremendous utility:
- for every $E \subset \mathbb{R}^n$,
$$\mathscr{L}^n(E) = \inf\{\mathscr{L}^n(U) : E \subset U,\ U \text{ open}\}\ ;$$
- for every Lebesgue-measurable set $E \subset \mathbb{R}^n$,
$$\mathscr{L}^n(E) = \sup\{\mathscr{L}^n(K) : K \subset E,\ K \text{ compact}\}\ .$$

1.6.1. More Lebesgue null sets: the middle-third Cantor set. In this subsection, we focus on the Lebesgue measure $\mathscr{L}^1$ on the real line $\mathbb{R}$. We have seen that $\mathscr{L}^1$ (and, more generally, $\mathscr{L}^n$) assigns zero mass to every countable set. In particular, we see here a first instance of the phenomenon of "lack of interaction"[22] between measure theory and topology: there are subsets of $\mathbb{R}$ (more generally, of $\mathbb{R}^n$) which are large from the topological point of view, namely dense, and yet negligible from the (Lebesgue) measure-theoretic perspective: an example is the set $\mathbb{Q}$ of rational numbers.

On the other hand, $\mathscr{L}^n$ gives positive mass to any non-empty open set. It turns out, as a sort of intermediate behaviour between the latter two cases, there are many "large" subsets of $\mathbb{R}$ with zero Lebesgue measure, where "large" here is intended with respect to cardinality. Spefically, there are sets $C \subset \mathbb{R}$ in bijection with $\mathbb{R}$, namely with the cardinality of the *continuum*, such that $\mathscr{L}^1(C) = 0$.

The classical example is the **middle-third Cantor set** C, whose construction we now present. We define a hierarchy of closed subintervals of the unit interval $[0, 1]$ in the following way. We start with $I_0 = [0, 1]$, which is the only interval appearing in the zero-th generation; the first generation is obtained from E_0 by removing the "middle-third" interval $(1/3, 2/3)$, so as to obtain the two intervals $I_{1,1} = [0, 1/3]$ and $I_{1,2} = [2/3, 1]$. We continue this procedure inductively: at every step $n \geq 1$, the n-th generation of intervals consists of 2^n disjoint closed intervals $I_{n,1}, \dots, I_{n,2^n}$, of length 3^{-n}, each of which gives rise to two new disjoint subintervals of length $3^{-(n+1)}$, by removing from it the "middle-third segment" from it. The middle-third Cantor set is then defined as the intersection
$$C = \bigcap_{n\geq 0} \bigcup_{1\leq k\leq 2^n} I_{n,k}\ .$$
As a decreasing intersection of non-empty compact sets (namely of the unions of all closed intervals at a fixed generation), C is non-empty[23]. In fact, C is quite large:

Lemma 1.6.8. *For every $x \in [0, 1]$, let $\sum_{n\geq 1} a_n 3^{-n}$ be the expansion of x in base 3, with $a_n \in \{0, 1, 2\}$ for all $n \geq 1$ and where, for each triadic rational $j/3^k$, $k \geq 1$, $0 \leq j \leq 3^k$, we choose the expansion with infinitely many 2's. Then*
$$C = \left\{ x = \sum_{n\geq 1} a_n 3^{-n} \in [0, 1] : a_n \neq 1 \text{ for all } n \geq 1 \right\}.$$

We summarize the properties of the middle-third Cantor set in the following proposition.

Proposition 1.6.9. *The middle-third Cantor set $C \subset \mathbb{R}$ satisfies the following properties:*

(1) C is compact, totally disconnected, with empty interior and without isolated points;

[22] In §?? we shall see that, from various other standpoints, topology and measure theory interact nicely for the Lebesgue measure, and more generally for a class of Borel measures known as *Radon measures*.

[23] Without appealing to topology, this fact is easily seen by observing that all endpoints of the intervals $I_{n.k}$ belong to C: the "removal" process never affects them. This already shows that C is infinite.

(2) C is in bijection with $\mathbb{R}$;
(3) $\mathscr{L}^1(C) = 0$.

PROOF. The first two assertions are part of the exercises.
As to the last, measure-theoretic claim, continuity of $\mathscr{L}^1$ from above yields

$$\mathscr{L}^1(C) = \lim_{n\to\infty} \mathscr{L}^1\Big(\bigcup_{1\le k\le 2^n} I_{n,k}\Big) = \lim_{n\to\infty} \sum_{1\le k\le 2^n} \mathscr{L}^1(I_{n,k}) = \lim_{n\to\infty} \frac{2^n}{3^n} = 0\,.$$

□

REMARK 1.6.10. Observe that non-emptiness of the interior of C can be deduced from the fact that $\mathscr{L}^1(C) = 0$: since every non-empty open set $O \subset \mathbb{R}$ has positive Lebesgue measure, the same is *a fortiori* true, by monotonicity, of any set with non-empty interior.

1.7. Lebesgue-Stieltjes measures on the real line

As a further application of Theorem 1.5.11, we now turn our attention to the class of Borel measures μ on the real line (that is, measures on the Borel space $(\mathbb{R}, \mathfrak{B}_\mathbb{R})$) with the property that $\mu(K)$ is finite for every compact $K \subset \mathbb{R}$. In the terminology to be introduced in later chapters, these are *Radon measures* on the real line.

To motivate our forthcoming discussion, suppose first μ is a Borel probability measure on $\mathbb{R}$. In probability theory, an object of interest is the *distribution function* of μ, namely the function $F_\mu : \mathbb{R} \to \mathbb{R}$ defined as

$$F_\mu(x) = \mu((-\infty, x])$$

for all $x \in \mathbb{R}$. Notice that F_μ completely determines the probability measure μ: if $F_\nu = F_\mu$ for a second Borel probability measure ν on $\mathbb{R}$, then μ and ν agree on the collection of closed half-lines $(-\infty, x]$, which is closed under finite intersections, and generates the Borel σ-algebra $\mathfrak{B}_\mathbb{R}$ by virtue of Proposition 1.3.21. The Monotone Class Lemma (Theorem 1.5.17), coupled with Lemma 1.5.16, delivers $\mu = \nu$.

Notice that the function F_μ is increasing, namely $F_\mu(x_1) \le F_\mu(x_2)$ for all $x_1 \le x_2 \in \mathbb{R}$, simply because of monotonicity of μ; furthermore, continuity from above of μ implies that F_μ is right-continuous: for every $x \in \mathbb{R}$ and every sequence $(x_n)_{n\ge 0}$ decreasing to x,

$$F_\mu(x) = \mu((-\infty, x]) = \mu\Big(\bigcap_{n\ge 0} (-\infty, x_n]\Big) = \lim_{n\to\infty} \mu((-\infty, x_n]) = \lim_{n\to\infty} F_\mu(x_n)\,. \tag{1.7.1}$$

Does every increasing, right-continuous function $F : \mathbb{R} \to [0, 1]$ arise as the distribution function of a Borel probability measure on $\mathbb{R}$? More generally, we would like to consider Borel measures μ which are only required to be finite on compact sets; however, if μ is infinite, then defining F_μ as in (1.7.1) makes little sense, as it might well happen that $F_\mu(x) = \infty$ for all x (such as with $\mu = \mathscr{L}^1$). The appropriate replacement in contained in the following theorem, giving a one-to-one correspondence between increasing, right-continuous functions $\mathbb{R} \to \mathbb{R}$, up to constant translation, and Borel measures on $\mathbb{R}$ which are finite on compact sets.

THEOREM 1.7.1. *(1) Let $F : \mathbb{R} \to \mathbb{R}$ be an increasing, right-continuous function. Then there exists a unique Borel measure μ_F on $\mathbb{R}$ satisfying*

$$\mu((a, b]) = F(b) - F(a)$$

for all $a < b \in \mathbb{R}$. If $G : \mathbb{R} \to \mathbb{R}$ is another such function, then $\mu_G = \mu_F$ if and only if $G - F$ is a constant function.

(2) Let μ be a Borel measure on $\mathbb{R}$ such that $\mu(I) < \infty$ for every bounded interval $I \subset \mathbb{R}$. Define a function $F : \mathbb{R} \to \mathbb{R}$ by

$$F(x) = \begin{cases} \mu((0, x]) & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -\mu((x, 0]) & \text{if } x < 0 \end{cases}.$$

Then F is increasing and right-continuous, and $\mu = \mu_F$.

For instance, the Lebesgue measure $\mathscr{L}^1$ corresponds, under the theorem, to all functions of the form $F(x) = x + C$, as C ranges in $\mathbb{R}$.

The measure μ_F determined, as in the theorem, by an increasing, right-continuous function $F\colon \mathbb{R} \to \mathbb{R}$ is called the **Lebesgue-Stieltjes** measure associated to F.

Proof of Theorem 1.7.1. It is relegated to the exercises. □

1.8. Metric outer measures and Hausdorff measures

If X is a topological space, a **Borel measure on** X is a measure on the measurable space $(X, \mathfrak{B}_X)$ where $\mathfrak{B}_X$ is the Borel σ-algebra on X.

Recall that, if (X, d) is a metric space and $A, B \subset X$, the distance between A and B is defined as

$$d(A, B) = \inf\{d(a, b) : a \in A,\ b \in B\} .$$

Definition 1.8.1 (Borel and metric outer measure). Let X be a topological space. A **Borel outer measure** on X is an outer measure μ^* on X with the property that every Borel set $A \subset X$ is measurable with respect to μ^*.

Let (X, d) be a metric space. A **metric outer measure** on X is an outer measure μ^* on X such that, for every $A, B \subset X$, the condition $d(A, B) > 0$ implies

$$\mu^*(A \cup B) = \mu^*(A) + \mu^*(B) .$$

Thus, an outer measure on a metric space is a metric outer measure if it is additive on subsets which are well separated by the distance function.

If μ^* is a Borel outer measure on a topological space X, then the restriction of μ^* to the Borel σ-algebra $\mathfrak{B}_X$ is a Borel measure on X.

Proposition 1.8.2. *Let (X, d) be a metric space, μ^* a metric outer measure on X. Then μ^* is a Borel outer measure on X, endowed with the topology induced by d.*

Proof. It suffices to show that every open set is μ^*-measurable. Let thus $O \subset X$ be open, and $E \subset X$ arbitrary; we need to show that

$$\mu^*(E) \geq \mu^*(E \cap O) + \mu^*(E \cap O^{\mathrm{c}}) .$$

Define, for every integer $n \geq 1$,

$$O_n = \{x \in O : d(x, O^{\mathrm{c}}) > 1/n\} .$$

Then $O_n \subset O_{n+1}$ for all n, and $\bigcup_{n\geq 1} O_n = O$, since O^{c} is closed and thus coincides with $\{x \in X : d(x, O^{\mathrm{c}}) = 0\}$. Observe that, by construction, for every fixed $n \geq 1$, the sets O_n and O^{c} are d-separated, since $d(O_n, O^{\mathrm{c}}) \geq 1/n > 0$, hence

$$\mu^*(E) \geq \mu^*(E \cap (O_n \cup O^{\mathrm{c}})) = \mu^*(E \cap O_n) + \mu^*(E \cap O^{\mathrm{c}}) ,$$

for μ^* is a metric outer measure. It follows that

$$\mu^*(E) \geq \mu^*(E \cap O^{\mathrm{c}}) + \lim_{n\to\infty} \mu^*(E \cap O_n) ,$$

where the limit exists by monotonicity of μ^* (and equals the supremum of the given quantity over all $n \geq 1$). The conclusion follows from the following lemma, whose proof is part of the exercises, and which gives a sort of continuity from below of metric outer measures on separated increasing sequences of sets.

Lemma 1.8.3. *Let μ^* be a metric outer measure on a metric space (X, d), $(E_n)_{n\geq 0}$ an increasing sequence of subsets of X, $E = \bigcup_{n\geq 0} E_n$. Suppose that*

$$d(E \setminus E_{n+1}, E_n) > 0 \tag{1.8.1}$$

for every $n \geq 0$. Then

$$\mu^*(E) = \lim_{n\to\infty} \mu^*(E_n) .$$

In our case, for $E_n = E \cap O_n$, it is clear by construction that (1.8.1) holds: if $x \in E_n$ and $y \in E \setminus E_{n+1}$, let $\varepsilon > 0$ be strictly smaller than $\frac{1}{n} - \frac{1}{n+1}$, and choose $z \in O^c$ such that

$$d(y,z) \le \frac{1}{n+1} + \varepsilon\,.$$

Then the reverse triangle inequality gives

$$d(x,y) \ge d(x,z) - d(y,z) > \frac{1}{n} - \frac{1}{n+1} - \varepsilon > 0\,,$$

so that (1.8.1) holds by taking the infimum on the left-hand side of the previous inequality, over all $x \in E_n$ and $y \in E \setminus E_{n+1}$. □

1.8.1. Hausdorff measures. Let (X,d) be a metric space. For every real number $s \ge 0$, we shall define the *s-dimensional Hausdorff measure* on $(X, \mathfrak{B}_X)$. The construction will require a limiting process; we thus fix a parameter $\delta > 0$, which we will ultimately let tend to 0, and define a function $\mathscr{H}^s_\delta \colon \mathcal{P}(X) \to [0,\infty]$ as

$$\mathscr{H}^s_\delta(A) = \inf\left\{\sum_{n=0}^{\infty} \operatorname{diam}(E_n)^s : A \subset \bigcup_{n\ge 0} E_n,\ \operatorname{diam}(E_n) \le \delta\ \forall n \in \mathbb{N}\right\},$$

where we recall that $\operatorname{diam}(E) = \sup\{d(x,y) : x,y \in E\}$ for all $E \subset X$.

By virtue of Proposition 1.5.3 applied to the collection $\mathcal{E}_\delta = \{E \subset X : \operatorname{diam}(E) \le \delta\}$ and to the function $\rho_{\delta,s}(E) = \operatorname{diam}(E)^s$, $E \in \mathcal{E}$, we deduce at once that $\mathscr{H}^s_\delta(A)$ is an outer measure on X. It is clear that $\delta_1 \le \delta_2$ implies $\mathscr{H}^s_{\delta_1}(A) \ge \mathscr{H}^s_{\delta_2}(A)$ for all $A \subset X$; we define a function $\mathscr{H}^s \colon \mathcal{P}(X) \to [0,\infty]$ by

$$\mathscr{H}^s(A) = \sup_{\delta > 0} \mathscr{H}^s_\delta(A) = \lim_{\delta \to 0} \mathscr{H}^s_\delta(A)\,.$$

We call $\mathscr{H}^s$ the **s-dimensional Hausdorff measure** on (X,d).

PROPOSITION 1.8.4. *For every $s \ge 0$, $\mathscr{H}^s$ is a metric outer measure on X; as a consequence, it is a Borel outer measure.*

PROOF. We need to show first that $\mathscr{H}^s$ is an outer measure, and then that it is additive on pairs of sets which are separated by the distance d. The second assertion is then a corollary of Proposition 1.8.2.

Monotonicity of $\mathscr{H}^s$ and the fact that $\mathscr{H}^s(\emptyset) = 0$ follow directly from the analogous properties for all $\mathscr{H}^s_\delta$, $\delta > 0$. As to σ-subadditivity, let $(A_n)_{n\ge 0}$ be a sequence of subsets of X, and fix $\delta > 0$. Then the fact that $\mathscr{H}^s_\delta$ is an outer measure yields

$$\mathscr{H}^s_\delta\Big(\bigcup_{n\ge 0} A_n\Big) \le \sum_{n=0}^{\infty} \mathscr{H}^s_\delta(A_n) \le \sum_{n=0}^{\infty} \mathscr{H}^s(A_n)\,,$$

the last bound following by definition of $\mathscr{H}^s$. Taking the supremum over $\delta > 0$ on the left-hand side of the last-displayed inequality, we deduce that

$$\mathscr{H}^s\Big(\bigcup_{n\ge 0} A_n\Big) \le \sum_{n=0}^{\infty} \mathscr{H}^s(A_n)\,,$$

as desired.

Finally, let $A, B \subset X$ be such that $d(A,B) > 0$. We need to show that $\mathscr{H}^s(A \cup B) \ge \mathscr{H}^s(A) + \mathscr{H}^s(B)$, as the reverse inequality holds automatically by the already established fact that $\mathscr{H}^s$ is an outer measure. Fix $\delta > 0$ such that $2\delta < d(A,B)$, and let $(E_n)_{n\ge 0}$ be a sequence of sets of diameter at most δ such that

$$A \cup B \subset \bigcup_{n\ge 0} E_n\,.$$

We may assume without loss of generality that $E_n \cap (A \cup B) \neq \emptyset$ for all $n \ge 0$. Due to our choice of δ, the family $(E_n)_{n\ge 0}$ can be partitioned into two countable subfamilies $(E_i)_{i\in I}$ and

$(E_j)_{j\in J}$ consisting, respectively, of those elements in (E_n) intersecting A and B non-trivially. We then have

$$\sum_{n=0}^{\infty} \operatorname{diam}(E_n)^s = \sum_{i\in I} \operatorname{diam}(E_i)^s + \sum_{j\in J} \operatorname{diam}(E_j)^s \geq \mathscr{H}^s_\delta(A) + \mathscr{H}^s_\delta(B) \ ,$$

where the last inequality follows from the obvious fact that $(E_i)_{i\in I}$ is a covering of A and $(E_j)_{j\in J}$ is a covering of B, both consisting of sets of diameter at most δ. Taking the infimum over all such coverings $(E_n)_{n\geq 0}$ of $A\cup B$, we deduce that

$$\mathscr{H}^s_\delta(A\cup B) \geq \mathscr{H}^s_\delta(A) + \mathscr{H}^s_\delta(B) \ ,$$

and taking the limit as $\delta \to 0$ on both sides of the previous inequality, the conclusion is achieved. □

Remark 1.8.5. If the metric space (X,d) is separable, then every subset A of X admits, for every $\delta > 0$, a countable cover by sets of diameter at most δ. It suffices to take balls of radius $\delta/2$ centered at points in a given countable dense subset. If (X,d) is not separable, on the other hand, there is typically a wealth of subsets A of X not admitting any such countable cover, for no choice of $\delta > 0$. For all those sets A, we have $\mathscr{H}^s_\delta(A) = \infty$, since the quantity is defined as the infimum over an empty family, and thus $\mathscr{H}^s(A) = \infty$.

This is the reason why Hausdorff measures furnish a good notion of size only on separable metric spaces.

Recall the notions of isometry and of C-Lipschitz function between metric spaces from §A.2.

Lemma 1.8.6. *Let (X,d_X) and (Y,d_Y) be metric spaces, $f\colon X\to Y$ a C-Lipschitz function, for some $C>0$. Then, for all $s\geq 0$ and all $A\subset X$,*

$$\mathscr{H}^s(f(A)) \leq C^s\mathscr{H}^s(A) \ .$$

If f is an isometry, then

$$\mathscr{H}^s(f(A)) = \mathscr{H}^s(A) \ .$$

Proof. Suppose first $f\colon X\to Y$ is C-Lipschitz for some $C>0$. Let $(E_n)_{n\geq 0}$ be a covering of A by sets of diameter at most δ. Then $(f(E_n))_{n\geq 0}$ is a covering of $f(E)$ by sets of diameter at most $C\delta$, whence

$$\mathscr{H}^s_{C\delta}(f(A)) \leq \sum_{n=0}^{\infty} \operatorname{diam}(f(E_n))^s \leq C^s \sum_{n=0}^{\infty} \operatorname{diam}(E_n)^s \ .$$

Taking the infimum on the right-hand side, over all possible such covers $(E_n)_{n\geq 0}$, we deduce that

$$\mathscr{H}^s_{C\delta}(f(A)) \leq C^s\mathscr{H}^s_\delta(A) \ ,$$

whence the conclusion follows by taking the limit as $\delta\to 0$ on both sides.

If f is an isometry, all previous inequalities become equalities, and the conclusion follows all the same. □

Proposition 1.8.7. *Let (X,d) be a metric space, A a subset of X, $0\leq r < s$.*

(1) If $\mathscr{H}^r(A) < \infty$, then $\mathscr{H}^s(A) = 0$.
(2) If $\mathscr{H}^s(A) > 0$, then $\mathscr{H}^r(A) = \infty$.

The second assertion is simply the contrapositive of the first one; we prefer, however, to emphasize its content in light of the forthcoming definition of Hausdorff dimension.

Proof. As just observed, it suffices to prove the first assertion. Suppose thus $\mathscr{H}^r(A) < \infty$, and let $\delta > 0$. Choose a covering $(E_n)_{n\geq 0}$ of A by sets of diameter at most δ such that

$$\sum_{n=0}^{\infty} \operatorname{diam}(E_n)^r < \mathscr{H}^r(A) + 1 \ .$$

Then

$$\sum_{n=0}^{\infty} \operatorname{diam}(E_n)^s = \sum_{n=0}^{\infty} \operatorname{diam}(E_n)^r \operatorname{diam}(E_n)^{s-r} \leq \delta^{s-r} \sum_{n=0}^{\infty} \operatorname{diam}(E_n) < (\mathscr{H}^r(A) + 1)\delta^{s-r} ,$$

whence, by definition of the Hausdorff s-measure,

$$\mathscr{H}^s_\delta(A) \leq (\mathscr{H}^r(A) + 1)\delta^{s-r} ;$$

taking the limit as $\delta \to 0$ of the above inequality achieves the conclusion. □

Definition 1.8.8 (Hausdorff dimension). Let (X, d) be a metric space, A a subset of X. The **Hausdorff dimension** of A is defined as

$$\dim_{\mathsf{H}}(A) = \sup\{r \geq 0 : \mathscr{H}^r(A) = \infty\} = \inf\{s \geq 0 : \mathscr{H}^s(A) = 0\} . \qquad (1.8.2)$$

Observe that equality holds in (1.8.2) between the supremum and the infimum due to Proposition 1.8.7. A further consequence of the latter is the alternative characterization

$$\dim_{\mathsf{H}}(A) = \sup\{r \geq 0 : \mathscr{H}^r(A) > 0\} = \inf\{s \geq 0 : \mathscr{H}^s(A) < \infty\} .$$

The following corollary follows readily from the definition of Hausdorff dimension and Lemma 1.8.6.

Corollary 1.8.9. *Let (X, d_X) and (Y, d_Y) be metric spaces, $f\colon X \to Y$ a Lipschitz function. Then, for all $A \subset X$,*

$$\dim_{\mathsf{H}}(f(A)) \leq \dim_{\mathsf{H}}(A) .$$

If f is an isometry, then

$$\dim_{\mathsf{H}}(f(A)) = \dim_{\mathsf{H}}(A) .$$

CHAPTER 2

Integration

After having examined measures in the opening chapter, we now turn to the Lebesgue theory of integration on abstract measure spaces. It vastly generalizes Riemann theory of integration on Euclidean spaces; on the other hand, as already alluded to in the foregoing chapter, it moves from the same geometric intuition of the integral of a positive function, say on the real line, as the area of the region situated between the real axis and the graph of the function. As it happens, this geometric adherence of the notion of integral furnishes equally the most intuitive notion of *expected value* (or average value) of a random variable, namely of integral of a measurable function defined over a probability space. Just as step functions are the building blocks of integration for the Riemann integral of the line, *simple functions*, which are a natural abstract generalization thereof, are the building blocks of integration according to Lebesgue.

We remind the reader that the following convention is in place. Whenever $(X, \mathfrak{M})$ is a Borel space and $f\colon X \to \overline{\mathbb{R}}$ or $f\colon X \to \mathbb{C}$ is a function, we say simply that f is measurable, or $\mathfrak{M}$-measurable, if it is $(\mathfrak{M}, \mathfrak{B}_{\overline{\mathbb{R}}})$ or $(\mathfrak{M}, \mathfrak{B}_{\mathbb{C}})$-measurable, respectively.

2.1. Measurable functions with values in number systems

Throughout this section, we fix a Borel space $(X, \mathfrak{M})$.

If $f\colon X \to \mathbb{C}$ is a function, we define the real and the imaginary parts of f as the functions $\Re f\colon X \to \mathbb{R}$, $\Im f\colon X \to \mathbb{R}$, given by

$$\Re f(x) = \Re(f(x))\,; \quad \Im f(x) = \Im(f(x))$$

for all $x \in X$.

LEMMA 2.1.1. *A function $f\colon X \to \mathbb{C}$ is measurable if and only if both $\Re f$ and $\Im f$ are $\mathfrak{M}$-measurable.*

PROOF. The set $\mathbb{C}$, as a topological space, is identified with $\mathbb{R}^2$ with the Euclidean topology; the claim thus follows from the fact that $\mathfrak{B}_{\mathbb{R}^2} = \mathfrak{B}_{\mathbb{R}} \otimes \mathfrak{B}_{\mathbb{R}}$ (see Corollary 1.3.19) and the universal property of the product σ-algebra. □

Sums, products and quotients of real-(or complex-)valued functions are defined pointwise: if $f, g\colon X \to \mathbb{R}$ are two functions, we set

$$f + g(x) = f(x) + g(x)\,, \quad fg(x) = f(x)g(x)\,, \quad \frac{f}{g}(x) = \frac{f(x)}{g(x)}$$

for all $x \in X$, the latter in case $g(x) \neq 0$ for all $x \in X$.

LEMMA 2.1.2. *If $f, g\colon X \to \mathbb{R}$ are measurable, the same is true of the sum $f + g$ and the product fg. If $g(x) \neq 0$ for all $x \in X$, the same assertion holds for f/g.*

PROOF. Both claims are consequences of the continuity of the field operations on $\mathbb{R}$. We detail the argument just for the sum, the one for products and quotients being identical.

If $f, g\colon X \to \mathbb{R}$ are measurable, then the function $X \to \mathbb{R} \times \mathbb{R}$, $x \mapsto (f(x), g(x))$ is measurable with respect to the product σ-algebra $\mathfrak{B}_{\mathbb{R}} \otimes \mathfrak{B}_{\mathbb{R}}$. Now $f + g$ is obtained as the composition of the previous function with the sum function $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $(x, y) \mapsto x + y$, which is continuous with respect to the Euclidean topologies on $\mathbb{R}^2$ and $\mathbb{R}$, and hence measurable with respect to $\mathfrak{B}_{\mathbb{R}} \otimes \mathfrak{B}_{\mathbb{R}}$, owing to Corollaries 1.3.15 and 1.3.19. The conclusion follows from Lemma 1.3.13. □

REMARK 2.1.3. An analogous statement holds, with the same proof, for complex-valued functions.

LEMMA 2.1.4. *Let* $(f_n)_{n\geq 0}$ *be a family of measurable functions* $f_n\colon X \to \overline{\mathbb{R}}$. *Then the functions*

$$\inf_{n\geq 0} f_n \ , \quad \sup_{n\geq 0} f_n \ , \quad \liminf_{n\to\infty} f_n \ , \quad \limsup_{n\to\infty} f_n$$

are measurable.

PROOF. By definition of the inferior and superior limit, it suffices to prove the claim for $\inf_{n\geq 0} f_n$ and $\sup_{n\geq 0} f_n$.

We start with $\inf_{n\geq 0} f_n$. By virtue of Corollary 1.3.23 (and its related Remark 1.3.24), it suffices to show that

$$\{x \in X : \inf_{n\geq 0} f_n(x) < a\}$$

is measurable for all $a \in \mathbb{R}$. The definition of infimum of a collection of real numbers readily yields that the last displayed set equals the union

$$\bigcup_{n\geq 0} \{x \in X : f_n(x) < a\} \ ,$$

which is a countable union of measurable sets. The measurability claim is thus proved.

As for $\sup_{n\geq 0} f_n$, we argue similarly, observing that, for every $a \in \mathbb{R}$, we have equality

$$\{x \in X : \sup_{n\geq 0} f_n(x) > a\} = \bigcup_{n\geq 0} \{x \in X : f_n(x) > a\} \ .$$

□

Since the limit of a sequence, when it exists, coincides with its inferior limit (and with its superior limit), we deduce at once the following corollary.

COROLLARY 2.1.5. *Let* $(f_n)_{n\geq 0}$ *be a family of measurable functions* $f_n\colon X \to \overline{\mathbb{R}}$. *Suppose that, for every* $x \in X$, *the limit*

$$\lim_{n\to\infty} f_n(x)$$

exists in $\overline{\mathbb{R}}$. *Then the function* $f(x) = \lim_{n\to\infty} f_n(x)$ *is measurable.*

The same assertion holds for the pointwise limit of a sequence of measurable $\mathbb{C}$*-valued functions.*

2.2. Integral of simple functions

Henceforth in this chapter, until explicit mention to the contrary, we work over a fixed measure space $(X, \mathfrak{M}, \mu)$. The characteristic function[1] of a subset $E \subset X$ is indicated with χ_E. We point out the properties

$$\chi_{E^c} = 1 - \chi_E \ , \quad \chi_{E\cap F} = \chi_E\, \chi_F \ , \quad \chi_E = \sum_{n\geq 0} \chi_{E_n} \quad \text{if } E = \bigsqcup_{n\geq 0} E_n \ ,$$

holding for all subsets E, F, E_n $(n \geq 0)$ of X.

Observe that χ_E is a measurable function if and only if E is a measurable set.

DEFINITION 2.2.1 (Simple function). Let $(X, \mathfrak{M})$ be a Borel space. A **simple function** on X is a measurable function $f\colon X \to \mathbb{C}$ such that there exist finitely many $c_i \in \mathbb{C}$, $i \in I$, and $E_i \in \mathfrak{M}$, $i \in I$, with

$$f = \sum_{i\in I} c_i \chi_{E_i} \ . \tag{2.2.1}$$

[1] Also known as *indicator function*, especially in probability theory, where the terminology "characteristic function" is reserved for the Fourier transform of the distrinbution of an $\mathbb{R}^d$-valued random variable.

In other words, a measurable function is simple if it is a linear combination, with complex coefficients, of measurable characteristic functions.

REMARK 2.2.2. We express it in yet more abstract linear algebraic terms: the set of simple functions $X \to \mathbb{C}$ is the $\mathbb{C}$-vector subspace of $\mathbb{C}^X$ generated by the characteristic functions χ_E, $E \in \mathfrak{M}$.

It is equivalent to define a simple function as a measurable function $X \to \mathbb{C}$ with finite image, namely such that $f(X)$ is a finite set. If this is the case, then clearly

$$f = \sum_{z \in f(X)} z \chi_{f^{-1}(z)} \ , \tag{2.2.2}$$

where $f^{-1}(z) \in \mathfrak{M}$ for all $z \in f(X)$ by measurability of f.

It is plain that a given simple function f admits infinitely many distinct representations of the form (2.2.1). The specific representation in (2.2.2) is called the **standard representation** of f as a simple function. It enjoys[2] the additional property that it expresses f as a linear combination of characteristic functions of pairwise disjoint measurable sets covering the whole space X.

DEFINITION 2.2.3 (Integral of a simple function). Let $(X, \mathfrak{M}, \mu)$ be a measure space, $f\colon X \to \mathbb{R}_{\geq 0}$ a simple function, with standard representation

$$f = \sum_{i \in I} c_i \chi_{E_i} \ .$$

The **integral of f with respect to** μ, indicated with

$$\int_X f \ \mathrm{d}\mu \ ,$$

is defined as the quantity

$$\sum_{i \in I} c_i \ \mu(E_i) \ .$$

Here a fundamental arithmetical convention is in place: if $c_i = 0$ and $\mu(E_i) = \infty$ for some $i \in I$, then $0 \cdot \infty$ is set to be equal to 0, the point being that μ-null sets should not affect the integral of any function. On the other hand, if $\mu(E_i) = \infty$ and $c_i \neq 0$, then $c_i \cdot \mu(E_i) = \infty$. Therefore, the integral of a simple function takes values in the extended half-line $[0, \infty]$.

Further widely adopted notation for the integral is any of the following:

$$\int_X f(x) \ \mathrm{d}\mu(x) \ , \quad \int_X f(x) \ \mu(\mathrm{d}x) \ , \quad \int_X f(x) \ \mu(x) \ , \quad \int f \ \mathrm{d}\mu \ , \quad \int f \ ,$$

the last two when either X or both X and μ are understood from the context. The notation

$$\mu(f)$$

is also frequently encountered, especially when f is a continuous function on a topological space and μ is a Borel measure on it. We reserve §**??** for the treatment of integration of continuous functions on locally compact topological spaces.

The same notational remark applies to integrals of more general measurable functions, to be discussed shortly.

REMARK 2.2.4. The following observation shall prove to be ofter relevant in the sequel. Let $f\colon X \to \mathbb{R}$ be a simple function, expressed as a linear combination

$$f = \sum_{i \in I} c_i \chi_{E_i}$$

[2] But it is clearly not the unique one enjoying such a property.

of characteristic functions of measurable sets which are pairwise disjoint. Upon enlarging the families $(E_i)_{i\in I}$ and $(c_i)_{i\in I}$, we can assume that the E_i's are a partition of X, namely that $X = \bigcup_{i\in I} E_i$. Notice that such a representation may not be the standard representation of f, since we might have $c_i = c_j$ for some $i \neq j \in I$. However, it still holds that

$$\int_X f \, \mathrm{d}\mu = \sum_{i\in I} c_i \mu(E_i) \, .$$

Indeed, it is clear that, if

$$\sum_{j\in J} d_j F_j$$

is the standard representation of f, then $\{d_j : j \in J\}$ is a subset of $\{c_i : i \in I\}$, and we can partition I as the disjoint union

$$I = \bigcup_{j\in J} I_j$$

with $F_j = \bigcup_{i\in I_j} E_i$ for all $j \in J$ and $d_j = c_i$ for all $j \in J$ and $i \in I_j$. Hence, finite additivity of μ yields

$$\int_X f \, \mathrm{d}\mu = \sum_{j\in J} d_j \mu(F_j) = \sum_{j\in J} d_j \sum_{i\in I_j} \mu(E_i) = \sum_{j\in J}\sum_{i\in I_j} c_i \mu(E_i) = \sum_{i\in I} c_i \mu(E_i) \, .$$

LEMMA 2.2.5. *Let $f, g\colon X \to \mathbb{R}_{\geq 0}$ be simple functions.*

(1) If $c \in \mathbb{R}_{\geq 0}$, then cf is a simple function[3] *and*

$$\int_X cf \, \mathrm{d}\mu = c \int_X f \, \mathrm{d}\mu \, .$$

(2) The sum $f + g$ is a simple function, and

$$\int_X f + g \, \mathrm{d}\mu = \int_X f \, \mathrm{d}\mu + \int_X g \, \mathrm{d}\mu \, .$$

(3) If $f \leq g$, then

$$\int_X f \, \mathrm{d}\mu \leq \int_X g \, \mathrm{d}\mu \, .$$

PROOF. The first assertion is obvious: if

$$f = \sum_{i\in I} c_i \chi_{E_i}$$

is a simple function expressed in standard representation, then

$$\sum_{i\in I} c c_i \chi_{E_i}$$

is a representation of cf of the type discussed in Remark 2.2.4 (if $c \neq 0$, it is actually the standard representation of cf), whence

$$\int_X cf \, \mathrm{d}\mu = \sum_{i\in I} c c_i \mu(E_i) = c \sum_{i\in I} c_i \mu(E_i) = c \int_X f \, \mathrm{d}\mu \, .$$

Let now

$$g = \sum_{j\in J} d_j \chi_{F_j}$$

be a second simple function, expressed in standard representation. Then

$$f + g = \sum_{i\in I,\, j\in J} (c_i + d_j) \chi_{E_i \cap F_j}$$

[3]Clearly, cf is a simple function for all $c \in R$; the restriction $c \geq 0$ is only in place to talk about the integral, which so far has been defined only for positive functions.

is a representation of $f + g$ as a finite linear combination of characteristic functions of a measurable partition of X. As established in Remark 2.2.4, we then have

$$\begin{aligned}\int_X f + g \, \mathrm{d}\mu &= \sum_{i\in I,\, j\in J} (c_i + d_j)\mu(E_i \cap F_j) = \sum_{i\in I,\, j\in J} c_i\mu(E_i \cap F_j) + \sum_{i\in I,\, j\in J} d_j\mu(E_i \cap F_j) \\ &= \sum_{i\in I} c_i \sum_{j\in J} \mu(E_i \cap F_j) + \sum_{j\in J} d_j \sum_{i\in I} \mu(E_i \cap F_j) = \sum_{i\in I} c_i\mu(E_i) + \sum_{j\in J} d_j\mu(F_j) \\ &= \int_X f \, \mathrm{d}\mu + \int_X g \, \mathrm{d}\mu \,,\end{aligned}$$

using finite additivity of μ in the second-to-last step.

Finally, if $f \le g$, then $c_i \le d_j$ for all $i \in I$ and $j \in J$ such that $E_i \cap F_j \ne \emptyset$, so that, again invoking Remark 2.2.4,

$$\begin{aligned}\int_X f \, \mathrm{d}\mu &= \sum_{i\in I,\, j\in J} c_i\mu(E_i \cap F_j) = \sum_{i\in I,\, j\in J,\, E_i\cap F_j\ne\emptyset} c_i\mu(E_i \cap F_j) \le \sum_{i\in I,\, j\in J,\, E_i\cap F_j\ne\emptyset} d_j\mu(E_i \cap F_j) \\ &= \sum_{i\in I,\, j\in J} d_j\mu(E_i \cap F_j) = \int_X g \, \mathrm{d}\mu \,.\end{aligned}$$

□

2.3. Integral of positive functions

The integral of positive simple functions extends to the integral of positive measurable functions via an approximation procedure from below.

Definition 2.3.1 (Integral of a positive function). Let $(X, \mathfrak{M}, \mu)$ be a measure space, $f\colon X \to [0, \infty]$ a measurable function. The **integral of f with respect to μ**, indicated with

$$\int_X f \, \mathrm{d}\mu \,,$$

is defined as the quantity

$$\sup\left\{\int_X \varphi \, \mathrm{d}\mu : 0 \le \varphi \le f,\ \varphi \text{ simple}\right\} . \tag{2.3.1}$$

Observe that the supremum in the definition is always taken over a non-empty family, since the constant function $\varphi = 0$ is simple and satisfies $\varphi \le f$. The integral $\int_X f \, \mathrm{d}\mu$ plainly takes values in $[0, \infty]$.

As should be expected, the definition agrees with Definition 2.2.3 when f is a simple function: one inequality follows from the fact that f itself belongs to the family of simple functions φ such that $0 \le \varphi \le f$, while the reverse one stems from monotonicity in Lemma 2.2.5.

It is also evident from the definition that, if $f, g\colon X \to [0, \infty]$ are measurable functions with $f \le g$, then

$$\int_X f \, \mathrm{d}\mu \le \int_X g \, \mathrm{d}\mu \,,$$

since the supremum defininning the latter integral is taken over a larger family with respect to the supremum defining the former.

The definition (2.3.1) involves the supremum over a typically huge collection of functions. We shall now see that it actually suffices to consider an increasing sequence of simple functions converging to f, in order to compute $\int_X f \, \mathrm{d}\mu$. The notion of pointwise and uniform convergence of a sequence of functions enters the next statement: we refer to §A.4.

Proposition 2.3.2. *Let $(X, \mathfrak{M})$ be a Borel space, $f\colon X \to [0, \infty]$ a measurable function. Then there exists a sequence $(\varphi_n)_{n\ge1}$ of simple functions $\varphi_n\colon X \to [0, \infty]$ such that*

$$\varphi_1 \le \varphi_2 \le \cdots \le \varphi_n \le \cdots$$

and

$$\varphi_n \overset{n\to\infty}{\longrightarrow} f$$

pointwise on X. Furthermore, the convergence is uniform over all subsets of X over which f is bounded.

PROOF. Fix an integer $n \geq 1$, and define a function $\varphi_n \colon X \to \mathbb{R}_{\geq 0}$ as follows. We partition X into measurable sets as

$$X = \{x \in X : f(x) \geq 2^n\} \cup \left(\bigcup_{0\leq k<2^{2n} \text{ integer}} \{x \in X : k2^{-n} \leq f(x) < (k+1)2^{-n}\} \right) ,$$

and define

$$\varphi_n(x) = \begin{cases} k2^{-n} & \text{if } k2^{-n} \leq f(x) < (k+1)2^{-n} \text{ for some } 0 \leq k < 2^{2n} \\ 2^n & \text{if } f(x) \geq 2^n \end{cases} .$$

Since φ_n takes on finitely many values, and the preimage under φ_n of each such value is measurable by construction, φ_n is a simple function.

Fix now a point $x \in X$; then $\varphi_n(x) = k2^{-n}$ for some integer $0 \leq k \leq 2^{2n}$, and so $k2^{-n} \leq f(x) < (k+1)2^{-n}$, with the understanding that $k+1 = \infty$ if $k = 2^{2n}$. It follows that either

$$2k \cdot 2^{-(n+1)} \leq f(x) < (2k+1)2^{-(n+1)}$$

or

$$(2k+1)2^{-(n+1)} \leq f(x) < (2k+2)2^{-(n+1)} ;$$

in the former case,

$$\varphi_{n+1}(x) = 2k \cdot 2^{-(n+1)} = \varphi_n(x) ,$$

and in the latter case

$$\varphi_{n+1}(x) = (2k+1)2^{-(n+1)} > \varphi_n(x)$$

which shows that the sequence $(\varphi_n)_{n\geq 1}$ is increasing.

As far as the pointwise convergence claim is concerned, it is obvious if $f(x) = \infty$, since then $\varphi_n(x) = 2^n$ for all $n \geq 1$. If $f(x) < \infty$, choose an integer $N \geq 1$ such that $f(x) < 2^N$. For every $n \geq N$, choose an integer $0 \leq k_n < 2^{2n}$ such that $k_n 2^{-n} \leq f(x) < (k_n+1)2^{-n}$. Then $\varphi_n(x) = k_n 2^{-n}$, and thus

$$|f(x) - \varphi_n(x)| = f(x) - k_n 2^{-n} < ((k_n+1) - k_n)2^{-n} = 2^{-n} ,$$

which shows that $\varphi_n(x)$ tends to $f(x)$ as $n \to \infty$.

Let now Y be a subset of X on which f is bounded: there is $M > 0$ such that $f(x) \leq M$ for all $x \in Y$. If $N \geq 1$ is an integer chosen so that $M < 2^N$, the previous argument[4] shows that

$$\sup_{x\in Y} |f(x) - \varphi_n(x)| \leq 2^{-n}$$

for all $n \geq N$, which establishes the desired uniform convergence over Y. □

REMARK 2.3.3. In general, even if $f(x)$ is finite for all $x \in X$, uniform convergence cannot be upgraded to the entire space. In the exercises, it is asked to find a counterexample.

EXERCISE 2.3.4. Let $(X, \mathfrak{M})$ be a measurable space, $f \colon X \to \mathbb{C}$ a measurable function. Show that there is a sequence $(\phi_n)_{n\geq 1}$ of simple functions $X \to \mathbb{C}$ such that

$$0 \leq |\phi_1| \leq \cdots \leq |\phi_n| \leq \cdots \leq |f|$$

and, as $n \to \infty$, $\phi_n \to f$ pointwise, and uniformly on every set on which f is bounded.

[4]Notice that, even though k_n depends on x, this doesn't affect the validity of the "uniform" conclusion.

Once we know that every measurable functions with values in $[0,\infty]$ can be approximated by an increasing sequence of positive simple functions, it remains to ascertain that taking the supremum over such a specific sequence amounts to taking the full supremum in (2.3.1). This is ensured by the first important convergence theorem in the theory of integration we will discuss, which will serve as a basis for all the others, and ultimately hinges upon the countable additivity property of a measure.

THEOREM 2.3.5 (Monotone Convergence Theorem). *Let* $(X,\mathfrak{M},\mu)$ *be a measure space,* $(f_n)_{n\geq 0}$ *a sequence of measurable functions* $f_n\colon X\to[0,\infty]$ *such that*

$$f_0 \leq f_1 \leq \cdots \leq f_n \leq \cdots .$$

Then

$$\int_X \lim_{n\to\infty} f_n(x)\,\mathrm{d}\mu(x) = \lim_{n\to\infty}\int_X f_n(x)\,\mathrm{d}\mu(x) .$$

Before delving into the proof, let us consider a special case, which will serve as an inspiration for the argument. If $f_n=\chi_{E_n}$ for an increasing sequence of measurable sets E_n, $n\geq 0$, then $\lim_{n\to\infty} f_n = \chi_E$ for $E=\bigcup_{n\geq 0} E_n$, and thus the statement of the Monotone Convergence Theorem amounts in this case to continuity from below of μ.

PROOF. Observe first that the limit

$$\lim_{n\to\infty} f_n(x)$$

exists in $[0,\infty]$ for all $x\in X$, since the sequence $(f_n(x))_{n\geq 0}$ is increasing. Furthermore, the assignment

$$x\mapsto f(x) \coloneqq \lim_{n\to\infty} f_n(x)$$

defines a measurable function $X\to[0,\infty]$, as results from Corollary 2.1.5. Since plainly $f\geq f_n$ for all $n\geq 0$, monotonocity of the integral for positive functions yields

$$\int_X f(x)\,\mathrm{d}\mu(x) \geq \sup_{n\geq 0}\int_X f_n(x)\,\mathrm{d}\mu(x) = \lim_{n\to\infty}\int_X f_n(x)\,\mathrm{d}\mu(x) .$$

We need to show the converse inequality. Let thus φ be a simple function with $0\leq\varphi\leq f$, and fix $0<\alpha<1$. Define, for all $n\geq 0$,

$$E_n = \{x\in X : f_n(x)\geq \alpha\varphi(x)\} .$$

Then E_n is measurable for all $n\geq 0$ since f_n and φ are measurable; furthermore, the sequence $(E_n)_{n\geq 0}$ is increasing and $X=\bigcup_{n\geq 0} E_n$, as follows from $\alpha\varphi<f$ and $f_n\to f$ as $n\to\infty$. Now

$$\int_X f_n\,\mathrm{d}\mu \geq \int_X f_n\chi_{E_n}\,\mathrm{d}\mu \geq \int_X \alpha\varphi\chi_{E_n}\,\mathrm{d}\mu . \tag{2.3.2}$$

Notice that $\varphi\chi_{E_n}$ is a simple function; also, if $\varphi=\sum_{i\in I} c_i\chi_{A_i}$ is the standard representation of φ, then by Remark 2.2.4 we have

$$\int_X \varphi\chi_{E_n}\,\mathrm{d}\mu = \sum_{i\in I} c_i\mu(A_i\cap E_n) \overset{n\to\infty}{\longrightarrow} \sum_{i\in I} c_i\mu(A_i) = \int_X \varphi\,\mathrm{d}\mu , \tag{2.3.3}$$

where the claimed convergence is given by continuity of μ from below. Combining (2.3.2) and (2.3.3), and using Lemma 2.2.5(1), we get

$$\lim_{n\to\infty}\int_X f_n\,\mathrm{d}\mu \geq \alpha\int_X \varphi\,\mathrm{d}\mu .$$

Taking the supremum over all φ and over all α achieves the conclusion, in light of (2.3.1). □

A first consequence of the Monotone Convergence Theorem is additivity of the integral of positive functions: if $f, g\colon X \to [0,\infty]$ are measurable, then

$$\int_X f + g \ \mathrm{d}\mu = \int_X f \ \mathrm{d}\mu + \int_X g \ \mathrm{d}\mu \ .$$

To see this, choose increasing sequences $(\varphi_n)_{n\geq 0}$ and $(\psi_n)_{n\geq 0}$ of simple functions such that $\varphi_n \to f$ and $\psi_n \to g$ as $n \to \infty$. The existence of such sequences is guaranteed by Proposition 2.3.2. Then the sequence of sums $(\varphi_n + \psi_n)_{n\geq 0}$ is increasing and converges to the sum $f+g$. By the Monotone Convergence Theorem,

$$\int_X f + g \ \mathrm{d}\mu = \lim_{n\to\infty} \int_X \varphi_n + \psi_n \ \mathrm{d}\mu = \lim_{n\to\infty} \int_X \varphi_n \ \mathrm{d}\mu + \lim_{n\to\infty} \int_X \psi_n \ \mathrm{d}\mu = \int_X f \ \mathrm{d}\mu + \int_X g \ \mathrm{d}\mu \ ,$$

using additivity of the integral for simple functions (cf. Lemma 2.2.5) in the second equality.

Corollary 2.3.6. *Let $(X, \mathfrak{M}, \mu)$ be a measure space, $(f_n)_{n\geq 0}$ a sequence of measurable functions $X \to [0,\infty]$. Then*

$$\int_X \sum_{n=0}^{\infty} f_n \ \mathrm{d}\mu = \sum_{n=0}^{\infty} \int_X f_n \ \mathrm{d}\mu \ .$$

We shall interpret this as a special case of *Fubini's theorem*, which concerns exchanging orders of integration, for the product of μ and the counting measure on $\mathbb{N}$: see §**??**.

Proof. First, the infinite sum $\sum_{n=0}^{\infty} f_n$ is a measurable function $X \to [0,\infty]$, as follows from Lemma 2.1.2 and Corollary 2.1.5. The claimed equality is a direct consequence of the Monotone Convergence Theorem applied to the sequence $(g_n)_{n\geq 0}$ of partial sums

$$g_n(x) = \sum_{k=0}^{n} f_k(x) \ , \quad x \in X,$$

in conjunction with (finite) additivity of the integral of positive functions, which has been discussed in the paragraph preceding the corollary. □

2.4. Integral of real, extended-real, and complex-valued functions

2.4.1. Integral of real-valued functions. Given a set X and a function $f\colon X \to \overline{\mathbb{R}}$, we define the **positive part** f^+ and the **negative part** f^- of f as

$$f^+(x) \coloneqq \sup\{f(x), 0\} \ , \quad f^-(x) \coloneqq \sup\{-f(x), 0\} = -\inf\{f(x), 0\} \ .$$

It is clear that both f^+ and f^- take values in $[0,\infty]$. Furthermore, a straightforward case-by-case analysis shows that

$$f = f^+ - f^- \ , \quad |f| = f^+ + f^- \ .$$

If $(X, \mathfrak{M})$ is a Borel space and f is measurable, then Lemma 2.1.4 shows that f^+ and f^- are measurable. It further results from Lemma 2.1.2 that $|f|$ is measurable.

Definition 2.4.1 (Integral of an $\overline{\mathbb{R}}$-valued function). Let $(X, \mathfrak{M}, \mu)$ be a measure space, $f\colon X \to \overline{\mathbb{R}}$ a measurable function such that either

$$\int_X f^+ \ \mathrm{d}\mu$$

or

$$\int_X f^- \ \mathrm{d}\mu$$

is finite. The **integral of f with respect to μ**, indicated with

$$\int_X f \ \mathrm{d}\mu \ ,$$

is defined as the difference

$$\int_X f^+ \, \mathrm{d}\mu - \int_X f^- \, \mathrm{d}\mu \, .$$

The assumption that at least one of $\int_X f^+ \, \mathrm{d}\mu$ and $\int_X f^- \, \mathrm{d}\mu$ is finite ensures that $\int_X f \, \mathrm{d}\mu$ is well defined as an element of $\overline{\mathbb{R}}$. Also, if f takes values in $[0, \infty]$, then the given definition agrees with Definition 2.3.1, since f^- vanishes everywhere.

We say that f is **integrable with respect to** μ, or μ**-integrable**, if

$$\int_X f \, \mathrm{d}\mu$$

is finite. From the definition and the relation between f, f^+, f^- and $|f|$, it follows immediately that:

LEMMA 2.4.2. *A function $f \colon X \to \overline{\mathbb{R}}$ is integrable if and only if both f^+ and f^- are integrable, which happens if and only if $|f|$ is integrable.*

2.4.2. Integral of complex-valued functions. We now turn to the integral of complex-valued functions. Let $(X, \mathfrak{M})$ be a Borel space, $f \colon X \to \mathbb{C}$ a complex-valued measurable function. The absolute value $|f| \colon X \to \mathbb{R}_{\geq 0}$, which is given by

$$|f(x)| = \left((\Re f(x))^2 + (\Im f(x))^2 \right)^{1/2} ,$$

is measurable, begin an algebraic expression[5] in the measurable functions $\Re f$ and $\Im f$ (cf. Lemma 2.1.1).

DEFINITION 2.4.3 (Integral of a complex valued function). Let $(X, \mathfrak{M} \mu)$ be a measure space, $f \colon X \to \mathbb{C}$ a measurable function such that $\int_X |f| \, \mathrm{d}\mu$ is finite. The **integral of** f **with respect to** μ, indicated with

$$\int_X f \, \mathrm{d}\mu \, ,$$

is defined as

$$\int_X \Re f \, \mathrm{d}\mu + i \int_X \Im f \, \mathrm{d}\mu \, .$$

Any measurable function $f \colon X \to \mathbb{C}$ satisfying $\int_X |f| \, \mathrm{d}\mu < \infty$, as in the previous definition, is called **integrable with respect to** μ, or μ**-integrable**. The assumption on finiteness of $\int_X |f| \, \mathrm{d}\mu$ ensures that both $\int_X \Re f \, \mathrm{d}\mu$ and $\int_X \Im f \, \mathrm{d}\mu$ are finite, which is seen by taking the elementary inequalities

$$|\Re f| \leq |f| \, , \quad |\Im f| \leq |f|$$

in conjunction with Lemma 2.4.2. Conversely, if both $\Re f$ and $\Im f$ are μ-integrable, then the triangle inequality

$$|f| \leq |\Re f| + |\Im f|$$

shows that $|f|$ is μ-integrable. We can thus upgrade Lemma 2.4.2 to complex-valued functions.

LEMMA 2.4.4. *A function $f \colon X \to \mathbb{C}$ is integrable*[6] *if and only if both $\Re f$ and $\Im f$ are integrable.*

If V is a real (resp. complex) vector space, a real (resp. complex) linear functional on V is a linear map $V \to \mathbb{R}$ (resp. $V \to \mathbb{C}$).

PROPOSITION 2.4.5. *Let $(X, \mathfrak{M}, \mu)$ be a measure space. The set of μ-integrable functions $X \to \mathbb{C}$ is a complex vector space, and the integral*

$$f \mapsto \int_X f \, \mathrm{d}\mu$$

is a complex linear functional on it.

[5]More precisely, it is given by the composition of products, sums and square-roots of measurable functions, which are measurable in view of Lemma 2.1.2 and Corollary 1.3.15.

[6]By definition, this is equivalent to integrability of $|f|$.

We shall denote the complex vector space of μ-integrable functions $X \to \mathbb{C}$ by $\mathcal{L}^1(X, \mathfrak{M}, \mu)$, a notation we shall vastly generalize in Chapter 3, together with its abridged, also widely employed versions $\mathcal{L}^1(X, \mu)$ and $\mathcal{L}^1(\mu)$; $\mathcal{L}^1(X, \mathfrak{M}, \mu)$ is a vector subspace of the $\mathbb{C}$-vector space of measurable functions $X \to \mathbb{C}$, which in turn is a subspace of the $\mathbb{C}$-vector space $\mathbb{C}^X$.

PROOF. Let $f, g\colon X \to \mathbb{C}$ be μ-integrable functions. We begin by showing that $f + g$ is μ-integrable and

$$\int_X f + g \, \mathrm{d}\mu = \int_X f \, \mathrm{d}\mu + \int_X g \, \mathrm{d}\mu \,. \tag{2.4.1}$$

When both f and g take positive values, we have already established both claims in the discussion succeeding Theorem 2.3.5. If f and g are real-valued, then write $h = f + g$, and decompose all three functions into the difference of their positive and negative parts:

$$h^+ - h^- = f^+ - f^- + g^+ - g^- \,.$$

Rearranging terms in the previous equality, and using the already established additivity for integrals of positive functions, we obtain

$$\int_X h^+ \, \mathrm{d}\mu + \int_X f^- \, \mathrm{d}\mu + \int_X g^- \, \mathrm{d}\mu = \int_X h^- \, \mathrm{d}\mu + \int_X f^+ \, \mathrm{d}\mu + \int_X g^+ \, \mathrm{d}\mu \,;$$

rearranging back terms, this yields (2.4.1).

At this point, the case of general complex-valued f and g can be inferred directly from the real case by decomposing them into real and imaginary parts.

We now prove that, if $c \in \mathbb{C}$, then cf is μ-integrable and

$$\int_X cf \, \mathrm{d}\mu = c \int_X f \, \mathrm{d}\mu \,. \tag{2.4.2}$$

If $c \in \mathbb{R}_{\geq 0}$ and f takes values in $\mathbb{R}_{\geq 0}$, then

$$\begin{aligned} \int_X cf \, \mathrm{d}\mu &= \sup\Big\{\int_X c\varphi \, \mathrm{d}\mu : 0 \leq \varphi \leq f, \ \varphi \text{ simple}\Big\} = \sup\Big\{c\int_X \varphi \, \mathrm{d}\mu : 0 \leq \varphi \leq f, \ \varphi \text{ simple}\Big\} \\ &= c\sup\Big\{\int_X \varphi \, \mathrm{d}\mu : 0 \leq \varphi \leq f, \ \varphi \text{ simple}\Big\} = c\int_X f \, \mathrm{d}\mu \,, \end{aligned} \tag{2.4.3}$$

where the second equality is taken from Lemma 2.2.5. In particular, the first term in the chain (3.2.3) is finite if and only if the last one is finite, and in this case (2.4.2) holds. It now follows that (2.4.2) holds for any real-valued f and any $c \in \mathbb{R}$, by splitting both c and f in their positive and negative parts and invoking the result for positive functions and positive scalars. If now f is complex-valued and $c \in \mathbb{C}$, then $|cf| = |c||f|$, whence $|cf|$ is integrable, and thus so is cf by definition. Furthermore,

$$\begin{aligned} \int_X cf \, \mathrm{d}\mu &= \int_X (\Re c + i\Im c)(\Re f + i\Im f) \, \mathrm{d}\mu = \int_X \Re c\Re f - \Im c\Im f + i(\Re c\Im f + \Im c\Re f) \, \mathrm{d}\mu \\ &= \int_X \Re c\Re f - \Im c\Im f \, \mathrm{d}\mu + i\int_X \Re c\Im f + \Im c\Re f \, \mathrm{d}\mu \\ &= \Re c\int_X \Re f \, \mathrm{d}\mu - \Im c\int_X \Im f \, \mathrm{d}\mu + i\Big(\Re c\int_X \Im f \, \mathrm{d}\mu + \Im c\int_X \Re f \, \mathrm{d}\mu\Big) \\ &= \Re c\Big(\int_X \Re f \, \mathrm{d}\mu + i\int_X \Im f \, \mathrm{d}\mu\Big) + i\Im c\Big(\int_X \Re f \, \mathrm{d}\mu + i\int_X \Im f \, \mathrm{d}\mu\Big) \\ &= (\Re c + i\Im c)\int_X f \, \mathrm{d}\mu = c\int_X f \, \mathrm{d}\mu \,, \end{aligned}$$

using the already established (2.4.2) for real-valued functions in the fourth equality, as well as the previously proven fact that the integral preserves sums. ☐

REMARK 2.4.6. Proposition 2.4.5 entails readily, by restriction, the analogous assertion with $\mathbb{R}$ in place of $\mathbb{C}$: the set of μ-integrable functions $X \to \mathbb{R}$ is a real vector space (more precisely, a subspace of $\mathbb{R}^X$), and the integral $f \mapsto \int_X f \, \mathrm{d}\mu$ is a real linear functional on it.

REMARK 2.4.7 (General version of monotonicity). If $f, g \in \mathcal{L}^1(X, \mathfrak{M}, \mu)$ are real-valued, then $f \leq g$ implies

$$\int_X f \, \mathrm{d}\mu \leq \int_X g \, \mathrm{d}\mu \, .$$

This has already been discussed for positive f, g, as a direct consequence of the definition. The general real case follows by decomposing f and g into their positive and negative parts, via a case-by-case analysis.

We now present a fundamental inequality for integrals, which in functional-analytic terms can be expressed as continuity of the linear functional in Proposition 2.4.5 with respect to the L^1-norm (to be defined in §**??**).

PROPOSITION 2.4.8. *Let $(X, \mathfrak{M}, \mu)$ be a measure space, $f \colon X \to \mathbb{C}$ a μ-integrable function. Then*

$$\left| \int_X f \, \mathrm{d}\mu \right| \leq \int_X |f| \, \mathrm{d}\mu \, .$$

In the proof we use the sign of a complex number $z \neq 0$, defined as $\operatorname{sgn}(z) = z/|z|$; it has unit modulus.

PROOF. We can assume $\int_X f \, \mathrm{d}\mu \neq 0$, else the claim is obvious. We use linearity of the integral, writing

$$\left| \int_X f \, \mathrm{d}\mu \right| = \alpha \int_X f \, \mathrm{d}\mu$$

for

$$\alpha = \operatorname{sgn}\Big(\int_X f \, \mathrm{d}\mu \Big)^{-1} \, .$$

We then have, by linearity,

$$\left| \int_X f \, \mathrm{d}\mu \right| = \int_X \alpha f \, \mathrm{d}\mu = \Re\Big(\int_X \alpha f \, \mathrm{d}\mu \Big) = \int_X \Re(\alpha f) \, \mathrm{d}\mu \leq \int_X |\Re(\alpha f)| \, \mathrm{d}\mu \leq \int_X |f| \, \mathrm{d}\mu \, ,$$

where the second equality holds since a real number, seen as a complex number, equals its real part, the third equality holds by definition of the integral of complex functions, and the last two inequalities are consequence of monotonicity for integrals of real-valued functions (Remark 2.4.7). □

As a first application of the foregoing fundamental inequality, we present a result which, in the language of §**??**, expresses the fact that simple functions are dense in the set of integrable functions, with respect to the L^1*-metric*.

PROPOSITION 2.4.9. *Let $(X, \mathfrak{M}, \mu)$ be a measure space, $f \in \mathcal{L}^1(X, \mathfrak{M}, \mu)$. For every $\varepsilon > 0$, there is a simple function $\varphi \in \mathcal{L}^1(X, \mathfrak{M}, \mu)$ such that*

$$\int_X |f - \varphi| \, \mathrm{d}\mu \leq \varepsilon \, .$$

PROOF. We start with the case of a positive function $f \in \mathcal{L}^1(X, \mathfrak{M}, \mu)$. Since, by definition, the finite quantity $\int_X f \, \mathrm{d}\mu$ equals

$$\sup\Big\{ \int_X \varphi \, \mathrm{d}\mu : 0 \leq \varphi \leq f, \ \varphi \text{ simple} \Big\} \, ,$$

for every $\varepsilon > 0$ there is a simple φ with $0 \leq \varphi \leq f$ and

$$\int_X \varphi \, \mathrm{d}\mu \geq \int_X f \, \mathrm{d}\mu - \varepsilon \, ,$$

which shows precisely that

$$\int_X |f-\varphi| \, \mathrm{d}\mu = \int_X f-\varphi \, \mathrm{d}\mu \le \varepsilon \ .$$

Suppose now f is real-valued, and write $f = f^+ - f^-$. Let $\varepsilon > 0$, and pick simple, integrable functions φ^+ and φ^- such that $0 \le \varphi^\pm \le f^\pm$ and $\int_X |f^\pm - \varphi^\pm| \, \mathrm{d}\mu \le \varepsilon/2$. The simple, integrable function $\varphi = \varphi^+ - \varphi^-$ satisfies

$$\int_X |f-\varphi| \, \mathrm{d}\mu = \int_X |(f^+ - \varphi^+) - (f^- - \varphi^-)| \, \mathrm{d}\mu \le \int_X |f^+ - \varphi^+| \, \mathrm{d}\mu + \int_X |f^- - \varphi^-| \le \varepsilon \ .$$

Lastly, if f is complex valued and $\varepsilon > 0$, then by the preceding argument there are simple integrable real-valued functions φ_1, φ_2 with $\int_X |\Re f - \varphi_1| \, \mathrm{d}\mu \le \varepsilon/2$ and $\int_X |\Im f - \varphi_2| \, \mathrm{d}\mu \le \varepsilon/2$. It is then clear that the simple function $\varphi = \varphi_1 + i\varphi_2$ is integrable and satisfies

$$\int_X |f-\varphi| \, \mathrm{d}\mu \le \int_X |\Re f - \varphi_1| \, \mathrm{d}\mu + \int_X |\Im f - \varphi_2| \le \varepsilon \ .$$

□

2.4.3. Integration on measurable subsets. Let $(X, \mathfrak{M}, \mu)$ be a measure space, E a measurable subset of X. We shall say that measurable function f, defined on X and taking values either in $\overline{\mathbb{R}}$ or $\mathbb{C}$, is **integrable on** E if the function $f\chi_E$, which equals f on E and vanishes outside of E, is μ-integrable. In this case we adopt the notation

$$\int_E f \, \mathrm{d}\mu$$

to indicate $\int_X f\chi_E \, \mathrm{d}\mu$.

It is plain that, if a function f as above is μ-integrable, then it is so on every measurable $E \subset X$: this follows from monotonicity of the integral for positive functions and the trivial inequality

$$|f\chi_E| \le |f| \ .$$

An immediate, but exceedingly useful observation is that, if $\mu(E) = 0$, then any measurable f as above is μ-integrable on E and

$$\int_E f \, \mathrm{d}\mu = 0 \ .$$

To see this, observe first that is suffices to check it for positive functions, upon decomposing a complex-valued function into its real and imaginary parts and an $\overline{\mathbb{R}}$-valued function into its positive and negative parts. Then notice that the property holds for any simple function

$$f = \sum_{i \in I} c_i \chi_{E_i} \ ,$$

expressed for convenience in standard representation; indeed, we have

$$f\chi_E = \sum_{i\in I} c_i \chi_{E_i} \chi_E = \sum_{i \in I} c_i \chi(E_i \cap E) \ ,$$

and thus, by definition of integral of a simple function,

$$\int_E f \, \mathrm{d}\mu = \sum_{i \in I} c_i \mu(E_i \cap E) = 0 \ ,$$

the last inequality holding since $E_i \cap E \subset E$ implies $\mu(E_i \cap E) = 0$. Lastly, the claimed property holds for an arbitrary positive function f since

$$\{\varphi \text{ simple} : 0 \le \varphi \le f\chi_E\} = \{\psi\chi_E : \psi \text{ simple}, \ 0 \le \psi \le f\} \ .$$

2.4.4. Further properties of integrable functions. In the next proposition, we summarize some useful properties that can be deduced on functions by knowledge of properties of their integrals.

We have already defined the notion of a σ-finite measure space. We generalize it to the following: given a measure space $(X, \mathfrak{M}, \mu)$, a subset $A \subset X$ is called **σ-finite with respect to μ** if there is a countable covering $(E_i)_{i\in I}$ of A consisting of measurable sets with $\mu(E_i) < \infty$ for all $i \in I$.

PROPOSITION 2.4.10. *Let $(X, \mathfrak{M}, \mu)$ be a Borel space.*

(1) If $f : X \to \overline{\mathbb{R}}$ is integrable, then $f(x)$ is finite for μ-almost every $x \in X$.

(2) If a measurable $f : X \to [0, \infty]$ satisfies

$$\int_X f \, \mathrm{d}\mu = 0 \, ,$$

then $f = 0$ μ-almost everywhere[7].

(3) If $f \in \mathcal{L}^1(X, \mathfrak{M}, \mu)$, then the set

$$\{x \in X : f(x) \neq 0\}$$

is σ-finite.

(4) If $f, g \in \mathcal{L}^1(X, \mathfrak{M}, \mu)$, then the following are equivalent:

(a) for all $E \in \mathfrak{M}$,

$$\int_E f \, \mathrm{d}\mu = \int_E g \, \mathrm{d}\mu \, ;$$

(b) $f = g$ μ-almost everywhere.

PROOF. We start with an integrable function $f : X \to \overline{\mathbb{R}}$, and let

$$Y = \{x \in X : f(x) \text{ is not finite}\} = \{x \in X : |f(x)| = \infty\} \, .$$

The inequality $|f| \geq \infty \cdot \chi_Y$ holds by definition of Y; as a consequence

$$\infty > \int_X |f| \, \mathrm{d}\mu \geq \int_X \infty \cdot \chi_Y \, \mathrm{d}\mu = \infty \cdot \mu(Y) \, , \tag{2.4.4}$$

where the last equality follows directly from the definition

$$\int_X \infty \cdot \chi_Y \, \mathrm{d}\mu = \sup\left\{ \int_X \varphi \, \mathrm{d}\mu : 0 \leq \varphi \leq \infty \cdot \chi_Y, \ \varphi \text{ simple} \right\} \, .$$

The only way in which (2.4.4) can hold is for $\mu(Y)$ to vanish, which is what we wanted to show.

Suppose now $f : X \to [0, \infty]$ is measurable and satisfies $\int_X f \, \mathrm{d}\mu = 0$. For every integer $n \geq 1$, let

$$X_n = \{x \in X : f(x) \geq 1/n \} \, .$$

Then we can estimate, for all $n \geq 1$,

$$0 = \int_X f \, \mathrm{d}\mu \geq \int_{X_n} f \, \mathrm{d}\mu \geq \int_{X_n} \frac{1}{n} \, \mathrm{d}\mu = \frac{1}{n} \, \mu(X_n) \, ,$$

where the third inequality follows from the fact that $f \geq 1/n$ on X_n, by construction. The last displayed chain of inequalities forces $\mu(X_n) = 0$ for all $n \geq 1$; therefore, $f > 0$ on a set which is μ-null, being the union $\bigcup_{n\geq 1} X_n$.

Suppose now $f \in \mathcal{L}^1(X, \mathfrak{M}, \mu)$. We write the (measurable) set $\{x \in X : f(x) \neq 0\}$ as the countable union $\bigcup_{n\in\mathbb{N}^*} X_n$ where

$$X_n = \{x \in X : |f(x)| \geq 1/n\}$$

[7] The converse being obviously true as well, as follows easily from the discussion in §2.4.3.

for all $n \geq 1$. Since $|f| \geq |f\chi_{X_n}|$, monotonicity of the integral yields

$$\infty > \int_X |f| \, \mathrm{d}\mu \geq \int_X |f\chi_{X_n}| \, \mathrm{d}\mu = \int_{X_n} |f| \, \mathrm{d}\mu \geq \int_{X_n} \frac{1}{n} \, \mathrm{d}\mu = \frac{1}{n} \, \mu(X_n) \ ,$$

This implies $\mu(X_n) < \infty$ for all $n \geq 1$, which establishes the σ-finite claim.

Finally, let $f, g \in \mathcal{L}^1(X, \mathfrak{M}, \mu)$, and define

$$N = \{x \in X : f(x) \neq g(x)\} \ ,$$

which is a measurable set. First, assume $\mu(N) = 0$, and let $E \in \mathfrak{M}$; decomposing $E = (E \cap N) \cup (E \setminus N)$, we get

$$\int_E f \, \mathrm{d}\mu = \int_X f\chi_E \, \mathrm{d}\mu = \int_X f(\chi_{E\cap N} + \chi_{E\setminus N}) \, \mathrm{d}\mu = \int_{E\cap N} f \, \mathrm{d}\mu + \int_{E\setminus N} f \, \mathrm{d}\mu = \int_{E\setminus N} f \, \mathrm{d}\mu \ ,$$

the last inequality being a consequence of the fact that $\mu(E \cap N) = 0$. Similarly, we get

$$\int_E g \, \mathrm{d}\mu = \int_{E\setminus N} g \, \mathrm{d}\mu \ ,$$

and since $f = g$ on $E \setminus N$, we deduce that

$$\int_E f \, \mathrm{d}\mu = \int_{E\setminus N} f \, \mathrm{d}\mu = \int_{E\setminus N} g \, \mathrm{d}\mu = \int_E g \, \mathrm{d}\mu \ .$$

Conversely, suppose $\int_E f \, \mathrm{d}\mu = \int_E g \, \mathrm{d}\mu$ for all $E \in \mathfrak{M}$, and set $h = f - g$. Upon splitting h into its real and imaginary part, we may assume h is real-valued. Then $N = N_1 \sqcup N_2$ with

$$N_1 = \{x \in X : h(x) > 0\} \ , \quad N_2 = \{x \in X : h(x) < 0\} \ .$$

The assumption

$$0 = \int_{N_1} h \, \mathrm{d}\mu = \int_X f\chi_{N_1} \, \mathrm{d}\mu$$

implies, by the first assertion of the proposition, that $h\chi_{N_1} = 0$ μ-almost everywhere, which ostensibly shows $\mu(N_1) = 0$ as $N_1 \subset \{x \in X : h\chi_{N_1} \neq 0\}$.

An entirely analogous argument delivers $\mu(N_2) = 0$, whence $h = 0$ μ-almost everywhere and thus $f = g$ μ-almost everywhere, as desired.

□

REMARK 2.4.11. The last two assertions of the proposition hold without modifications, and with the same proof, for integrable functions $f, g \colon X \to \overline{\mathbb{R}}$.

2.4.5. The Dominated Convergence Theorem and its consequences. It is now the appropriate moment to present two more fundamental convergence theorems for integrals, which improve upon the Monotone Convergence Theorem (Theorem 2.3.5) by relaxing its assumptions. The first of those theorems is a straightforward application of Theorem 2.3.5.

THEOREM 2.4.12 (Fatou's Lemma). *Let $(X, \mathfrak{M}, \mu)$ be a measure space, $(f_n)_{n\geq 0}$ a sequence of measurable functions $f_n \colon X \to [0, \infty]$. Then*

$$\int_X \liminf_{n\to\infty} f_n \, \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_X f_n \, \mathrm{d}\mu \ .$$

Notice that no integrability assumption is in place; both sides of the inequality may be infinite.

PROOF. First of all, Lemma 2.1.4 yields measurability of the function $\liminf_{n\to\infty} f_n$. Define now a new sequence $(g_n)_{n\geq 0}$ of measurable functions $g_n \colon X \to [0, \infty]$ by

$$g_n = \inf_{m\geq n} f_m$$

for all $n \geq 0$. Then it is clear that $(g_n)_{n\geq 0}$ is increasing, whence by the Monotone Convergence Theorem

$$\int_X \liminf_{n\to\infty} f_n \, \mathrm{d}\mu = \int_X \lim_{n\to\infty} g_n \, \mathrm{d}\mu = \lim_{n\to\infty} \int_X g_n \, \mathrm{d}\mu \, . \tag{2.4.5}$$

Now, for every fixed $n \geq 0$, $g_n \leq f_m$ for all $m \geq n$, and monotonicity of the integral yields

$$\int_X g_n \, \mathrm{d}\mu \leq \inf_{m\geq n} \int_X f_m \, \mathrm{d}\mu \, . \tag{2.4.6}$$

Combining (2.4.5) and (2.4.6) delivers the claim. □

The prime application of Fatou's Lemma is to the second fundamental limit theorem for integrals of this subsection, which allows to take the limit under the integral sign under a "uniform integrability" assumption. To motivate the need for such an assumption, let us present an example where taking limits under the integral sign is not justified, and indeed would yield a false equality.

Example 2.4.13. For every integer $n \geq 1$, define a function $f_n\colon [0,1] \to \mathbb{R}_{\geq 0}$ by

$$f_n(x) = \begin{cases} 2n^2 x & \text{if } 0 \leq x < \frac{1}{2n} \\ 2n - 2n^2 x & \text{if } \frac{1}{2n} \leq x < \frac{1}{n} \\ 0 & \text{if } x \geq \frac{1}{n} \end{cases} .$$

Thus, f_n is the continuous function whose graph interpolates linearly between the points $(0,0), (\frac{1}{2n}, n), (\frac{1}{n}, 0)$ and $(1,0)$ in $\mathbb{R}^2$. It is clear that $f_n \to 0$ as $n \to \infty$ pointwise, since $f_n(0) = 0$ for all n and, if $x > 0$, $f_n(x) = 0$ for all n such that $1/n < x$. Admitting for the time being the (entirely expected) fact that the Lebesgue integral on compact intervals of $\mathbb{R}$ yields the standard Riemann integral for continuous (more generally, Riemann-integrable) functions[8], we easily compute

$$\int_{[0,1]} f_n \, \mathrm{d}\mathscr{L}^1 = 1$$

for all n, whence

$$\int_{[0,1]} \lim_{n\to\infty} f_n \, \mathrm{d}\mathscr{L}^1 = 0 \neq 1 = \lim_{n\to\infty} \int_{[0,1]} f_n \, \mathrm{d}\mathscr{L}^1 \, .$$

The issue in the previous example is that the sequence $(f_n)_{n\geq 1}$ is not bounded from above, uniformly in n, by an integrable function; there is a sort of "escape of mass" which occurs close to the point 0. Ruling out such instances allows for a neat convergence theorem.

Theorem 2.4.14 (Dominated Convergence Theorem). *Let $(X, \mathfrak{M}, \mu)$ be a measure space, $(f_n)_{n\geq 0}$ a sequence of measurable functions $f_n\colon X \to \mathbb{C}$ converging pointwise to a (measurable) function f. Suppose there exists an integrable function g such that*

$$|f_n(x)| \leq g(x)$$

for all $x \in X$ and $n \geq 0$. Then f is integrable and

$$\int_X f \, \mathrm{d}\mu = \lim_{n\to\infty} \int_X f_n \, \mathrm{d}\mu \, . \tag{2.4.7}$$

Proof. We first show that f is integrable. Fatou's lemma applied to the sequence $(|f_n)_{n\geq 0}$ yields

$$\int_X |f| \, \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_X |f_n| \, \mathrm{d}\mu \leq \int_X g \, \mathrm{d}\mu \, ,$$

where the last inequality follows from monotonicity of integrals and the fact that $|f_n| \leq g$ for all n.

[8] This fact will be discussed at length in §??.

We now establish (2.4.7). Upon decomposing f into $\Re f$ and $\Im f$, it suffices to prove equality for real-valued functions. We apply first Fatou's Lemma to the sequence

$$h_n = g + f_n\,, \quad n \geq 0,$$

which consists of positive functions as $g \geq |f_n| \geq -f_n$ for all n. Using also finite additivity of integrals, we get

$$\int_X g\ \mathrm{d}\mu + \int_X f\ \mathrm{d}\mu = \int_X \lim_{n\to\infty} h_n\ \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_X h_n\ \mathrm{d}\mu = \int_X g\ \mathrm{d}\mu + \liminf_{n\to\infty} \int_X f_n\ \mathrm{d}\mu\,,$$

which delivers

$$\int_X f\ \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_X f_n\ \mathrm{d}\mu\,. \tag{2.4.8}$$

We now apply Fatou's lemma to a new sequence

$$h'_n = g - f_n\,, \quad n \geq 0,$$

which consists of positive functions as $g \geq |f_n| \geq f_n$ for all $n \geq 0$. Exploiting again finite additivity, we get

$$\int_X g\ \mathrm{d}\mu - \int_X f\ \mathrm{d}\mu = \int_X \lim_{n\to\infty} h'_n\ \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_X h'_n\ \mathrm{d}\mu = \int_X g\ \mathrm{d}\mu - \limsup_{n\to\infty} \int_X f_n\ \mathrm{d}\mu\,,$$

which yields

$$\limsup_{n\to\infty} \int_X f_n\ \mathrm{d}\mu \leq \int_X f\ \mathrm{d}\mu\,. \tag{2.4.9}$$

Taking (2.4.8) and (2.4.9) together delivers (2.4.7). □

We now reformulate Theorems 2.3.5 and 2.4.14 to allow for almost-everywhere holding assumptions on the converging sequence of measurable functions.

Theorem 2.4.15 (Monotone and Dominated Convergence Theorems, almost-everywhere version). *Let $(X, \mathfrak{M}, \mu)$ be a measure space.*

(1) (Monotone Convergence) Let $(f_n)_{n\geq 0}$ be a sequence of measurable functions $X \to [0,\infty]$ such that, for all $n \geq 0$,

$$f_n(x) \leq f_{n+1}(x) \quad \text{for } \mu\text{-almost every } x \in X.$$

Then $(f_n)_{n\geq 0}$ converges μ-almost everywhere to a measurable function $f\colon X \to [0,\infty]$ and

$$\int_X f\ \mathrm{d}\mu = \lim_{n\to\infty} \int_X f_n\ \mathrm{d}\mu\,.$$

(2) (Dominated Convergence) Let $(f_n)_{n\geq 0}$ be a sequence of measurable functions $X \to \mathbb{C}$ convering μ-almost everywhere to a measurable function f. Suppose that there is a μ-integrable function g such that, for all $n \geq 0$,

$$|f_n(x)| \leq g(x) \quad \text{for } \mu\text{-almost every } x \in X.$$

Then

$$\int_X f\ \mathrm{d}\mu = \lim_{n\to\infty} \int_X f_n\ \mathrm{d}\mu\,.$$

Proof. For both assertions, it suffices to disregard the μ-negligible sets over which the assumptions of Theorem 2.3.5 and 2.4.14 are not satisfied.

We now give the details of for the case of monotone convergence, leaving the analogous argument for dominated convergence to the reader. Define, for all $n \geq 0$,

$$E_n = \{x \in X : f_n(x) \leq f_{n+1}(x)\}\,,$$

and set also

$$F_n = \bigcup_{0\leq k\leq n} E_k$$

for all $n \geq 0$. The union $E = \bigcup_{n\geq 0} E_n = \bigcup_{n\geq 0} F_n$ is μ-null by the assumption and σ-subadditivity of μ. On E^c, the sequence $(f_n)_{n\geq 0}$ is pointwise increasing, and thus admits a measurable limit f. Extend f to a measurable function on X by setting $f = 0$ on E. Then $(f_n)_{n\geq 0}$ converges to f μ-almost everywhere.

The sequence of measurable functions $(f_n \chi_{F_n})_{n\geq 0}$ is easily seen to be increasing, in a way which allows us to apply Theorem 2.3.5: we infer that

$$\int_X f\chi_E \, \mathrm{d}\mu = \lim_{n\to\infty} \int_X f_n \chi_{F_n} \, \mathrm{d}\mu \, .$$

Since the complements of E and of F_n are μ-null for all $n \geq 0$, we have

$$\int_X f\chi_E \, \mathrm{d}\mu = \int_X f \, \mathrm{d}\mu \, , \qquad \int_X f_n \chi_{F_n} \, \mathrm{d}\mu = \int_X f_n \, \mathrm{d}\mu \, ,$$

achieving the conclusion. □

A consequence of the (general version of) Dominated Convergence Theorem for series is the following:

COROLLARY 2.4.16. *Let $(X, \mathfrak{M}, \mu)$ be a measure space, $(f_n)_{n\geq 0}$ a sequence of measurable functions $f_n \colon X \to \mathbb{C}$ such that*

$$\sum_{n=0}^{\infty} \int_X |f_n| \, \mathrm{d}\mu < \infty$$

Then there is an integrable function f such that the infinite sum $\sum_{n=0}^{\infty} f_n$ converges μ-almost everywhere to f, and

$$\int_X f \, \mathrm{d}\mu = \sum_{n=0}^{\infty} \int_X f_n \, \mathrm{d}\mu \, . \tag{2.4.10}$$

PROOF. Define a measurable function $f \colon X \to [0, \infty]$ by

$$g(x) = \sum_{n=0}^{\infty} |f_n(x)| \, .$$

By the hypothesis and Corollary 2.3.6,

$$\int_X g \, \mathrm{d}\mu = \sum_{n=0}^{\infty} \int_X |f_n| \, \mathrm{d}\mu < \infty \, ,$$

which in particular shows that $g(x)$ is finite for μ-almost every $x \in X$ (see the first assertion in Proposition 2.4.10). As absolute convergence of series of complex numbers implies simple convergence, we deduce that $\sum_{n=0} f_n(x)$ converges for all x in the complement of a μ-null subset N. Define

$$f(x) = \begin{cases} \sum_{n=0}^{\infty} f_n(x) & \text{if } x \notin N \\ 0 & \text{if } x \in N \end{cases} .$$

Then it is a relatively straightforward matter to show that f is measurable. Moreover,

$$\int_X |f| \, \mathrm{d}\mu = \int_{N^c} |f| \, \mathrm{d}\mu = \int_{N^c} \left| \sum_{n=0}^{\infty} f_n \right| \mathrm{d}\mu \leq \int_{N^c} \sum_{n=0}^{\infty} |f_n| \, \mathrm{d}\mu = \int_X g \, \mathrm{d}\mu \, ,$$

whence f is integrable. Finally, apply the Dominated Convergence Theorem to the sequence of partial sums

$$\sum_{k=0}^{n} f_k \, , \quad n \geq 0,$$

which verifies $|\sum_{k=0}^{n} f_k| \leq g$ for all $n \geq 0$, to get equality in (2.4.10). □

As a final, classical application of the Dominated Convergence Theorem, we discuss regularity properties of functions defined by integrals.

Proposition 2.4.17 (Integrals depending on a parameter). *Let (T, d) be a metric space, $(X, \mathfrak{M}, \mu)$ a measure space, $f\colon T \times X \to \mathbb{C}$ a function with the property that, for every $t \in T$, the map*

$$X \to \mathbb{C}\,, \quad x \mapsto f(t, x)$$

is in $\mathcal{L}^1(X, \mathfrak{M}, \mu)$. Define a function $F\colon T \to \mathbb{C}$ by

$$F(t) = \int_X f(t, x)\, \mathrm{d}\mu(x)\,.$$

Let t_0 be a basepoint in T.

(1) Assume there is a neighborhood V of t_0 in T and function $g \in \mathcal{L}^1(X, \mathfrak{M}, \mu)$ such that, for all $x \in X$, the function

$$\mathbb{R}^n \to \mathbb{C}\,, \quad t \mapsto f(t, x)$$

is continuous on V and satisfies

$$|f(t, x)| \leq g(x)$$

for all $t \in V$. Then F is continuous at t_0.

(2) Suppose now $T = \mathbb{R}^n$ for some integer $n \geq 1$, and d is the Euclidean metric. Assume further that, for a given vector $v \in \mathbb{R}^n$, there is a neighborhood V of t_0 in $\mathbb{R}^n$ and a function $h \in \mathcal{L}^1(X, \mathfrak{M}, \mu)$ such that, for all $x \in X$, the function

$$\mathbb{R}^n \to \mathbb{C}\,, \quad t \mapsto f(t, x)$$

admits a partial derivative $\partial_v f(t, x)$ in direction v on V satisfying

$$|\partial_v f(t, x)| \leq h(x)$$

for all $t \in V$. Then F admits a partial derivative in direction v at t_0, and

$$\partial_v F(t_0) = \int_X \partial_v f(t_0, x)\, \mathrm{d}\mu(x)\,.$$

In a way, the second assertion expresses the fact that, under appropriate assumptions, it is possible to "differentiate under the integral sign".

Proof. Both assertions are fairly straightforward consequences of the Dominated Convergence Theorem (Theorem 2.4.14).

We only prove the first assertion, leaving the second one for the exercises. By the sequential characterization of continuity in metric spaces[9], it suffices to show that, given any sequence $(t_n)_{n\geq 1}$ converging to t_0,

$$\int_X f(t_0, x)\, \mathrm{d}\mu(x) = \lim_{n\to\infty} \int_X f(t_n, x)\, \mathrm{d}\mu(x)\,. \tag{2.4.11}$$

Let $N \in \mathbb{N}$ be such that $t_n \in V$ for all $n \geq N$. Theorem 2.4.14 applied to the sequence of functions

$$X \to \mathbb{C}\,, \quad x \mapsto f(t_n, x)$$

for $n \geq N$ yields directly the convergence claim in (2.4.11). $\square$

[9]This is a crucial point; a similar argument wouldn't work on a general topological space. It works, more generally, for first countable topological spaces.

2.5. The integrals of Riemann and Lebesgue on the real line

In our opening considerations of §1.1 we discussed how the quest to generalize the traditional Riemann integral on the real line led to the development of Lebesgue's theory of integration. Now we compare the Lebesgue integral on $\mathbb{R}$ and the RIemann integral, showing that the former indeed subsumes the latter.

We start with a reminder on the notion of Riemann integral. Let $[a, b]$ be a compact interval of $\mathbb{R}$. A subdivision of $[a, b]$ is a finite sequence

$$\mathcal{S} = \{a = a_0 < a_1 < \cdots < a_n = b\}$$

of points in $[a, b]$. A step function on $[a, b]$ is a function $\varphi\colon [a, b] \to \mathbb{R}$ such that there is a subdivision $\mathcal{S}$ of $[a, b]$ as above with $\varphi|_{(a_{i-1}, a_i)}$ constant for all $1 \le i \le n$. If $f|_{(a_{i-1}, a_i)} = c_i$ for all $1 \le i \le n$, we define the Riemann integral of φ as

$$\int_a^b \varphi(x)\, \mathrm{d}x := \sum_{i=1}^n c_i(a_i - a_{i-1})\ .$$

A function $f\colon [a, b] \to \mathbb{R}$ is Riemann-integrable if

$$\sup\left\{\int_a^b \varphi(x)\, \mathrm{d}x : \varphi \le f,\ \varphi \text{ step function}\right\} = \inf\left\{\int_a^b \psi(x)\, \mathrm{d}x : f \le \psi,\ \psi \text{ step function}\right\} \tag{2.5.1}$$

and the common value is finite.

In particular, a Riemann-integrable function is necessarily bounded on $[a, b]$, as both the supremum and the infimum in (2.5.1) must be taken over non-empty families.

THEOREM 2.5.1. *Let $[a, b]$ be a compact interval of $\mathbb{R}$, $f\colon [a, b] \to \mathbb{R}$ a function.*

(1) If f is Riemann-integrable, then f is Lebesgue-measurable, and is integrable with respect to $\mathscr{L}^1|_{[a,b]}$. Furthermore,

$$\int_a^b f(x)\, \mathrm{d}x = \int_{[a,b]} f\, \mathrm{d}\mathscr{L}^1\ ;$$

(2) f is Riemann-integrable if and only if

$$\{x \in [a, b] : f \textit{ is discontinuous at } x\}$$

has zero Lebesgue measure.

PROOF. We only prove the first statement, while the second one is relegated to the exercise sheets. Suppose f is Riemann-integrable; notice that, once we show that f is Lebesgue-measurable, then integrability with respect to $\mathscr{L}^1|_{[a,b]}$ is automatic since $\mathscr{L}^1([a, b])$ is finite and f is bounded on $[a, b]$. For every integer $n \ge 1$, there are step functions $\varphi_n \le f \le \psi_n$ such that

$$\int_a^b \psi_n(x)\, \mathrm{d}x - \int_a^b \varphi_n\, \mathrm{d}x \le \frac{1}{n}$$

Notice that step functions are trivially measurable (in fact, they are simple functions even with respect to the Borel σ-algebra on $[a, b]$), and for them equality between the Riemann integral and the Lebesgue integral is obvious from the definition of both. Thus

$$\int_{[a,b]} \psi_n\, \mathrm{d}\mathscr{L}^1 - \int_{[a,b]} \varphi_n\, \mathrm{d}\mathscr{L}^1 \le \frac{1}{n}\ .$$

Upon taking common refinements of subdivisions, we may assume that the sequence $(\varphi_n)_n$ is $\mathscr{L}^1$-a.e. increasing and that the sequence $(\psi_n)_n$ is $\mathscr{L}^1$-a.e. decreasing. Let g and G denote, respectively, the pointwise a.e. limit of (φ_n) and (ψ_n) so that

$$g \le f \le G$$

$\mathscr{L}^1$-a.e. By the Dominated Convergence Theorem (all functions under consideration are uniformly bounded, $\mathscr{L}^1$-a.e., by the bounded step function $\sup\{|\varphi_0|, |\psi_0|\}$),

$$\int_{[a,b]} G \, \mathrm{d}\mathscr{L}^1 - \int_{[a,b]} g \, \mathrm{d}\mathscr{L}^1 = \lim_{n\to\infty} \int_{[a,b]} \psi_n \, \mathrm{d}\mathscr{L}^1 - \int_{[a,b]} \varphi_n \, \mathrm{d}\mathscr{L}^1 = 0 \, ,$$

whence by Proposition 2.4.10 $G(x) = g(x)$ for $\mathscr{L}^1$-almost every $x \in [a,b]$. Thus $f = g = G$ L^1-almost everywhere; since $\mathscr{L}^1$ is complete on the Lebesgue σ-algebra $\mathfrak{M}_{\mathscr{L}^1}$, it follows that f is Lebesgue measurable.

Finally,

$$\int_a^b f(x) \, \mathrm{d}x = \lim_{n\to\infty} \int_a^b \varphi_n(x) \, \mathrm{d}x = \lim_{n\to\infty} \int_{[a,b]} \varphi_n \, \mathrm{d}\mathscr{L}^1 = \int_{[a,b]} f \, \mathrm{d}\mathscr{L}^1 \, ,$$

showing the sought after equality. □

2.6. Measure-theoretic notions of convergence

We have already encountered three notions of convergence for sequences of measurable functions, which we briefly recapitulate. If $(X, \mathfrak{M}, \mu)$ is a measure space, a sequence $(f_n)_{n\geq 0}$ of measurable functions $f_n \colon X \to \mathbb{C}$ is said to converge towards a measurable function $f \colon X \to \mathbb{C}$

- uniformly, if
$$\sup_{x\in X} |f(x) - f_n(x)| \xrightarrow{n\to\infty} 0 \, ,$$
- pointwise, if
$$|f(x) - f_n(x)| \xrightarrow{n\to\infty} 0$$
for every $x \in X$, and
- pointwise almost everywhere if
$$|f(x) - f_n(x)| \xrightarrow{n\to\infty} 0$$
for μ-almost every $x \in X$.

It is obvious that uniform convergence implies pointwise convergence, which in turn implies pointwise almost everywhere convergence. We shall now exhibit examples showing that the converse implications do not hold.

Example 2.6.1. Consider the sequence $(f_n)_{n\geq 1}$ of Example 2.4.13. It converges pointwise to the zero function, but not in a uniform fashion: in fact,

$$\sup_{[0,1]} |f_n| = n$$

for all n, which does not tend to 0 as $n \to \infty$.

Example 2.6.2. For every integer $n \geq 1$, define a function $f_n \colon [0,1] \to \mathbb{R}_{\geq 0}$ by

$$f_n(x) = \begin{cases} 1 - nx & \text{if } 0 \leq x < \frac{1}{n} \\ 0 & \text{if } x \geq \frac{1}{n} \end{cases} .$$

In other words, f_n is the continuous function whose graph interpolates linearly between the points $(0,1), (\frac{1}{n}, 0)$ and $(1,0)$ in $\mathbb{R}^2$. Then $f_n(x) \to 0$ as $n \to \infty$ for all $x \in (0,1]$, whence $f_n \to 0$ as $n \to \infty$ almost everywhere with respect to the Lebesgue measure on $[0,1]$. However, $f_n(0) = 1$ for all n, whence (f_n) does not converge pointwise to 0.

A fundamental notion of convergence, of measure-theoretic origin, is known as *convergence in L^1* (the reason for the terminology will manifest itself in §3).

DEFINITION 2.6.3 (L^1-convergence). let $(X, \mathfrak{M}, \mu)$ be a measure space. We say that a sequence $(f_n)_{n\geq 0}$ of measurable functions $X \to \mathbb{C}$ **converges** towards a measurable function $f\colon X \to \mathbb{C}$ **in** L^1 if

$$\lim_{n\to\infty} \int_X |f - f_n| \, \mathrm{d}\mu = 0 \, . \tag{2.6.1}$$

It is thus a sort of averaged notion of convergence. In particular, since the condition in (2.6.1) is unchanged by altering the values of f_n and f on μ-null sets, it is plain that L^1-convergence has nothing to do with the purely topological notions of uniform and pointwise convergence. More striking is the fact that it is also unrelated to almost-everywhere convergence, meaning that it doesn' imply nor is implied by it.

EXAMPLE 2.6.4. The sequence $(f_n)_{n\geq 1}$ in Example 2.4.13 converges to 0 pointwise, hence *a fortiori* it does so Lebesgue-almost everywhere. However,

$$\int_{[0,1]} |f_n| \, \mathrm{d}x = 1$$

for all $n \geq 1$, whence (f_n) does not converge to 0 in L^1.

EXAMPLE 2.6.5. The sequence $(f_n)_{n\geq 0}$ of functions $[0,1] \to \mathbb{R}_{\geq 0}$ defined by

$$f_0 = \chi_{[0,1/2]}, \; f_1 = \chi_{[1/2,1]}, \; f_2 = \chi_{[0,1/4]}, \; f_3 = \chi_{[1/4,1/2]}, \; f_4 = \chi_{[1/2,3/4]}, \; f_5 = \chi_{[3/4,1]}, \; f_6 = \chi_{[0,1/8]}$$

and so forth, is easily seen to satisfy

$$\int_{[0,1]} |f_n| \, \mathrm{d}x \overset{n\to\infty}{\longrightarrow} 0 \, ,$$

that is, it converges to the zero function in L^1. However, it does not converge to it in a Lebesgue-almost everywhere sense, since for every $x \in [0,1]$ it holds that $f_n(x) = 1$ for infinitely many n.

The Dominated Convergence Theorem, on the other hand, shows that convergence in L^1 holds provided a domination assumption is added to almost everywhere convergence. If $(f_n)_{n\geq 0}$ is a sequence converging μ-almost everywhere to f, and there is an integrable g with $|f_n| \leq g$ μ-almost everywhere for all $n \geq 0$, then not only $\int_X f_n \, \mathrm{d}\mu \to \int_X f \, \mathrm{d}\mu$ as $n \to \infty$, but in fact (f_n) converges to f in L^1. To establish the latter fact, it suffices to apply the Dominated Convergence Theorem to the sequence $(|f_n - f|)_{n\geq 0}$, which converges to 0 almost everywhere and satisfies $|f_n - f| \leq 2|g|$ for all $n \geq 0$.

In terms of the converse implication, we shall shortly see that, although L^1-convergence does not imply almost-everywhere convergence, it does imply the latter upon restricting considerations to a subsequence.

DEFINITION 2.6.6 (Convergence in measure). Let $(X, \mathfrak{M}, \mu)$ be a measure space. We say that a sequence $(f_n)_{n\geq 0}$ of measurable functions $X \to \mathbb{C}$ **converges in measure** towards a measurable function $f\colon X \to \mathbb{C}$ if, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \mu(\{x \in X : |f(x) - f_n(x)| \geq \varepsilon\}) = 0 \, .$$

We say that $(f_n)_{n\geq 0}$ is **Cauchy in measure** if, for every $\varepsilon > 0$,

$$\lim_{m,n\to\infty} \mu(\{x \in X : |f_m(x) - f_n(x)| \geq \varepsilon \;\}) = 0 \, .$$

For the reader convenience, we spell out the condition of being Cauchy in measure: for every $\varepsilon > 0$ and for every $\delta > 0$, there is $N \in \mathbb{N}$ such that, for all $m, n \geq N$,

$$\mu(\{x \in X : |f_m(x) - f_n(x)| \geq \varepsilon\}) \leq \delta \, .$$

Just as a convergent sequence in a metric space is a Cauchy sequence, a sequence of functions which converges in measure is Cauchy in measure; this follows from a trivial application of subadditivity of μ.

We will shortly prove the converse, namely that sequences which are Cauchy in measure converge in measure. Before doing that, we present a fundamental inequality relating measures of sets on which a function is "large" to its integral, which shall imply, in particular, that convergence in measure is a weaker notion with respect to L^1-convergence.

THEOREM 2.6.7 (Markov's inequality). *Let $(X, \mathfrak{M}, \mu)$ be a measure space, $f\colon X \to [0, \infty]$ an integrable function. Then, for every real $\alpha > 0$,*

$$\mu(\{x \in X : f(x) \geq \alpha\}) \leq \frac{\int_X f \, \mathrm{d}\mu}{\alpha} .$$

PROOF. If $E_\alpha = \{x \in X : f(x) \geq \alpha\}$, then we have $f \geq \alpha \chi_{E_\alpha}$, whence

$$\int_X f \, \mathrm{d}\mu \geq \int_{E_\alpha} \alpha \, \mathrm{d}\mu = \alpha \mu(E_\alpha) ;$$

the claim follows by dividing the first and last term of the previous chain of inequalities by α. □

COROLLARY 2.6.8. *If a sequence $(f_n)_{n \geq 0}$ converges to f in L^1, then it converges to f in measure.*

PROOF. For every $\varepsilon > 0$, Markov's inequality yields

$$\mu(\{x \in X : |f(x) - f_n(x)| \geq \varepsilon\}) \leq \frac{\int_X |f - f_n| \, \mathrm{d}\mu}{\varepsilon} ;$$

as $n \to \infty$; the right-hand side is infinitesimal by assumption, whence so is the left-hand side as well. □

We now turn to the abovementioned converse of the fact that converging sequences in measure are Cauchy. For the proof, we we avail ourselves of a fundamental result in measure theory[10], which is routinely employed in arguments aimed at showing that a certain sequence of functions converges almost everywhere with respect to a certain measure.

THEOREM 2.6.9 (The Borel-Cantelli Lemma). *Let $(X, \mathfrak{M}, \mu)$ be a measure space, $(E_n)_{n \geq 0}$ a sequence of measurable sets. Suppose that*

$$\sum_{n=0}^{\infty} \mu(E_n) < \infty .$$

Then

$$\mu\Big(\limsup_{n \to \infty} E_n\Big) = 0 .$$

PROOF. By definition,

$$\limsup_{n \to \infty} E_n = \bigcap_{n \geq 0} \bigcup_{k \geq n} E_k ;$$

hence, by continuity from above of μ (and since $\mu(\bigcup_{n \geq 0} E_n)$ is finite, as given by subadditivity of μ and the hypothesis), it suffices to show that

$$\lim_{n \to \infty} \mu\Big(\bigcup_{k \geq n} E_n\Big) = 0 .$$

Now σ-subadditivity of μ yields

$$\mu\Big(\bigcup_{k \geq n} E_k\Big) \leq \sum_{k \geq n} \mu(E_k) \overset{n \to \infty}{\longrightarrow} 0 ,$$

the last convergence claim being a consequence of the assumption $\sum_{n \geq 0} \mu(E_n) < \infty$. □

[10] Arguably even more in probability theory, in which context it admits two complementary formulations, one of which (the one absent from these notes) involves the notion of *independent events*.

We can now state and prove:

PROPOSITION 2.6.10. *Suppose a sequence $(f_n)_{n\geq 0}$ is Cauchy in measure.*

(1) There exists a measurable function f such that $(f_n)_{n\geq 0}$ converges towards f in measure.
(2) There exists a subsequence $(f_{n_k})_{k\geq 0}$ which converges towards f μ-almost everywhere.
(3) If g is a measurable function such that $(f_n)_{n\geq 0}$ converges towards g in measure, then $f = g$ μ-almost everywhere.

PROOF. For the first assertion, it suffices to show that there is a subsequence $(f_{n_k})_{k\geq 1}$ converging in measure to some measurable function f. Indeed, if this is so, then for every $\varepsilon, \delta > 0$ there is an integer $N \geq 1$ such that both

$$\mu(\{x \in X : |f_m(x) - f_n(x)| \geq \varepsilon/2\}) \leq \delta/2$$

for all $m, n \geq N$ and

$$\mu(\{x \in X : |f_{n_k}(x) - f(x)| \geq \varepsilon/2\}) \leq \delta/2$$

for all $n_k \geq N$. Since, for all $n, n_k \geq 1$,

$$\{x \in X : |f_n(x) - f(x)| \geq \varepsilon\} \subset \{x \in X : |f_n(x) - f_{n_k}(x)| \geq \varepsilon/2\} \cup \{x \in X : |f_{n_k}(x) - f(x)| \geq \varepsilon/2\} ,$$

it follows from finite subadditivity of μ that, for $n \geq N$,

$$\mu(\{x \in X : |f_n(x) - f(x)| \geq \varepsilon\}) \leq \delta ,$$

which shows that the full sequence $(f_n)_{n\geq 0}$ converges to f.

Now, by assumption, for any integer $k \geq 1$, there is an integer $N = N(k) \geq 1$ such that, for all $m, n \geq N$,

$$\mu(\{x \in X : |f_m(x) - f_n(x)| \geq 2^{-k}\}) \leq 2^{-k} .$$

As a consequence, there is a subsequence $(f_{n_k})_{k\geq 1}$ such that, for all $k \geq 0$,

$$\mu(\{x \in X : |f_{n_{k+1}}(x) - f_{n_k}(x)| \geq 2^{-k}\}) \leq 2^{-k} ;$$

call the latter measurable set E_k. The Borel-Cantelli Lemma (Theorem 2.6.9) shows that, outside of a μ-null set N of exceptions, $x \in E_k^{\mathrm{c}}$ for all but finitely many k, whence $(f_{n_k}(x))_{k\geq 0}$ is Cauchy in $\mathbb{C}$, and thus converges to some $f(x) \in \mathbb{C}$. Set $f = 0$ on N. Then f is measurable as it is so when restricted to N (plainly) and to N^{c} (where it is a pointwise limit of measurable functions).

The previous argument already delivers that $(f_{n_k})_{k\geq 0}$ converges μ-almost everywhere to f. Let us show that it also converges to f in measure. Fix $\varepsilon, \delta > 0$, and choose $k_0 \geq 1$ such that $2^{-(k_0-1)} \leq \varepsilon$ and

$$\mu\Big(\bigcup_{\ell\geq k_0} E_\ell\Big) \leq \delta \tag{2.6.2}$$

(the second one being guaranteed, for all k_0 sufficiently large, by Theorem 2.6.9). Let $k \geq k_0$, and suppose $x \in \bigcap_{\ell\geq k} E_\ell^{\mathrm{c}}$; then, for all integers $j > k$,

$$|f_{n_j}(x) - f_{n_k}(x)| \leq \sum_{\ell=k}^{j-1} |f_{n_{\ell+1}}(x) - f_{n_\ell}(x)| \leq \sum_{\ell=k}^{j-1} 2^{-\ell} < 2^{-k} \sum_{t\in\mathbb{N}} 2^{-t} = 2^{-(k-1)} .$$

Taking the limit as $j \to \infty$, we conclude that $|f(x) - f_{n_k}(x)| \leq 2^{-(k-1)}$. We have thus shown that

$$\{x \in X : |f(x) - f_{n_k}(x)| > \varepsilon\} \subset \bigcup_{\ell\geq k} E_\ell ,$$

which combined with (2.6.2) gives

$$\mu(\{x \in X : |f(x) - f_{n_k}(x)| > \varepsilon\}) \leq \delta ,$$

as desired.

Finally, if g is a measurable function such that $(f_n)_{n\geq 0}$ converges to g in measure, then, by what has already been established, there is a subsequence $(f_{n_k})_{k\geq 0}$ converging to g μ-almost everywhere; by the same token, one can extract a further subsequence $(f_{n_{k_\ell}})_{\ell\geq 0}$ converging to f μ-almost everywhere, since $(f_{n_k})_{k\geq 0}$ converges to f in measure. Uniqueness of limits in $\mathbb{C}$ then yields $f = g$ μ-almost everywhere. □

On finite measure spaces, almost-everywhere convergence implies convergence in measure[11].

PROPOSITION 2.6.11. *Suppose $\mu(X) < \infty$. Let $(f_n)_{n\geq 0}$ be a sequence converging to f μ-almost everywhere. Then $(f_n)_{n\geq 0}$ converges to f in measure.*

In the exercises it is asked to provide a proof of the proposition.

We conclude this section with a fundamental result result which allows to upgrade, on finite measure spaces, almost-everywhere convergence to uniform convergence on arbitrarily large (in the measure-theoretic sense) subsets.

THEOREM 2.6.12 (Egoroff's Theorem). *Suppose $\mu(X) < \infty$. Let $(f_n)_{n\geq 0}$ be a sequence converging to f μ-almost everywhere. Then, for every $\varepsilon > 0$, there is a measurable subset $Y \subset X$ with $\mu(X \setminus Y) < \varepsilon$ such that $(f_n)_{n\geq 0}$ converges uniformly to f on Y.*

PROOF. Fix $\varepsilon > 0$. Define, for all integers $n, k \geq 1$,

$$E_{n,k} = \{x \in X : |f(x) - f_n(x)| \leq 1/k\} .$$

For every fixed k, the union of the increasing sequence of measurable sets

$$\bigcap_{n\geq N} E_{n,k} , \quad N \geq 1,$$

is a set whose complement is μ-null, by assumption. By continuity of μ from below, and since $\mu(X)$ is finite, we deduce that there is $N_0(k) \geq 1$ such that

$$\mu\Big(X \setminus \bigcap_{n\geq N_0(k)} E_{n,k}\Big) < \varepsilon/2^k . \tag{2.6.3}$$

Define

$$Y = \bigcap_{k\geq 1} \bigcap_{n\geq N_0(k)} E_{n,k} .$$

Then $\mu(X \setminus Y) < \varepsilon$ by (2.6.3) and subadditivity of μ; moreover, for if $x \in Y$, then for all $k \geq 1$ and $n \geq N_0(k)$ one has $|f(x) - f_n(x)| \leq 1/k$. Reformulating it, this means that, for all $k \geq 1$, for all $n \geq N_0(k)$,

$$\sup_{x\in Y} |f(x) - f_n(x)| \leq 1/k ,$$

which shows uniform convergence of $(f_n)_{n\geq 0}$ to f on Y. □

In the literature, the sort of uniform convergence which appears as conclusion of Egoroff's theorem is called at times *almost uniform convergence*, though we shall avoid this terminology for the possible risk of confusion with the case of sequences with converge uniformly on a full-measure set.

REMARK 2.6.13. The finiteness assumption on $\mu(X)$ is indispensable. The exercises ask to provide a counterexample if it is not fulfilled.

In terms of the conclusion of the theorem, it is not possible to deduce the existence of a set Y with μ-null complement, over which the sequence converges uniformly. Again, counterexamples are to be provided in the exercises.

[11] The most fundamental limit theorem in probability theory is the Law of Large Numbers, about convergence to the expectation of empirical means of independent identically distributed real random variables with finite first moment. It comes in two formulations, known as the Strong and the Weak Law of Large Numbers; the first claims almost-sure convergence, the second one convergence in measure, or as one says in that context, *in probability*. The reason for the terminology "strong" and "weak" is given in Proposition 2.6.11.

2.7. Some further constructions of measures

In this section we discuss a couple more ways to construct measures on Borel spaces, via *pushforwards* and *densities*.

2.7.1. Measures with densities.

We begin with densities.

PROPOSITION 2.7.1. *Let $(X, \mathfrak{M}, \mu)$ be a measure space.*

(1) *Let $\rho\colon X \to \mathbb{R}_{\geq 0}$ be a measurable function. The function $\mu_\rho\colon \mathfrak{M} \to [0, \infty]$ defined by*

$$\mu_\rho(E) = \int_E \rho \,\mathrm{d}\mu \tag{2.7.1}$$

is a measure on $(X, \mathfrak{M})$.

(2) *If $f\colon X \to [0, \infty]$ is a measurable function, then*

$$\int_X f \,\mathrm{d}\mu_\rho = \int_X f\rho \,\mathrm{d}\mu \,.$$

(3) *If $f\colon X \to \mathbb{C}$ is a measurable function, then f is μ_ρ-integrable if and only if $f\rho$ is μ-integrable, in which case*

$$\int_X f \,\mathrm{d}\mu_\rho = \int_X f\rho \,\mathrm{d}\mu \,.$$

The proof is part of the exercises.

We say that the measure μ_ρ has **density** ρ with respect to the measure μ. Observe that μ_ρ has the property that, whenever $E \in \mathfrak{M}$ satisfies $\mu(E) = 0$, then $\mu_\rho(E) = 0$; this defines, as we will discuss in §**??**, the class of *absolutely continuous measures* (with respect to the given starting measure μ), of which measures with densities will be the prime (and often only, in light of the *Radon-Nikodym* theorem) example.

REMARK 2.7.2. If we define a function $\mu_\rho\colon \mathfrak{M} \to [0, \infty]$ as in (2.7.1) with ρ a μ-integrable function $X \to \mathbb{C}$, the outcome is a *complex measure*, a notion we shall introduce in §**??**.

EXAMPLE 2.7.3 (Gaussian distributions on the real line). Let $\mathscr{L}^1$ be the Lebesgue measure on the Borel space $(\mathbb{R}, \mathfrak{B}_\mathbb{R})$, and fix two parameters $m \in \mathbb{R}$, $\sigma \in \mathbb{R}_{\geq 0}$. The **Gaussian distribution** (a.k.a. **Gaussian measure**, or **Normal distribution**) **of mean** 0 **and variance** σ^2 is the probability measure μ_{m,σ^2} with density

$$\rho_{m,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

with respect to $\mathscr{L}^1$. The normalizing constant in front of the exponential function ensures that $\int_\mathbb{R} \rho_{m,\sigma^2} \,\mathrm{d}\mathscr{L}^1 = 1$.

The Gaussian distribution (and its higher-dimensional generalizations) is ubiquitous in probability theory, mainly in light of the *Central Limit Theorem*, which asserts that, in a *distributional sense*, empirical means of a sequence of independent identically distributed real-valued random variables *with finite second moment* converge to a Gaussian distribution, irrespective of the specific common law of the original variables.

2.7.2. Pushforward of a measure.

Measures with densities are constructed on the same Borel space of the original measure. We now see how to "transfer" an original measure to another Borel space via a measurable map.

PROPOSITION 2.7.4. *Let $(X, \mathfrak{M}_X)$, $(Y, \mathfrak{M}_Y)$ be Borel spaces, $T\colon X \to Y$ a measurable map, μ a measure on $(X, \mathfrak{M}_X)$.*

(1) *The function $T_*\mu\colon \mathfrak{M}_Y \to [0, \infty]$ defined by*

$$T_*\mu(E) = \mu(T^{-1}(E))$$

is a measure on $(Y, \mathfrak{M}_Y)$.

(2) If $f\colon Y \to [0,\infty]$ is a measurable function, then

$$\int_Y f \, \mathrm{d}T_*\mu = \int_X f \circ T \, \mathrm{d}\mu \, .$$

(3) If $f\colon Y \to \mathbb{C}$ is a measurable function, then f is $T_*\mu$-integrable if and only if $f \circ T$ is μ-integrable, in which case

$$\int_Y f \, \mathrm{d}T_*\mu = \int_X f \circ T \, \mathrm{d}\mu \, .$$

The exercises ask to give a proof of the proposition.

The measure $T_*\mu$ is called the **pushforward** of μ by T.

Remark 2.7.5 (Distribution of a random variable). If $(X, \mathfrak{M}_X, \mu)$ is a probability space, $(Y, \mathfrak{M}_Y)$ is a Borel space and $f\colon X \to Y$ is a Y-valued random variable, the pushforward of μ under f is called the *distribution*, or *law* of the random variable f.

2.8. Product measures and Fubini's theorem

Let $(X, \mathfrak{M}, \mu)$ and $(Y, \mathfrak{N}, \nu)$ be measure spaces. As discussed in Example 1.5.9, the collection

$$\mathcal{E} = \{A \times B : A \in \mathfrak{M},\ B \in \mathfrak{N}\}$$

is a semiring on the product set $X \times Y$ which generates the product σ-algebra $\mathfrak{M} \otimes \mathfrak{N}$. We define a function $\rho\colon \mathcal{E} \to [0,\infty]$ by

$$\rho(A \times B) = \mu(A)\nu(B)$$

for all $A \times B \in \mathcal{E}$, with the usual convention that $0 \cdot \infty = 0$.

Recall from §1.5.3 the notion of finitely additive and σ-subadditive function on a semiring.

Lemma 2.8.1. *The function ρ defined above is finitely additive, σ-subadditive and satisfies $\rho(\emptyset) = 0$.*

Proof. The last claim is obvious. We show finite additivity. Let $(A_i \times B_i)_{i \in I}$ be a finite collection of disjoint elements of $\mathcal{E}$ with $\bigcup_{i \in I} A_i \times B_i = A \times B \in \mathcal{E}$. Then, for all $(x,y) \in X \times Y$,

$$\sum_{i \in I} \chi_{A_i}(x)\chi_{B_i}(y) = \sum_{i \in I} \chi_{A_i \times B_i}(x,y) = \chi_{A \times B}(x,y) = \chi_A(x)\chi_B(y) \, .$$

Fix $x \in X$, and integrate the previous equality (which then becomes an equality between positive measurabile functions on Y) with respect to ν, so as to obtain

$$\sum_{i \in I} \chi_{A_i}(x)\nu(B_i) = \chi_A(x)\nu(B)$$

by finite additivity of the integral. Now the last displayed one is an equality between positive measurable functions on X; we can integrate it with respect to μ, which yields

$$\sum_{i \in I} \mu(A_i)\nu(B_i) = \mu(A)\nu(B) \, ,$$

using again finite additivity of the integral, which is the desired finite additivity of ρ.

We now turn to σ-subadditivity. Let $(A_n \times B_n)_{n \geq 0}$ be an $\mathcal{E}$-cover of some $A \times B \in \mathcal{E}$. This implies that, for all $(x,y) \in X \times Y$,

$$\chi_A(x)\chi_B(y) = \chi_{A \times B}(x,y) \leq \chi_{\bigcup_{n \geq 0} A_n \times B_n}(x,y) \leq \sum_{n \geq 0} \chi_{A_n \times B_n}(x,y) = \sum_{n \geq 0} \chi_{A_n}(x)\chi_{B_n}(y) \, .$$

As above, we first fix $x \in X$, and integrate the previous inequality with respect to ν: we obtain, by virtue of Corollary 2.3.6,

$$\chi_A(x)\nu(B) \leq \sum_{n \geq 0} \chi_{A_n}(x)\nu(B_n) \, .$$

Integrating now with respect to μ, and invoking Corollary 2.3.6 once more, we get

$$\mu(A)\nu(B) \leq \sum_{n\geq 0} \mu(A_n)\nu(B_n) ,$$

which is the sough after subadditivity inequality. □

We shall henceforth operate under the assumption that both $(X, \mathfrak{M}, \mu)$ and $(Y, \mathfrak{N}, \nu)$ are σ-finite, which ensures that ρ is σ-finite. Theorems 1.5.11 and 1.5.13 imply that there is a unique measure on the product Borel space $(X \times Y, \mathfrak{M} \otimes \mathfrak{N})$ extending ρ.

Definition 2.8.2 (Product measure). Let $(X, \mathfrak{M}, \mu)$ and $(Y, \mathfrak{N}, \nu)$ be σ-finite measure spaces. The **product measure** of μ and ν, denoted $\mu \times \nu$, is the unique measure on the product space $(X \times Y, \mathfrak{M} \otimes \mathfrak{N})$ satisfying

$$\mu \times \nu(A \times B) = \mu(A)\nu(B) \tag{2.8.1}$$

for all $A \in \mathfrak{M}$ and $B \in \mathfrak{N}$.

If $(X, \mathfrak{M}, \mu)$ and $(Y, \mathfrak{N}, \nu)$ are arbitrary measure spaces, we still define the product of μ and ν, indicated with $\mu \times \nu$, as the measure on the product measurable space arising from ρ as in Theorem 1.5.11; it still satisfies (2.8.1), but it may not be the unique extension of ρ to $\mathfrak{M} \otimes \mathfrak{N}$ with such a property.

When μ and ν are σ-finite, the same clearly holds for $\mu \times \nu$ (and if both μ and ν are finite, the same applied to $\mu \times \nu$). We shall confine ourselves to the σ-finite case in all our remaining considerations of the present section.

The construction extends immediately to the case of finitely many σ-finite measure spaces $(X_i, \mathfrak{M}_i, \mu_i)$, $1 \leq i \leq r$, yielding the notion of the product measure $\mu_1 \times \cdots \times \mu_r$. The classical uniqueness argument, invoking the Monotone Class Lemma on appropriate collections of measurable subsets, shows then that associativity holds for the product operation, which for the case of three factors is expressed by

$$(\mu_1 \times \mu_2) \times \mu_3 = \mu_1 \times (\mu_2 \times \mu_3) = \mu_1 \times \mu_2 \times \mu_3 .$$

Example 2.8.3. As a fundamental example of product measure, the Lebesgue measure $\mathscr{L}^n$ on $(\mathbb{R}^n, \mathfrak{B}_{\mathbb{R}^n})$ is the n-fold product of the Lebesgue measure $\mathscr{L}^1$ on $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$[12]. Equality can be checked, by σ-finiteness, on any semiring generating $\mathfrak{B}_{\mathbb{R}^n}$, for instance on half-open rectangles, where it holds trivially.

Remark 2.8.4 (Product measures and independence). In the context of probability theory, given two random variables $f, g\colon X \to \mathbb{R}$ defined on a probability space $(X, \mathfrak{M}, \mu)$, if μ_f and μ_g denote the *laws* of f and g respectively, namely the pushforwards of μ under f and g, then f and g are *independent* if and only if the law $\mu_{(f,g)}$ of the pair $(f, g)\colon X \to \mathbb{R}^2$ is the product $\mu_f \times \mu_g$.

Given a set $E \subset X \times Y$, we define the **x-section** of E, for every $x \in X$, as the set

$$E_x = \{y \in Y : (x, y) \in E\} ,$$

and the **y-section** of E, for every $y \in Y$, as the set

$$E_y = \{x \in X : (x, y) \in E\} .$$

Similarly, if $f\colon X \times Y \to Z$ is a function with values in a set Z, we define the **x-section** of f, for all $x \in X$, as the function

$$f_x\colon Y \to Z , \quad y \mapsto f(x, y)$$

and the **y-section** of f for all $y \in Y$, as the function

$$f_y\colon X \to Z , \quad x \mapsto f(x, y) .$$

[12]More generally, for any pair (r, s) of positive integers with $r + s = n$, $\mathscr{L}^n$ is the product $\mathscr{L}^r \times \mathscr{L}^s$ for the canonical identification of $\mathbb{R}^n$ with $\mathbb{R}^r \times \mathbb{R}^s$.

If $E \subset X \times Y$, we thus have

$$(\chi_E)_x = \chi_{E_x} \ , \quad (\chi_E)_y = \chi_{E_y} \tag{2.8.2}$$

for all $x \in X$ and $y \in Y$.

LEMMA 2.8.5. *Let $(X, \mathfrak{M}, \mu)$ and $(Y, \mathfrak{N}, \nu)$ be measure spaces.*

(1) If $E \in \mathfrak{M} \otimes \mathfrak{N}$, then $E_x \in \mathfrak{N}$ for all $x \in X$ and $E_y \in \mathfrak{M}$ for all $y \in Y$.
(2) If $f : X \times Y \to \mathbb{C}$ is measurable with respect to $\mathfrak{M} \otimes \mathfrak{N}$, then f_x is $\mathfrak{N}$-measurable for all $x \in X$ and f_y is $\mathfrak{M}$-measurable for all $y \in Y$.

PROOF. We start with the first assertion, and consider the family

$$\mathscr{M} = \{E \in \mathfrak{M} \otimes \mathfrak{N} : E_x \text{ is } \mathfrak{N}\text{-measurable for all } x \in X\} \ .$$

Then $\mathscr{M}$ contains the semiring

$$\mathcal{E} = \{A \times B : A \in \mathfrak{M}, \ B \in \mathfrak{N}\} \ ,$$

since for $E = A \times B$ we have $E_x \in \{\emptyset, B\}$ for all $x \in X$. Also, $\mathscr{M}$ is a σ-algebra. We have already showed that it contains $\emptyset$ and $X \times Y$; if $E \in \mathscr{M}$, then $(E^{\mathrm{c}})_x = E_x^{\mathrm{c}}$, and is thus $\mathfrak{N}$-measurable, for all $x \in X$, whence $\mathscr{M}$ is closed under complements. Lastly, if $(E_i)_{i \in I}$ is a countable collection of elements of $\mathscr{M}$, then

$$\left(\bigcup_{i \in I} E_i\right)_x = \bigcup_{i \in I} (E_i)_x$$

for all $x \in X$, so that $\bigcup_{i \in I} E_i \in \mathscr{M}$.

It follows that $\mathscr{M} = \mathfrak{M} \otimes \mathfrak{N}$, since the latter is generated by $\mathcal{E}$, and an entirely analogous argument shows that

$$\{E \in \mathfrak{M} \otimes \mathfrak{N} : E_y \text{ is } \mathfrak{M}\text{-measurable for all } y \in Y\} = \mathfrak{M} \otimes \mathfrak{N} \ .$$

As far as the second assertion is concerned, notice that it boils down to the first one for $f = \chi_E$ with $E \in \mathfrak{M} \otimes \mathfrak{N}$, on account of (2.8.2). Taking sections of a function is an operation that clearly commutes with finite sums and limits, which shows that the statement is readily extended to finite linear combinations of characteristic functions, namely to simple functions, then to positive functions via approximation by positive simple functions, and finally to real-valued and complex-valued functions by decomposition, respectively, into positive and negative part and into real and imaginary part. □

LEMMA 2.8.6. *Let $(X, \mathfrak{M}, \mu)$ and $(Y, \mathfrak{N}, \nu)$ be σ-finite measure spaces, $E \in \mathfrak{M} \otimes \mathfrak{N}$. Then the function*

$$X \to [0, \infty] \ , \quad x \mapsto \nu(E_x)$$

is $\mathfrak{M}$-measurable and the function

$$Y \to [0, \infty] \ , \quad y \mapsto \mu(E_y)$$

is $\mathfrak{N}$-measurable.

PROOF. We proof $\mathfrak{M}$-measurability of $X \to [0, \infty]$, $x \mapsto \nu(E_x)$; the second assertion is proved analogously. Let

$$\mathscr{M} = \{E \in \mathfrak{M} \otimes \mathfrak{N} : \text{the map } x \mapsto \nu(E_x) \text{ is } \mathfrak{M}\text{-measurable}\} \ .$$

Observe first that $\mathscr{M}$ contains the semiring

$$\mathcal{E} = \{A \times B : A \in \mathfrak{M}, \ B \in \mathfrak{N}\} \ ,$$

for when $E = A \times B$ as above one has

$$\nu(E_x) = \begin{cases} \nu(B) & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} \ , \tag{2.8.3}$$

which is clearly an $\mathfrak{M}$-measurable (actually simple) function. Secondly, $\mathscr{M}$ is a σ-algebra. The fact that it contains $\emptyset$ and $X \times Y$ has already been established. Suppose first that $(Y, \mathfrak{N}, \nu)$ is a finite measure space. Then, if $E \in \mathscr{M}$, we have

$$\nu((E^c)_x) = \nu(E_x^c) = \nu(Y) - \nu(E_x)$$

for all $x \in X$, whence $x \mapsto \nu((E^c)_x)$ is $\mathfrak{M}$-measurable as difference of $\mathfrak{M}$-measurable functions. This shows that $E^c \in \mathscr{M}$. Finally, if $(E_i)_{i \in I}$ is a countable sequence of pairwise disjoint sets in $\mathscr{M}$, then countable additivity of ν gives

$$\nu\Big(\Big(\bigcup_{i \in I} E_i\Big)_x\Big) = \nu\Big(\bigcup_{i \in I} (E_i)_x\Big) = \sum_{i \in I} \nu((E_i)_x)$$

for all $x \in X$, so that the function $x \mapsto \nu\big(\big(\bigcup_{i \in I} E_i\big)_x\big)$ is $\mathfrak{M}$-measurable as limit of sums of $\mathfrak{M}$-measurable functions, which implies $\bigcup_{i \in I} E_i \in \mathscr{M}$.

Since the semiring $\mathcal{E}$ generates the product σ-algebra $\mathfrak{M} \otimes \mathfrak{N}$, we conclude that $\mathscr{M} = \mathfrak{M} \otimes \mathfrak{N}$, which is the desired claim when $(Y, \mathfrak{N}, \nu)$ is finite.

Suppose now $Y = \bigcup_{n \in \mathbb{N}} Y_n$ with $\mu(Y_n)$ finite for all $n \geq 0$. Without loss of generality, we can assume that the sequence $(Y_n)_{n \geq 0}$ is increasing. Let $E \in \mathfrak{M} \otimes \mathfrak{N}$; then we write, for every $x \in X$, using continuity of ν from below,

$$\nu(E_x) = \nu\Big(\Big(\bigcup_{n \geq 0} E \cap (X \times Y_n)\Big)_x\Big) = \nu\Big(\bigcup_{n \geq 0} (E \cap (X \times Y_n))_x\Big) = \sum_{n \geq 0} \nu((E \cap (X \times Y_n))_x)\ . \quad (2.8.4)$$

For every fixed $n \geq 0$, the function

$$x \mapsto \nu((E \cap (X \times Y_n))_x)$$

is $\mathfrak{M}$-measurable, applying the previously established result to the finite measure space $(Y, \mathfrak{N}, \nu_n)$ with $\nu_n(B) = \nu(B \cap Y_n)$ for all $B \in \mathfrak{N}$. It follows from (2.8.4) that $x \mapsto \nu(E_x)$ is $\mathfrak{M}$-measurable. ☐

PROPOSITION 2.8.7. *Let $(X, \mathfrak{M}, \mu)$ and $(Y, \mathfrak{N}, \nu)$ be σ-finite measure spaces, $E \in \mathfrak{M} \otimes \mathfrak{N}$. Then*

$$\mu \times \nu(E) = \int_X \nu(E_x)\, \mathrm{d}\mu(x) = \int_Y \mu(E_y)\, \mathrm{d}\nu(y)\ .$$

PROOF. We shall prove that

$$\mu \times \nu(E) = \int_X \nu(E_x)\, \mathrm{d}\mu(x)\ , \quad (2.8.5)$$

the other equality being established similarly. Notice that the integral on the right-hand side is well defined, as the integrand is positive and $\mathfrak{M}$-measurable by Lemma 2.8.6.

We first treat the case of finite measure spaces $(X, \mathfrak{M}, \mu)$ and $(Y, \mathfrak{N}, \nu)$. Consider the collection

$$\mathscr{M} = \{E \in \mathfrak{M} \otimes \mathfrak{N} :\ (2.8.5) \text{ holds}\}\ .$$

First, notice that $\mathscr{M}$ contains the semiring

$$\mathcal{E} = \{A \times B : A \in \mathfrak{M},\ B \in \mathfrak{N}\}\ ;$$

indeed, if $E = A \times B$ as above, then (cf. (2.8.3))

$$\int_X \nu(E_x)\, \mathrm{d}\mu(x) = \int_X \nu(B) \chi_A(x)\, \mathrm{d}\mu(x) = \mu(A)\nu(B) = \mu \times \nu(E)\ .$$

We now show that $\mathscr{M}$ is a monotone class. It contains $\emptyset$, as already established. If $E \subset F$ are elements of $\mathscr{M}$, then

$$\int_X \nu((F \setminus E)_x)\, \mathrm{d}\mu(x) = \int_X \nu(F_x \setminus E_x)\, \mathrm{d}\mu(x) = \int_X \nu(F_x)\, \mathrm{d}\mu(x) - \int_X \nu(E_x)\, \mathrm{d}\mu(x)$$
$$= \mu \times \nu(F) - \mu \times \nu(E) = \mu \times \nu(F \setminus E)\ ,$$

where we have crucially used finiteness of μ and ν. This shows that $F \setminus E \in \mathscr{M}$. Lastly, if $(E_n)_{n\geq 0}$ is an increasing sequence of elements of $\mathscr{M}$, then

$$\begin{aligned}\int_X \nu\Big(\Big(\bigcup_{n\geq 0} E_n\Big)_x\Big)\, \mathrm{d}\mu(x) &= \int_X \nu\Big(\bigcup_{n\geq 0} (E_n)_x\Big)\, \mathrm{d}\mu(x) = \int_X \lim_{n\geq 0} \nu((E_n)_x)\, \mathrm{d}\mu(x) \\ &= \lim_{n\to\infty} \int_X \nu((E_n)_x)\, \mathrm{d}\mu(x) = \lim_{n\to\infty} \mu\times\nu(E_n) \\ &= \mu\times\nu\Big(\bigcup_{n\geq 0} E_n\Big) ,\end{aligned}$$

using, in successive order, continuity from below of ν, the Monotone Convergence Theorem, the assumption that $E_n \in \mathscr{M}$ for all $n \geq 0$, and continuity from below of $\mu \times \nu$.

At this point, the Monotone Class Lemma delivers $\mathscr{M} = \mathfrak{M} \otimes \mathfrak{N}$, which is the sought after statement for finite measure spaces.

Suppose now $X = \bigcup_{n\geq 0} X_n$ and $Y = \bigcup_{n\geq 0} Y_n$ for increasing sequences $(X_n)_{n\geq 0}$ and $(Y_n)_{n\geq 0}$ with $\mu(X_n)$ and $\nu(Y_n)$ finite for all $n \geq 0$. If $E \in \mathfrak{M} \otimes \mathfrak{N}$, then

$$\begin{aligned}\int_X \nu(E_x)\, \mathrm{d}\mu(x) &= \int_X \nu\Big(\bigcup_{n\geq 0} (E\cap (X_n\times Y_n))_x\Big)\, \mathrm{d}\mu(x) = \int_X \lim_{n\to\infty} \nu((E\cap (X_n\times Y_n))_x)\, \mathrm{d}\mu(x) \\ &= \lim_{n\to\infty} \int_X \nu((E\cap (X_n\times Y_n))_x)\, \mathrm{d}\mu(x) = \lim_{n\to\infty} \mu\times\nu(E\cap (X_n\times Y_n)) \\ &= \mu\times\nu(E) ,\end{aligned}$$

using, in successive order, continuity of ν from below, the Monotone Convergence Theorem, the already established equality (2.8.5) for the finite restrictions $\mu|_{X_n}$ and $\nu|_{Y_n}$, and continuity of $\mu \times \nu$ from below. This completes the proof. □

THEOREM 2.8.8 (The Fubini-Tonelli Theorem). *Let $(X, \mathfrak{M}, \mu)$ and $(Y, \mathfrak{N}, \nu)$ be σ-finite measure spaces.*

(1) (Tonelli's theorem) *Let $f\colon X\times Y \to [0,\infty]$ be measurable with respect to $\mathfrak{M}\otimes\mathfrak{N}$. Then the function*

$$x \mapsto \int_Y f(x,y)\, \mathrm{d}\nu(y) \tag{2.8.6}$$

is $\mathfrak{M}$-measurable, the function,

$$y \to \int_X f(x,y)\, \mathrm{d}\mu(x) \tag{2.8.7}$$

is $\mathfrak{N}$-measurable, and

$$\int_{X\times Y} f(x,y)\, \mathrm{d}\mu\times\nu(x,y) = \int_X \Big(\int_Y f(x,y)\, \mathrm{d}\nu(y)\Big)\, \mathrm{d}\mu(x) = \int_Y \Big(\int_X f(x,y)\, \mathrm{d}\mu(x)\Big)\, \mathrm{d}\nu(y) . \tag{2.8.8}$$

(2) (Fubini's theorem) *Let $f\colon X\times Y \to \mathbb{C}$ be a $(\mu\times\nu)$-integrable function. Then the section f_x is ν-integrable for μ-almost every $x \in X$, the section f_y is μ-integrable for ν-almost every $y \in Y$. Furthermore, the functions*

$$x \mapsto \int_Y f(x,y)\, \mathrm{d}\nu(y) , \quad y \mapsto \int_X f(x,y)\, \mathrm{d}\mu(x)$$

extend, respectively, to an μ-integrable function on X and a ν-integrable function on Y, and

$$\int_{X\times Y} f(x,y)\, \mathrm{d}\mu\times\nu(x,y) = \int_X \Big(\int_Y f(x,y)\, \mathrm{d}\nu(y)\Big)\, \mathrm{d}\mu(x) = \int_Y \Big(\int_X f(x,y)\, \mathrm{d}\mu(x)\Big)\, \mathrm{d}\nu(y) . \tag{2.8.9}$$

Notice that, if $f\colon X\times Y\to\mathbb{C}$ is measurable with respect to $\mathfrak{M}\otimes\mathfrak{N}$, then the x-section of f is $\mathfrak{N}$-measurable for every $x\in X$ and the y-section of f is $\mathfrak{M}$-measurable for every $y\in Y$, as of Lemma 2.8.5; in particular, the functions in (2.8.6) and (2.8.7) are well defined.

Proof. We start with Tonelli's theorem. When $f=\chi_E$ for some $E\in\mathfrak{M}\otimes\mathfrak{N}$, a moment's thought reveals that the statement is a combination of Lemma 2.8.6 and Proposition 2.8.7. Suppose now f is simple, expressed in standard representation as $\sum_{i\in I}c_i\chi_{E_i}$. Then the functions in (2.8.6) and (2.8.7) are measurable as finite linear combinations of measurable functions; furthermore, linearity of integrals on $(X,\mathfrak{M},\mu)$ and $(Y,\mathfrak{N},\nu)$ gives

$$\begin{aligned}\int_X\Big(\int_Y f(x,y)\,\mathrm{d}\nu(y)\Big)\,\mathrm{d}\mu(x)&=\int_X\Big(\int_Y\sum_{i\in I}c_i\chi_{E_i}(x,y)\,\mathrm{d}\nu(y)\Big)\,\mathrm{d}\mu(x)\\&=\int_X\sum_{i\in I}c_i\Big(\int_Y\chi_{E_i}(x,y)\,\mathrm{d}\nu(y)\Big)\,\mathrm{d}\mu(x)\\&=\sum_{i\in I}c_i\int_X\Big(\int_Y\chi_{E_i}(x,y)\,\mathrm{d}\nu(y)\Big)\,\mathrm{d}\mu(x)\\&=\sum_{i\in I}c_i\int_{X\times Y}\chi_{E_i}(x,y)\,\mathrm{d}\mu\times\nu(x,y)\\&=\sum_{i\in I}c_i\mu\times\nu(E_i)=\int_{X\times Y}f\,\mathrm{d}\mu\times\nu\,.\end{aligned}$$

For a general positive f, approximate it via an increasing sequence $(\varphi_n)_{n\geq1}$ of positive simple functions; then the functions in (2.8.6) and (2.8.7) are measurable since, by the Monotone Convergence Theorem, they can be expressed as limits of measurable functions (with f replaced by φ_n). By the same token, applying the Monotone Convergence Theorem three times, we get

$$\begin{aligned}\int_X\Big(\int_Y f(x,y)\,\mathrm{d}\nu(y)\Big)\,\mathrm{d}\mu(x)&=\int_X\Big(\lim_{n\to\infty}\int_Y\varphi_n(x,y)\,\mathrm{d}\nu(y)\Big)\,\mathrm{d}\mu(x)\\&=\lim_{n\to\infty}\int_X\Big(\int_Y\varphi_n(x,y)\,\mathrm{d}\nu(y)\Big)\,\mathrm{d}\mu(x)\\&=\lim_{n\to\infty}\int_{X\times Y}\varphi_n\,\mathrm{d}\mu\times\nu=\int_{X\times Y}f\,\mathrm{d}\mu\times\nu\,.\end{aligned}$$

The second equality in (2.8.8) is proved analogously.

We now turn to Fubini's theorem. We apply Tonelli's theorem to the measurable function $|f|$, and deduce that (2.8.8) holds for it, with the integrals being finite by assumption. If follows at once that

$$\int_Y|f(x,y)|\,\mathrm{d}\nu(y)$$

is finite for all x is in a set X' whose complement is μ-null. On X' there is thus a well-defined function

$$x\mapsto\int_Y f(x,y)\,\mathrm{d}\nu(y)\,,$$

which can be shown to be measurable by decomposing f into real and imaginary part, decomposing each of these in turn into their positive and negative part, and applying the measurability claim in Tonelli's theorem. Extend now the function measurably to X, for instance by setting it to be constantly 0 on the complement of X'.

The first equality in (2.8.9) follows then again by the routine decomposition of f, which reduces matters to the case of a positive f, for which in turn Tonelli's theorem can be applied.

The second equality in (2.8.9), as well as the associated first part of the statement, is established by the same argument. This concludes the proof of the theorem. □

We shall henceforth omit parenthesis in (2.8.8) and (2.8.9), and simply write

$$\int_X \int_Y f(x,y)\ \mathrm{d}\nu(y)\ \mathrm{d}\mu(x) \quad \text{and} \quad \int_Y \int_X f(x,y)\ \mathrm{d}\mu(x)\ \mathrm{d}\nu(y)\ .$$

We now make several comments about Fubini-Tonelli's theorem.

(1) Tonelli's theorem and Fubini's theorem are typically used in combination, in the following way. One is routinely confronted with the task of proving that the order of integration can be validly exchanged, that is, one wants to show that

$$\int_X \int_Y f(x,y)\ \mathrm{d}\nu(y)\ \mathrm{d}\mu(x) = \int_Y \int_X f(x,y)\ \mathrm{d}\mu(x)\ \mathrm{d}\nu(y) \tag{2.8.10}$$

for a certain measurable $f\colon X \times Y \to \mathbb{C}$. First one applies Tonelli's theorem to $!f!$ verify that f is integrable with respect to $\mu \times \nu$; to this effect, it is equivalent by (2.8.8) to compute one of the two integrals

$$\int_X \int_Y |f(x,y)|\ \mathrm{d}\nu(y)\ \mathrm{d}\mu(x)\ , \quad \int_Y \int_X |f(x,y)|\ \mathrm{d}\mu(x)\ \mathrm{d}\nu(y)\ ,$$

which is then equal to $\int_{X\times Y} |f|\ \mathrm{d}\mu \times \nu$. At this point, one can then resort to Fubini's theorem, which justifies (2.8.10).

(2) The σ-finiteness assumption on $(X, \mathfrak{M}, \mu)$ and $(Y, \mathfrak{N}, \nu)$ has been fundamentally exploited at several stages of the proof, and is absolutely crucial for the validity of the statement. For instance, take the Borel space $([0,1], \mathfrak{B}_{\mathbb{R}}|_{[0,1]})$ equipped with $\mu = \mathscr{L}^1|_{[0,1]}$ and with the counting measure $\nu = \mathrm{c}$. Consider the Borel measurable function $f = \chi_\Delta$ where $\Delta = \{(x,x) : x \in [0,1]\}$ is the diagonal in $[0,1] \times [0,1]$. Then

$$\int_{[0,1]} \int_{[0,1]} f(x,y)\ \mathrm{dc}(y)\ \mathrm{d}x) \int_{[0,1]} 1\ \mathrm{d}x = 1\ ,$$

but

$$\int_{[0,1]} \int_{[0,1]} f(x,y)\ \mathrm{d}x\ \mathrm{dc}(y) = \int_{[0,1]} 0\ \mathrm{dc}(y) = 0\ ,$$

so that equality in (2.8.8) fails.

(3) Integrability of f with respect to the product $\mu \times \nu$ is essential in Fubini's theorem. The exercises ask to show that (2.8.9) can fail when such hypothesis is dropped.

(4) Even if $(X, \mathfrak{M}, \mu)$ and $(Y, \mathfrak{N}, \nu)$ are complete measure spaces, $(X \times Y, \mathfrak{M} \otimes \mathfrak{N}, \mu \times \nu)$ is almost never complete[13]. For instance, assuming $\mathfrak{N} \neq \mathcal{P}(Y)$, let $B \subset X$ be a set which is not in $\mathfrak{N}$, and let $A \subset X$ be a μ-null set. Then $A \times B$ is contained in the $(\mu \times \nu)$-null set $A \times Y$, and yet is not in $\mathfrak{M} \otimes \mathfrak{N}$, else otherwise B would be in $\mathfrak{N}$ by Lemma 2.8.5.

2.9. Change of variable formula for Lebesgue integrals on Euclidean spaces

In introductory calculus courses one learns the change-of-variable-formula for one-dimensional Riemann integrals

$$\int_{\varphi(a)}^{\varphi(b)} f(x)\ \mathrm{d}x = \int_a^b f(\varphi(y))\varphi'(y)\ \mathrm{d}y\ ,$$

where φ is, say an increasing continuously differentiable function $[a,b] \to [\varphi(a), \varphi(b)]$, and f is a continuous function on $[\varphi(a), \varphi(b)]$. In this section, we shall see how the formula extends to the Lebesgue integral in any Euclidean space $\mathbb{R}^n$.

In this section, a Lebesgue-integrable function is a function $f\colon \mathbb{R}^n \to \mathbb{C}$ which is measurable with respect to $\mathfrak{B}_{\mathbb{R}^n}$ and integrable with respect to the n-dimensional Lebesgue measure.

[13] On the other hand, we already know that Theorem 1.5.11 yields a complete measure on a σ-algebra containing $\mathfrak{M} \otimes \mathfrak{N}$.

We start with the case of linear changes of coordinates. If $T\colon \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation, $\det T$ indicates its determinant, a real number which vanishes if and only if T is not invertible.

PROPOSITION 2.9.1. *Let $T\colon \mathbb{R}^n \to \mathbb{R}^n$ be an invertible linear transformation.*

(1) If $f\colon \mathbb{R}^n \to [0,\infty]$ is $\mathfrak{B}_{\mathbb{R}^n}$-measurable, then so is $f\circ T$ and

$$\int_{\mathbb{R}^n} f(x)\,\mathrm{d}(x) = \int_{\mathbb{R}^n} f(T(y))|\det T|\,\mathrm{d}y\ . \tag{2.9.1}$$

(2) A function $f\colon \mathbb{R}^n \to \mathbb{C}$ is Lebesgue-integrable if and only $f\circ T$ is Lebesgue-integrable, in which case

$$\int_{\mathbb{R}^n} f(x)\,\mathrm{d}(x) = \int_{\mathbb{R}^n} f(T(y))|\det T|\,\mathrm{d}y\ .$$

In particular, specializing to the case $f = \chi_E$ for a Borel set $E \subset \mathbb{R}^n$ (and using the result for T^{-1}), we obtain the transformation formula

$$\mathscr{L}^n(T(E)) = |\det T|\mathscr{L}^n(E)\ , \tag{2.9.2}$$

from which we deduce the linear counterpart of the fact that the Lebesgue measure is invariant under translations:

COROLLARY 2.9.2. *The Lebesgue measure $\mathscr{L}^n$ is invariant under linear Euclidean isometries: for every Euclidean isometry $T\colon \mathbb{R}^n \to \mathbb{R}^n$ and every Borel set $E \subset \mathbb{R}^n$,*

$$\mathscr{L}^n(T(E)) = \mathscr{L}^n(E)\ .$$

This follows readily from (2.9.2) as a linear isometry T satisfies $T^*\circ T = \mathrm{id}_{\mathbb{R}^n}$, where T^* is the adjoint of T, and thus $1 = \det T^* \det T = (\det T)^2$.

PROOF OF PROPOSITION 2.9.1. Let $(e_1,\dots,e_n)$ be the canonical basis of $\mathbb{R}^n$. Upon employing the classical decomposition arguments for complex-valued functions, and thereby reducing matters to positive functions, it suffices to establish the first statement. For it, one can use the familiar linear-algebraic fact, stemming Gauss'algorithm for the row-reduction of a square matrix, that every invertible linear $T\colon \mathbb{R}^n \to \mathbb{R}^n$ is a product of finitely many elementary transformations, namely of invertible linear maps of the following form:

(1) there is $1 \le j \le n$ and $\alpha \in \mathbb{R}^\times$ such that $T(e_j) = \alpha e_j$ and $T(e_i) = e_i$ for all $i \neq j$;
(2) there are $1 \le i \neq j \le n$ such that $T(e_i) = e_j$, $T(e_j) = e_i$ and $T(e_k) = e_k$ for all $k \notin \{i,j\}$;
(3) there are $1 \le i \neq j \le n$ and $\alpha \in \mathbb{R}$ such that $T(e_i) = e_i + \alpha e_j$ and $T(e_k) = e_k$ for all $k \neq i$.

Since the determinant is a group homomorphism $\mathrm{GL}_n(\mathbb{R}) \to \mathbb{R}^\times$, it suffices to prove (2.9.1) for elementary transformations, which is rather straightforward. Details are to be provided in exercises. □

We now indicate how the change-of-variable formula just presented admits a vast generalization to continuously differentiable change of coordinates between open sets. Let $U, V \subset \mathbb{R}^n$ be open sets. A $\mathscr{C}^1$-diffeomorphism between U and V is an invertible map $F\colon U \to V$ admitting continuous first-order partial derivatives on U, such that $F^{-1}\colon V \to U$ admits continuous first-order partial derivatives on V; recall that this implies, in particular, that both F and F^{-1} are differentiable on their domains of definition. For every $y \in U$, the Jacobian of F at y, denoted $\mathrm{Jac}_F(y)$, is the determinant of the differential $\mathrm{d}F_y\colon \mathbb{R}^n \to \mathbb{R}^n$ of F at y.

THEOREM 2.9.3 (Change of variable formula). *Let $U, V \subset \mathbb{R}^n$ be open sets, $F\colon U \to V$ a $\mathscr{C}^1$-diffeomorphism.*

(1) If $f\colon V \to [0,\infty]$ *is measurable with respect to* $\mathfrak{B}_{\mathbb{R}^n}|_V$*, then* $f \circ F$ *is measurable with respect to* $\mathfrak{B}_{\mathbb{R}^n}|_U$ *and*

$$\int_V f(x)\,\mathrm{d}x = \int_U f(F(y))|\operatorname{Jac}_F(y)|\,\mathrm{d}y\,.$$

(2) A function $f\colon V \to \mathbb{C}$ *is Lebesgue-integrable on* V *if and only if* $f \circ F$ *is Lebesgue-integrable on* U*, in which case*

$$\int_V f(x)\,\mathrm{d}x = \int_U f(F(y))|\operatorname{Jac}_F(y)|\,\mathrm{d}y\,.$$

Proof. The exercises guide the interested reader to the (tedious) argument. □

Theorem 2.9.3 admits an equivalent formulation in terms of pushforwards and densities. It states nothing but the fact that the pushforward of $\mathscr{L}^n|_U$ under a $\mathscr{C}^1$-diffeomorphism $F\colon U \to V$ is the measure on V having density

$$\rho(y) = |\operatorname{Jac}_{F^{-1}}(y)|$$

with respect to $\mathscr{L}^n|_V$.

CHAPTER 3

L^p spaces

This chapter is devoted to the study of a fundamental class of function spaces, defined in terms of integrals, which are ubiquitous in modern analysis: they are known as L^p spaces, and provide a generalization of the space $\mathcal{L}^1$ of integrable functions on a measure space, which has been already introduced in previous chapters.

3.1. Functional-analytic preliminaries

This preliminary section introduces a few fundamental concepts in functional analysis we shall avail ourselves of in the sequel.

Definition 3.1.1 (Normed vector space). Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. A **normed vector space over** $\mathbb{K}$ is a pair $(E, \|\cdot\|)$ consisting of a $\mathbb{K}$-vector space E and a function $\|\cdot\| : E \to \mathbb{R}_{\geq 0}$, called **norm**, satisfying the following properties:

(1) $\|v\| = 0$ if and only if $v = 0$;
(2) (homogeneity) $\|\lambda v\| = |\lambda| \, \|v\|$ for all $v \in V$ and $\lambda \in \mathbb{K}$;
(3) (triangle inequality) $\|v + w\| \leq \|v\| + \|w\|$ for all $v, w \in V$.

In the definition, the notation $|\cdot|$ refers to the ordinary absolute value for complex numbers. Notice that, given the homogeneity property (2),

$$\|0\| = \|0v\| = 0 \, \|v\| = 0$$

for all $v \in E$, whence (1) is tantamount to the non-degeneracy implication

$$\|v\| = 0 \implies v = 0 \, . \tag{3.1.1}$$

A **seminorm** on a $\mathbb{K}$-vector space E is a function $\|\cdot\| : E \to \mathbb{R}_{\geq 0}$ which is homogeneous and satisfies the triangle inequality. Thus, a seminorm on E is a norm if and only if (3.1.1) holds.

For a seminorm $\|\cdot\|$ on E, the following **reverse triangle inequality** holds:

$$| \, \|u\| - \|v\| \, | \leq \|u - v\| \tag{3.1.2}$$

for all $u, v \in E$. It is obtained by applying the triangle inequality to the pairs $(u - v, v)$ and $(v - u, u)$.

There is a canonical way of manufacturing a norm out of a seminorm, which shall be relevant in the forthcoming construction of L^p spaces. In what follows, if E is a $\mathbb{K}$-vector space and F is a vector subspace of E, we denote elements of the quotient E/F either by $[u]$ or by $u + F$, with $u \in E$.

Lemma 3.1.2. *Let $\|\cdot\|$ be a seminorm on a $\mathbb{K}$-vector space E. Let*

$$F = \{v \in E : \|v\| = 0\} \, .$$

Then F is a $\mathbb{K}$ vector subspace of E, and the assignment

$$\|\cdot\|_{E/F} : E/F \to \mathbb{R}_{\geq 0} \, , \quad [u] \mapsto \|u\|$$

defines a norm on the quotient $\mathbb{K}$-vector space E/F.

Proof. We first need to show that $\|\cdot\|_{E/F}$ is well defined. Suppose thus that

$$u_1 + F = u_2 + F$$

for some $u_1, u_2 \in E$; we need to show that $\|u_1\| = \|u_2\|$. Now there is $v \in F$ such that $u_2 = u_1 + v$, and the triangle inequality gives

$$\|u_2\| \le \|u_1\| + \|v\| = \|u_1\| \ .$$

On the other hand, the reverse triangle inequality (3.1.2) yields

$$\|u_2\| = \|u_2 - v + v\| \ge \|u_2 - v\| - \|v\| = \|u_1\| \ .$$

Taking the two last displayed inequalities together yields the claim.

Homogeneity and the triangle inequality for $\|\cdot\|_{E/F}$ follow directly from the analogous properties of $\|\cdot\|$; we omit the very simple details. Suppose now that $u \in E$ satisfies $\|u + F\|_{E/F} = 0$; then by definition this means that $\|u\| = 0$, whence $u \in F$ and thus $u + F$ is the zero coset in E/F. This shows that $\|\cdot\|_{E/F}$ is a norm. □

If $(E, \|\cdot\|)$ is a normed vector space over $\mathbb{K}$, then $\|\cdot\|$ endowes E with the structure of a metric space, by defining a distance function $d \colon E \times E \to \mathbb{R}_{\ge 0}$ as

$$d(u, v) = \|u - v\| \ .$$

Verification of the axioms of a distance function is routine, starting from the axioms of a norm.

Definition 3.1.3 (Banach space). Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. A **Banach space** over $\mathbb{K}$ is a normed vector space $(E, \|\cdot\|)$ such that the induced metric d on E is complete.

Example 3.1.4. The field $\mathbb{K}$ itself is a Banach space for the norm given by its ordinary absolute value. More generally, for every integer $n \ge 1$, the Euclidean space $\mathbb{K}^n$ is a Banach space for the norm

$$\|(u_1, \dots, u_n)\| = (|u_1|^2 + \cdots |u_n|^2)^{1/2} \ , \quad (u_1, \dots, u_n) \in \mathbb{K}^n \ .$$

Example 3.1.5. Let X be a set. We denote by $\mathrm{B}(X, \mathbb{K})$ the $\mathbb{K}$-vector space[1] of bounded functions $X \to \mathbb{K}$, namely those $f \colon X \to \mathbb{K}$ such that the **supremum norm**

$$\|f\| \coloneqq \sup_{x \in X} |f(x)|$$

is finite. Then $(\mathrm{B}(X, \mathbb{K}), \|\cdot\|)$ is easily seen to be a normed vector space over $\mathbb{K}$; it is actually a Banach space.

Example 3.1.6. In the previous example, take X to be a topological space, and consider the subset of $\mathrm{B}(X, \mathbb{K})$ consiting of continuous functions, which we denote by $\mathrm{C_b}(X, \mathbb{K})$. Then $\mathrm{C_b}(X, \mathbb{K})$ is a vector subspace of $\mathrm{B}(X, \mathbb{K})$, essentially because the algebraic operations of sum and product are continuous on $\mathbb{K}$. With the induced norm, $\mathrm{C_b}(X, \mathbb{K})$ is a Banach space over $\mathbb{K}$; this follows at once from the fact that $\mathrm{B}(X, \mathbb{K})$ is a Banach space, and $\mathrm{C_b}(X, \mathbb{K})$ is a closed subspace thereof[2], for the topology induced by the supremum norm.

A fundamental class of Banach spaces is given by *Hilbert spaces*, which we now set out to introduce.

Definition 3.1.7 (Inner product). Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, E a vector space over $\mathbb{K}$. An **inner product** on E is a map $\langle \cdot, \cdot \rangle \colon E \times E \to \mathbb{K}$ satisfying the following properties:

(1) (sesquilinear property) for all $u, v, w \in E$ and $\lambda \in \mathbb{K}$,

$$\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle \ , \quad \langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$$
$$\langle \lambda u, v \rangle = \lambda \langle u, v \rangle \ , \quad \langle u, \lambda v \rangle = \overline{\lambda} \langle u, v \rangle \ ;$$

(2) (hermitian property) for all $u, v \in E$,

$$\langle v, u \rangle = \overline{\langle u, v \rangle} \ ;$$

[1] More precisely, it is a $\mathbb{K}$-subspace of the space of all functions $X \to \mathbb{K}$.

[2] To see this, recall simply that the limit of a uniformly converging sequence of continuous functions is continuous.

(3) (positive definiteness property) $\langle u, u\rangle \geq 0$ for all $u \in E$, and $\langle u, u\rangle = 0$ if and only if $u = 0$.

The pair $(E, \langle\cdot,\cdot\rangle)$ is called an **inner product space**, or a **pre-Hilbert space**, over $\mathbb{K}$.

If $\mathbb{K} = \mathbb{R}$, then the sesquilinear property becomes the **bilinear property** of the inner product, and the hermitian property becomes the **symmetric property**.

Notice that the hermitian property yields already $\langle u, u\rangle \in \mathbb{R}$ for all $u \in \mathbb{R}$, upon which the positive definiteness property imposes a further constraint.

Pre-Hilbert spaces are naturally vector spaces, in the following way which generalizes the familiar definition of the Euclidean norm on $\mathbb{K}^n$ in terms of the Euclidean inner product. If $(E, \langle\cdot,\cdot\rangle)$ is an inner product space over $\mathbb{K}$, define a function $\|\cdot\| : E \to \mathbb{R}_{\geq 0}$ by

$$\|u\| = \langle u, u\rangle^{1/2} .$$

It is immediate from the properties of $\langle\cdot,\cdot\rangle$ that $\|u\| = 0$ if and only if $u = 0$ and $\|\lambda u\| = |\lambda| \, \|u\|$ for all $u \in E$. To complete the verification that $\|\cdot\|$ is, as indicated by the notation, a norm on E, we need to check the validity of the triangle inequality, which follows from the following ubiquitous inequality.

THEOREM 3.1.8 (The Cauchy-Schwarz inequality). *Let $(E, \langle\cdot,\cdot,\rangle)$ be a pre-Hilbert space over $\mathbb{K}$, $\|\cdot\| : E \to \mathbb{R}_{\geq 0}$ the function defined above. Then, for all $u, v \in V$,*

$$|\langle u, v\rangle| \leq \|u\| \, \|v\| \; ; \tag{3.1.3}$$

moreover, equality holds if and only if u and v are linearly dependent.

PROOF. Without loss of generality, we can assume $\langle u, v\rangle \neq 0$, else there is nothing to prove. Let

$$\alpha = \operatorname{sgn}(\langle u, v\rangle) = \frac{1}{|\langle u, v\rangle|} \langle u, v\rangle \; ;$$

for every $t \in \mathbb{R}$, we have

$$0 \leq \|u + t\alpha v\|^2 = \langle u + t\alpha v, u + t\alpha v\rangle = \|u\|^2 + 2\Re\langle u, \alpha v\rangle t + |\alpha|^2 \|v\|^2 t^2 . \tag{3.1.4}$$

Regarding the last expression as a polynomial function in the real variable t, it follows that the (reduced) discriminant of the associated polynomial is negative, that is,

$$(\Re\langle u, \alpha v\rangle)^2 - |\alpha|^2 \|u\|^2 \|v\|^2 \leq 0 \; ;$$

since $|\alpha| = 1$ and $\langle u, \alpha v\rangle = |\langle u, v\rangle|$, the previous inequality amounts to

$$|\langle u, v\rangle|^2 \leq \|u\|^2 \|v\|^2 \; ,$$

which gives (3.1.3) upon taking square roots. From (3.1.4), equality occurs if and only if there is $t \in \mathbb{R}$ such that

$$\|u + t\alpha v\|^2 = 0 \; ,$$

that is, if and only if $u + t\alpha v = 0$ for some $t \in \mathbb{R}$, which imples that u and v are linearly dependent. Conversely, if $v = \lambda u$ for some $\lambda \in \mathbb{K}$, it is a straightforward matter to check that equality holds in (3.1.3). □

Armed with the Cauchy-Schwarz inequality, we can now conclude the verification that $\|u\| = \langle u, u\rangle^{1/2}$ is actually a norm on the pre-Hilbert space $(E, \langle\cdot,\cdot\rangle)$.

COROLLARY 3.1.9. *Let the assumptions be as in Theorem 3.1.8. Then, for all $u, v \in E$,*

$$\|u + v\| \leq \|u\| + \|v\| \; ,$$

with equality if and only if there is $\lambda \in \mathbb{R}_{\geq 0}$ such that $v = \lambda u$.

PROOF. Let u, v be vectors in E. Then we expand

$$\|u+v\|^2 = \|u\|^2 + \|v\|^2 + 2\Re\langle u, v\rangle \ .$$

Since, by Theorem 3.1.8,

$$|\Re\langle u, v\rangle| \leq |\langle u, v\rangle| \leq \|u\| \, \|v\| \ ,$$

we deduce that

$$\|u+v\|^2 \leq \|u\|^2 + \|v\|^2 + 2\,\|u\|\,\|v\| = (\|u\| + \|v\|)^2 \ ,$$

from which the triangle inequality follows by taking square roots.

Regarding the equality case, it is immediate to check that equality holds when $v = \lambda u$ for some real $\lambda \geq 0$. Conversely, suppose $\|u+v\| = \|u\| + \|v\|$ for some $u, v \in E$. Then, in particular, equality holds in the Cauchy-Schwarz inequality, which yields the existence of $\lambda \in \mathbb{K}$ such that $v = \lambda u$. Moreover, we must have

$$\Re\langle u, v\rangle = |\langle u, v\rangle| \ ,$$

that is,

$$\|u\|^2 \, \Re\overline{\lambda} = |\lambda| \, \|u\|^2 \ .$$

If $u = 0$, then $v = 0$ and λ can be taken to be a positive real number; else, the above implies $|\lambda| = \Re\overline{\lambda}$, which can only occur if $\lambda \in \mathbb{R}_{\geq 0}$. □

It is a consequence of the previous corollary that $\|\cdot\|$ is a norm on E, so that every inner product space is naturaly a normed vector space.

DEFINITION 3.1.10 (Hilbert space). Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. A **Hilbert space over** $\mathbb{K}$ is an inner product space $(H, \langle\cdot,\cdot\rangle)$ which is a Banach space for the norm $\|\cdot\|$ induced by the inner product.

3.1.1. Projections onto closed convex sets. The geometry of Hilbert spaces ensures existence and uniqueness of projections onto closed convex sets, a fact which has a number of useful consequences. Let us first introduce the notion of *projection* on a subset of a metric space. If (X, d) is a metric space and $Y \subset X$ is a subset, a **projection** of a point $x \in X$ onto Y is an element $a \in Y$ such that

$$d(x, a) = d(x, Y) \ ,$$

where the latter quantity is, we recall, defined as $\inf\{d(x, y) : y \in Y\}$. In general, there is no reason for a projection to exist, nor for it to be unique, assuming existence. It is left to the reader to find appropriate counterexamples.

We recall that a subset C of a $\mathbb{K}$-vector space E is convex if, for all $x, y \in C$ and all $0 \leq t \leq 1$, $tx + (1-t)y \in C$.

PROPOSITION 3.1.11 (Projection onto closed convex sets). *Let $(H, \langle\cdot,\cdot\rangle)$ be a Hilbert space, $C \subset H$ a non-empty closed convex set. Then every $u \in H$ admits a unique projection onto C.*

We denote the unique projection of an element u onto C by $\pi_C(u)$.

PROOF. We show here existence of the projection; uniqueness will follows from the characterization given in the forthcoming proposition. Let $(v_n)_{n\geq 1}$ be a sequence in C such that

$$\|u - v_n\| \leq d(u, C) + \frac{1}{n} \ ,$$

whose existence is assured by the definition of $d(u, C)$. We claim that $(v_n)_{n\geq 1}$ is a Cauchy sequence; this yields the existence claim, as by completeness of H the sequence converges to a vector v, which lies in C since C is closed. Furthermore, continuity of the distance function yields

$$d(u, C) \leq \|u - v\| = \lim_{n\to\infty} \|u - v_n\| \leq \liminf_{n\to\infty} \|u - v_n\| \leq d(u, C) \ ,$$

that is, v is a projection of u onto C.

We show that $(v_n)_{n\geq 1}$ is a Cauchy sequence. Fix integers $m, n \geq 1$: then we apply the *parallelogram's law* to the vectors $u - v_m$ and $u - v_n$, which is the equality

$$\|v_m - v_n\|^2 + \|2u - (v_m + v_n)\|^2 = 2(\|u - v_m\|^2 + \|u - v_n\|^2) ,$$

verifiable by an elementary expansion of both sides in terms of inner products. Since $2u-(v_m+v_n) = 2(u - \frac{v_m+v_n}{2})$ with $\frac{v_m+v_n}{2} \in C$ by convexity, it follows that

$$\begin{aligned}\|v_m - v_n\|^2 &\leq 2\Big(\Big(d(u,C) + \frac{1}{m}\Big)^2 + \Big(d(u,C) + \frac{1}{n}\Big)^2\Big) - 4\left\|u - \frac{v_m + v_n}{2}\right\|^2 \\ &\leq 2\Big(\Big(d(u,C) + \frac{1}{m}\Big)^2 + \Big(d(u,C) + \frac{1}{n}\Big)^2\Big) - 4d(u,C)^2 .\end{aligned}$$

It is clear that, for every $\varepsilon > 0$, one can choose an integer $N \geq 1$ such that the last displayed expression does not exceed ε for all $m, n \geq N$; this achieves the result. □

PROPOSITION 3.1.12 (Characterization of the projection). *Let the setup be as in Proposition 3.1.11, u an element of H. For a vector $v \in C$, the following are equivalent:*

(1) v is a projection of u onto C;
(2) for all $w \in C$,

$$\Re\langle u - v, w - v\rangle \leq 0 . \tag{3.1.5}$$

Before turning to the proof of the proposition, let us see how it implies the uniqueness claim in Proposition 3.1.11. If $v_1, v_2 \in C$ are projections of a vector u onto C, then from (3.1.5) we have

$$\Re\langle u - v_1, v_2 - v_1\rangle \leq 0 , \quad \Re\langle u - v_2, v_1 - v_2\rangle \leq 0 . \tag{3.1.6}$$

Now

$$\|v_1 - v_2\|^2 = \langle v_1-v_2, v_1-v_2\rangle = \langle v_1-u, v_1-v_2\rangle+\langle u-v_2, v_1-v_2\rangle = \langle u-v_1, v_2-v_1\rangle+\langle u-v_2, v_1-v_2\rangle ;$$

the last displayed complex number is a real number with negative real part, by (3.1.6). It follows from the above that $\|v_1 - v_2\|^2 \leq 0$, whence $v_1 = v_2$.

PROOF OF PROPOSITION 3.1.12. Suppose first that (3.1.5) holds. Then, for every $w \in C$,

$$\begin{aligned}\|u - w\|^2 &= \langle u - w, u - w\rangle = \langle u - v + v - w, u - v + v - w\rangle \\ &= \|u - v\|^2 + \|v - w\|^2 - 2\Re\langle u - v, w - v\rangle \geq \|u - v\|^2 ;\end{aligned}$$

it follows that $d(u,v) \leq d(u,C)$, whence v is a projection of u onto C.

Conversely, assume v is a projection of u onto C. Fix $w \in C$; then, for all $0 \leq t \leq 1$, $tv + (1-t)w \in C$, whence

$$\begin{aligned}\|u - v\|^2 &\leq \|u - (tv + (1-t)w)\|^2 = \langle u - tv - (1-t)w, u - tv - (1-t)w\rangle \\ &= \langle u - v + (1-t)(v - w), u - v + (1-t)(v - w)\rangle \\ &= \|u - v\|^2 + (1-t)^2 \|v - w\|^2 - 2(1-t)\Re\langle u - v, w - v\rangle .\end{aligned}$$

As a consequence,

$$2\Re\langle u - v, w - v\rangle \leq (1-t)\|v - w\|^2 ,$$

from which (3.1.5) follows by taking the limit as $t \to 1$ in the last displayed inequality. □

We shall now focus on the particular case of a closed vector subspace V of H, where the previous characterization takes on a neater form. We say that two vectors $u, v \in H$ are **orthogonal** if $\langle u, v\rangle = 0$; if V is a subspace of H, we say that u is orthogonal to V if it is orthogonal to every $v \in V$.

COROLLARY 3.1.13. *Let $(H, \langle\cdot,\cdot\rangle)$ be a Hilbert space, V a closed subspace of H, u an element of H, $\pi_V(u)$ its projection onto V. Then $\pi_V(u)$ is the only element of V such that $u - \pi_V(u)$ is orthogonal to V.*

PROOF. By Proposition 3.1.12, and since V is a subspace, we know that $\pi_V(u)$ is the only element of V such that

$$\Re\langle u - \pi_V(u), w\rangle \leq 0$$

for all $w \in V$. Replacing w with $-w$ gives that

$$\Re\langle u - \pi_V(u), w\rangle = 0$$

for all $w \in V$, and replacing w with iw yields

$$\Im\langle u - \pi_V(u), w\rangle = 0$$

for all $w \in V$. It follows that $u - \pi_V(u)$ is orthogonal to V.

Conversely, if $u - \pi_V(u)$ is orthogonal to V, then (3.1.5) is obviously verified (with equality), and thus $\pi_V(u)$ is the unique projection of u onto V. □

Given a subspace V of H, the **orthogonal complement** of V in H is the closed subspace

$$V^\perp = \{u \in H : u \text{ is orthogonal to } V\} .$$

Linearity of the inner product ensures that $V^\perp$ is a subspace; if $(v_n)_{n\geq 0}$ is a sequence in $V^\perp$ converging to v and if $u \in V$, then

$$\langle u, v\rangle = \lim_{n\to\infty} \langle u, v_n\rangle = 0 ,$$

where the second equality is given by continuity of the inner product (see §3.1.2); the argument proves that $V^\perp$ is closed.

Corollary 3.1.13 shows that, when V is a closed subspace, we can write every vector $u \in H$ as

$$u = \pi_V(u) + u - \pi_V(u)$$

with $\pi_V(u) \in V$ and $u - \pi_V(u) \in V^\perp$; since V and $V^\perp$ have necessarily trivial interesection (for $v \in V \cap V^\perp$ gives $0 = \langle v, v\rangle = \|v\|^2$), we derive the following corollary.

COROLLARY 3.1.14. *Let $(H, \langle\cdot,\cdot\rangle)$ be a Hilbert space, V a closed subspace of H, $V^\perp$ its orthogonal complement. Then*

$$H = V \oplus V^\perp .$$

One says that H is the **orthogonal direct sum** of the subspaces V and $V^\perp$ (each of which, being closed, is a Hilbert space on its own for the induced inner product). Orthogonal decompositions are a crucial feature of the geometry of Hilbert spaces, and will enable us to characterize continuous linear functionals on Hilbert spaces as inner products against fixed vectors. This is the subject of the next subsection.

3.1.2. The Riesz-Fréchet representation theorem. We start by introducing the relevant notions. Given a normed vector space $(E, \|\cdot\|_E)$, the **topological dual** of E, denoted E^*, is the set

$$\{\varphi\colon E \to \mathbb{K} : \varphi \text{ is linear and continuous}\} .$$

It is thus the subset of the *algebraic dual* of E consisting of continuous linear functionals, where as usual the topology on E is the norm topology. Since the field operations on $\mathbb{K}$ are continuous, it is straightforward to realize that E^* is a $\mathbb{K}$-subspace of the algebraic dual, and thus a $\mathbb{K}$-vector space on its own. We shall endow it with an appropriate norm, by means of the following criterion for continuity of linear operators between normed vector spaces.

LEMMA 3.1.15. *Let $(E, \|\cdot\|_E), (F, \|\cdot\|_F)$ be normed vector spaces over $\mathbb{K}$, $T\colon E \to F$ a $\mathbb{K}$-linear map. The following are equivalent:*

(1) T is continuous;
(2) T is continuous at the origin 0_E;

(3) T is bounded, that is, there is $C > 0$ such that

$$\|T(u)\|_F \leq C \|u\|_E \tag{3.1.7}$$

for all $u \in E$.

PROOF. Continuity implies plainly continuity at every single point, in particular at 0_E. Suppose T is continuous at 0_E, and let $\delta > 0$ be such that the image under T of the open ball of radius δ, centered at 0_E, is contained in the open ball of radius 1 centered at 0_F: thus,

$$\|u\|_E < \delta \implies \|T(u)\|_F < 1$$

for all $u \in E$. Let now u be an arbitrary vector in E. If $u = 0$, then (3.1.7) is trivially satisfied; else, homogeneity of the norm gives that the vector

$$v = \delta \frac{1}{2\|u\|_E} u$$

satisfies $\|v\|_E < \delta$, whence

$$1 > \|T(v)\|_F = \left\| T\left(\delta \frac{1}{2\|u\|_E} u\right) \right\|_F = \delta \frac{1}{2\|u\|_E} \|T(u)\|_F \ ,$$

which gives (3.1.7) for $C = 2/\delta$.

Finally, if T is bounded in the sense described above, then it is Lipschitz: for all $u, v \in E$,

$$\|T(v) - T(u)\|_F = \|T(v-u)\|_F \leq C \|v - u\|_E \ .$$

It is a general fact that Lipschitz maps between metric spaces are continuous. □

On the topological dual E^* of a normed vector space $(E, \|\cdot\|_E)$, we define the **operator norm**

$$\|\varphi\|_{E^*} = \sup_{0 < \|u\|_E \leq 1} |\varphi(u)| = \inf\{C \geq 0 : |\varphi(u)| \leq C \|u\|_E \text{ for all } u \in E\} \ ,$$

which is a finite quantity by the previous lemma. Verification that $\|\cdot\|_{E^*}$ is a norm is routine, and left as an exercise.

We shall now establish a canonical identification of any Hilbert space H with its topological dual[3] H^*, the two considered as normed vector spaces.

First, if $(H, \langle\cdot,\cdot\rangle)$ is a Hilbert space over $\mathbb{K}$, we observe that there is a natural map

$$\Phi \colon H \to H^* \ ,$$

which is obtained assigning to every vector $u \in H$ the linear functional

$$H \to \mathbb{K} \ , \quad v \mapsto \langle v, u \rangle \ .$$

The Cauchy-Schwarz inequality gives

$$|\Phi(u)(v)| = |\langle v, u \rangle| \leq \|u\| \|v\| \tag{3.1.8}$$

for all $v \in H$, whence $\Phi(u)$ is continuous (thus indeed an element of the topological dual H^*) and

$$\|\Phi(u)\|_{H^*} \leq \|u\| \ .$$

Actually, applying (3.1.8) with $v = u$ and the definition of the operator norm, we immediately see that

$$\|\Phi(u)\|_{H^*} = \|u\| \ ,$$

that is, Φ is an isometry for the distances induced by the norms on the two spaces. This in particular implies that Φ is injective: if $u, v \in H$ satisfy $\Phi(u) = \Phi(v)$, then

$$0 = \|\Phi(v) - \Phi(u)\|_{H^*} = \|\Phi(v-u)\|_{H^*} = \|v - u\| \ ,$$

whence $u = v$.

[3] In general, an arbitrary Banach space admits only a canonical identification, that is, an isometric linear isomorphism, with a closed subspace of its topological *bidual*.

Notice that the map Φ is only *semilinear*: it preserves the sum but acts via conjugation on scalar multiplication (it is linear for $\mathbb{K} = \mathbb{R}$).

THEOREM 3.1.16 (The Riesz-Frechét representation theorem). *Let $(H, \langle \cdot, \cdot \rangle)$ be a Hilbert space over $\mathbb{K}$. For any $\varphi \in H^*$, there is a vector $u \in H$ such that*

$$\varphi(v) = \langle v, u \rangle$$

for all $v \in H$.

Before proving the theorem, we directly infer from it the aforementioned identification of H and H^*.

COROLLARY 3.1.17. *The assignment $\Phi \colon H \to H^*$ defined above establishes a $\mathbb{K}$-semilinar isometric isomorphism between H and H^*.*

PROOF OF THEOREM 3.1.16. If $\varphi = 0$, then we can take $u = 0$. Assume thus $\varphi \neq 0$, so that its kernel $\ker \varphi$ is a closed (by continuity of φ) proper subspace of H. In view of Corollary 3.1.14, we have the direct sum decomposition

$$H = \ker \varphi \oplus \ker \varphi^\perp ,$$

where $\ker \varphi^\perp$ is one-dimensional since φ induces a linear isomorphism between $H/\ker \varphi$ and $\mathbb{K}$, and the restriction to $\ker \varphi^\perp$ of the canonical projection $H/\ker \varphi$ induces a linear isomorphism between $\ker \varphi^\perp$ and $H/\ker \varphi$.

Choose $u' \in \ker \varphi^\perp$ with $\varphi(u) = 1$ (it can always be guaranteed by conveniently renormalizing any non-zero vector in the orthogonal complement of the kernel). If v is any vector in H, then $v = \pi_{\ker \varphi}(u) + \pi_{\ker \varphi^\perp}(u)$ with

$$\pi_{\ker \varphi^\perp}(u) = \frac{\langle v, u' \rangle}{\|u'\|^2} u' ,$$

as follows directly from imposing the orthogonality condition $\langle v - \alpha u, u \rangle = 0$ to find α. Thus

$$\varphi(v) = \varphi(\pi_{\ker \varphi}(v)) + \frac{\langle v, u' \rangle}{\|u'\|^2} \varphi(u') = \frac{1}{\|u'\|^2} \langle v, u' \rangle = \langle v, \frac{1}{\|u'\|^2} u' \rangle ,$$

which delivers the conclusion with $u = \frac{1}{\|u'\|^2} u$. □

REMARK 3.1.18. When H is finite-dimensional, Theorem 3.1.16 (and its corollary) is a well known fact from linear algebra: in this case surjectivity of the map Φ follows from injectivity and the fact that H^*, which *coincides with the linear algebraic dual in finite dimension*, has the same dimension of H. As we have seen, the case of an arbitrary Hilbert space requires considerably more work.

3.2. The inequalities of Hölder and Minkowski

Having dealt with the necessary functional-analytic preliminaries we now turn L^p spaces. Hereinafter, we will work with a fixed underlying measure space $(X, \mathfrak{M}, \mu)$. Fix a real number $p > 0$. If $f \colon X \to \mathbb{C}$ is a measurable function, we define

$$\|f\|_p = \left(\int_X |f|^p \, \mathrm{d}\mu \right)^{1/p}$$

with the understanding that $\infty^{1/p} = \infty$. We define

$$\mathcal{L}^p(X, \mathfrak{M}, \mu) = \{ \, f \colon X \to \mathbb{C} \text{ measurable} : \|f\|_p \text{ is finite} \} ,$$

and abbreviate it to $\mathcal{L}^p$ whenever $(X, \mathfrak{M}, \mu)$ is understood from the context.

The set $\mathcal{L}^p$ is a vector subspace of the $\mathbb{C}$-vector space of measurable functions $X \to \mathbb{C}$, as follows from the fact that

$$|\alpha f + \beta g|^p \leq 2^p \sup\{|\alpha|^p |f|^p, |\beta|^p |g|^p\}$$

for all $f, g \in \mathcal{L}^p$ and $\alpha, \beta \in \mathbb{C}$. The adopted notation suggests that $\|\cdot\|_p$ is a norm on $\mathcal{L}^p(X, \mathfrak{M}, \mu)$: now it holds true, and is easily verified by linearity of the integral, that $\|\cdot\|_p$ is homogeneous, namely

$$\|\alpha f\|_p = |\alpha| \, \|f\|_p$$

for all $f \in \mathcal{L}^p$ and all $\alpha \in \mathbb{C}$. However, $\|\cdot\|_p$ fails to be a norm on $\mathcal{L}^p$ for two reasons:

- if $\|f\|_p = 0$ for some $f \in \mathcal{L}^p$, then this only implies that $f = 0$ μ-almost everywhere;
- the triangle inequality

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p$$

can be shown to fail for $0 < p < 1$.

The first deficiency is easily remedied, by considering the quotient of $\mathcal{L}^p$ by the subspace of functions which vanish μ-almost everywhere on X, in the spirit of Lemma 3.1.2. The second is a more serious issue, which is why we shall only consider, henceforth, the case $p \geq 1$.

The fact that $\|\cdot\|_p$ satisfies the triangle inequality is a consequence of the pillar of the whole theory of L^p spaces, namely *Hölder's inequality*, which we shall establish momentarily. Before doing that, we complete the picture and define a further space $\mathcal{L}^p$ for $p = \infty$. If $f \colon X \to \mathbb{C}$ is measurable, we define

$$\|f\|_\infty = \inf\{M \in \mathbb{R}_{\geq 0} : |f(x)| \leq M \text{ for } \mu\text{-almost every } x \in X\} , \tag{3.2.1}$$

with the usual understanding that $\inf \emptyset = +\infty$, and set

$$\mathcal{L}^\infty(X, \mathfrak{M}, \mu) = \{f \colon X \to \mathbb{C} \text{ measurable} : \|f\|_\infty \text{ is finite}\} .$$

Notice that, for $f \in \mathcal{L}^\infty$ (omitting the measure space for notation when no confusion arises, as for all other values of p), the infimum in (3.2.1) is attained, that is,

$$|f(x)| \leq \|f\|_\infty$$

for μ-almost every $x \in X$: to see this, observe that the set

$$\{x \in X : |f(x)| \geq \|f\|_\infty + 1/n\}$$

is μ-null for all integers $n \geq 1$, whence so is the union of those over all n, which is precisely

$$\{x \in X : |f(x)| > \|f\|_\infty \} .$$

An immediate pointwise application of the triangle inequality gives that $\mathcal{L}^\infty$ is a vector subspace of the $\mathbb{C}$-vector space of measurable functions $X \to \mathbb{C}$. Homogeneity and the triangle inequality for $\|\cdot\|_\infty$ are this time both straightforward. Quotienting out by the subspace of μ-almost everywhere vanishing functions would then yield a normed vector space by virtue of Lemma 3.1.2, but we postpone until after we establish the triangle inequality for $\|\cdot\|_p$ for all other values of p.

We introduce a fundamental notion: two numbers $1 \leq p, q \leq \infty$ are said to be **conjugate exponents** (also, **Hölder-conjugate exponents**) if

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

or equivalently $p + q = pq$, with the understanding that $\frac{1}{\infty} = 0$. The two most important cases are $p = 1$, $q = \infty$, and $p = q = 2$.

THEOREM 3.2.1 (Hölder's inequality). *Let $(X, \mathfrak{M}, \mu)$ be a measure space, $1 \leq p, q \leq \infty$ conjugate exponents. If $f, g \colon X \to \mathbb{C}$ are measurable functions, then*

$$\|fg\|_1 \leq \|f\|_p \, \|g\|_q \, . \tag{3.2.2}$$

In particular, if $f \in \mathcal{L}^p$ and $g \in \mathcal{L}^q$, then $fg \in \mathcal{L}^1$. In this case, equality holds

(1) *when* $1 < p < \infty$, *if and only if there are* $\alpha, \beta \geq 0$, *not both vanishing, such that*

$$\alpha|f|^p = \beta|g|^q$$

μ-almost everywhere,

(2) *and in the case* $p = 1$, *if and only if* $|g|$ *is constant μ-almost everywhere on the set* $\{x \in X : f(x) \neq 0\}$.

PROOF. We start with the easier case $p = 1$, $q = \infty$. If $\|g\|_\infty = \infty$, there is nothing to prove, unless $\|f\|_p = 0$, in which case $f = 0$ μ-almost everywhere, and then the same holds for fg and (3.2.2) is established. Suppose thus $\|g\|_\infty$ is finite; then

$$|fg| \leq |f| \, \|g\|_\infty$$

μ-almost everywhere, both sides being potentially infinite. Monotonicity of the integral then yields

$$\|fg\|_1 = \int_X |fg| \, \mathrm{d}\mu \leq \int_X |f| \, \|g\|_\infty \, \mathrm{d}\mu = \|g\|_\infty \int_X |f| \, \mathrm{d}\mu = \|f\|_1 \, \|g\|_\infty \, .$$

This also shows that $fg \in \mathcal{L}^1$ provided that $f \in \mathcal{L}^1$. In this case, equality holds in (3.2.2) if and only if $|g| = \|g\|_\infty$ μ-almost everywhere on $\{x \in X : f(x) \neq 0\}$, that is, if and only if $|g|$ is constant μ-almost everywhere on such a set.

Consider now the case of finite p and q. We may assume both $\|f\|_p$ and $\|g\|_q$ to be finite; else, either both are infinite, in which case there is nothing to prove, or only one of them is infinite, in which case the right-hand side of (3.2.2) is infinite (leaving us once more with nothing to prove) unless $\|f\|_p = 0$, in which case $f = 0$ μ-almost everywhere and thus the left-hand side in (3.2.2) vanishes as well. Notice further that Hölder's inequality is bi-homogeneous, that is, once it holds for a fixed pair (f, g), it holds automatically for all pairs $(\alpha f, \beta g)$ with $\alpha, \beta \in \mathbb{C}$. Upon renormalization, we may thus assume that $\|f\|_p = 1 = \|g\|_q$.

We rely on the following fundamental inequality for pairs of positive real numbers.

LEMMA 3.2.2. *Let* $1 < p, q < \infty$ *be Hölder-conjugate exponents,* $a, b \in \mathbb{R}_{\geq 0}$. *Then*

$$ab \leq \frac{1}{p} \, a^p + \frac{1}{q} \, b^q \, ,$$

with equality if and only if $a^p = b^q$.

Notice that for $p = q = 2$ the assertion is nothing but the fundamental inequality between the geometric and the arithmetic mean of two positive real numbers.

We start by showing the lemma, observing that if $ab = 0$ there is nothing to prove. Thus we can assume $a > 0, b > 0$. Strict convexity of the exponential function $t \mapsto e^t$ then yields

$$ab = e^{\log ab} = e^{\frac{1}{p} \log a^p + \frac{1}{q} \log b^q} \leq \frac{1}{p} e^{\log a^p} + \frac{1}{q} e^{\log b^q} = \frac{1}{p} \, a^p + \frac{1}{q} \, b^q \, ,$$

with equality if and only if $\log a^p = \log b^q$, that is, if and only if $p \log a = q \log b$.

Let us now go back to the proof of Hölder's inequality. We apply the lemma pointwise, and obtain

$$|f(x)g(x)| \leq \frac{1}{p} |f(x)|^p + \frac{1}{q} |g(x)|^q$$

for all $x \in X$. Integrating on both sides, and applying monotonicity and linearity of the integral, we get

$$\int_X |fg| \, \mathrm{d}\mu \leq \frac{1}{p} \int_X |f|^p \, \mathrm{d}\mu + \frac{1}{q} \int_X |g|^q \, \mathrm{d}\mu = \frac{1}{p} + \frac{1}{q} = 1 \, , \tag{3.2.3}$$

where the second-to-last equality is given by the unrestrictive assumption $\|f\|_p = \|g\|_q = 1$. We have thus shown (3.2.2); furthermore, equality holds if and only if there is equality in the first step of (3.2.3), which occurs if and only if

$$|f(x)g(x)| = \frac{1}{p} |f(x)|^p + \frac{1}{q} |g(x)|^q$$

for μ-almost every $x \in X$. By the lemma, this happens in turn if and only if

$$|f(x)|^p = |g(x)|^q$$

for μ-almost every $x \in X$. Taking into account our initial our initial rescaling of f and g, we have thus shown the equality case in the assertion. ☐

THEOREM 3.2.3 (Minkowski's inequality). *Let $(X, \mathfrak{M}, \mu)$ be a measure space, $1 \le p \le \infty$. If $f, g\colon X \to \mathbb{C}$ are measurable functions, then*

$$\|f + g\|_p \le \|f\|_p + \|g\|_p \ . \tag{3.2.4}$$

PROOF. We have already discussed this for $p = \infty$, hence we focus on the case of p finite. We write

$$|f + g|^p \le (|f| + |g|)^p = |f|(|f| + |g|)^{p-1} + |g|(|f| + |g|)^{p-1} \ ;$$

monotonicity of the integral and Hölder's inequality, for q the Hölder-conjugate of p, yield

$$\int_X |f + g|^p \, \mathrm{d}\mu \le \left(\int_X (|f| + |g|)^{q(p-1)} \, \mathrm{d}\mu \right)^{1/q} (\|f\|_p + \|g\|_p) \ .$$

Since $q(p - 1) = p$ and $1/q = 1 - 1/p$, the latter is precisely tantamount to (3.2.4). ☐

The exercises ask to figure out the equality case.

Minowski's inequality shows that $\|\cdot\|_p$ satisfies the triangle inequality, and is thus a seminorm on $\mathcal{L}^p(X, \mathfrak{M}, \mu)$, Applying Lemma 3.1.2, we obtain a normed vector space by considering the quotient of $\mathcal{L}^p(X, \mathfrak{M}, \mu)$ by the vector subspace

$$\mathcal{N} = \{f \in \mathcal{L}^p : \|f\|_p = 0\} = \{f\colon X \to \mathbb{C} \text{ measurable} : f = 0 \ \mu\text{-almost everywhere}\}$$

DEFINITION 3.2.4 (L^p spaces). Let $(X, \mathfrak{M}, \mu)$ be a measure space, $1 \le p \le \infty$. The L^p **space** on $(X, \mathfrak{M}, \mu)$ is defined as the normed vector space $(L^p(X, \mathfrak{M}, \mu), \|\cdot\|_p)$ where

$$L^p(X, \mathfrak{M}, \mu) = \mathcal{L}^p(X, \mathfrak{M}, \mu)/\mathcal{N}$$

and

$$\|f + \mathcal{N}\|_p = \|f\|_p$$

for all $f \in \mathcal{L}^p(X, \mathfrak{M}, \mu)$.

If the underlying measure space, or some element of it, is clear from context, we shall typically write L^p, $L^p(X)$, $L^p(\mu)$ or $L^p(X, \mu)$ in place of $L^p(X, \mathfrak{M}, \mu)$. Adhering to the tradition, we keep denoting equivalence classes of functions up to equality μ-almost everywhere in functional notation, and thus routinely write expressions such as $f \in L^p$ instead of the formally more appropriate $[f] \in L^p$. For a function $f \in L^p$, the quantity $\|f\|_p$ is known as the L^p**-norm** of f; if $p = \infty$, it is also referred to as the **essential supremum** of f, and f is said to be **essentially bounded**.

When μ is the counting measure c on a discrete Borel space $(A, \mathcal{P}(A))$, then the L^p-space on $(A, \mathcal{P}(A), \mathrm{c})$ is customarily denoted $\ell^p(A)$; thus,

$$\ell^p(X) = \left\{ (z_a)_{a \in A} \in \mathbb{C}^A : \sum_{a \in A} |z_a|^p \text{ is finite} \right\}$$

for p finite, and

$$\ell^\infty(A) = \left\{ (z_a)_{a \in A} : \sup_{a \in A} |z_a| \text{ is finite} \right\}$$

for $p = \infty$.

Consider the case $p = 2$; then there is a natural inner product on L^2 which induces the L^2-norm: it is given by

$$\langle f, g \rangle_2 = \int_X f \overline{g} \, \mathrm{d}\mu \ ,$$

where finiteness of the integral is given by Hölder's inequality. All axioms of an inner product are easilly verified.

PROPOSITION 3.2.5. *Let $(X, \mathfrak{M}, \mu)$ be a measure space. For every $1 \leq p \leq \infty$, $(L^p(X, \mathfrak{M}, \mu), \|\cdot\|_p)$ is a Banach space; when $p = 2$, $(L^2(X, \mathfrak{M}, \mu), \langle\cdot,\cdot\rangle_2)$ is a Hilbert space.*

PROOF. It suffices to prove the first assertion; the second one follows automatically from the previous considerations on L^2 and from the definition of Hilbert space. We separate two cases.

Suppose first $p = \infty$, and let $(f_n)_{n\geq 0}$ be a Cauchy sequence in L^∞. As countable unions of μ-null sets are μ-null, we can find a μ-null set $N \subset X$ such that the restrictions $(f_n|_{X\setminus N})_{n\geq 0}$ form a Cauchy sequence in the uniform norm: for every $\varepsilon > 0$ there in an integer $N \geq 1$ such that, for all $m, n \geq N$,

$$\sup_{x\in X\setminus N} |f_m(x) - f_n(x)| \leq \varepsilon \ .$$

The $\mathbb{C}$-vector space of bounded functions $X \setminus N \to \mathbb{C}$, equipped with the uniform norm, is a Banach space; thus there is a bounded function $f\colon X \setminus N \to \mathbb{C}$ which is the uniform limit of $(f_n|_{X\setminus N})_{n\geq 0}$. In particular, f is measurable. Extend it arbitrarily, in a measurable way, to N. Then it is clear that $(f_n)_{n\geq 0}$ converges to f in the L^∞-norm.

Let now p be finite, and let $(f_n)_{n\geq 0}$ be a Cauchy sequence in L^p. It suffices to show that there is a subsequence $(f_{n_k})_{k\geq 0}$ converging in L^p to some $f \in L^p$. Markov's inequality yields, for all $\varepsilon > 0$ and all integers $m, n \geq 0$,

$$\mu(\{x \in X : |f_m - f_n| \geq \varepsilon\}) \leq \frac{\int_X |f_m - f_n|^p \, d\mu}{\varepsilon^p} \ ,$$

from which it readily follows that $(f_n)_{n\geq 0}$ is Cauchy in measure, and thus subsequence $(f_{n_k})_{k\geq 0}$ of it converges μ-almost everywhere to a measurable function $f\colon X \to \mathbb{C}$. Now the Cauchy property for $(f_n)_{n\geq 0}$ with the L^p-norm implies, in particular, that $\sup_{n\geq 0} \|f_n\|_p$ is finite; by Fatou's lemma,

$$\int_X |f|^p \, d\mu \leq \liminf_{k\to\infty} \int_X |f_{n_k}|^p \, d\mu \leq \sup_{n\geq 0} \int_X |f_n|^p \, d\mu \ ,$$

which shows that $f \in L^p$. By judiciusly selecting the subsequence $(f_{n_k})_{k\geq 0}$, as in the proof of Proposition **??** (to which we refer for the details), for instance to satisfy

$$\int_X |f_{n_{k+1}} - f_{n_k}|^p \, d\mu \leq 2^{-k}$$

for all $k \geq 0$, we can make sure that there is $h \in L^1$ with $|f - f_{n_k}|^p \leq h$ for all $k \geq 0$. The Dominated Convergence Theorem, applied to the sequence $g_k = |f - f_{n_k}|^p$, delivers then convergence of $(f_{n_k})_{k\geq 0}$ to f in the L^p-norm. □

REMARK 3.2.6. Observe that, for $p = \infty$, the argument in the proof illustrates the validity of the following characterization: for a sequence $(f_n)_{n\geq 0}$ in L^∞ and a function $f \in L^\infty$, convergence of $(f_n)_{n\geq 0}$ towards f in the L^∞-norm occurs if and only if there is a μ-null set $N \subset X$ such that the sequence $(f_n)_{n\geq 0}$ converges to f uniformly on $X \setminus N$.

We state separately the partial result, about almost-everywhere convergence along a subsequence, appearing in the proof of the foregoing proposition.

PROPOSITION 3.2.7. *Let $1 \leq p \leq \infty$. If $(f_n)_{n\geq 0}$ is a Cauchy sequence in L^p, then there is a subsequence $(f_{n_k})_{k\geq 0}$ converging μ-almost everywhere to a measurable function f.*

REMARK 3.2.8. For $p = 1$, it clearly holds that a sequence $(f_n)_{n\geq 0}$ in L^1 converges to an element $f \in L^1$ in the L^1-norm if and only if it converges to it in L^1 in the way formulated in Definition **??**.

We now generalize Proposition 2.4.9 to arbitrary exponents p.

PROPOSITION 3.2.9. *For any $1 \le p \le \infty$, the set of simple functions $X \to \mathbb{C}$ with finite L^p-norm is dense in L^p.*

Notice that every simple function is, directly from the definition, uniformly bounded, and thus in L^∞. It may not lie, however, in any other L^p, such as the constant function 1 on $(\mathbb{R}, \mathfrak{B}_\mathbb{R}, \mathscr{L}^1)$.

PROOF. When p is finite, the argument for Proposition 2.4.9 applies almost *verbatim*, with the appropriate modifications. We omit the details.

When $p = \infty$, the claim follows at once from Exercise 2.3.4, using the characterization of L^∞-convergence in Remark 3.2.6. □

3.3. Relations between L^p spaces and some useful inequalities

For an arbitrary measure space $(X, \mathfrak{M}, \mu)$, there is no general containment relation between L^p spaces of different exponents.

EXAMPLE 3.3.1. Let $X = (0, +\infty)$, with the induced Borel σ-algebra, and μ the restriction of $\mathscr{L}^1$ to X. Suppose $1 \le p < p' < \infty$ (they are not meant to be Hölder-conjugate). Take $\alpha \in (1/p', 1/p)$; then the function

$$f(x) = x^{-\alpha} \chi_{(0,1)}(x)$$

is in L^p but not in $L^{p'}$, whereas the function

$$f(x) = x^{-\alpha} \chi_{(1,+\infty)}(x)$$

is in $L^{p'}$ but not in L^p. Similar examples can be found with $p' = \infty$.

A double issue emerged in the previous counterexample, preventing inclusion of different L^p spaces in either direction; on the one hand, the function x^α may grow too fast to infinity in a neighborhood of 0, on the other it may decrease too slowly to zero, both conditions possibly preventing the sought after integrability.

It stands to reason to expect that, by ruling out the former two occurrences via the imposition of additional constraints on the underlying measure space, some general containment relations between L^p spaces can actually be established. This is indeed the case, but before stating conditional (on the measure space) results of this sort, we provide a general inclusion involving intersections of L^p spaces.

PROPOSITION 3.3.2. *Let $1 \le p < q < r \le \infty$. Then*

$$L^p \cap L^r \subset L^q$$

and, for every $f \in L^p \cap L^r$,

$$\|f\|_q \le \|f\|_p^\lambda \, \|f\|_r^{1-\lambda}$$

where $\lambda \in (0,1)$ satisfies

$$\frac{1}{q} = \lambda \, \frac{1}{p} + (1 - \lambda) \, \frac{1}{r} \, .$$

PROOF. It suffices to prove the inequality about norms, as the first assertion then follows automatically. First, assume $r = \infty$. Then

$$\int_X |f|^q \, \mathrm{d}\mu = \int_X |f|^p |f|^{q-p} \, \mathrm{d}\mu \le \|f\|_\infty^{q-p} \int_X |f|^p \, \mathrm{d}\mu \, ,$$

whence

$$\|f\|_q \le \|f\|_\infty^{1-p/q} \, \|f\|_p^{p/q} \, ,$$

as desired.

Let now r be finite. Let $\lambda' \in (0,1)$ be such that $q = \lambda' p + (1-\lambda')r$. Applying Hölder's inequality to the pair of conjugate exponents $1/\lambda'$, $1/(1-\lambda')$ we obtain

$$\int_X |f|^q \, \mathrm{d}\mu = \int_X |f|^{\lambda' p} |f|^{(1-\lambda')r} \, \mathrm{d}\mu \le \left(\int_X |f|^p\right)^{\lambda'} \left(\int_X |f|^r\right)^{1-\lambda'} ,$$

thus

$$\|f\|_q \le \|f\|_p^{\lambda' p} \|f\|_r^{(1-\lambda')r} ,$$

and to conclude it suffices to notice that $\lambda = \lambda' p$ and $1-\lambda = (1-\lambda')r$. □

The following result presents a containment relation whenever the possibility of "non-integrability at infinity" is excluded by the nature of the measure space.

Proposition 3.3.3. *Suppose $(X, \mathfrak{M}, \mu)$ is a finite measure space, and let $1 \le p \le q \le \infty$. Then*

$$L^q \subset L^p$$

and, for every $f \in L^q$,

$$\|f\|_p \le \mu(X)^{1/p-1/q} \|f\|_q .$$

Proof. It is unrestrictive to assume $p < q$, else there is nothing to prove. If $q = \infty$, then

$$\int_X |f|^p \, \mathrm{d}\mu \le \int_X \|f\|_\infty^p \, \mathrm{d}\mu = \mu(X) \|f\|_\infty^p ,$$

which gives

$$\|f\|_p \le \mu(X)^{1/p} \|f\|_\infty .$$

Suppose now q is finite. We apply Hölder's inequality with to $|f|^p$ and the constant function $g = 1$, with exponents q/p and $(q/p)'$: we obtain

$$\int_X |f|^p \, \mathrm{d}\mu \le \left(\int_X |f|^q \, \mathrm{d}\mu\right)^{p/q} \left(\int_X 1 \, \mathrm{d}\mu\right)^{1/(q/p)'} = \mu(X)^{\frac{q-p}{p}} \|f\|_q^p ,$$

from which we deduce

$$\|f\|_p \le \mu(X)^{1/p-1/q} \|f\|_q .$$

□

At the other extreme, we have the case in which functions cannot "blow-up" in the proximity of a point.

Proposition 3.3.4. *Let A be a set, $1 \le p \le q \le \infty$. Then*

$$\ell^p(A) \subset \ell^q(A)$$

and, for every $f \in \ell^p(A)$,

$$\|f\|_q \le \|f\|_p .$$

Proof. Without loss of generality we take $p < q$. If $q = \infty$, then summability of the family $(|f(a)|^p)_{a \in A}$ implies that $|f(a)| \le 1$ for all but finitely many $a \in A$, from which we obviously infer that $f \in \ell^\infty(A)$. Also,

$$\|f\|_\infty = \sup_{a \in A} |f(a)| = \left(\sup_{a \in A} |f(a)|^p\right)^{1/p} \le \left(\sum_{a \in A} |f(a)|^p\right)^{1/p} = \|f\|_p , \tag{3.3.1}$$

where notice that all suprema are attained in the previous chain of inequalities.

Suppose now q is finite, and let $f \in \ell^p(A)$. Since we already know that $f \in \ell^\infty(A)$, we can invoke Proposition 3.3.2 and thereby deduce that $f \in \ell^q(A)$; furthermore, the same proposition gives that, for a certain explicit $\lambda \in [0,1]$,

$$\|f\|_q \le \|f\|_p^\lambda \|f\|_\infty^{1-\lambda} \le \|f\|_p^\lambda \|f\|_p^{1-\lambda} = \|f\|_p ,$$

using (3.3.1) in the second inequality. □

We conclude this section with a couple of measure-theoretic inequalities of common use in analysis (and beyond).

THEOREM 3.3.5 (Chebychev's inequality). *Let $1 \leq p < \infty$, $f \in L^p$. Then, for all real $\alpha > 0$,*

$$\mu(\{x \in X : |f(x)| \geq \alpha\}) \leq \frac{\|f\|_p^p}{\alpha^p} .$$

Notice that, the higher the exponent p, the finer the bound becomes; in other words, the higher the exponent p, the less likely it is that $|f|$ takes on large values.

PROOF. The inequality follows readily by applying Markov's inequality to the positive, μ-integrable function $|f|^p$, observing that

$$\{x \in X : |f(x)| \geq \alpha\} = \{x \in X : |f(x)|^p \geq \alpha^p\} .$$

□

REMARK 3.3.6. Textbooks on probability theory typically present the L^2-version of Chebychev's inequality, in the following form: if $(X, \mathfrak{M}\, \mu)$ is a probability space and $f\colon X \to \mathbb{C}$ is a random variable in L^2[4], then

$$\mu(\{x \in X : |f(x) - \mathbf{E}[f]| \geq \alpha\}) \leq \frac{\mathbf{Var}[f]}{\alpha^2} ,$$

where

$$\mathbf{E}[f] = \int_X f(x)\, \mathrm{d}\mu$$

is the *expected value* of f and

$$\mathbf{Var}(f) = \int_X (f(x) - \mathbf{E}[f])^2\, \mathrm{d}\mu$$

is the *variance* of f.

THEOREM 3.3.7 (Jensen's inequality). *Let $(X, \mathfrak{M}, \mu)$ be a probability space, $-\infty \leq a < b \leq +\infty$, $f\colon X \to (a, b)$ a μ-integrable function, $\varphi\colon (a, b) \to \mathbb{R}$ a convex function such that $\varphi \circ f$ is μ-integrable. Then*

$$\varphi\Big(\int_X f\, \mathrm{d}\mu\Big) \leq \int_X \varphi \circ f\, \mathrm{d}\mu . \tag{3.3.2}$$

Thus, for instance, if $f\colon X \to (0, \infty)$ is in L^q for some $1 \leq q < \infty$, then ($f \in L^1$ and)

$$\Big(\int_X f\, \mathrm{d}\mu\Big)^q \leq \int_X f^q\, \mathrm{d}\mu .$$

PROOF. To begin with, observe that the left-hand side of (3.3.2) is well defined, that is,

$$\int_X f\, \mathrm{d}\mu \in (a, b) ;$$

this follows immediately by integrating the pointwise inequality $a < f(x) < b$, valid for all $x \in X$, and using the fact that $\mu(X) = 1$.

Suppose first φ is an affine function, that is, there are $\alpha, \beta \in \mathbb{R}$ such that

$$\varphi(x) = \alpha x + \beta$$

for all $x \in \mathbb{R}$. Then, by linearity of the integral and the assumption that $\mu(X) = 1$, we have

$$\varphi\Big(\int_X f\, \mathrm{d}\mu\Big) = a \int_X f\, \mathrm{d}\mu + b = a \int_X f\, \mathrm{d}\mu + \int_X b\, \mathrm{d}\mu = \int_X af + b\, \mathrm{d}\mu = \int_X \varphi \circ f\, \mathrm{d}\mu ,$$

thus (3.3.2) holds with equality.

[4]A terminology employed frequently in probability theory is that f has *finite second moment*.

Let now $\varphi\colon (a,b) \to \mathbb{R}$ be an arbitrary convex function, then

$$\varphi(x) = \sup_{\psi \le \varphi,\ \psi \text{ affine}} \psi(x) \tag{3.3.3}$$

for all $x \in (a,b)$; for every affine $\psi\colon (a,b) \to \mathbb{R}$ with $\psi \le \varphi$, we have, by monotonicity of the integral,

$$\psi\Big(\int_X f \,\mathrm{d}\mu\Big) = \int_X \psi \circ f \,\mathrm{d}\mu \le \int_X \varphi \circ f \,\mathrm{d}\mu\ .$$

Taking the supremum over all ψ on the left-hand side of the last displayed inequality, and using (3.3.3), we achieve the desired inequality. □

3.4. Convolutions and regularization

The need often emerges to approximate integrable functions with more regular representatives; for instance, on Euclidean spaces, one would like to approximate L^p functions (with respect to the Lebesgue measure) with differentiable, or even smooth functions. This process of approximation, typically referred to as *regularization*, is the subject of the present section, throughout which we confine ourselves to the measure spaces $(\mathbb{R}^n, \mathfrak{B}_{\mathbb{R}^n}, \mathscr{L}^n)$ without further comment.

Regularization of integrable functions on $\mathbb{R}^n$ is achieved by means of a fundamental operation which combines algebra and analysis: the *convolution* of measurable functions. In order to motivate the definition, we present a probabilistic framework in which the notion appears naturally.

Suppose μ and ν are Borel probability measures on $\mathbb{R}^n$ with densities ρ_μ and ρ_ν, respectively, with respect to $\mathscr{L}^n$. Suppose you sample first a point $x \in \mathbb{R}^n$ according to μ, and then a second point $y \in \mathbb{R}^n$ according to ν, *independently* of the sampling of x; what is the resulting distribution of the sum $x+y$ of the two sampled points? Formally, the two sampling processes are encoded by two independent random variables X, Y, defined on a common underlying probability space $(\Omega, \mathcal{F}, \mathbf{P})$, with laws $\mathbf{P}_X = \mu$ and $\mathbf{P}_Y = \nu$, respectively[5]. Independence of X and Y is tantamount to the fact that the law $\mathbf{P}_{(X,Y)}$ of the pair (X,Y), which is an $\mathbb{R}^n \times \mathbb{R}^n$-valued random variable defined on $(\Omega, \mathcal{F}, \mathbf{P})$, is the product $\mu \times \nu$ of the laws of X and Y (referred to as the *marginals*, in probability theory). The distribution of the sum $X+Y$ is then nothing but the pushforward of the law $\mathbf{P}_{(X,Y)}$ under the sum map $\mathbb{R}^n \times \mathbb{R}^n (x,y) \mapsto x+y \in \mathbb{R}^n$. We thus have, for every Borel set $E \subset \mathbb{R}^n$,

$$\begin{aligned}
\mathbf{P}_{X+Y}(E) &= \int_{\mathbb{R}^n} \chi_E \,\mathrm{d}\mathbf{P}_{X+Y} = \int_{\mathbb{R}^n \times \mathbb{R}^n} \chi_E(x+y) \,\mathrm{d}\mu \times \nu(x,y) \\
&= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \chi_E(x+y) \rho_\mu(x) \,\mathrm{d}x\ \rho_\mu(y) \,\mathrm{d}y \\
&= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \chi_E(z) \rho_\mu(z-y) \,\mathrm{d}z\ \rho_\mu(y) \,\mathrm{d}y \\
&= \int_{\mathbb{R}^n} \chi_E(z) \int_{\mathbb{R}^n} \rho_\mu(z-y) \rho_\nu(y) \,\mathrm{d}y \,\mathrm{d}z\ ,
\end{aligned}$$

where we have used Fubini-Tonelli's theorem twice, and applied the change of variable $z = x+y$ in the third equality. The previous chain of equalities shows that the distribution $\mathbf{P}_{X+Y}$ of the sum $X+Y$ has density

$$\rho(z) = \int_{\mathbb{R}^n} \rho_\mu(z-y) \rho_\nu(y) \,\mathrm{d}y$$

with respect to $\mathscr{L}^n$.

[5]The law $\mathbf{P}$ of a random variable X defined on $(\Omega, \mathcal{F}, \mathbf{P})$ is the pushforward of $\mathbf{P}$ under X.

DEFINITION 3.4.1 (Convolution of functions on $\mathbb{R}^n$). Let $f, g\colon \mathbb{R}^n \to \mathbb{C}$ be Borel-measurable functions. The **convolution** of f and g is the function $f * g$ defined as

$$f * g(x) = \int_{\mathbb{R}^n} f(x-y)g(y)\,\mathrm{d}y$$

for all $x \in \mathbb{R}^n$ for which the integral is defined.

We shall now examine conditions on f and g ensuring that the convolution $f * g$ is defined for (Lebesgue-almost) all $x \in \mathbb{R}^n$. To begin with, observe that, at the very least, the integrand $y \mapsto f(x-y)g(y)$ appearing in the definition of convolution is a Borel-measurable function, since the difference map $\mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}^n$, $(x,y) \mapsto x-y$ is $(\mathfrak{B}_{\mathbb{R}^n} \otimes \mathfrak{B}_{\mathbb{R}^n}, \mathfrak{B}_{\mathbb{R}^n})$-measurable.

If X is a topological space, the **support** of a function $f\colon X \to \mathbb{C}$ is the set

$$\operatorname{supp} f = \overline{\{x \in X : f(x) \neq 0\}}\,.$$

If A and B are subsets of $\mathbb{R}^n$, we let

$$A + B = \{a + b : a \in A,\ b \in B\}$$

be their *sumset.*

PROPOSITION 3.4.2. *Let $f, g, h\colon \mathbb{R}^n \to \mathbb{C}$ be Borel-measurable functions. Then the following equalities holds whenever both sides are defined:*

(1) $f * g = g * f$;
(2) $f * (g * h) = (f * g) * h$.

Furthermore, if $f * g$ is defined on $\mathbb{R}^n$, then

$$\operatorname{supp} f * g \subset \overline{\operatorname{supp} f + \operatorname{supp} g}\,.$$

PROOF. We first show commutativity. Consider the $\mathscr{C}^1$-diffeomorphism $T\colon \mathbb{R}^n \to \mathbb{R}^n$, $y \mapsto x-y$; then $|\operatorname{Jac}_T(y)| = 1$ for all $y \in \mathbb{R}^n$, and thus the change of variable formula yields[6]

$$g * f(x) = \int_{\mathbb{R}^n} g(x-y)f(y)\,\mathrm{d}y = \int_{\mathbb{R}^n} g(x-T(u))f(T(u))\,\mathrm{d}u = \int_{\mathbb{R}^n} g(u)f(x-u)\,\mathrm{d}u = f * g(x)\,.$$

We now turn to associativity. Using Fubini-Tonelli's Theorem, and a change of variable analogous to the one above, we get

$$\begin{aligned} f * (g * h)(x) &= \int_{\mathbb{R}^n} f(x-y)g * h(y)\,\mathrm{d}y = \int_{\mathbb{R}^n} f(x-y)\int_{\mathbb{R}^n} g(y-z)h(z)\,\mathrm{d}z\,\mathrm{d}y \\ &= \int_{\mathbb{R}^n} h(z)\int_{\mathbb{R}^n} f(x-y)g(y-z)\,\mathrm{d}y\,\mathrm{d}z = \int_{\mathbb{R}^n} h(z)\int_{\mathbb{R}^n} f(x-z-u)g(u)\,\mathrm{d}u\,\mathrm{d}z \\ &= \int_{\mathbb{R}^n} h(z)f * g(x-z)\,\mathrm{d}z = (f * g) * h(x)\,. \end{aligned}$$

Finally we show the last claim about the supports. Let x be a point in the complement of $\overline{\operatorname{supp} f + \operatorname{supp} g}$; we need to show that there is an open neighborhood V of x such that $f * g$ vanishes on V. Take V to be the complement of $\overline{\operatorname{supp} f + \operatorname{supp} g}$; then, for all $z \in V$, if $g(y) \neq 0$ for some $y \in \mathbb{R}^n$, necessarily $f(z-y) = 0$, else z would lie in $\operatorname{supp} f + \operatorname{supp} g$. Thus $f(z-y)g(y)$ vanishes identically, whence $f * g(z) = 0$. □

We now formulate two sets of conditions under which the convolution is almost everywhere defined, providing in each case a corresponding estimate for the relevant L^p-norm.

[6]More pedantically, the claimed chain of equalities holds with all integrands taken in absolute value, which shows that $g * f(x)$ is well defined whenever $f * g(x)$ is, and then one can pass to equality of the integrals without absolute values. Similar considerations apply to the rest of the proof: we argue directly without absolute values, for the sake of conciseness.

THEOREM 3.4.3 (Young's inequality). *Let $1 \le p \le \infty$, and suppose $f \in L^1$ and $g \in L^p$. Then $f*g$ is well defined Lebesgue-almost everywhere and extends to a Borel-measurable function on $\mathbb{R}^n$, which we also denote $f * g$. Furthermore, $f * g \in L^p$ and*

$$\|f * g\|_p \le \|f\|_1 \|g\|_p$$

In the case p finite, the proof shall rely upon the following far-reaching generalization of Minkowski's inequality, whose proof we postpone to §??.

PROPOSITION 3.4.4 (Minkowski's inequality for integrals). *Let $(X, \mathfrak{M}, \mu)$ and $(Y, \mathfrak{N}, \nu)$ be σ-finite measure spaces, $f\colon X \times Y \to [0, \infty]$ an $(\mathfrak{M} \otimes \mathfrak{N})$-measurable function. Then, for every $1 \le p < \infty$,*

$$\left(\int_X \left(\int_Y f(x,y)\, \mathrm{d}\nu(y)\right)^p \mathrm{d}\mu(x)\right)^{1/p} \le \int_Y \left(\int_X f(x,y)^p\, \mathrm{d}\mu(x)\right)^{1/p} \mathrm{d}\nu(y)\,. \tag{3.4.1}$$

Both sides of (3.4.1) are allowed to be infinite. The interpretation of the inequality as a generalization of the standard Minkowski inequality is given by regarding the left-hand side as the L^p-norm of an integral, with respect to a second parameter y, of functions defined on $(X, \mathfrak{M}, \mu)$ (generalizing the L^p-norm of the sum of two functions), and the right-hand side as the integral of the L^p-norms (generalizing the sum of the L^p-norms of two functions).

PROOF OF THEOREM 3.4.3. We start with the case $p = \infty$. For all $x \in \mathbb{R}^n$, we have

$$\int_{\mathbb{R}^n} |f(x-y)g(y)|\, \mathrm{d}y \le \int_{\mathbb{R}^n} |f(x-y)|\, \|g\|_\infty\, \mathrm{d}y = \|g\|_\infty \int_{\mathbb{R}^n} |f(x-y)|\, \mathrm{d}y = \|g\|_\infty \|f\|_1\;,$$

applying implicitly the change of variable $z = x - y$ for the last equality. It follows that $f * g(x)$ is well defined, and

$$|f * g(x)| = \left|\int_{\mathbb{R}^n} f(x-y)g(y)\, \mathrm{d}y\right| \le \int_{\mathbb{R}^n} |f(x-y)g(y)|\, \mathrm{d}y \le \|f\|_1 \|g\|_\infty\;,$$

whence $f * g \in L^\infty$ (it is actually uniformly bounded) and

$$\|f * g\|_\infty \|f\|_1 \|g\|_\infty\;.$$

We now turn to the case of a finite exponent p. Applying Minkowski's inequality for integrals, we have

$$\begin{aligned}\left(\int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} |g(x-y)f(y)|\, \mathrm{d}y\right)^p \mathrm{d}x\right)^{1/p} &\le \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} |g(x-y)|^p |f(y)|^p\, \mathrm{d}x\right)^{1/p} \mathrm{d}y \\ &= \int_{\mathbb{R}^n} |f(y)| \left(\int_{\mathbb{R}^n} |g(x-y)|^p\, \mathrm{d}x\right)^{1/p} \mathrm{d}y = \int_{\mathbb{R}^n} |f(y)|\, \|g\|_p\, \mathrm{d}y = \|f\|_1 \|g\|_p\;,\end{aligned} \tag{3.4.2}$$

using implicitly the substitution $z = x - y$ in the second-to-last equality. It follows that the integral

$$\int_{\mathbb{R}^n} |g(x-y)f(y)|\, \mathrm{d}y$$

is finite for Lebesgue-almost every $x \in \mathbb{R}^n$, whence $f * g(x)$ $(= g * f(x))$ is well-defined for all such x. Measurability of the almost-everywhere defined function $f * g$ is then proven exactly as in the context of Fubini's theorem; $f * g$ admits then a Borel-measurable extension to $\mathbb{R}^n$. Finally, from (3.4.2) we deduce

$$\begin{aligned}\|f * g\|_p = \|g * f\|_p = \left(\int_{\mathbb{R}^n} \left|\int_{\mathbb{R}^n} g(x-y)f(y)\, \mathrm{d}y\right|^p \mathrm{d}x\right)^{1/p} &\le \left(\int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} |g(x-y)f(y)|\, \mathrm{d}y\right)^p \mathrm{d}x\right)^{1/p} \\ &\le \|f\|_1 \|g\|_p\;,\end{aligned}$$

which concludes the proof. □

PROPOSITION 3.4.5. *Let $1 \leq p, q \leq \infty$ be Hölder-conjugate exponents, and suppose $f \in L^p$ and $g \in L^q$. Then $f * g$ is well defined everywhere and is a uniformly bounded function on $\mathbb{R}^n$. Furthermore,*

$$\sup_{x\in\mathbb{R}^n} |f * g(x)| \leq \|f\|_p \|g\|_q \ .$$

Building upon later results of this section, we shall further derive that, under the assumptions of the proposition, $f * g$ is uniformly continuous (in particular, Borel-measurable) and, if $1 < p, q < \infty$, that it *vanishes at infinity.*

PROOF. Hölder's inequality delivers immediately, for all $x \in \mathbb{R}^n$,

$$\int_{\mathbb{R}^n} |f(x-y)||g(y)| \,\mathrm{d}y \leq \left(\int_{\mathbb{R}^n} |f(x-y)|^p \,\mathrm{d}y\right)^{1/p} \left(\int_{\mathbb{R}^n} |g(y)|^q \,\mathrm{d}y\right)^{1/q} = \|f\|_p \|g\|_q \ ,$$

from which it follows at once that $f * g(x)$ is well defined and

$$|f * g(x)| \leq \|f\|_p \|g\|_q \ .$$

□

After having established the basic properties of convolutions, we turn to the quest or regularizing functions in L^p spaces. We begin by some notation concerning differential calculus in $\mathbb{R}^n$. A **multi-index** is a tuple $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$, its **length** is $|\alpha| = \sum_{1\leq i\leq n} \alpha_i$. To each multi-index α, we associate the differential operator ∂^α which acts on a function $f \in \mathscr{C}^k(\mathbb{R}^n)$, with $k = |\alpha|$, via

$$\partial^\alpha f = \partial_{x_1}^{\alpha_1} \cdots \partial_{x_n}^{\alpha_n} f \ ,$$

where $\partial_{x_j}^{\alpha_j}$ denotes the partial derivative in the direction of the j-th vector e_j of the canonical basis $(e_1, \ldots, e_n)$ of $\mathbb{R}^n$, iterated α_j times. If $|\alpha| = 0$, we declare by convention that $\partial^\alpha f = f$.

The following proposition expresses an instance of a general phenomenon, according to which the convolution operation preserves the regularity of each of the factors.

PROPOSITION 3.4.6. *Let $f \in L^1(\mathbb{R}^n)$, $g \in \mathscr{C}^k(\mathbb{R}^n)$. Assume that, for every multi-index α with $|\alpha| \leq k$, the partial derivative $\partial_\alpha g$ is uniformly bounded on $\mathbb{R}^n$. Then $f * g \in \mathscr{C}^k(\mathbb{R}^n)$*[7] *and*

$$\partial^\alpha(f * g) = f * \partial^\alpha g$$

for all $|\alpha| \leq k$.

PROOF. Notice first that $f * g$ is well defined everywhere since $f \in L^1$ and g is uniformly bounded on $\mathbb{R}^n$. If $|\alpha| = 0$, then the claim in the statement is simply that $f * g$ is continuous; this follows from a direct application of Proposition 2.4.17.

Suppose now $|\alpha| = 1$, so that $\partial^\alpha = \partial_{x_j}$ for some $1 \leq j \leq n$; then $f * \partial^\alpha g$ is well defined everywhere, since $f \in L^1$ and $\partial^\alpha g$ is uniformly bounded. It follows then, resorting once more to Proposition 2.4.17, that $\partial^\alpha(f * g)$ exists everywhere, and

$$\partial^\alpha(f*g)(x) = \partial_{x_j}(f*g)(x) = \partial_{x_j}(g*f)(x) = \int_{\mathbb{R}^n} \partial_{x_j} g(x-y) f(y) \,\mathrm{d}y = \partial^\alpha g * f(x) = f * \partial^\alpha g(x)$$

for all $x \in \mathbb{R}^n$.

Arguing by induction on $|\alpha|$, one obtains the statement for all multi-indices in a similar fashion. □

Let $U \subset \mathbb{R}^n$ be an open set. We define, for all integers $k \geq 0$,

$$\mathscr{C}^k(U) = \{f \colon U \to \mathbb{C} : f \text{ admits continuous partial derivatives of order } k \text{ on } U\} \ ,$$

with the convention that $\mathscr{C}^0(U)$ is the set of continuous functions $U \to \mathbb{C}$; we also set

$$\mathscr{C}^\infty(U) = \bigcap_{k\geq 1} \mathscr{C}^k(U) \ ,$$

[7] Implicitly, this states in particular that $f * g$ is well defined everywhere.

and let

$$\mathscr{C}^k_{\mathrm{c}}(U) = \{f \in \mathscr{C}^k(U) : \operatorname{supp} f \text{ is compact}\}\ ,$$

for all $k \in \mathbb{N} \cup \{\infty\}$.

If $x \in \mathbb{R}^n$ and $\varepsilon > 0$, we denote by $B(x, \varepsilon)$ the closed Euclidean ball centered at x of radius ε.

DEFINITION 3.4.7 (Mollifiers). A family of **mollifiers**, also known as an **approximate identity**, on $\mathbb{R}^n$ is a family $(\rho_\varepsilon)_{\varepsilon>0}$ of functions $f_\varepsilon \in \mathscr{C}^\infty_{\mathrm{c}}(\mathbb{R}^n)$ satisfying the following properties:

(1) $\rho_\varepsilon \geq 0$ for all $\varepsilon > 0$;
(2) $\operatorname{supp} \rho_\varepsilon \subset B(0, \varepsilon)$ for all $\varepsilon > 0$;
(3) $\int_{\mathbb{R}^n} \rho_\varepsilon \, \mathrm{d}x = 1$ for all $\varepsilon > 0$.

Convolving with mollifiers shall produce the desired approximation of L^p functions by regular functions. The terminology "approximate identity" refers precisely to this feature: as we shall prove in this section, the convolution operator with the function ρ_ε provides, on appropriate function spaces and for ε close to 0, an an approximation to the identity operator.

Let us show how to construct families of mollifiers. We start with a positive function $\rho \in \mathscr{C}^\infty_{\mathrm{c}}(\mathbb{R}^n)$ with $\int_{\mathbb{R}^n} \rho \, \mathrm{d}x = 1$ and support contained in $B(0, 1)$. In what follows, $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^n$, unless otherwise specified.

LEMMA 3.4.8. *(1) The function $\varphi\colon \mathbb{R} \to \mathbb{R}$ defined by*

$$\varphi(t) = \begin{cases} e^{-1/t} & \text{if } t > 0 \\ 0 & \text{if } t \leq 0 \end{cases} .$$

is in $\mathscr{C}^\infty(\mathbb{R})$.

(2) The function $\rho_0\colon \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ defined by

$$\rho_0(x) = \begin{cases} e^{1/(\|x\|^2 - 1)} & \text{if } \|x\| < 1 \\ 0 & \text{if } \|x\| \geq 1 \end{cases}$$

is in $\mathscr{C}^\infty(\mathbb{R}^n)$, with support contained in $B(0, 1)$.

PROOF. (1) It is clear that $\varphi \in \mathscr{C}^\infty(\mathbb{R} \setminus \{0\})$; at $t = 0$, one verifies easily by induction on k, that

$$\lim_{t \to 0^+} \varphi^{(k)}(t) = 0\ ,$$

from which the right derivative of order k exists at $t = 0$, and vanishes. This shows that $\varphi \in \mathscr{C}^\infty(\mathbb{R})$.

(2) We have

$$\rho_0(x) = \varphi(1 - \|x\|^2)\ ,$$

with φ the function of the previous item. Since $x \mapsto 1 - \|x\|^2$ is a polynomial function, we deduce at once that $\rho_0 \in \mathscr{C}^\infty(\mathbb{R}^n)$. As ρ_0 vanishes on the complement of the closed ball $B(0, 1)$, it is clear that $\operatorname{supp} \rho_0 \subset B(0, 1)$.

□

Let ρ_0 be the function of the previous proposition; notice that $\rho_0 \in L^1$, being continuous of compact support, and

$$\int_{\mathbb{R}^n} \rho_0 \, \mathrm{d}x > 0\ .$$

To see this, let $x \in \mathbb{R}^n$ with $\rho_0(x) > 0$; then by continuity $\rho_0 > \rho_0(x)/2$ on a neighborhood V of x, and thus

$$\int_{\mathbb{R}^n} \rho_0 \, \mathrm{d}y \geq \int_V \rho_0 \, \mathrm{d}y \geq \int_V \frac{\rho_0(x)}{2} \, \mathrm{d}y = \frac{\mathscr{L}^n(V)\rho_0(x)}{2} > 0\ ,$$

the last inequality stemming from the fact that $\mathscr{L}^n$ assigns positive mass to every non-empty open set. We can thus renormalize ρ to obtain a function

$$\rho = \frac{1}{\int_{\mathbb{R}^n} \rho_0 \, \mathrm{d}x} \, \rho_0 \ ,$$

which is in $\mathscr{C}^\infty(\mathbb{R}^n)$, has support contained in $B(0,1)$, and verifies $\int_{\mathbb{R}^n} \rho \, \mathrm{d}x = 1$.

Define now, for all $\varepsilon > 0$, a function $\rho_\varepsilon \colon \mathbb{R} \to \mathbb{R}$ by

$$\rho_\varepsilon(x) = \frac{1}{\varepsilon^n} \rho\left(\frac{x}{\varepsilon}\right) .$$

Then it is plain that $\rho_\varepsilon \geq 0$ and $\operatorname{supp} \rho_\varepsilon \subset B(0,\varepsilon)$ for all $\varepsilon > 0$. Also,

$$\int_{\mathbb{R}^n} \rho_\varepsilon \, \mathrm{d}x = \frac{1}{\varepsilon^n} \int \rho\left(\frac{x}{\varepsilon}\right) \mathrm{d}x = \frac{1}{\varepsilon^n} \int_{\mathbb{R}^n} \varepsilon^n \rho(y) \, \mathrm{d}y = 1 \ ,$$

using the change of variable $y = x/\varepsilon$ in the second equality. We have thus obtained a family $(\rho_\varepsilon)_{\varepsilon>0}$ of mollifiers.

For the remainder of this section, we fix an approximate identity $(\rho_\varepsilon)_{\varepsilon>0}$ on $\mathbb{R}^n$. Observe that, if $1 \leq p \leq \infty$ and $f \in L^p$, then $\rho_\varepsilon * f$ is well defined almost everywhere and is in L^p for all $\varepsilon > 0$, since $\rho_\varepsilon \in L^1$ (see Theorem 3.4.3).

THEOREM 3.4.9. *Let $(\rho_\varepsilon)_{\varepsilon>0}$ be an approximate identity on $\mathbb{R}^n$.*

(1) Let $f \in L^p$ for some $1 \leq p < \infty$. Then $\rho_\varepsilon * f \xrightarrow{\varepsilon\to 0} f$ in L^p.

(2) Let $f \in L^\infty$, continuous on an open set $U \subset \mathbb{R}^n$. Then $\rho_\varepsilon * f \xrightarrow{\varepsilon\to 0} f$ uniformly on compact subsets of U.

If, moreover, f is uniformly continuous on U, then $\rho_\varepsilon * f \xrightarrow{\varepsilon\to 0} f$ uniformly on U.

Recall that a collection $(f_\varepsilon)_{\varepsilon>0}$ of complex-valued functions defined on an open set $U \subset \mathbb{R}^n$ is said to converge to a bounded function $f \colon U \to \mathbb{C}$ uniformly on compact subsets of U if

$$\sup_{x\in K} |f(x) - f_\varepsilon(x)| \xrightarrow{\varepsilon\to 0} 0$$

for every compact subset $K \subset U$.

The remainder of this section is devoted to the proof of Theorem 3.4.9, which shall require a considerable deal of preparation.

APPENDIX A

Topology

A.1. Topological and metric spaces

Standard references for the material in this chapter are [**11**, **16**, **2**, **3**, **5**].

DEFINITION A.1.1 (Topological space). Let X be a set. A **topology** on X is a collection $\tau \subset \mathcal{P}(X)$ satisfying the following properties:

(1) $\emptyset \in \tau$, $X \in \tau$;
(2) if $(O_i)_{i\in I}$ is an arbitrary family of elements of τ, then $\bigcup_{i\in I} O_i \in \tau$;
(3) if $(O_j)_{j\in J}$ is a finite family of elements of τ, then $\bigcap_{j\in J} O_j \in \tau$.

The pair (X, τ) is called a **topological space**.

If (X, τ) is a topological space, elements of τ are called **open sets**. A set $F \subset X$ is **closed** if the complement F^{c} is open. A **neighborhood** of a point $x \in X$ is a set $V \subset X$ containing an open set O to which x belongs[1].

DEFINITION A.1.2 (Continuous map). Let (X, τ_X) and (Y, τ_Y) be topological spaces. A map $f\colon X \to Y$ is **continuous** if, for every open set $O \subset Y$, $f^{-1}(O)$ is open in X.

As is well known, continuity is a local property:

LEMMA A.1.3. *Let (X, τ_X), (Y, τ_Y) be topological spaces. A function $f\colon X \to Y$ is continuous if and only if, for every $x \in X$ and every neighborhood V of $f(x)$, there is a neighborhood U of x such that $f(U) \subset V$.*

PROOF. Suppose f is continuous. Let $x \in X$, V a neighborhood of $f(x)$; then there is an open set $O \subset V$ such that $f(x) \in O$. The inverse image $U = f^{-1}(O)$ is an open set containing x, thus in particular an open neighborhood of x, and satisfies $f(U) = f(f^{-1}(O)) \subset O \subset V$.

Conversely, suppose the local condition holds for all $x \in X$, and let $O \subset Y$ be an open set. If $x \in f^{-1}(O)$, then there is a neighborhood U_x of x, which we may assume open without loss of generality, such that $f(U_x) \subset O$, that is, $U_x \subset f^{-1}(O)$. We have thus expressed $f^{-1}(O)$ as a union of the open sets U_x, $x \in f^{-1}(O)$; it follows that $f^{-1}(O)$ is open. As O is arbitrary, this shows continuity of f. □

If $E \subset X$ is any subset, we define:

- the **interior** of E, denoted E°, as the union of all open sets contained in E;
- the **closure** of E, denoted $\overline{E}$, as the intersection of all closed sets containing E;
- the **boundary** of E, denoted ∂E, as the intersection $\overline{E} \cap \overline{X \setminus E}$.

The interior E° is open, and is the largest open set contained in E; the closure $\overline{E}$ is closed, and is the smalles closed set containing E. The boundary ∂E is closed, and X can be partitioned as

$$X = E^\circ \sqcup \partial E \sqcup (X \setminus E)^\circ \ ,$$

as follows immediately from the descriptions

$$E^\circ = \{x \in E : \text{there is a neighborhood } V \text{ of } x \text{ s.t. } V \subset E\} \ ,$$

$$\overline{E} = \{x \in X : \text{for every neighborhood } V \text{ of } x,\ V \subset E \neq \emptyset\} \ ,$$

[1]Observe that, in many textbooks on disparate mathematical disciplines, a neighborhood is meant to be an open set; we adopt a more general notion here, which is equally well established in the literature.

and the ensuing

$$\partial E = \{x \in X : \text{for every neighborhood } V \text{ of } x, V \cap E \neq \emptyset \text{ and } V \cap (X \setminus E) \neq \emptyset\} .$$

A subset $Y \subset X$ is **dense** if $\overline{Y} = X$; equivalently, if $Y \cap O \neq \emptyset$ for every non-empty open set $O \subset X$.

Arguably the most important source of examples of topological spaces comes from the notion of a metric, or distance function on a set.

DEFINITION A.1.4 (Metric space). Let X be a set. A **metric** on X is a function $d\colon X \times X \to \mathbb{R}_{\geq 0}$ satisfying the following properties:

(1) $d(x, y) = 0$ if and only if $x = y$;
(2) (symmetry) $d(x, y) = d(y, x)$ for all $x, y \in X$;
(3) (triangle inequality) $d(x, y) \leq d(x, z) + d(z, y)$ for all $x, y, z \in X$.

The pair (X, d) is called a **metric space**.

EXAMPLE A.1.5. The **Euclidean metric** $d_{\mathbb{R}^n}$ on $\mathbb{R}^n$ is defined as follows. Let $|\cdot|_{\mathbb{R}^n} : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ be defined via

$$|x|_{\mathbf{R}^n} = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2}$$

for all $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Observe that, for $n = 1$, $|\cdot|_{\mathbb{R}}$ is nothing but the ordinary absolute value of real numbers.

We then set

$$d_{\mathbb{R}^n}(x, y) = |x - y|_{\mathbb{R}^n}$$

for all $x, y \in \mathbb{R}^n$. The topology induced by $d_{\mathbb{R}^n}$ is called the **Euclidean topology** on $\mathbb{R}^n$.

Let (X, d) be a metric space. If $x \in X$ and $r \geq 0$, the **open ball centered at x of radius** r, with respect to the metric d, is the set

$$B(x, r) = \{y \in X : d(x, y) < r\} .$$

Clearly $B(x, 0) = \emptyset$ for all $x \in X$.

If (X, d) is a metric space, the metric d induces a topology τ_d on X defined as follows: a set $O \subset X$ is open for τ_d if and only if, for every $x \in O$, there is $r > 0$ such that $B(x, r) \subset O$. The fact that the collection τ_d thus defined is closed under finite intersections is an immediate consequence of the fact that

$$B(x, r_1) \cap B(x, r_2) = B(x, \inf\{r_1, r_2\})$$

for all $x \in X$ and $r_1, r_2 > 0$. Closure under arbitrary unions is trivial, whence τ_d is indeed a topology.

It is easy to show that every open ball is an open set for the topology induced by the metric, which justifies the terminology[2].

If X is a set, a topology τ on X is said to be **metrizable** if there exists a metric d on X inducing τ.

If (X, d) is a metric space and $E \subset X$, the **diameter** of E is defined as

$$\operatorname{diam}(E) = \sup\{d(x, y) : x, y \in E\} .$$

A set $E \subset X$ is **bounded** if its diameter is finite; equivalently, if it is contained in an open (or closed) ball (of finite radius).

[2]The converse is obviously not true: for instance, $(0, +\infty)$ is open for the Euclidean topology on $\mathbb{R}$, but is not an open ball; the latter are indeed of the form (a, b) for $a < b \in R$.

A.1.1. On the extended real number system $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ we shall always consider the topology that naturally extends the Euclidean topology on $\mathbb{R}$, namely the one generated by the collection

$$\{[-\infty, a) : a \in \mathbb{R}\} \cup \{(a, +\infty] : a \in \mathbb{R}\} .$$

It can be checked that such a topology is metrizable; for instance, the metric

$$d(x, y) = |\arctan x - \arctan y| , \quad x, y \in \overline{\mathbb{R}},$$

with the convention that $\arctan \pm\infty = \pm\pi/2$, is easily seen to induce the topology just defined.

A.1.2. Let X be a set. Given an arbitrary collection $\mathcal{E}$ of subsets of X, the smallest topology $\tau_{\mathcal{E}}$ on X containing $\mathcal{E}$ is called the **topology generated by** $\mathcal{E}$; it is obtained as the intersection of all topologies on X containing $\mathcal{E}$, using the fact that arbitrary intersections of topologies (on a fixed set) are topologies, and that $\mathcal{P}(X)$ is always a topology on X containing $\mathcal{E}$, so that the aforementioned intersection is always over a non-empty family. The collection $\mathcal{E}$ is called a **subbasis** of the topology $\tau_{\mathcal{E}}$.

A **basis** of a topology τ is a subcollection $\mathcal{B} \subset \tau$ with the property that every $O \in \tau$ can be expressed as the union of elements of $\mathcal{B}$.

EXAMPLE A.1.6. If (X, d) is a metric space and τ_d is the topology on X induced by the metric d, then the collection of open balls for d is, essentially by definition of τ_d, a basis of τ_d.

A topological space (X, τ) is **Hausdorff** if, for all $x_1 \neq x_2 \in X$, there are disjoint open sets V_1, V_2 such that $x_1 \in V_1$ and $x_2 \in V_2$.

A topological space (X, τ) is **second countable** if τ admits a countable basis. It is **separable** if there is a countable dense subset $Y \subset X$.

LEMMA A.1.7. *Let (X, d) be a metric space, τ_d the topology on X induced by d. The following are equivalent:*

(1) (X, τ_d) is second countable;
(2) (X, τ_d) is separable.

PROOF. Second countability implies separability for every topological space: if $\mathcal{B}$ is a countable basis for the topology, pick an element $x_B \in B$ for each non-empty $B \in \mathcal{B}$. Then the countable set $Y = \{x_B : B \in \mathcal{B}\}$ is dense, for if $O \subset$ is a non-empty open set, there is a non-empty $B \in \mathcal{B}$ such that $B \subset O$, and thus $x_B \in Y \cap O$.

Conversely, let (X, τ_d) be the topological space associated to a metric space (X, d), and let $Y \subset X$ be a countable dense set. We claim that the countable collection

$$\mathcal{B} = \{B(x, 1/n) : x \in Y,\ n \in \mathbb{N}^*\}$$

is a basis of τ_d. If $O \subset X$ is open and $x \in O$, then there is $r > 0$ such that $B(x, r) \subset O$; choose now $n \in \mathbb{N}^*$ such that $1/n < r/2$, and pick $y \in Y \cap B(x, 1/n)$, whose existence is guaranteed by the fact that Y is dense. Then by symmetry $x \in B(y, 1/n)$, the latter being an element of $\mathfrak{B}$, and $B(y, 1/n)$ is contained in $B(x, r) \subset O$, since $d(z, y) < 1/n$ implies

$$d(z, x) \leq d(z, y) + d(y, x) < \frac{1}{n} + \frac{1}{n} < r .$$

We have thus shown that $\mathcal{B}$ is a basis for τ_d. □

In particular, the Euclidean topology on $\mathbb{R}^n$ is second countable for all $n \geq 1$: the subset $\mathbb{Q}^n$ is countable and dense.

A.1.3. Product and induced topologies. Let $(X_\alpha, \tau_\alpha)_{\alpha\in\mathsf{A}}$ be a family of topological spaces. On the product set $X = \prod_{\alpha\in\mathsf{A}} X_\alpha$ we define a topology τ, called the **product topology** of the τ_α's, as the coarsest[3] topology on X for which all canonical projection maps $\pi_\alpha\colon X \to X_\alpha,\ \alpha \in \mathsf{A}$ are continuous. Equivalently, τ is the topology on X generated by the collection

$$\{\pi_\alpha^{-1}(O_\alpha) : O_\alpha \subset X_\alpha \text{ open},\ \alpha \in \mathsf{A}\}\ .$$

The topological space (X, τ), called the **product of the topological spaces** (X_α, τ_α), satisfies the following universal property. For every topological space (Z, τ_Z) and every continuous map $f\colon Z \to X$, the compositions $\pi_\alpha\colon f\colon Z \to X_\alpha$ are continuous for all $\alpha \in \mathsf{A}$; conversely, if (Z, τ_Z) is a topological space and $f_\alpha\colon Z \to X_\alpha,\ \alpha \in \mathsf{A}$ is a family of continuous maps, there exists a unique continuous map $f\colon Z \to X$ such that $f_\alpha = \pi_\alpha \circ f$ for all $\alpha \in \mathsf{A}$.

Let (X, τ_X) be a topological space, $Y \subset X$ a subset. The **induced topology** τ_Y on Y is the coarsest topology on Y making the canonical inclusion map $\iota\colon Y \to X$ continuous. It can be explicitly described as the collection

$$\tau_Y = \{O \cap Y : O \subset X \text{ open for } \tau_X\}\ .$$

If (Z, τ_Z) is a topological space and $f\colon Z \to Y$ is a function, then f is continuous with respect to τ_Y if and only if the composition $\iota \circ f\colon Z \to X$ is continuous with respect to τ_X.

Remark A.1.8. If (X, d_X) is a metric space, there is a notion of **induced metric** d_Y on any subset Y, simply by restricting the domain of d to $Y \times Y$. Then the topology on Y determined by the induced metric d_Y coincides with the induced topology τ_Y from the topology τ_X on X determined by the metric d_X.

A.2. Functions between metric spaces

If (X, d_X) and (Y, d_Y) are metric spaces, a **Lipschitz function** between X and Y is a function $f\colon X \to Y$ with the property that there is $C > 0$ such that

$$d_Y(f(x_1), f(x_2)) \leq C d_X(x_1, x_2)$$

for all $x_1, x_2 \in X$. In this case, we also say that f is **C-Lipschitz**.

If (X, d_X) and (Y, d_Y) are metric spaces, an **isometry** is a map $h\colon X \to Y$ such that

$$d_Y(h(x_1), h(x_2)) = d_X(x_1, x_2)$$

for all $x_1, x_2 \in X$. Every isometry is injective, but not every isometry is surjective.

If $X = \mathbb{R}^n$, an **Euclidean isometry** of X is an isometry $h\colon \mathbb{R}^n \to \mathbb{R}^n$ where $\mathbb{R}^n$ is endowed with the standard Euclidean metric.

We denote by $\mathrm{O}_n(\mathbb{R})$ the group of linear isometries of the vector space $\mathbb{R}^n$ endowed with the standard inner product.

Proposition A.2.1. *Let $h\colon \mathbb{R}^n \to \mathbb{R}^n$ be a Euclidean isometry. Then h is an affine bijection. More precisely, there exist $f \in \mathrm{O}_n(\mathbb{R})$ and $a \in \mathbb{R}^n$ such that*

$$h(x) = f(x) + a$$

for all $x \in \mathbb{R}^n$.

A.3. Various topological properties

Definition A.3.1 (Compactness). A topological space (X, τ) is compact if every cover $(O_\alpha)_{\alpha\in\mathsf{A}}$ of X by open sets admits a finite subcover, namely there is a finite $\mathsf{B} \subset \mathsf{A}$ such that $X = \bigcup_{\alpha\in\mathsf{B}} O_\alpha$.

[3]That is, the smalles, in the sense of inclusion.

When (X, τ) is a topological space and we speak of a compact subset $Y \subset X$, we always tacitly endow Y with the induced topology from τ.

If (X, τ) is a compact topological space, then every closed subset $C \subset X$ is compact. To see this, let $(O_\alpha)_{\alpha \in \mathsf{A}}$ be an open cover of C; then adjoining to it the open set $X \setminus C$ gives an open cover of X, which admits a finite subcover $(O_\alpha)_{\alpha \in \mathsf{B}}$. The latter is, in particular, a finite open cover of C.

PROPOSITION A.3.2 (Tychonoff's theorem). *Let $(X_\alpha, \tau_\alpha)_{\alpha \in \mathsf{A}}$ be a family of compact topological spaces. Endow the product $X = \prod_{\alpha \in \mathsf{A}} X_\alpha$ with the product topology τ. Then (X, τ) is compact.*

A.4. Convergence of functions

Let X be a set, E a topological space. A sequence $(f_n)_{n \geq 0}$ of functions $f \colon X \to E$ is said to **converge pointwise** to a function $f \colon X \to E$ if, for every $x \in X$, the sequence $(f_n(x))_{n \geq 0}$ converges towards $f(x)$ in E.

Let X be a set, (E, d) a metric space. A sequence $(f_n)_{n \geq 0}$ of functions $f \colon X \to E$ is said to **converge uniformly** to a function $f \colon X \to E$ if, for every $\varepsilon > 0$, there is $N \in \mathbb{N}$ such that for all integers $n \geq N$ and for all $x \in X$,

$$d(f(x), f_n(x)) \leq \varepsilon \;;$$

said differently, $f_n \to f$ uniformly as $n \to \infty$ if

$$\sup_{x \in X} d(f(x), f_n(x)) \overset{n \to \infty}{\longrightarrow} 0 \;.$$

If Y is a subset of X, we say that the sequence $(f_n)_{n \geq 0}$ converges uniformly to f on Y if the sequence of restrictions $(f_n|_Y)_{n \geq 0}$ converges uniformly to the restriction $f|_Y$.

It is clear that uniform convergence implies pointwise convergence[4]: if $(f_n)_{n \geq 0}$ converges uniformly towards f, then, for every fixed $x_0 \in X$,

$$d(f(x_0), f_n(x_0)) \leq \sup_{x \in X} d(f(x), f_n(x)) \overset{n \to \infty}{\longrightarrow} 0 \;,$$

whence $f_n(x_0) \to f(x_0)$ as $n \to \infty$.

[4]Where, as usual, the topology on (E, d) is the one induced from the metric.

APPENDIX B

Infinite sums

B.1. Sums of families of real numbers

A reference for the material in this chapter is [3].

DEFINITION B.1.1 (Summable family of real numbers). A family $(x_\alpha)_{\alpha\in\mathsf{A}}$ of real numbers is said to be **summable** if there is a real number x with the property that, for every open neighborhood V of x, there is a finite set $\mathsf{B}_0 \subset \mathsf{A}$ such that, for every finite set $\mathsf{B} \supset \mathsf{B}_0$,

$$\sum_{\alpha\in\mathsf{B}} x_\alpha \in V \ .$$

In this case, x is called the **sum** of the family $(x_\alpha)_{\alpha\in\mathsf{A}}$. We write $x = \sum_{\alpha\in\mathsf{A}} x_\alpha$.

Clearly, if A is finite, then $(x_\alpha)_{\alpha\in\mathsf{A}}$ is summable with sum the ordinary sum of a finite collection of real numbers.

REMARK B.1.2. We conform to the traditional (and useful) convention that the sum of an empty family of numbers equals 0.

The definition of summable family does not involve any ordering of the elements of the collection. More precisely, we have the following "commutativity" result, whose proof is left as an exercise.

LEMMA B.1.3. *Let $(x_\alpha)_{\alpha\in\mathsf{A}}$ be a summable family of real numbers, with sum x. For every bijection $\phi\colon \mathsf{A}\to\mathsf{A}$, the family $(x_{\phi(\alpha)})_{\alpha\in\mathsf{A}}$ is summable, with sum x.*

PROPOSITION B.1.4 (Cauchy criterion for summable families). *Let $(x_\alpha)_{\alpha\in\mathsf{A}}$ be a family of real numbers. The following are equivalent:*

(1) $(x_\alpha)_{\alpha\in\mathsf{A}}$ is summable;
(2) $(x_\alpha)_{\alpha\in\mathsf{A}}$ satisfies the Cauchy criterion: for every neighborhood V of 0, there is a finite set $\mathsf{B}\subset\mathsf{A}$ such that, for every finite set $\mathsf{C}\subset\mathsf{A}$ with $\mathsf{C}\cap\mathsf{B}=\emptyset$,

$$\sum_{\alpha\in\mathsf{C}} x_\alpha \in V \ .$$

We deduce the following consequences.

COROLLARY B.1.5. *Let $(x_\alpha)_{\alpha\in\mathsf{A}}$ be a summable family.*

(1) For every neighborhood V of 0, the set

$$\{\alpha\in\mathsf{A} : x_\alpha\notin V\}$$

is finite.

(2) The set

$$\{\alpha\in\mathsf{A} : x_\alpha\neq 0\}$$

is countable.

PROOF. For the first assertion, it suffices to apply the Cauchy criterion in Proposition B.1.4 with C varying all singletons disjoint from B. As to the second assertion, for every integer $n\geq 1$, the set

$$\mathsf{B}_n = \{\alpha\in\mathsf{A} : x_\alpha\notin(-1/n,1/n)\ \}$$

is finite by the first statement; the set of interest is the union over all $n\geq 1$ of the B_n, and is thus countable. □

Corollary B.1.6. *If* $(x_\alpha)_{\alpha\in\mathsf{A}}$ *is a summable family of real numbers, and* B *is a subset of* A*, then the subfamily* $(x_\alpha)_{\alpha\in\mathsf{B}}$ *is summable.*

Proof. It is clear that the Cauchy criterion for the full family implies the same criterion for any subfamily. □

Theorem B.1.7 (Associativity of the sum). *Let* $(x_\alpha)_{\alpha\in\mathsf{A}}$ *be a summable family of real numbers. Suppose* $(\mathsf{A}_\lambda)_{\lambda\in\Lambda}$ *is a partition of the index set* A*. Then the family of partial sums*

$$\sum_{\alpha\in\mathsf{A}_\lambda} x_\alpha\ , \quad \lambda\in\Lambda$$

is summable, and

$$\sum_{\alpha\in\mathsf{A}} x_\alpha = \sum_{\lambda\in\Lambda}\sum_{\alpha\in\mathsf{A}_\lambda} x_\alpha\ .$$

Specializing the theorem to the case when A is a product $\mathsf{A}_1\times\mathsf{A}_2$, and $(A_\lambda)_{\lambda\in\mathsf{A}_1}$ is the partition $A_\lambda = \{(\lambda,\eta) : \eta\in\mathsf{A}_2\}$ or $(A_\eta)_{\eta\in\mathsf{A}_2}$ is the partition $A_\eta = \{(\lambda,\eta) : \lambda\in\mathsf{A}_1\}$, we obtain the formula of change of order of summation.

$$\sum_{\lambda\in\mathsf{A}_1,\eta\in\mathsf{A}_2} x_{(\lambda,\eta)} = \sum_{\lambda\in\mathsf{A}_1}\sum_{\eta\in\mathsf{A}_2} x_{(\lambda,\eta)} = \sum_{\eta\in\mathsf{A}_2}\sum_{\lambda\in\mathsf{A}_1} x_{(\lambda,\eta)}\ .$$

B.2. Sums of extended real numbers

Definition B.2.1. If $(x_\alpha)_{\alpha\in\mathsf{A}}$ is a family in $[0,\infty]$, its **sum**, denoted $\sum_{\alpha\in\mathsf{A}} x_\alpha$, is defined as

$$\sup_{\mathsf{B}\subset\mathsf{A},\ \mathsf{B}\text{ finite}} \sum_{\alpha\in\mathsf{B}} x_\alpha\ .$$

Proposition B.2.2. *A family* $(x_\alpha)_{\alpha\in\mathsf{A}}$ *of positive real numbers is summable according to Definition B.1.1 if and only if its sum, as in Definition B.2.1, is finite, in which case it coincides with the sum of the family* $(x_\alpha)_{\alpha\in\mathsf{A}}$ *as in Definition B.1.1.*

Lemma B.1.3 and Theorem B.1.7 apply *verbatim* to sums of positive extended real numbers.

Bibliography

[1] P. Billingsley. *Probability and measure.* Anniversary edition. With a foreword by Steve Lalley and a brief biography of Billingsley by Steve Koppes. wiley Ser. Probab. Stat. John Wiley & Sons, Inc., Hoboken, NJ, 2012. xviii+624 pp.

[2] N. Bourbaki. *General topology. Chapters 1-4* (English summary). Translated from the French. Reprint of the 1989 English translation. Elem. Math. (Berlin). Springer-Verlag, Berlin, 1998. vii+437 pp.

[3] N. Bourbaki. *General topology. Chapters 5-10* (English summary). Translated from the French. Reprint of the 1989 English translation. Elem. Math. (Berlin). Springer-Verlag, Berlin, 1998. iv+363 pp.

[4] L.C. Evans, R.F. Gariepy. *Measure theory and fine properties of functions.* Revised edition. Textb. Math. CRC Press, Boca Raton, FL, 2015. xiv+299 pp.

[5] R. Engelking. *General topology.* Translated from the Polish by the author. Second edition. Sigma Ser. Pure Math., **6**. Heldermann Verlag, Berlin, 1989. viii+529 pp.

[6] K. Falconer. *Fractal geometry. Mathematical foundations and applications.* Third edition. John wiley & Sons, Ltd., Chichester, 2014. xxx+368 pp.

[7] K. Falconer. *Techniques in fractal geometry.* (English summary) John Wiley & Sons, Ltd., Chichester, 1997. xviii+256 pp.

[8] H. Federer. *Geometric measure theory.* Die Grundlehren der mathematischen Wissenschaften, Band **153**. Springer-Verlag New York, Inc., New York, 1969. xiv+676 pp.

[9] G.B. Folland, *Real analysis. Modern techniques and their applications.* Second edition. Pure Appl. Math. (N.Y.). Wiley-Intersci. Publ. John Wiley & Sons, Inc., New York, 1999. xvi+386 pp.

[10] P.R. Halmos. *Naive set theory.* Reprint of the 1960 edition. Undergrad. Texts Math. Springer-Verlag, New York-Heidelberg, 1974. vii+104 pp.

[11] A. Hatcher. *Notes on introductory point-set topology.* Available on the author's webpage, at https://pi.math.cornell.edu/ hatcher/Top/TopNotes.pdf

[12] H. Klenke. *Probability theory - a comprehensive course.* Third edition. Universitext. Springer, Cham, 2020. xiv+716 pp.

[13] E. Kowalski. *Measure and integral.* Available on the author's webpage, at

[14] S. MacLane. *Categories for the working mathematician.* Second edition. Grad. Texts in Math., **5**. Springer-Verlag, New York, 1998. xii+314 pp.

[15] P. Mattila. *Geometry of sets and measures in Euclidean spaces. Fractals and rectifiability* . (English summary) Cambridge Stud. Adv. Math., **44**. Cambridge University Press, Cambridge, 1995. xii+343 pp.

[16] J.R. Munkres. *Topology.* Second edition. Prentice Hall, Inc., Upper Saddle River, NJ, 2000. xvi+537 pp.

[17] W. Rudin. *Real and complex analysis.* Third edition. McGraw-Hill Book Co., New York, 1987. xiv+416 pp.

[18] T. Tao, https://terrytao.wordpress.com/2009/01/08/245b-notes-2-amenability-the-ping-pong-lemma-and-the-banach-tarski-paradox-optional/.