Real Analysis

Nicola Arcozzi

Typeset in LaTeX with Overleaf

Preface

These are the lecture notes of a one-semester course Advanced Analysis I taught for several years. Since the students come from different universities and different countries, and have different backgrounds, the material partially overlaps with that of courses taught at the Bachelor level at Unibo. However, exercises are on average more challenging than the ones many students were assigned in earlier courses. Each topic in the notes was covered, as it is in the notes, or otherwise, in one of the years I have taught the course. Honestly, covering the whole material with all the proofs would require more than one semester.

Sources for these notes include:

- Rudin, Walter Real and complex analysis. Third edition. McGraw-Hill Book Co., New York, 1987. xiv+416 pp. ISBN: 0-07-054234-1
- Folland, Gerald B. Real analysis. Modern techniques and their applications. 2nd ed. Pure and Applied Mathematics. A Wiley-Interscience Series of Texts, Monographs, and Tracts. New York, NY: Wiley. xiv, 386 p. (1999).
- Real Analysis for Graduate Students, version 5.0, by Richard F. Bass (2024). The paper version is here: Real Analysis for Graduate Students, version 2.1 (2014)
- An Introduction to Measure Theory (a draft) by Terence Tao. The definite version is An introduction to measure theory by Terence Tao, 2011; 206 pp; hardcover ISBN-10: 0-8218-6919-1 ISBN-13: 978-0-8218-6919-2 Graduate Studies in Mathematics, vol. 126.

Other references are given in the text. I am the only responsible, of course, for the mistakes, for unnecessarily long arguments, etcetera.

There is an expanded version of these lecture notes, which includes a number of supplements and more topics. If you are interested, write me at nicola.arcozzi@unibo.it and I will send you the file.

I wish to thank students, and colleagues (Annalisa Baldi, Nikolaos Chalmoukis, Giovanni Dore, Davide Guidetti, and many others), with whom I have discussed the material of the course over the years. Especially precious was the work of Dr. Nicola Zavatta, who went through much of the notes, pointed out mistakes, suggested improvements.

Bologna 2025

Contents

CHAPTER 1

A review of metric spaces

A basic problem in mathematics and its applications is measuring how much two points in space, or two point configurations, or two functions, signals, etc., are close to each other. A way to do this is by means of a *distance function*. A more flexible and general way to do the same is by introducing a *topology*. In most of these lectures, however, we will deal with distance functions only, and this chapter contains the basic results which are needed in the sequel.

We will soon see that we can define meaningful and useful distances between functions, which brings us straight away into the world of *functional analysis* ("functions of functions").

1.1. Generalities on metric spaces

1.1.1. Definition and basic examples. A *metric space* (X, d) is a set X endowed with a *distance function* $d\colon X \times X \to [0, \infty)$ satisfying, for all $x, y, z \in X$,

(i) $d(x, y) = 0$ if and only if $x = y$;

(ii) $d(x, y) = d(y, x)$;

(iii) $d(x, y) \leqslant d(x, z) + d(z, y)$, the *triangle inequality*.

EXERCISE 1.1. *Let (X, d) be a metric space and let $\phi\colon [0, \infty) \to [0, \infty)$ be concave, strictly increasing, $\phi(0) = 0$. Define $\delta(x, y) = \phi(d(x, y))$. Show that δ is a distance on X.*

If ϕ is continuously differentiable, you can use integrals to provide a slick proof.

Some examples of metric spaces are:

(i) $X = \mathbb{R}^n$ or $X = \mathbb{C}^n$ with the distance

$$d_p(x, y) = \begin{cases} \left(\sum_{j=1}^n |x_j - y_j|^p\right)^{1/p} & \text{if } 1 \leqslant p < \infty \\ \max_{j=1,\dots,n} |x_j - y_j| & \text{if } p = \infty \end{cases}$$

We will see that each d_p satisfies (i-iii). The proof of the triangle inequality depends on Minkowsky's inequality, which we will discuss later.

The distance d_2 is called **Euclidean distance**. For $n = 1$, $d_p(x, y) = |x - y|$ for all $1 \le p \le \infty$.

(ii) $X = C[a, b]$, the space of the functions which are continuous on the interval $[a, b]$, with the distance

$$\delta_p(f, g) = \begin{cases} \left(\int_a^b |f(x) - g(x)|^p\right)^{1/p} & \text{if } 1 \leqslant p < \infty \\ \max_{a \le x \le b} |f(x) - g(x)| & \text{if } p = \infty \end{cases}$$

(iii) If (X, d) is a metric space and $Y \subset X$, then $(Y, d|_Y)$ is a metric space. Here, $d|_Y$ is the restriction of d to Y. Unless it leads to ambiguities, we simply write $(Y, d) = (Y, d|_Y)$.

(iv) Let X be a set and defined $d(x, y) = 1$ in $x \neq y$, and $d(x, x) = 0$. Then, (X, d) is a metric space.

(v) If (X, d) and (Y, δ) are metric spaces, then $X \times Y$ becomes a metric space under the distance

$$D((x_1, y_1), (x_2, y_2)) = \max\{d(x_1, x_2), \delta(y_1, y_2)\}.$$

There are several useful ways to produce new metric spaces from old ones: (v) being just one of them. Also, metric spaces arise in a number of contexts. A notable class is that of *Riemannian manifolds*, which we will not discuss.

EXERCISE 1.2. *Show the triangle inequality for d_p when $p = 1, 2, \infty$.*

EXERCISE 1.3. *Draw in $\mathbb{R}^2$ $\{(x, y) : d_p((x, y), (0, 0)) = 1\}$ for $p = 1, 2, \infty$.*

EXERCISE 1.4. (i) *Show that $\lim_{p\to\infty} d_p(x, y) = d_\infty(x, y)$ and that $\lim_{p\to\infty} \delta_p(f, g) = \delta_\infty(f, g)$.*

(ii) *Let $X_p = \{f \in C(0, 1] : \delta_p(f, 0) < \infty\}$. Show that $\bigcap_{1 \leqslant p < \infty} X_p \supsetneqq X_\infty$.*

(iii) *Show that for all $1 \le p < \infty$ there is $f \in X_p$ such that $f \notin X_q$ for all $q > p$.*

A subset Y of a metric space (X, d) is *dense in* X if for all x in X and all $\epsilon > 0$ there is y in Y such that $d(x, y) < \epsilon$.

1.1.2. Normed linear spaces. An especially important family of metric spaces is that of the normed linear spaces. A *normed linear space* is a vector space X over $\mathbb{C}$ (or over $\mathbb{R}$) endowed with a *norm*: a function

$$\| \cdot \| \colon X \to [0, \infty),$$

satisfying the properties:

(i) $\|x\| = 0$ if and only if $x = 0$;

(ii) $\|\lambda x\| = |\lambda|\|x\|$ if $x \in X$ and $\lambda \in \mathbb{C}$;

(iii) $\|x + y\| \leq \|x\| + \|y\|$.

In practice, when we introduce a perspective norm on some concrete vector space X, we have to verify that it has a finite value for each x in X, and we have to verify the conditions above, (iii) being sometimes subtle.

A normed linear space becomes a metric space when endowed with the distance

$$d(x, y) := \|x - y\|.$$

In addition to (i-iii), this distance satisfies properties linking it with the algebraic structure.

(iv) $d(\lambda x, \lambda y) = |\lambda| d(x, y)$ if $x, y \in X$ and $\lambda \in \mathbb{C}$;

(v) $d(x + a, y + a) = d(x, y)$ if $x, y, a \in X$.

The map $x \mapsto \lambda x$ is a *(complex) homothety*, and $x \mapsto x + a$ is a *translation.*

Exercise 1.5. *Show that if a metric d on a vector space X satisfies (iv-v), then $\|x\| := d(x, 0)$ defines a norm, and that $d(x, y) = \|x - y\|$.*

The distances in examples (i) and (ii) in Section 1.1.1 come from a norm. For a continuous function $f\colon [a, b] \to \mathbb{C}$ we define:

$$\|f\|_{L^p} = \left(\int_a^b |f(x)|^p \right)^{1/p} \text{ for } 1 \leq p < \infty, \text{ the } L^p \textit{ norm},$$
$$\|f\|_u = \max_{x \in [a,b]} |f(x)| = \sup_{x \in [a,b]} |f(x)|, \text{ the } \textit{uniform norm.}$$

They make $(C[a, b], \|\cdot\|_{L^p})$ and $(C[a, b], \|\cdot\|_u)$ into normed linear spaces.

1.2. Complete metric spaces

A basic problem in applications is the convergence of a sequence of objects to an object. Consider for instance the convergence of an algorithm. This notion can be formalized as the convergence of a sequence of points to a point in a metric space. Often, however, the nature of the objects in the sequence is clear, but so is not that of the limiting object. Think of the approximation of $\sqrt{2}$ by means of decimal (or binary) numbers having finitely many digits. The notion of "familiar" objects converging to a "nonfamiliar", "ghost" one is encoded in the notion of Cauchy sequence. The conceptual tool to make "ghosts" into "real" objects is the completion of a metric space.

A sequence $\{x_n\}_{n=1}^\infty$ in (X, d) *converges* to the *limit* $a \in X$, $\lim_{n\to\infty} x_n = a$, if for all $\epsilon > 0$ there is $n(\epsilon) > 0$ such that if $n > n(\epsilon)$, then $d(x_n, a) \leqslant \epsilon$. The sequence $\{x_n\}_{n=1}^\infty$ is *Cauchy* if for all $\epsilon > 0$ there is $n(\epsilon) > 0$ such that if $n > n(\epsilon)$ and $j > 0$ one has $d(x_n, x_{n+j}) \leqslant \epsilon$.

EXERCISE 1.6. *Show that, if* $\lim_{n\to\infty} x_n = x$ *and* $\lim_{n\to\infty} y_n = x$ *in* (X, d)*, then* $\lim_{n\to\infty} d(x_n, y_n) = d(x, y)$ *exists in* $\mathbb{R}$ *in the usual sense.*

All convergent sequences are Cauchy, but the opposite implication generally fails. The space (X, d) is *complete* if all Cauchy sequences in it converge.

For instance, if $\mathbb{Q} \subset \mathbb{R}$ is the set of the rational numbers and $\{x_n\}_{n=1}^{\infty}$ is a sequence in $\mathbb{Q}$ converging to $\sqrt{2}$, then (with respect to the usual distance function $d(x, y) = |x - y|$) the sequence $\{x_n\}_{n=1}^{\infty}$ is Cauchy, but not convergent, in $\mathbb{Q}$. Familiar examples of complete metric spaces are $\mathbb{R}$ and $\mathbb{C}$, with respect to the distance $d(x, y) = |y - x|$.

1.2.1. The completion of a metric space. Any metric space can be canonically imbedded in a complete one.

THEOREM 1.1. *Let* (X, d) *be a metric space. Then there exists a complete metric space* $(\tilde{X}, \tilde{d})$*, the* **completion** *of* (X, d)*, and an injective map* $i \colon X \to \tilde{X}$ *such that:*

(i) $\tilde{d}(i(x), i(y)) = d(x, y)$;

(ii) $i(X)$ *is dense in* $\tilde{X}$.

The completion is unique in the following sense. For any other complete metric space (Z, δ) *endowed with an injection* $j \colon X \to Z$ *satisfying properties (i)* $\delta(j(x), j(y)) = d(x, y)$*, and (ii)* $j(X)$ *is dense in* Z*, there is a unique map* $F \colon \tilde{X} \to Z$ *which is a surjective isometry,* $\delta(F(\tilde{x}), F(\tilde{y})) = \tilde{d}(\tilde{x}, \tilde{y})$.

PROOF. We start with the construction of $(\tilde{X}, \tilde{d})$. Let C be the set of all Cauchy sequences in X and for $\{x_n\}, \{y_n\} \in C$ set

$$\{x_n\} \sim \{y_n\} \Leftrightarrow \lim_{n\to\infty} d(x_n, y_n) = 0.$$

The relation $\sim$ is an equivalence relation (check it!). Write $[\{x_n\}]$ for the equivalence class of the Cauchy sequence $\{x_n\}$, and let $\tilde{X} = C/\sim$, be the corresponding quotient space. Define

$$\tilde{d}([\{x_n\}], [\{y_n\}]) = \lim_{n\to\infty} d(x_n, y_n).$$

It is easy to see that $\tilde{d}$ is well defined, and that it defines a distance on $\tilde{X}$. (check that it is independent of the representatives). Finally, for $x \in X$ set $i(x) = [\{x_n = x\}]$, the class of the corresponding constant function. Properties (i) and (ii) are easily verified (exercise). Some work is needed to prove completeness, using a diagonal trick.

Let $\{\tilde{x}_n\}$, $\tilde{x}_n = [\{x_m^n\}_{m=1}^\infty]$, be a Cauchy sequence in $\tilde{X}$. For $k \geqslant 1$ select integer $n(k) > n(k-1)$ (if $n(k-1)$ was already selected) such that, for $n \geq n(k)$ and $j \geqslant 1$,

$$\frac{1}{2^k} \geqslant \tilde{d}(\tilde{x}_n, \tilde{x}_{n+j}) := \lim_{m\to\infty} d(x_m^n, x_m^{n+j}).$$

Select then $m(k) > m(k-1)$ such that, for $i = 0, 1$, $m \geqslant m(k)$, and $j \geqslant 1$,

$$(*) \quad d(x_m^{n(k+i)}, x_{m+j}^{n(k+i)}) \leqslant \frac{1}{2^k},$$

which we can ask because each $\{x_m^n\}_{m=1}^\infty$ is Cauchy, and

$$(**) \quad d(x_{m(k)}^{n(k)}, x_{m(k)}^{n(k+1)}) \leqslant \frac{2}{2^k},$$

which we can ask because $\frac{1}{2^k} \geqslant \tilde{d}(\tilde{x}_n, \tilde{x}_{n+j})$. Set then $a_k = x_{m(k)}^{n(k)}$.

The sequence $\{a_k\}_{k=1}^\infty$ is Cauchy in X:

$$\begin{aligned} d(a_k, a_{k+1}) &= d(x_{m(k)}^{n(k)}, x_{m(k+1)}^{n(k+1)}) \\ &\leqslant d(x_{m(k)}^{n(k)}, x_{m(k)}^{n(k+1)}) + d(x_{m(k)}^{n(k+1)}, x_{m(k+1)}^{n(k+1)}) \\ &\leqslant \frac{2}{2^k} + \frac{1}{2^k} \end{aligned}$$

by $(*)$ and $(**)$, and, by geometric sums, $d(a_k, a_{k+j}) \leqslant \frac{6}{2^k}$. Let $a = [\{a_k\}]$.

We want to prove that $\tilde{d}(\tilde{x}_{n(k)}, a) \to 0$. Using the fact that $\lim_{n\to\infty} d(x_n, y_n) = \lim_{m\to\infty} d(x_{j_m}, y_{k_m})$ for all subsequences $\{j_m\}$ and $\{k_m\}$ of the positive integers, provided the initial sequences $\{x_n\}$ and $\{y_n\}$ are Cauchy (check it), we have

$$\begin{aligned} \tilde{d}(\tilde{x}_{n(k)}, a) &= \lim_{l\to\infty} d(x_{m(l)}^{n(k)}, x_{m(l)}^{n(l)}) \\ &\leqslant \limsup_{l\to\infty} d(x_{m(l)}^{n(k)}, x_{m(k)}^{n(k)}) + \lim_{l\to\infty} d(x_{m(k)}^{n(k)}, x_{m(l)}^{n(l)}) \\ &\leqslant \frac{1}{2^k} + \frac{6}{2^k}. \end{aligned}$$

by $(*)$ and the fact that $\{a_k = x_{m(k)}^{n(k)}\}$ is Cauchy (the precise value 6 is indeed unimportant).

We now come to uniqueness. Let (Z, δ), j as in the hypothesis, and let $\tilde{x} = [\{x_n : \ n \geq 1\}]$ be an element in $\tilde{X}$, where $\{x_n\}$ is a Cauchy sequence in X. Then, $\{i(x_n) : \ n \geq 1\}$ is Cauchy in $\tilde{X}$ and $\{j(x_n) : \ n \geq 1\}$ is Cauchy in Z. Define $F(\tilde{x}) = \lim_{n\to\infty} j(x_n)$, which exists because Z is complete. The definition is well posed, since for two equivalent Cauchy sequences $\{x_n\}$, $\{y_n\}$ in Z we have

$$\delta(j(x_n), j(y_n)) = d(x_n, y_n) \to 0 \text{ as } n \to \infty,$$

hence, they have the same limit in (Z, δ). Moreover, for $\tilde{x} = [\{x_n\}]$ and $\tilde{y} = [\{y_n\}]$ in $\tilde{X}$,

$$\begin{aligned}
\delta(F(\tilde{x}, \tilde{y})) &= \delta(F([\{x_n\}]), F([\{y_n\}])) \\
&= \delta(\lim_{n\to\infty} j(x_n), \lim_{n\to\infty} j(y_n)) \\
&= \lim_{n\to\infty} \delta(j(x_n), j(y_n)) = \lim_{n\to\infty} d(x_n, y_n) = \lim_{n\to\infty} \tilde{d}(i(x_n), i(y_n)) \\
&= \tilde{d}(\tilde{x}, \tilde{y}).
\end{aligned}$$

The equalities from second to third, and third to fourth line, follow from Exercise 1.6 applied to δ and $\tilde{d}$, respectively.

The isometry $F\colon \tilde{X} \to Z$ is surjective. If $z \in Z$, by hypothesis there is a sequence $\{x_n\}$ in X such that $\delta(j(x_n), z) \to 0$, so that $\{x_n\}$ is Cauchy in X. Hence, $\{i(x_n)\}$ is Cauchy in $\tilde{X}$ and $\lim_{n\to\infty} F(i(x_n)) = \lim_{n\to\infty} j(x_n) = z$. □

1.2.2. Banach spaces. A normed linear space is a *Banach space* if it is complete with respect to the distance induced by the norm.

Let $(X, \|\cdot\|)$ be a normed linear space, and let $(\tilde{X}, \tilde{d})$ be its completion with respect to the distance $d(x, y) = \|x - y\|$ on X. We want to introduce on $(\tilde{X}, \tilde{d})$ a structure of normed, linear space, in such a way that $\tilde{X}$ becomes a Banach space. To this aim, we define algebraic operations between equivalence classes of Cauchy sequences,

$$[\{x_n\}] + [\{y_n\}] := [\{x_n + y_n\}], \ \lambda[\{x_n\}] := [\{\lambda x_n\}];$$

and of the norm:

$$\|[\{x_n\}]\| := \lim_{n\to\infty} \|x_n\|.$$

PROPOSITION 1.1. *Sum, multiplication times scalar, and norm are well defined, and make $\tilde{X}$ into a normed linear space. Moreover, if z, w belong to $\tilde{X}$, then*

$$\tilde{d}(z, w) = \|z - w\|,$$

so that, in particular, $(\tilde{X}, \|\cdot\|)$ is a Banach space. Also, the imbedding map $i\colon X \to \tilde{X}$ is linear.

PROOF. The statement can be split in a number of statements, whose proof is left to the reader.

(a) About the sum, we have to show that (i) $\{x_n + y_n\}$ is Cauchy if $\{x_n\}$ and $\{y_n\}$ are; (ii) that $[\{x_n + y_n\}] = [\{z_n + w_n\}]$ if $[\{x_n\}] = [\{z_n\}]$ and $[\{y_n\}] = [\{w_n\}]$.

(b) Similar statements must be verified for the product of a vector with a scalar.

(c) We have to verify that for $[\{x_n\}]$ in $\tilde{X}$, $\lim_{n\to\infty} \|x_n\|$ exists and does not depend on the particular representative.

(d) For z, w in $\tilde{X}$, $\tilde{d}(z, w) = \|z - w\|$.

(e) The imbedding i is linear.

□

EXERCISE 1.7. *Prove the assertions above.*

1.2.3. Metric and topology. In a metric space (X, d) the *open ball* with center $x \in X$ and radius r is $B(x, r) = \{y \in X : d(x, y) < r\}$. A subset O of X is *open* if for each of its points, it contains an open ball centered at it. More generally, a point a of a subset E of X is an *interior point* of E if it is contained in E together with a ball centered at a, so that E is open exactly when all its points are interior points. A *neighborhood* N of $x \in X$ is a subset of X containing an open ball centered at x. A subset F of X is *closed* if and only if $X \setminus F$ is open.

An important and useful fact is that, in metric spaces, many topological notions can be equivalently expressed in terms of sequences. A point $a \in X$ is a *limit point* of a subset F of X if each neighborhood of a intersects F.

PROPOSITION 1.2. *A subset F of X is closed if and only if it contains all its limit points. Equivalently, if and only if any sequence $\{x_n\}$ in F which has limit, has limit in F.*

PROOF. Suppose F is closed and $\{x_n\}$ is a sequence in F converging to some a in $X\setminus F$, which is open. Then there is $r > 0$ such that $B(a, r) \subseteq X\setminus F$. Thus, $d(x_n, a) \geq r$ for all n, which is absurd.

In the other direction, suppose that F is not closed. Hence, there is $a \in X \setminus F$ (which is not open) such that for all integer $n \geq 1$ there is $x_n \in F$ with $d(x_n, a) < 1/n$. Thus, $F \ni x_n \to a \notin F$. □

COROLLARY 1.1. *Let $F \subseteq X$, (X, d) being complete. If F is closed, then (F, d) is complete.*

PROOF. Let $\{x_n\}$ be a Cauchy sequence in F, hence in X. Then the limit $\lim_{n\to\infty} x_n = a$ exists in X. But F is closed, hence $a \in F$. □

Exercise 1.8. *Let (X,d) be a metric space.*

(i) *Let $\tau = \{O \subseteq X : O$ is open$\}$. Then, τ is a* **topology**:

(a) $X, \emptyset \in \tau$;

(b) *if* $O_1, \ldots, O_n \in \tau$, *then* $O_1 \cap \ldots \cap O_n \in \tau$;

(c) *if* $\{O_a\}_{a\in A}$ *is a family of elements of* τ, *then* $\cup_{a\in A} O_a \in \tau$.

(ii) *Moreover, the topology is* **Hausdorff**: *for $x \neq y$ in X there are disjoint open sets O_x, O_y containing x, y, respectively.*

(iii) *Furthermore, the topological space (X,τ) is* **first countable**: *for each x in X there is a countable family on neighborhoods $N_n(x)$ of x such that any other neighborhood N of x contains some $N_n(x)$.*

It is natural to ask whether for any first countable, Hausdorff topological space (X,τ) there is a metric d on X having τ as the class of its open sets. The answer is negative, but the good news is that a characterization of *metrizable* topological spaces exists (Smirnov's metrization theorem), and that a simple sufficient condition was proved by Urysohn. See e.g. the lecture notes Metrizability theorems by Marius Crainic.

It follows from (c) and the definition of closed set that, if $\{F_\alpha\}_{\alpha\in I}$ is a family of closed subsets of X, then $\cap_{\alpha\in I} F_\alpha$ is closed. For a subset E of X we define its *closure* to be

$$\overline{E} = \cap_{F\supseteq E,\ F \text{ closed}} F.$$

Exercise 1.9. *Let (X,d) be a metric space (but the assertions hold for any topological space).*

(i) *Show that $\overline{E} \supseteq E$, $\overline{E}$ is closed, and in fact it is the smallest closed subset containing E: if $H \supseteq E$ is closed, then $H \supseteq \overline{E}$.*

(ii) *Show that* $\overline{\overline{E}} = \overline{E}$.

(iii) *Show that* $\overline{E} = \{a \in X :\ \exists\{x_n\}$ *in E such that* $a = \lim_{n\to\infty} x_n\}$.

A function $f\colon X \to Y$ between metric spaces (X, d_X) and (Y, d_Y) is *continuous* at $a \in X$ if for all $\epsilon > 0$ there is $\delta = \delta(a,\epsilon) > 0$ such that

$$\text{if } d_X(x,a) \leq \delta, \text{ then } d_Y(f(x, f(a))) \leq \epsilon.$$

This is the same as requiring that for each neighborhoods $N(f(a))$ of $f(a)$ there exists a neighborhood $N(a)$ of a such that $f(N(a)) \subseteq N(f(a))$.

If f is continuous at all x in X, we simply say that it is continuous and write $f \in C(X,Y)$. We write $C(X)$ to denote the class of the continuous functions having values in $\mathbb{R}$ (or, when it is clear from the context, in $\mathbb{C}$).

Proposition 1.3. *Let $f\colon X \to Y$ a function between metric spaces.*

(a) *[Sequential definition of continuity at a] f is continuous at $a \in X$ if and only if for all $x_n \to a$ in X, we have that $f(x_n) \to f(a)$ in Y.*

(b) *[Topological characterization of continuity on X] f is continuous if and only if, for any O open in Y, $f^{-1}(O)$ is open in X.*

Proof. (a) [Only if] Let $\{x_n\}$ be a sequence converging to a in X, fix $\epsilon > 0$, and let $\delta > 0$ such that $d_Y(f(x), f(a)) \leq \epsilon$ if $d_X(x, a) \leq \delta$. For $n \geq n(\epsilon)$ we have that $d_X(x_n, a) \leq \delta$, hence that $d_Y(f(x_n), f(a)) \leq \epsilon$. [If] If f is not continuous at a, there is some $\epsilon_0 > 0$ such that, for all $n \geq 1$, there is x_n in X with $d_X(x_n, a) \leq 1/n$, yet $d(f(x_n), f(a)) \geq \epsilon_0$. This exhibits a sequence $\{x_n\}$ which fails the test of sequential continuity.

(b) [Only if] Let $a \in f^{-1}(O)$, let $\epsilon > 0$ be such that $B(f(a), \epsilon) \subseteq O$, and $\delta > 0$ such that for $d_X(x, a) < \delta$ one has that $d_Y(f(x), f(y))\epsilon$. Then, $B_X(a, \delta) \subset f^{-1}(O)$. [If] Let a be a point in X and $\epsilon > 0$. By hypothesis $f^{-1}(B(f(a), \epsilon))$ is open in X, hence, it contains a ball $B(a, \delta)$. □

Exercise 1.10. *Show that $f\colon X \to Y$ is continuous at $a \in X$ if and only if for all open $O \ni f(a)$ in Y there is $A \ni a$ open in X such that $f(A) \subseteq O$ (i.e. $A \subseteq f^{-1}(O)$).*

Preimages of open sets under continuous maps are open, but so is not for images.

Exercise 1.11. *Find a map $f\colon \mathbb{R} \to \mathbb{R}$ which is continuous, yet there is an open subset O in $\mathbb{R}$ such that $f(O)$ is not open.*

The basic properties of continuous, real valued functions of a real variable, continue to hold in the general framework of metric spaces, with the same proofs.

- If $f\colon X \to Y$ and $g\colon Y \to Z$ are maps between metric spaces, f is continuous at a, and g is continuous at $f(a)$, then $g \circ f$, their composition, is continuous at a.
- Constant functions are continuous.
- If X is a metric space and $f, g\colon X \to \mathbb{R}$ (or $\mathbb{C}$) are continuous at a, the $f + g$, $\lambda \cdot f$, $f \cdot g$ and f/g (if $g(a) \neq 0$) are continuous at a.

Exercise 1.12. *Prove these properties.*

1.2.3.1. *Bounded continuous functions as a metric space.* A subset $A \subseteq X$ of a metric space (X, d) is *bounded* if there is a number $R > 0$ such that for all x, y in A we have that $d(x, y) \leq R$. This is the same as asking that there are $a \in X$ and $Q > 0$ such that $d(a, x) \leq Q$ for all x in A. The set $C_b(X, Y)$ contains the functions in $C(X, Y)$ which are *bounded*, i.e. such that

$f(X)$ is bounded in (Y,δ). As in the case of continuous functions defined on an interval, we can define the *uniform distance* d_u in $C_b(X,Y)$ by

$$d_u(f,g) = \sup_{x\in X} \delta(f(x), g(x)).$$

Observe that here we use the supremum instead of the maximum, because the maximum might not exist. Below, we will study continuous functions on metric spaces in some detail. Here we just need the following.

THEOREM 1.2. *Let (X,d) and (Y,δ) be metric spaces. If (Y,δ) is complete, then $(C_b(X,Y), d_u)$ is complete.*

The proof consists of a simple a "3ϵ-argument".

PROOF. Suppose $\{f_n\}$ is a Cauchy sequence in $C_b(X,Y)$ and fix x in X. Since $\delta(f_n(x), f_{n+j}(x)) \le d_u(f_n, f_{n+j})$, we have that $\{f_n(x)\}$ is Cauchy in Y, hence, it converges to some $f(x)$ because Y is complete. We only have to show that $\lim_{n\to\infty} d_u(f_n, f) = 0$ and that $f \in C_b(X,Y)$. The first statement is clear,

$$\delta(f(x), f_n(x)) = \lim_{m\to\infty} \delta(f_m(x), f_n(x)) \le \limsup_{m\to\infty} d_u(f_m, f_n) \le \epsilon,$$

provided $n \ge n(\epsilon)$, because $\{f_n\}$ is Cauchy. Passing to sup on the left, $d_u(f, f_n) \le \epsilon$ if $n \ge n(\epsilon)$.

We now show continuity. Fix $\epsilon > 0$. Then,

$$\begin{aligned} \delta(f(x), f(y)) &\le \delta(f(x), f_n(x)) + \delta(f_n(x), f_n(y)) + \delta(f_n(y), f(y)) \\ &\le d_u(f, f_n) + \delta(f_n(x), f_n(y)) + d_u(f_n, f) \\ &\le \epsilon + \delta(f_n(x), f_n(y)) + \epsilon \\ &\quad \text{if } n \text{ is chosen to be greater than } n(\epsilon) \text{ as above} \\ &\le 3\epsilon \end{aligned}$$

if $d(x,y) \le \delta = \delta(\epsilon)$, since f_n is continuous at x. □

The first part of the proof can be formalized as a general statement.

EXERCISE 1.13. *Let X be a set, and (Y,δ) a complete metric space. Then, $B(X,Y)$, the set of the bounded functions $f\colon X \to Y$, is complete metric space with respect to the distance d_u.*

In particular,

COROLLARY 1.2. *$C_b(X,\mathbb{C})$ is Banach with respect to $\|\cdot\|_u$ and $B(X,\mathbb{C})$ is Banach with respect to the norm $\|f\|_u = \sup_{x\in X} |f(x)|$.*

In the case of the bounded functions, we used the label u (*uniform*). We did so to avoid any confusion with the L^∞ norm we will meet in measure theory, where the supremum is taken on "almost all" points in X, allowing a "negligeable" set of exceptions.

The second part of the proof shows that the following holds.

COROLLARY 1.3. *If (X, d) and (Y, δ) are metric spaces, $f_n \in C(X, Y)$ for $n \geq 1$, and $f_n \to f$ uniformly on X, then f is continuous.*

1.2.4. Banach Fixed Point Theorem. It is an old and fruitful idea writing an equation, e.g.

$$x^2 - x - 1 = 0, \tag{1.2.1}$$

in the form of a fixed point problem,

$$x = f(x) = 1 + \frac{1}{x}, \tag{1.2.2}$$

and observe that (1.2.2) implies that

$$x = 1 + \frac{1}{x} = 1 + \frac{1}{1 + \frac{1}{x}} = 1 + \frac{1}{1 + \frac{1}{1+\frac{1}{x}}} = \dots \tag{1.2.3}$$

suggesting to turn the equation into a recursive scheme,

$$x_{n+1} = f(x_n) = 1 + \frac{1}{x_n}, \ n \geq 0. \tag{1.2.4}$$

If we are smart enough in choosing x_0, we might hope that $\lim_{n\to\infty} x_n = a \in \mathbb{R}$, and the continuity of f and (1.2.2) imply that $x = a$ is one of two solutions to (1.2.1). Of course, there are problems: (i) there are many ways to turn the equation into a fixed point problem ($x = g(x) = x^2 - 1$ would work as well); (ii) we have to choose x_0 in such a way that x_n belongs to the domain of f for all $n \geq 0$; (iii) it can be shown that for almost all admissible choices of x_0 the recursive scheme (1.2.4) will converge to the positive solution of (1.2.1) (which is an *attractive fixed point of f*), unless we are so lucky to choose $x = \frac{1-\sqrt{5}}{2}$, the negative solution (which is a *repulsive fixed point of f*). The study of the *dynamics* of functions $f\colon X \to X$ is an enormously vast subject. We consider here one of its simplest instances.

Let (X, d) be a metric space. A map $f\colon X \to X$ is a *contraction* if $d(f(x), f(y)) \leq \lambda d(x, y)$ for some $0 \leq \lambda < 1$ independent of x, y in X. A contraction is continuous, in the sense that if $\lim_{n\to\infty} x_n = x$, then $\lim_{n\to\infty} f(x_n) = f(x)$. In fact, if $d(x_n, x) \to 0$, then

$$d(f(x_n), f(x)) \leq \lambda d(x_n, x) \to 0.$$

THEOREM 1.3. *[Banach Fixed Point Theorem] Let* $f: X \to X$ *be a contraction on a complete metric space* (X,d). *Then, there exists a unique point* x *in* X *such that* $f(x)=x$.

PROOF. Uniqueness is clear, since for two fixed points x,y we have $d(x,y) = d(f(x),f(y)) \le \lambda d(x,y)$, so $d(x,y)=0$. About existence, let x_0 be any point in X and inductively define $x_{n+1}=f(x_n)$. We want to show that $\{x_n\}_{n=0}^{\infty}$ is a Cauchy sequence. In fact, for $n \ge 0$ and $j \ge 1$,

$$\begin{aligned} d(x_{n+j},x_n) &= d(f(x_{n-1+j}),f(x_{n-1})) \\ &\le \lambda d(x_{n-1+j},x_{n-1}) \\ &\le \lambda^n d(x_j,x_0) \\ &\le \lambda^n[d(x_j,x_{j-1})+\cdots+d(x_1,x_0)] \\ &\le \lambda^n[\lambda^{j-1}+\cdots+\lambda+1]d(x_1,x_0) \\ &\le \frac{\lambda^n}{1-\lambda}d(x_1,x_0) \to 0 \end{aligned}$$

as $n\to\infty$. Hence, $x=\lim_{n\to\infty}x_n$ exists in X, by completeness. Also,

$$x = \lim_{n\to\infty} x_n = \lim_{n\to\infty} f(x_{n-1}) = f(\lim_{n\to\infty} x_{n-1}) = f(x).$$

□

The examples $T(x)=x+1$ on $\mathbb{R}$, or $T(e^{it})=e^{i(t+1)}$ on the *torus* (or unit circle) $\mathbb{T} := \{e^{it}: \ t\in[0,2\pi)\} \subset \mathbb{C}$, which are 1-Lipschitz and have no fixed point, show that the hypothesis $0\le\lambda<1$ can not be relaxed. The starting example, (1.2.2), does not wholly fall within the scope of Banach's theorem: we have *two* fixed points. However, after restricting the domain of the function $f(x)=1+\frac{1}{x}$, Banach's theorem applies.

EXERCISE 1.14. *Show that there exist* $p<\frac{1+\sqrt{5}}{2}<q$ *such that* $f(x)=1+\frac{1}{x}$ *is a contraction of* $[p,q]$ *into itself. Find two explicit values of* p *and* q.

The hypothesis in Banach's theorem can be usefully relaxed.

EXERCISE 1.15 (Banach's fixed point theorem for power-contraction). *Let* $(X.d)$ *be a complete metric space, and let* $f: X\to X$ *be a continuous map such that* $f^{\circ n} = f\circ\ldots\dot{f}$, *the composition with itself* n *times, is a contraction. Then,* f *has a unique fixed point in* X. **Hint.** *Use continuity to show that the fixed point of* $f^{\circ n}$ *is a fixed point of* f.

1.3. Compact sets

Basically, and imprecisely, a metric space (X,d) (or, more generally, a subset K of X) is compact if it "looks like" a finite set at all metric scales. The following properties of a set X are obviously equivalent:

(a) X is finite;

(b) any sequence $\{x_n\}_{n=1}^{\infty}$ in X takes a value a infinitely often ("pigeon's principle");

(c) X has finitely many subsets;

(d) the intersection of nested, nonempty subsets of X is nonempty.

In this section we see four useful equivalent characterizations of compact sets which somehow mirror (a-d): that a metric space (X, d) is sequentially compact (b); that the intersection of nested, nonempty closed sets is nonempty (d); that it is totally bounded and complete (a); that it is compact in the topological sense (c). It is nice that many simple properties of finite sets and functions defined on them have counterparts in the context of compact metric spaces. For instance,

(i) "if $f\colon X \to Y$ and X is finite, then $f(X)$ is finite in Y (but Y might be finite, and X infinite) becomes "if $f\colon X \to Y$ is a continuous function between metric spaces and X is compact, than $f(X)$ is compact in Y" (but Y might be compact, and X not);

(ii) "if $f\colon X \to \mathbb{R}$ and X is finite, then f has a maximum value" becomes Weierstrass Theorem.

(iii) "if $f\colon X \to \mathbb{R}$, the value at any point x is well defined" might be read as "if $f\colon X \to Y$ is a continuous function between metric spaces and X is compact, then f is uniformly continuous".

Indeed, this intuition, like all intuitions, has to be used with care. Any uniformly bounded sequence of functions $f_n\colon X \to \mathbb{R}$ defined on a finite set X has a subsequence converging to a function $f\colon X \to \mathbb{R}$. If we replace "finite" by "compact" and "function" by "continuous function", however, the statement is false. We will see in the next section that we also need "equicontinuity", a notion which is hard to detect without leaving the finite set case.

1.3.1. Equivalent definitions of compactness. The easiest to manipulate notion of compactness says that no sequence in a compact can "fade away". A metric space (X, d) is *sequentially compact* if all sequences $\{x_n\}_{n=1}^{\infty}$ of X admit a converging subsequence $\{x_{n_j}\}_{j=1}^{\infty}$, $\lim_{j\infty} x_{n_j} = a$ for some a in X. We say that (X, d) is *totally bounded* if for all $\epsilon > 0$ there a finite ϵ-net: a sequence $a_1, \ldots, a_m$ such that $X = \cup_{j=1}^{m} B(a_j, \epsilon)$.

LEMMA 1.1. *Let (X, d) be totally bounded and $Z \subset X$. Then, (Z, d) is totally bounded.*

PROOF. Fix $\epsilon > 0$. By assumption, $X = B(x_1, \epsilon) \cup \cdots \cup B(x_n, \epsilon)$, hence, $Z = \cup_{j \in J = \{j:\ B(x_j,\epsilon)\cap Z \neq \emptyset\}} (B(x_j, \epsilon) \cap Z)$. For each ball in the last union, pick $z_j \in B(x_j, \epsilon) \cap Z$, so that, for $x \in B(x_j, \epsilon)$, $d(z_j, x) < 2\epsilon$. Thus, $\{z_j :\ j \in J\}$ is a 2ϵ-net in Z. □

THEOREM 1.4. *A metric space is sequentially compact if and only if it is totally bounded and complete.*

PROOF. Suppose X is sequentially compact and let $\{x_n\}$ be a Cauchy sequence. Then it has a converging subsequence $\{x_{n_j}\}$ converging to some a in X and it is easy to see that the whole sequence $\{x_n\}$ converges to a, so X is complete.

Suppose that X is not totally bounded: there is an $\epsilon > 0$ such that no finite family of balls $B(x_j, \epsilon)$ covering X. Pick $x_1 \in X$. There is $x_2 \in X$ such that $d(x_1, x_2) \geq \epsilon$, otherwise X is covered by $B(x_1, \epsilon)$ alone. Inductively we find $x_1, x_2, \ldots, x_n, \ldots$ such that $d(x_n, x_j) \geq \epsilon$ for $j < n$: this sequence does not have converging subsequences.

Viceversa, suppose that X is totally bounded, and let $S_0 = \{x_n\}$ be a sequence in X and consider a finite 1/2-net. There is some a_1 in the net so that $B(a_1, 1/2)$ contains an infinite subsequence $S_1 = \{x_j^1\}$ of S_0. Set $y_1 = x_1^1$. Since $B(a_1, 1/2)$ is itself totally bounded, it has a finite $1/2^2$-net, and so there is a_2 in the net so that $B(a_2, 1/2^2)$ contains an infinite subsequence $S_2 = \{x_j^2\}$ of S_1: choose in it y_2 which comes after y_1 in the original sequence. Iterating, we find a subsequence $\{y_m\}_{m=1}^{\infty}$, which satisfies

$$\begin{aligned} d(y_{m-1}, y_m) &\leq d(y_{m-1}, a_{m-1}) + d(a_{m-1}, y_m) \\ &\leq 1/2^{m-1} + 1/2^{m-1} = 4/2^m. \end{aligned}$$

It follows that $\{y_m\}$ is a Cauchy sequence,

$$d(y_{n+j}, y_n) \leq 4(1/2^n + \ldots 1/2^{n+j-1}) \leq 8/2^n,$$

which converges to some b in X because X is complete. □

A subset A of a metric space (X, d) is *sequentially precompact* if its closure $\overline{A}$ is compact.

EXERCISE 1.16. *Assume that (X, d) is complete. The subset A of X is sequentially precompact (each sequence in A has a subsequence converging to some point of X) if and only if it is totally bounded.*

THEOREM 1.5. *A metric space (X, d) is sequentially compact if and only if, given a decreasing sequence $F_n \subseteq F_{n-1}$ of closed subsets of X, the family $\{F_n\}_{n=1}^{\infty}$ has nonempty intersection.*

Proof. Suppose X is sequentially compact, and pick a sequence $\{x_n\}$ with $x_n \in F_n$. It ha a subsequence $\{x_{n_j}\}$ converging to some a in X. Since $\{x_{n_j}\}$ definitely lies in F_m, for each m, and F_m is closed, then a belongs to all sets of the sequence.

Viceversa, suppose that the intersection property holds, and let $\{x_n\}$ be a sequence in X. Set $F_m = \overline{\{x_n : \ n \geq m\}} \subseteq F_{m-1}$. By hypothesis there is some a contained in the intersection of the F_m's. Since $a \in \overline{F_1}$, we can choose n_1 such that $d(x_{n_1}, a) < 1/2$. Then, as $a \in \overline{F_{n_1}}$, we can choose $n_2 > n_1$ such that $d(x_{n_2}, a) < 1/2^2$. Iterating, we find a subsequence $\{x_{n_j}\}$ converging to a. □

Lemma 1.2. *[Lebesgue Lemma] Let (X, d) be sequentially compact, and let $X = \cup_{\alpha \in I} A_\alpha$ be a covering by open sets. Then, there is $\epsilon > 0$ such that, for all x in X, $B(x, \epsilon) \subseteq A_\alpha$ for some α.*

Proof. Suppose the thesis does not hold, and construct a sequence as follows. For $n \geq 1$, find x_n such that $B(x_n, 1/2^n)$ is not contained in any A_α. If X were sequentially compact, a subsequence $\{x_{n_j}\}$ would converge to some a in X: $\epsilon_j = d(x_{n_j}, a) \to 0$ as $j \to \infty$. By the triangle inequality, $B(a, \epsilon_j + 1/2^{n_j}) \supseteq B(x_{n_j}, 1/2^{n_j})$, which is not contained in any A_α. But this means that a is not contained in any A_α (if it were, it would be together with an open ball centered at it), hence the A_α's do not cover X. □

A metric space (X, d) is *compact* if for every open covering $\{A_\alpha : \ \alpha \in I\}$ of X, $X = \cup_{\alpha \in I} A_\alpha$, there is a finite subcover $\{A_{\alpha_i} : \ i = 1, \ldots, n\}$, $X = \cup_{i=1}^n A_{\alpha_i}$.

Theorem 1.6. *A metric space (X, d) is sequentially compact if and only if it is compact.*

Proof. If X be compact, we show that it has the intersection property, hence that it is sequentially compact. By contradiction, let $\{F_n : \ n \geq 1\}$ be a decreasing sequence of nonempty closed subsets of X and suppose that their intersection is empty. Then, $G_n = X \setminus F_n$ defines an increasing sequence of open subsets of X, and

$$\cup_n G_n = \cup_n (X \setminus F_n) = X \setminus (\cap_n F_n) = X.$$

By compactness, for some n, $X = G_1 \cup \cdots \cup G_n = G_n = X \setminus F_n$, hence F_n is empty, contradicting our assumption.

Viceversa, suppose X is sequentially compact and consider an open cover $\{A_\alpha : \ \alpha \in I\}$. By total boundedness, for all $\epsilon > 0$ there are points $x_1, \ldots, x_n$ such that $B(x_1) \cup \cdots \cup B(x_n, \epsilon) = X$. On the other hand, if ϵ is chosen as in Lebesgue Lemma, there are A_{α_j} in the cover so that $B(x_j, \epsilon) \subseteq A_{\alpha_j}$. Hence, $X = A_{\alpha_1} \cup \cdots \cup A_{\alpha_j}$. □

EXERCISE 1.17. *Prove Weierstrass Theorem. If $f \in C(X, \mathbb{R})$ and X is compact, then f has maximum and minimum on X.*

EXERCISE 1.18. *Show that any compact metric space is separable: it has a countable, dense subset.*

EXERCISE 1.19. *Prove that the image of a compact set under a continuous map between metric spaces is compact. In fact, you might provide a different proof for each of the four equivalent definitions of compactness we have surveyed.*

1.3.2. The Cantor set. The *Cantor set* is the simplest example of a "fractal": a set having a rich structure, but which is at the same time highly irregular. In the case of the Cantor set, the irregularity consist in the fact that it is totally disconnected; but its structure is rich enough to contain copies similar to itself at any small scale. These properties make it the standard example of what "decent" sets which are intermediate between a discrete family of points and a union of intervals might look like. We will meet it again when we discuss measures on the real line.

The *Cantor set* $C \subset \mathbb{R}$ is defined through the following algorithm, which is best understood by drawing pictures. Let $C_0 = I^0 = [0, 1]$. Removing the "middle third" $(1/3, 2/2)$ from it, we are left with $C_1 = I_0^1 \cup I_1^1$, where $I_j^1 = [2j/3, 2j/3+1/3]$. Remove now the middle thirds from I_0^1 and I_1^1. What is left is $C^2 = \cup_{j_1,j_2 \in \{0,1\}} I_{j_1 j_2}^2$, where

$$I_{j_1 j_2}^2 = \left[\frac{2j_1}{3} + \frac{2j_2}{3^2}, \frac{2j_1}{3} + \frac{2j_2}{3^2} + \frac{1}{3^2}\right].$$

The expression $j_1 j_2$ denotes here a binary string, which might be 00, 01, 10, 11.

We iterate the procedure, and obtain $C_0 \supset C_1 \supset \dots C_n$, where

$$C_n = \cup_{(j_1,\dots,j_n) \in \{0,1\}^n} I_{j_1 j_2 \dots j_n}^n, \tag{1.3.1}$$

with

$$I_{j_1 j_2 \dots j_n}^n = \left[\sum_{l=1}^n \frac{2j_l}{3^l}, \sum_{l=1}^n \frac{2j_l}{3^l} + \frac{1}{3^n}\right]. \tag{1.3.2}$$

The Cantor set is $C = \cap_{n=1}^\infty C_n$.

We denote $\mathbb{N}_* = \mathbb{N} \setminus \{0\} = \{1, 2, \dots, n, n+1, \dots\}$.

LEMMA 1.3. *The map $\psi : \{0,1\}^{\mathbb{N}_*} \to C$ defined by*

$$\psi\left(\{j_n\}_{n=1}^\infty\right) = \cap_{n=1}^\infty I_{j_1 j_1 \dots j_n}^n, \tag{1.3.3}$$

is a bijection. In particular, C is uncountable.

Moreover, the map ψ can be written in the form

$$\psi\left(\{j_n\}_{n=1}^{\infty}\right) = \sum_{n=1}^{\infty} \frac{2j_n}{3^n}. \tag{1.3.4}$$

Elements in $\{0,1\}^{\mathbb{N}_*}$ might be interpreted as infinite binary strings $j_1 j_2 \dots j_n \dots$.

PROOF. Given a binary string $j_1 j_2 \dots$, $\cap_{n=1}^{\infty} I^n_{j_1 j_1 \cdots j_n} \neq \emptyset$, and it can not contain more then one element because the diameter of $I^n_{j_1 j_1 \cdots j_n}$ is $1/3^n \to 0$ as $n \to \infty$. Different binary strings clearly produce disjoint intersection sets, hence the map is injective. We show surjectivity. Given $\zeta \in C$, for each $n \geq 1$ there is $I^n_{j_1 \cdots j_n}$ containing it, and $I^n_{j_1 \cdots j_n} \supset I^n_{j_1 \cdots j_n j_{n+1}}$, so we have a sequence $j_1 \dots j_n \dots$ such that $\zeta \in \cap_n I^n_{j_1 \cdots j_n}$, $\psi(j_1 \dots j_n \dots) = \zeta$.

The set of the infinite binary strings is uncountable, hance so is C.

The expression (1.3.4) immediately follows from (1.3.2) and (1.3.3). □

EXERCISE 1.20. *The set of the endpoints of the intervals $I^n_{j_1 \cdots j_n}$ is countable: "most" of the points of the Cantor set are not endpoints of intervals! In fact, endpoints of intervals correspond to binary strings whose digits which are definitely* 0 *or* 1.

PROPOSITION 1.4. *The Cantor set is (i) compact; (ii) any of its points is an accumulation point for it; (iii) totally disconnected.*

PROOF. (i) C is the intersection of compact sets, hence it is compact. (ii) If $\zeta = \psi(j_1 j_2 \dots)$, then the set of the endpoints of the intervals $I^n_{j_1 \cdots j_n}$ contains infinitely many points different from ζ, which belong to C and that can be arranged in a sequence converging to ζ. (iii) Suppose $\zeta < \xi$, and let n be largest so that both belong to the same connected component of C_n. The interval $[\zeta, \xi]$ is not contained in C_{n+1}, hence in C. □

1.4. Continuous functions

1.4.1. Spaces of continuous functions. Let $\{f_n\}$ be a sequence of functions on a set $f_n \colon X \to Y$, with X a set and (Y, d_Y) a metric space. We say that $f_n \to f$ *uniformly* if for all $\epsilon > 0$ there is $n(\epsilon) > 0$ such that $d_Y(f_n(x), f(x)) \leq \epsilon$ for all x in X and $n > n(\epsilon)$. With notation introduced earlier,

$$\lim_{n\to\infty} \delta_\infty(f_n, f) = 0.$$

If $Y = \mathbb{C}$, endowed with the Euclidean metric, We can phrase uniform convergence in terms of the uniform norm. A sequence $\{f_n\}$ of complex valued functions on a set X converges uniformly to a function f if and only if $\lim_{n\to\infty} \|f - f_n\|_u = 0$.

A function $f\colon X \to Y$ between metric spaces (X, d_X) and (Y, d_Y) is *uniformly continuous* if for all $\epsilon > 0$ there is $\delta > 0$, only depending on ϵ, such that $d_Y(f(x), f(y)) \le \epsilon$ whenever $d_X(x, y) \le \delta$.

THEOREM 1.7. *Let $f\colon K \to Y$ be continuous on K, a compact metric space. Then, f is uniformly continuous on K.*

PROOF. Fix $\epsilon > 0$ and for each x in K find $\delta_x > 0$ such that if $d_X(x, y) < 3\delta_x$, then $d_Y(f(x), f(y)) < \epsilon$. By compactness there exist $x_1, \dots, x_n$ such that $K = \cup_{j=1}^n B(x_j, \delta_{x_j})$. Let $\delta = \min\{\delta_{x_1}, \dots, \delta_{x_n}\}$.

Let $x, y \in K$, and find x_1 such that $d_X(x, x_1) < \delta_{x_1}$ and $d_X(y, x_2) < \delta_{x_2}$. If $d_X(x, y) < \delta$, then $d_X(x_1, x_2) < \delta_{x_1} + \delta + \delta_{x_2} < 3\max\{\delta_{x_1}, \delta_{x_2}\}$, so that $d_Y(f(x_1), f(x_2)) < \epsilon$. Hence,

$$\begin{aligned} d_Y(f(x), f(y)) &\le d_Y(f(x), f(x_1)) + d_Y(f(x_1), f(x_2)) + d_Y(f(x_2), f(y)) \\ &\le \epsilon + \epsilon + \epsilon. \end{aligned}$$

□

The space $C_0(X, \mathbb{R})$ is the space of the continuous functions f such that, for all $\epsilon > 0$, there is a compact set K_ϵ such that $|f(x)| < \epsilon$ for all $x \in X \setminus K_\epsilon$.

EXERCISE 1.21. *If $f \in C_0(X, \mathbb{R})$, then f is uniformly continuous.*

EXERCISE 1.22. *If (Y, d_Y) is complete, then the space $(C_0(X, Y), \delta_\infty)$ is complete.* **Hint.** *It suffices to show that the uniform limit of functions in $C_0(X, Y)$ belongs to $C_0(X, Y)$, then using the fact that a closed set of a complete space is complete.*

EXERCISE 1.23. *Show by examples that the previous properties do not hold if we just require pointwise convergence.*

1.4.2. Some spaces of complex valued continuous functions. Let (X, d) be a metric space. We define here some spaces of real (or complex) valued functions on X.

(i) We have already met $C_b(X)$, the space of the *bounded* continuous functions, endowed with the *uniform norm* $\|f\|_u = \sup_{x \in X} |f(x)|$.

(ii) By narrowing the definition just above, we have introduced $C_0(X) \subseteq C_b(X)$, the space of the functions $f : X \to \mathbb{C}$ which *vanish at infinity*: for all $\epsilon > 0$ there is a compact set K in X such that $|f(x)| \le \epsilon$ if $x \in X \setminus K$.

(iii) Recall that the *support* of a function $f : X \to \mathbb{C}$ is

$$\mathrm{supp}(f) = \overline{\{x \in X : f(x) \neq 0\}}.$$

The space $C_c(X)$ contains those $f \in C(X)$ which have compact support in X.

It is clear that $C_c(X) \subseteq C_0(X) \subseteq C_b(X)$.

EXERCISE 1.24. (i) *Show that X is compact if and only if $1 \in C_c(X)$, if and only if $1 \in C_0(X)$.*

(ii) *Find $f \in C_b((0,1))$ which is not uniformly continuous.*

(iii) *Show that real valued functions in $C_0(X)$ attain maximum and minimum on X.*

After we prove Urysohn lemma, we will see that $C_c(X)$ is in fact *dense* in $C_0(X)$.

1.5. Compactness: equicontinuity and the Ascoli-Arzelà Theorem

A family $\mathcal{F}$ of functions in $C((X, d_X), (Y, d_Y))$ is *equicontinuous* if for all x in X and $\epsilon > 0$ there is $\delta > 0$ such that for all f in $\mathcal{F}$ and all y in X, if $d_X(x, y) \leq \delta$, then $d_Y(f(x), f(y)) \leq \epsilon$. It is *uniformly equicontinuous* if δ is independent of x.

THEOREM 1.8. *[Ascoli-Arzelà] Let $\mathcal{F}$ be a uniformly equicontinuous family of functions $f: X \to Y$ from a compact space (X, d_X) to a compact space (Y, d_Y). Then $\mathcal{F}$ is precompact in $C(X, Y)$.*

In particular, if $\{f_n\}$ is a sequence in $\mathcal{F}$, then there exists a subsequence $\{f_{n_k}\}$ which converges uniformly on X.

PROOF. We show that $\mathcal{F}$ is totally bounded in the uniform norm.

Fix $\epsilon > 0$ and let $\delta = \delta(\epsilon) > 0$ as in the definition of equicontinuity. Cover X by finitely many balls of radius δ, and Y by finitely many balls of radius ϵ,

$$\begin{aligned} X &= B(x_1, \delta) \cup \ldots B(x_M, \delta), \\ Y &= B(y_1, \epsilon) \cup \ldots B(y_N, \epsilon). \end{aligned}$$

To each $f \in \mathcal{F}$ we associate a map $\varphi_f: \{1, \ldots, M\} \to \{1, \ldots, N\}$ as follows. For $i = 1, \ldots, M$, $f(x_i)$ belongs to some ball $B(y_j)$. Choose one, and let $\varphi_f(i) = j$.

Suppose that $\varphi_f = \varphi_g = \varphi$, and let $x \in X$ and suppose that $x \in B(x_i, \delta)$. Then,

$$\begin{aligned} d_Y(f(x), g(x)) &\leq d_Y(f(x), f(x_i)) + d_Y(f(x_i), g(x_i)) + d_Y(g(x_i), g(x)) \\ &\leq \epsilon + d_Y(f(x_i), g(x_i)) + \epsilon \\ &\leq 2\epsilon + d_Y(f(x_i), y_{\varphi(i)}) + d_Y(y_{\varphi(i)}, g(x_i)) \\ &< 4\epsilon. \end{aligned}$$

Thus, $\mathcal{F}$ is covered by a number of 4ϵ-balls in the uniform metric which does not exceed the number of functions $\varphi: \{1, \ldots, M\} \to \{1, \ldots, N\}$, i.e. no more that N^M of them. ☐

1.5.1. Spaces of functions defined by their derivatives. We consider here some spaces of real (or complex) valued functions defined on the interval $[0,1]$, endowed with the uniform norm.

(i) $C[0,1]$ is the space of the continuous functions.

(ii) $\mathrm{Lip}[0,1]$ is the space of the *Lipschitz* functions,

$$|f(x) - f(y)| \leq L|x - y|,$$

where $L \geq 0$. The *Lipschitz norm* of f is

$$\|f\|_{\mathrm{Lip}} = \max\left(\sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|}, \sup_x |f(x)|\right).$$

(iii) $C^1[0,1]$ is the space of the differentiable functions with continuous derivative, normed by

$$\|f\|_{C^1} = \max\{\|f\|_u, \|f'\|_u\}.$$

Clearly $C^1[0,1] \subset \mathrm{Lip}[0,1] \subset C[0,1]$.

Exercise 1.25. (i) *Let $B = \{f \in C[0,1] : \|f\|_u \leq 1\}$. Show that B is not compact in $C[0,1]$.*

(ii) *Let $B_1 = \{f \in C^1[0,1] : \|f\|_{C^1} \leq 1\}$. Show that B is not compact in $C^1[0,1]$.*

(iii) *Show that $C^1[0,1]$ is complete with respect to $\|\cdot\|_{C^1}$.* **Hint.** *You have to show that if $f_n \to f$ and $f_n' \to g$ uniformly on $[0,1]$, then $g = f'$. You might use the fundamental theorem of calculus, or Lagrange mean theorem. At the end of the proof, you will realize that you just need the assumptions $f_n(a) \to b$ at some point $a \in [0,1]$ and $f_n' \to g$ uniformly on $[0,1]$ to deduce that there is $f \in C^1[0,1]$ such that $f' = g$ and $f_n \to f$ uniformly (which is a stronger result).*

(iv) *Use the Ascoli-Arzelà theorem to show that B_1 is precompact in $C[0,1]$.*

(v) *Show that B_1 is not closed in $C[0,1]$.* **Hint.** *Find an explicit example of a sequence $\{f_n\}$ in B' which converges in the norm of C to a Lipschitz function which is not in C^1.*

(vi) *Show that the closure of B_1 in $C[0,1]$ is contained in the unit ball of Lip$[0,1]$.*

(vii) *Use Ascoli-Arzelà therem to show that the unit ball of Lip$[0,1]$ is closed in the unit ball of $C[0,1]$.*

In fact, one could prove a stronger version of (vi). See theorem 3.9.

1.6. Continuous functions with prescribed properties

Our function spaces are defined by imposing conditions, hence "rigidity", on the functions belonging to them. Sometimes it is not clear if these spaces contain many functions, and there have been cases where a function space was studied, before it was found that it only contained few functions, sometimes just the zero constant. In applications as in theory, it is often important to prove that the functions in the space can perform some tasks. For instance, assuming specified values at specified points. We will see below some results of this kind, which will be extensively used later on.

1.6.1. Urysohn Lemma. Let (X, d) be a metric space and $E \subseteq X$, nonempty. Define

$$d_E(x) = d(x, E) := \inf\{d(x, y) : \ y \in E\},$$

the *distance function* associated to E.

PROPOSITION 1.5. *If K is compact in X, then the infimum is achieved.*

PROOF. Pick $\{y_n\}$ in K so that $d(y_k, x) \to d(K, x)$. Find a converging subsequence $y_{n_j} \to y \in K$. Then, $d(x, y) = \lim_j d(x, y_{n_j}) = d(x, K)$. □

EXERCISE 1.26. (i) *Suppose that E is closed. Show that $d_E(x) = 0$ if and only if $x \in E$.*

(ii) *Let $A \subseteq X$ and let $\overline{A}$ be its closure. Show that $d_A = d_{\overline{A}}$.*

THEOREM 1.9. *If $E \subseteq X$ and $x, y \in X$, then $|d_E(x) - d_E(y)| \leq d(x, y)$. In particular, d_E is continuous.*

PROOF. For all z in E we have:

$$d(x, E) \leq d(x, z) \leq d(x, y) + d(y, z),$$

and passing to inf on z, $d(x, E) \leq d(x, y) + d(y, E)$. Exchange the roles of x and y to obtain the desired inequality. □

THEOREM 1.10. *[Urysohn Lemma in Metric Spaces] Let B, C be closed, disjoint subsets of X, not both empty. Then, there is $f \in C(X, [0, 1])$ such that $f(x) = 0$ for $x \in B$ and $f(x) = 1$ for $f \in C$.*

PROOF. If both sets are nonempty, set $f(x) = \frac{d(x,B)}{d(x,B)+d(x,C)}$. It has all desired properties provided the denominator does not vanish. But if $d(x, B) = 0 = d(x, C)$, then $x \in B \cap C$, contradiction. □

1.6.2. Locally compact spaces and partitions of unity. A metric space (X, d) is *locally compact* if each point x has a compact neighborhood K_x.

Lemma 1.4. *Show that this is the same as asking that for each x there is $r_x > 0$ such that $\overline{B(x,r)}$ is compact for all $0 < r < r_x$.*

Proof. Suppose X is locally compact and let K be one of its compact neighborhoods, $x \in B(x,r) \subseteq \overline{B(x,r)} \subseteq K$ for some $r > 0$. Since $\overline{B(x,r)}$ is a closed subset of a compact, it is compact. □

Let K and V be respectively a compact and an open subset of X. We write $f \prec V$ if $f \in C_c(X)$ has support $\text{supp}(f) = \overline{\{x \in X : f(x) \neq 0\}}$ contained in V, and $0 \le f(z) \le 1$ on X; and $K \prec f$ if $f \in C_c(X)$ satisfies $f(x) = 1$ for $x \in K$ and $0 \le f(z) \le 1$ on X.

Lemma 1.5. *Let $K \subseteq V$, K compact and V open in a locally compact metric space. Then, there exists U open with $\overline{U}$ compact, such that $K \subseteq U \subseteq \overline{U} \subseteq V$.*

Proof. For each x in K choose B_x, an open ball with compact closure $\overline{B_x} \subseteq V$ One can cover K by finitely many such balls,

$$K \subseteq B(x_1, r_1) \cup \dots \cup B(x_n, r_n) =: U \subseteq \overline{U} = \overline{B(x_1, r_1) \cup \dots \cup B(x_n, r_n)} \subseteq V,$$

and $\overline{U}$ is compact. □

Theorem 1.11. *[Urysohn Lemma in locally compact metric spaces] Let $K \subseteq V$, K compact and V open. Then, there exists $f \in C_c(X)$ such that $K \prec f \prec V$.*

Proof. Let U be as in the preceding Lemma, and apply Urysohn's Lemma to the closed sets K and $X \setminus U$. □

Exercise 1.27. *Let X be a locally compact, but not compact space. Show that $C_c(X)$ is not complete with respect to the uniform norm $\|\cdot\|_C$.* **Hint.** *Do it first with $X = \mathbb{R}$.*

Exercise 1.28. *Show that, if X is locally compact, then the closure of $C_c(X)$ in $C_b(X)$ is $C_0(X)$.*

Theorem 1.12. *[Partition of unity in metric spaces] Let K be compact, $K \subseteq V_1 \cup \dots V_n$, where each V_j is open. Then, there exist $h_i \prec V_i$ such that*

$$h_1 + \dots + h_n = 1$$

on K.

PROOF. For each x consider a ball B_x centered at x and having compact closure, $x \in B_x \subseteq \overline{B_x} \subset V_j$, for some V_j (pick first a ball, it has intersect some V_j, pick a smaller ball with closure contained in V_j). By compactness, $K \subseteq B_{x_1} \cup \cdots \cup B_{x_m}$. Set

$$H_l = \cup_{B_{x_i} \subseteq V_l} B_{x_i} \subseteq V_l,$$

so that $K \subseteq \cup_l H_l$. By Urysohn Lemma there are $\overline{H_l} \prec g_l \prec V_l$. Observe that

$$1 - (1 - g_1)(1 - g_2) \dots (1 - g_n) = \begin{cases} 1 \text{ on } \cup_i \overline{H_i} \\ 0 \text{ on } X \setminus \cup_i V_i. \end{cases}$$

Set now

$$h_i = (1 - g_1) \dots (1 - g_{i-1}) g_i.$$

It is easy to verify that $h_i \prec V_i$. On the other hand,

$$\begin{aligned} 1 - (1 - g_1) \dots (1 - g_n) &= 1 - (1 - g_1) \dots (1 - g_{n-1}) + h_n \\ &\dots \\ &= h_1 + \dots h_n, \end{aligned}$$

proving the theorem. □

1.6.3. Tietze Extension theorem. Tietze extension theorem is a far reaching generalization of Urysohn lemma. The proof below only depends on the lemma itself: the statement is valid in all topological spaces where the Urysohn lemma holds.

THEOREM 1.13. *[Tietze Extension Theorem] Let X be a metric space (or, more generally, a locally compact, Hausdorff topological space), let $A \subset X$ be closed, and $f \in C(A, [0, 1])$. Then, f has an extension $F \in C(X, [0, 1])$, $F(x) = f(x)$ on A.*

PROOF. The idea is recursively using step functions to approximate f, extending them by means of Urysohn Lemma to obtain F in the limit.
Step I Let $h \colon A \to [0, k]$ be continuous, and let $B = \{x \in A : h(x) \in [0, k/3]\}$ and $C = \{x \in A : h(x) \in [2k/3, k]\}$, which are closed and disjoint in A, hence in X. By Urysohn Lemma, there exists $g \in C(X, [0, k/3])$ such that $g(x) = 0$ on B and $g(x) = k/3$ on C. In case both B and C are empty, set $g = 1/2$, constant. By considering the possible, different cases, $h - g \in C(A, [0, 2/3k])$.

Here, g should be considered a rough approximation of h on A, which is defined and continuous on X.

Step II We see now how the gain in control from h to $h-g$ develops under iteration.

$$\begin{array}{lcl} h_0 & := & f : A \to [0,1] = [0,(2/3)^0], \\ g_0 & \in & C(X,[0,1/3(2/3)^0]), \\ h_1 & := & h_0 - g_0 : A \to [0,(2/3)^1], \\ g_1 & \in & C(X,[0,1/3(2/3)^1]), \\ & \cdots & \\ h_n & := & h_{n-1} - g_{n-1} : A \to [0,(2/3)^n], \\ g_n & \in & C(X,[0,1/3(2/3)^n]), \\ & \cdots & \end{array}$$

Step III Construction of F. Set $s_n = g_0 + \cdots + g_{n-1} \in C(X)$, so that $h_n = f - (g_0 + \cdots + g_{n-1}) = f - s_n \in C(A)$. We have the estimate

$$\begin{array}{rcl} \|s_{n+j} - s_n\|_C & \leq & \|g_n\|_C + \|g_{n+1}\|_C + \cdots + \|g_{n+j-1}\|_C \\ & \leq & 1/3[(2/3)^n + \cdots + (2/3)^{n+j-1}] \\ & \leq & K(2/3)^n \to 0 \text{ as } n \to \infty, \end{array}$$

showing that $\{s_n\}$ is Cauchy in $C(X)$, hence it converges to $F \geq 0$ with

$$\|F\|_{C(X)} \leq \frac{1}{3}\sum_{n=0}^{\infty}(2/3)^n = \frac{1}{3}\frac{1}{1-2/3} = 1,$$

and

$$\|f - F\|_{C(A)} \leq \lim_{n\to\infty} \|f - s_n\|_{C(A)} = \lim_{n\to\infty} \|h_n\|_{C(A)} \leq \lim_{n\to\infty} (2/3)^n = 0,$$

as wished. □

EXERCISE 1.29. *Let (X,d) be locally compact, $K \subset X$ compact, and $V \supseteq K$ be open. Let $f : K \to [a,b]$ be continuous. Show that there exists $F \in C_c(X,[a,b])$ extending f, and such that* $\operatorname{supp}(F) \subseteq V$.

1.7. More exercises

EXERCISE 1.30.
(a) *Let d be the Euclidean metric in the plane $\mathbb{R}^2$. Find a subset E of $\mathbb{R}^2$ such that (i) the metric space (E,d) is locally compact, and (ii) there is a disc $B(x,r)$ in (E,d) such that $\overline{B(x,r)} \subsetneq \{y \in E : |x-y| \leq r\}$.*
(b) *Find E,x,r as in (a), such that neither $\{y \in E : |x-y| \leq r\}$ nor $\overline{B(x,r)}$ are compact.*
(c) *Let (X,d) be a locally compact metric space. Show that for each x in X there is $r_x > 0$ such, that for $0 < r \leq r_x$, $\overline{B(x,r)}$ is compact.*

Exercise 1.31. (i) *Show that the closed unit ball in* $C[0,1]$, $B=\{f\in C[0,1]:\|f\|_u\le 1\}$, *is not pre-compact.* **Hint.** *Show that* B *is not totally bounded.*

(ii) *Do the same for the unit ball* B' *in*

$$C^1[0,1]:=\{f\colon [0,1]\to\mathbb{R}:\ f \text{ is differentiable and } f'\in C[0,1]\},$$

with $\|f\|_{C^1}:=\|f\|_u+\|f'\|_u$.

(iii) *Use Ascoli-Arzelà Theorem to show that* B' *is precompact in* $C[0,1]$ *(although, by (ii), is not precompact in* $C^1[0,1]$*: we say in these cases that the immersion* $C^1[0,1]\subset C[0,1]$ *is compact).*

(iv) *Find an explicit example of a sequence* $\{f_n\}$ *in* B' *which converges in the norm of* C*, but with no converging subsequence in that of* C^1.

Exercise 1.32. *Show that* $\mathbb{Q}$ *with the Euclidean metric is not locally compact.*

Exercise 1.33. *Show that, if* X *is locally compact, then the closure of* $C_c(X)$ *in* $C_b(X)$ *is* $C_0(X)$.

1.8. Summary

This chapter provides an introduction to metric spaces with a strong bias towards real analysis and its applications. In §1.1, we define a metric space and, as special, but relevant cases, *normed linear spaces*. Among the examples of metric spaces in the section, some are spaces of functions. That is, the "points" of the space are functions defined on points of other spaces. For instance, a point in $C[0,1]$ is a continuous, complex valued function $x\mapsto f(x)$, which is, in turn, defined on points x of the interval $[0,1]$. A *distance* on $C[0,1]$ measures how far two functions are. This circle of ideas is the basis of *functional analysis*.

The fundamental notion of *completeness* is considered in §1.2. A metric space is complete when all Cauchy sequences in it converge. To wit, all convergent algorithms living in the space, converge to an element of the space. By Theorem 1.1 in §1.2.1, any metric space (X,d) can be "densely and isometrically imbedded" in a complete metric space $(\tilde{X},\tilde{d})$, and, §1.2.2, any normed linear space can be densely and isometrically imbedded in a Banach space (linear, normed, and complete). Useful as it is, the *metric completion* of a metric space provided by the theorem is not very explicit; its points are equivalence classes of Cauchy sequences: not the most visible objects in mathematics. In concrete situations, it is relevant deciding whether a metric space is already complete and, if it is not, to have a good "model" for its completion, possibly one on which it is easy to perform calculations. In §1.2.3, together with some topological notions, we find some complete *function spaces*.

A special instance of the principle "convergent algorithms do actually converge" is provided by Banach's Fixed Point Theorem in §1.2.4.

In §1.3 we introduce the other topological notion which is foundational in real and functional analysis, that of *compactness*. Four relevant, equivalent definitions of compactness are introduced.

Continuous functions and their spaces are further studied in §1.4. There we see that *uniform convergence* preserves continuity (which is the basis for completeness of spaces of continuous functions with the uniform norm); that continuous functions on compact sets are in fact *uniformly continuous*, and that this property extends to functions in $C_0(X)$, "vanishing at infinity".

The main result of the section, however, is the compactness theorem of Ascoli-Arzelà in §1.5, which shows that an *equicontinuous family* of continuous functions from a compact metric space to a compact metric space, is *pre-compact*: it contains a sequence which converges uniformly.

In §1.6 we consider *locally compact metric spaces*. The function *distance to a set* is introduced. It is a very interesting object on its own, but here it is just used to provide a quick proof of Urysohn's Lemma, which is a central tool in the approximation of characteristic functions of sets by means of continuous functions. Urysohn's Lemma, in turn, provides the main step in the proof of the Partition of Unity.

Extending functions from a smaller to a larger set without changing their main properties is a recurrent theme in mathematics. In §1.6.3 we present as an example Tietze's Theorem on the continuous extension of continuous functions.

1.8.1. Some function spaces. We have seen a number of linear spaces of functions $f\colon X \to \mathbb{R}$ based on a metric space (X, d):

(i) $B(X)$, the space of the bounded functions, with norm $\|f\|_u := \sup_{x\in X} |f(x)|$, which is a Banach space;

(ii) $C_b(X) \subseteq B(X)$, the space of the bounded, continuous functions, with the same norm, which is closed in $B(X)$, hence complete, with respect to the same norm;

(iii) when X is locally compact, $C_c(X)$, the space of functions with compact support, which is *not* complete with respect to $\|\cdot\|_u$ (unless X is compact, in which case $C_c(X) = C_b(X) = C(X)$, the space of all continuous functions);

(iv) when X is locally compact, $C_0(X)$, the space of functions vanishing at infinity, which is the completion in $C_b(X)$ of $C_c(X)$ with respect to $\|f\|_u$; hence, it is a Banach space;

(v) the space $C^1[0,1]$ is complete with respect to the norm $\|f\|_{C^1} = \max(\|f\|_u, \|f'\|_u)$;

(vi) intermediate between $C[0,1]$ and $C^1[0,1]$, we introduced the Lipschitz class $\mathrm{Lip}[0,1]$.

We also saw an interesting phenomenon on which we will return later. The closed unit balls in $(C[0,1], \|\cdot\|_u)$ and $(C^1[0,1], \|\cdot\|_{C^1})$ are not compact (hence, they are not pre-compact) in the respective norms; but, by Ascoli-Arzelà, the closed unit ball of $(C^1[0,1], \|\cdot\|_{C^1})$ is pre-compact with respect to the uniform norm $\|\cdot\|_u$.

The definitions of the spaces and the properties they satisfy extend without effort to complex valued functions, once we have observed that the complex valued function $f = u + iv$ is continuous if and only if u and v are.

CHAPTER 2

Abstract measure theory

In this chapter the reader finds the basic facts of "abstract" measure theory. The abstraction here mostly consists in the fact that we will not have interesting measures to work with until the next chapter, unless the reader already had an early exposure to Lebesgue measure on the real line. Actually, the choice of developing the abstract theory with little examples to work with is based on the hope that most advanced undergraduates and early graduate students are typically familiar with Lebesgue measure. In any case, we will provide below a motivation for the theory we are going to see.

While in the previous chapter we had a set X with *metric structure*, (X, d), here we will deal with a set X *measurable structure*, $(X, \mathcal{F})$, where $\mathcal{F} \subseteq 2^X$ is a *set-algebra* which is closed under countable unions (a *σ-algebra*). We will also be interested in *measurable functions* $f : X \to Y$, where Y is a metric space. Finally, we will consider *measures* on X, which will be used to *integrate* positive, measurable functions f, and more. These two structures will come together in the next chapter, where we start investigating in some depth *Borel measures*, which are defined on the ("Borel") σ-algebra generated by the open sets of a metric space X.

2.1. Motivation

2.1.1. Riemann's integral. Riemann's definition of integral is based on finer and finer partitions of the x-axis. We recall its definition here, in a version which is equivalent to the usual one, but notationally less cumbersome.

Let $f\colon [0, 1] \to \mathbb{R}$ be a bounded function, e.g. assume $0 \leq f(x) \leq 1$. For $n \geq 1$ set

$$S_n(f) := \sum_{j=1}^{2^n} \sup_{x\in[(j-1)/2^n, j/2^n]} f(x)\frac{1}{2^n} \geq s_n(f) := \sum_{j=1}^{2^n} \inf_{x\in[(j-1)/2^n, j/2^n]} f(x)\frac{1}{2^n}.$$

One readily verifies that for all $m.n \geq 1$,

$$S_n(f) \geq S_{n+1}(f) \geq s_{m+1}(f) \geq s_m(f),$$

hence that the *upper and lower integrals* of f on $[0,1]$ are well defined,

$$\overline{\int_0^1} f(x)dx := \lim_{n\to\infty} S_n(f) \geq \lim_{n\to\infty} s_n(f) =: \underline{\int_0^1} f(x)dx.$$

The function f is *Riemann integrable* when $\overline{\int_0^1} f(x)dx = \underline{\int_0^1} f(x)dx =: \int_0^1 f(x)dx$, a number which we call the *Riemann integral of* f. (The definition of Riemann integral allows for more freedom in the choice of the intervals, but it turns out to be equivalent to the one we have given here).

It is easy to see that continuous functions are Riemann integrable, and that so are monotone, bounded functions, which can have a dense set of discontinuities. In general, in order to be Riemann integrable, a function must not oscillate "too much, too often". Define the oscillation of f on a closed interval f to be

$$\mathrm{Osc}(f, I) := \sup_{x\in I} f(x) - \inf_{x\in I} f(x).$$

Then, f is Riemann integrable if and only if

$$0 = \lim_{n\to\infty} [S_n(f) - s_n(f)] = \lim_{n\to\infty} \frac{1}{2^n} \sum_{j=1}^{2^n} \mathrm{Osc}(f, [(j-1)/2^n, j/2^n]).$$

Riemann integrability is equivalent, that is, to the vanishing of the *average oscillation* which appears on the right.

A desired feature of integrals is that they exhibit some stability with respect to the integrand,

$$f_n \to f \text{ in some reasonable sense implies } \int_I f_n(x)dx \to \int_I f(x)dx \text{ in } \mathbb{R}.$$

Riemann's definition of integral, motivated by his research on trigonometric series, was a major leap, but a host of troubling examples showed some of its limitations. A famous one is *Dirichlet's function*,

$$D(x) = \begin{cases} 0 \text{ if } x \in [0,1] \setminus \mathbb{Q}, \\ 1 \text{ if } x \in [0,1] \cap \mathbb{Q}, \end{cases}$$

which is not integrable since $\overline{\int_0^1} D(x)dx = 1 > 0 = \underline{\int_0^1} D(x)dx$. The embarrassing point here is that D is not especially exotic: it is the monotone limit of a sequence of Riemann integrable function (with vanishing integral). Let $\{q_n\}_{n=1}^\infty$ be an enumeration of the rationals in $[0,1]$, and set

$$D_n(x) = \begin{cases} 0 \text{ if } x \in [0,1] \setminus \{q_1, \dots, q_n\}, \\ 1 \text{ if } x \in \{q_1, \dots, q_n\}. \end{cases}$$

Then, each $D_n \le D_{n+1}$ is Riemann integrable with $\int_0^1 D_n(x)dx = 0$, but $\lim_{n\to\infty} D_n = D$ is not.

EXERCISE 2.1. *Here we want to find* **continuous** *functions* $0 \le f_1 \le \cdots \le f_n \le f_{n+1} \le \cdots \le 1$ *such that* $f := \lim_{n\to\infty} f_n$ *is not Riemann integrable.*

(i) *For* $n = 1, 2, \ldots$ *let* $\phi_n \colon \mathbb{R} \to [0,1]$ *be a function supported in* $[-l_n, l_n]$*, whose graph is an isosceles triangle having base* $[-l_n, l_n]$ *and* $\phi_n(0) = 1$*.*

(ii) *Let* $\psi_n \colon [0,1] \to \mathbb{R}$ *be the sum of translates of* ϕ_n*,*

$$\psi_n(x) = \sum_{j=0}^{2^n} \phi_n(x - j/2^n),$$

and show that if the numbers l_n *are small enough, then* $0 \le \psi_n(x) \le 1$*.*

(iii) *Let* $f_n(x) = \max\{\psi_1(x), \ldots, \psi_n(x)\} \le f_{n+1}$*. Verify that* $0 \le f_n \le 1$ *and* f_n *is continuous.*

(iv) *Let* $f(x) = \lim_{n\to\infty} f_n(x)$*,* $f \colon [0,1] \to [0,1]$*. Show that, for some choice of the parameters* $\{l_n\}$*,*

$$\overline{\int_0^1} f(x)dx = 1, \text{ but } \underline{\int_0^1} f(x)dx \le 1/2.$$

Similar examples and the need of ”passing the limit under the integral sign” which emerged in many different situations, led to various attempts to extend the notion of integral on the one hand, and to find characterizations of the functions which could be integrated in Riemann's sense on the other. These two lines of research were intertwined, as we will see in the chapter on the construction of measures, where Lebesgue's full definition of integral is given, and the comparison with Riemann's integral is analyzed in detail.

Some starting readings on the history of the subject:

- Principia Mathematatica Historallis Integratus, by Saul Foresta and Lawrence Goldman,
- Review by William Dunham of Lebesgue's Theory of Integration: Its Origins and Development by Thomas Hawkins,

2.1.2. Lebesgue's definition of integral. Henri Lebesgue turned things upside down and gave a definition of integral by taking finer and finer partitions of the y-axis instead. As above, let $f \colon [0,1] \to [0,1]$ and for $n \ge 1$, $j = 1, \ldots, 2^n$, consider

$$E_{n,j} = \{x \in [0,1] : \ (j-1)/2^n < f(x) \le j/2^n\},$$

and consider the corresponding approximations from below and above of f,

$$s_n(x) := \sum_{j=1}^{2^n} \frac{j-1}{2^n} \chi_{E_{n,j}}(x) \leq f(x) \leq S_n(x) := \sum_{j=1}^{2^n} \frac{j}{2^n} \chi_{E_{n,j}}(x).$$

Clearly $s_m \leq s_{m+1} \leq S_{n+1} \leq S_n$ for all m, n. Both s_n and S_n are *simple function*, i.e. they have the form

$$g = \sum_{i=1}^{m} a_i \chi_{A_i},$$

where $A_1, \dots, A_m$ are subsets of $[0,1]$. To each set A_i we associate its "length" (the "measure of A_i") $m(A_i) \geq 0$, and define the integral of g to be the obvious one:

$$\int_0^1 g(x)dx = \sum_{i=1}^{m} a_i m(A_i).$$

Going back to our approximations, we have:

$$\int_0^1 s_m(x)dx \leq \int_0^1 S_n(x)dx$$

for all m, n. We can then define lower and upper (Lebesgue) integrals as in Riemann's theory:

$$\underline{\int_0^1} f(x)dx := \lim_{m\to\infty} \int_0^1 s_m(x)dx \leq \lim_{n\to\infty} \int_0^1 S_n(x)dx =: \overline{\int_0^1} f(x)dx.$$

The function f might be called "Lebesgue integrable" if lower and upper integral coincide (we will have different terms to express this concept). The nice thing is that oscillations are not a problem anymore:

$$0 \leq \int_0^1 S_n(x)dx - \int_0^1 s_n(x)dx = \frac{1}{2^n} \sum_{j=1}^{2^n} m(E_{n,j}) \leq \frac{1}{2^n} \to 0 \text{ as } n \to \infty,$$

i.e. $\underline{\int_0^1} f(x)dx = \overline{\int_0^1} f(x)dx$. Here we used the "intuitive" fact that the sum of the lengths of disjoint subsets is no more than the length of their union, which is less than the length of $[0,1]$, $m([0,1]) = 1$.

All of this is very nice and convincing, but (and that's the elephant in the room) we do not yet have a definition of "length" for general subsets of the real line, such as the $E_{n,j}$ can be. A theory of " vanishing length" had been developed by Borel, Lebesgue and others to characterize which functions are Riemann integrable. Its development was the basis of Lebesgue integration theory. The main obstruction, we will see, is that we can not assign a length to all sets, if we want to preserve the properties of length we need to work with integrals.

2.2. Basic measure theory

In order to make sense of the definition of integral, the class of the "measurable" subsets of the line, those for which a "length" can be defined, must be closed under a number of Boolean manipulations. We develop this in a general, abstract setting.

First, we introduce some algebraic convention on the use of ∞ which turns to be adequate to the theory of integration. We keep the usual conventions in place in the theory of limits, plus $+\infty \cdot 0 = 0$, which is not here considered indefinite. All other indefinite expressions, such as $+\infty - \infty$, remain such.

2.2.1. σ-algebras, measures, and measurable functions. A *σ-algebra* on a set Ω and $\mathcal{F}$ be a family of subsets of Ω such that:

(i) $\emptyset, \Omega \in \mathcal{F}$;

(ii) if $A \in \mathcal{F}$, then $\Omega \setminus A \in \mathcal{F}$;

(iii) if $\{A_n\}_{n=1}^{\infty}$ is a sequence of sets in $\mathcal{F}$, then $\cup_{n=1}^{\infty} A_n \in \mathcal{F}$.

By (i) and (ii), the countable intersection of elements in $\mathcal{F}$ belongs to $\mathcal{F}$, too. We say that $(\Omega, \mathcal{F})$ is a *measurable space* and the elements of $\mathcal{F}$ are *measurable sets*.

A *(positive) measure* on Ω is a map $\mu\colon \mathcal{F} \to [0, +\infty]$ defined on a σ-algebra $\mathcal{F}$, satisfying $\mu(\emptyset) = 0$ and, for any family $\{A_n\}_{n=1}^{\infty}$ of *disjoint* sets in $\mathcal{F}$,

$$\mu\left(\cup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n) \text{ (countable additivity).}$$

We further require a *non-degeneracy condition*: if $\mu(A) = \infty$, then there is $B \subset A$ such that $0 < \mu(B) < \infty$ (there are no atoms of infinite mass). We say that $(\Omega, \mathcal{F}, \mu)$ is a *measure space.*

In order of increasing generality, the measure μ is a *probability measure* if $\mu(\Omega) = 1$, it is *finite* if $\mu(\Omega) < \infty$, it is *σ-finite* if $\Omega = \cup_{n=1}^{\infty} \Omega_n$ with $X_n \in \mathcal{F}$ and $\mu(\Omega_n) < \infty$.

A function $f\colon \Omega \to X$ defined from a measurable space $(\Omega, \mathcal{F})$ to a metric (or just topological) space X is *measurable* if $f^{-1}(A) \in \mathcal{F}$ is measurable for all open sets A in X. Recall that collection of all open subsets of X is called the *topology* of X. By definitions, it follows that if $f\colon \Omega \to X$ is measurable, and $g\colon X \to Y$ is continuous (Ω measurable, X and Y topological), then $g \circ f\colon \Omega \to Y$ is measurable: the preimage of an open subset in Y is measurable in Ω.

2.2.1.1. *Properties of measurable sets and functions.* Given a family $\mathcal{A}$ of subsets of Ω, the power set 2^Ω is a σ-algebra containing $\mathcal{A}$. On the other hand, the intersection of an arbitrary family of σ-algebras is a σ-algebra (Exercise: verify this), hence there is a smallest σ-algebra $\mathcal{F}(\mathcal{A})$ containing $\mathcal{A}$:

$$\mathcal{F}(\mathcal{A}) = \bigcap_{\mathcal{G} \supseteq \mathcal{A}} \mathcal{G},$$

where $\mathcal{G}$ ranges among the σ-algebras in Ω. We call $\mathcal{F}(\mathcal{A})$ the *σ-algebra generated by $\mathcal{A}$*.

An important case is when (X, τ) is a topological space with topology τ. In this case, we call $\mathcal{F}(\tau) =: \mathcal{B}(X)$ the *Borel σ-algebra* of X (assuming a topology τ on X was chosen, of course). If X, Y are topological spaces, and $f : X \to Y$ is measurable with respect to the Borel σ-algebra in X, then we say that f is *Borel measurable.*

Proposition 2.1. *If $f, g : \Omega \to \mathbb{R}$ are measurable and $\Phi : \mathbb{R}^2 \to \mathbb{R}$ is continuous, then $\Phi(f, g) : \Omega \to \mathbb{R}$ is measurable.*

Proof. It suffices to show that $(f, g) : \Omega \to \mathbb{R}^2$ is measurable. Let $(a, b) \times (c, d)$ be an open rectangle in $\mathbb{R}^2$: $x \in (f, g)^{-1}((a, b) \times (c, d))$ if and only if $(f(x), g(x)) \in (a, b) \times (c, d)$, if and only if $f(x) \in (a, b)$ and $g(x) \in (b, c)$, i.e. $x \in f^{-1}((a, b)) \cap g^{-1}((c, d))$, which is measurable, since it is the intersection of two measurable sets.

Any open subset A of $\mathbb{R}^2$ is the union of countably many rectangles R_n, $n = 1, \dots$, hence $(f, g)^{-1}(A) = \cup_{n=1}^{\infty} (f, g)^{-1}(R_n)$ is measurable, as wished. □

Exercise 2.2. *Show that any open set in $\mathbb{R}^n$ is the countable union of open squares (hence, of open rectangles).*

Corollary 2.1. (i) *If $u, v : \Omega \to \mathbb{R}$ are measurable, then $u + iv : \Omega \to \mathbb{C}$ is measurable.*

(ii) *If $f = u + iv : \Omega \to \mathbb{C}$ is measurable, then $u, v, |f| : \Omega \to \mathbb{R}$ are measurable.*

(iii) *Constant functions are measurable. If $u, v : \Omega \to \mathbb{R}$ are measurable, then $u + v, uv : \Omega \to \mathbb{R}$ are measurable.*

(iv) *Let $E \subseteq \Omega$. Then E is measurable as a set if and only if χ_E is measurable as a function.*

(v) *If $f : \Omega \to \mathbb{C}$ is measurable, then there exists a measurable $\alpha : \Omega \to \mathbb{C}$ such that $|\alpha(x)| = 1$ for all x and $f = \alpha|f|$.*

EXERCISE 2.3. *Prove the corollary (i-iv).*

PROOF. of (v). The set $E = f^{-1}(\{0\})$ is measurable, since it is the preimage of a closed set. Define

$$\alpha := \chi_E + \frac{f}{|f|}\chi_{X\setminus E}.$$

The image of α is contained in the unit circle. The preimage through α of an open disc in $\mathbb{C}$ coincides with the preimage of an open circular arc $I = \{e^{it} : \ a < t < b\}$, $\alpha^{-1}(D) = \alpha^{-1}(I)$, and

$$\alpha^{-1}(I) = \{x : \ f(x) \in \{re^{it} : \ r > 0, \ t \in I\}\} \cup F,$$

where $F = f^{-1}(\{0\})$ if $1 \in I$, and $F = \emptyset$ if $1 \notin I$. In both cases, $\alpha^{-1}(I)$ is measurable. □

Given a σ-algebra $\mathcal{F}$ on Ω and $f\colon \Omega \to X$, the σ-algebra $f_*(\mathcal{F})$ on X is the set of those $E \subseteq X$ such that $f^{-1}(E) \in \mathcal{F}$.

EXERCISE 2.4. *Show that $f_*(\mathcal{F})$ is in fact a σ-algebra.*

PROPOSITION 2.2. *Let $(\Omega, \mathcal{F})$ be a measurable space, and X be a topological space.*

(i) *If $f\colon \Omega \to X$ is measurable, then $f^{-1}(E) \in \mathcal{F}$ for all Borel measurable E in X.*

(ii) *If $h\colon \Omega \to [-\infty, +\infty]$, then h is measurable if and only if $h^{-1}((-\infty, a])$ is measurable for all real a.*

(iii) *If $f\colon \Omega \to X$ is measurable and $g\colon X \to Y$ is Borel measurable, where Y is another topological space, then $g \circ f\colon \Omega \to Y$ is measurable.*

PROOF. (i) By definition of Borel measurable function, $f_*(\mathcal{F})$ contains all open subsets of X, hence it contains the Borel σ-algebra $\mathcal{B}(X)$. i.e., if E is Borel in X, then $f^{-1}(E)$ is measurable in Ω.

(ii) The "only if" follows by the definition. Viceversa, if $h^{-1}((-\infty, a])$ is measurable for all real a, then

$$(a, b) = (a, +\infty] \cap \left(\bigcup_{n\geq 1} [\mathbb{R} \setminus (b - \frac{1}{n}, +\infty]] \right),$$

then

$$h^{-1}((a,b)) = h^{-1}((a,+\infty]) \cap \left(\bigcup_{n\geq 1}[\mathbb{R} \setminus h^{-1}\left((b-\frac{1}{n},+\infty]\right)]\right),$$

which belongs to $\mathcal{F}$.

(iii) Use (i).

□

2.2.1.2. *Sup and limsup of measurable functions, and approximation by simple functions.*

THEOREM 2.1. *Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of measurable functions $f_n\colon \Omega \to [-\infty,+\infty]$. Then,*

$$\sup_n f_n, \ \inf_n f_n, \ \liminf_n f_n, \ \textit{and} \ \limsup_n f_n$$

are measurable.

PROOF. Fix $a \in \mathbb{R}$. Then,

$$\begin{aligned}\sup_n f_n(x) > a \quad &\Longleftrightarrow \quad \exists n: \ f_n(x) > a \\ &\Longleftrightarrow \quad x \in \bigcup_n f_n^{-1}((a,\infty)),\end{aligned}$$

and the latter is measurable. Same holds for $\sup_n f_n$. As a consequence,

$$\limsup_n f_n(x) = \inf_m(\sup_{n\geq m} f_n(x))$$

is measurable. □

For a sequence $\{A_n\}_{n=1}^{\infty}$ of subsets of Ω, one defines by analogy

$$\limsup_{n\infty} A_n := \bigcap_{k\geq 1}\left(\bigcup_{n\geq k} A_n\right), \ \liminf_{n\infty} A_n := \bigcup_{k\geq 1}\left(\bigcap_{n\geq k} A_n\right).$$

EXERCISE 2.5. *Prove the following.*

(i) *We have that $x \in \limsup_n A_n$ if and only if there are infinitely many n's such that $x \in A_n$ (we say that $x \in A_n$ infinitely often).*

(ii) *We have that $x \in \liminf_n A_n$ if and only if there exists $n(x)$ such that $x \in A_n$ for all $n \geq n(x)$ (we say that $x \in A_n$ definitely).*

(iii) $\liminf_n A_n \subseteq \limsup_n A_n$.

As a reward for all this effort, we have a useful approximation procedure for measurable functions. A *simple function* $s\colon \Omega \to \mathbb{C}$ is one having the form

$$s(x) = \sum_{j=1}^{n} a_j \chi_{E_j},$$

where $E_1, \dots, E_n \subseteq \Omega$ and $a_1, \dots, a_n \in \mathbb{C}$.

THEOREM 2.2. (i) *Let $f\colon \Omega \to [0, +\infty]$ be a measurable function. Then, there exist measurable, nonnegative simple functions $0 \le s_0 \le \dots s_n \le s_{n+1} \le \dots$, such that $f(x) = \sup_n f_n(x)$ for $x \in \Omega$.*

(ii) *Viceversa, if $s_0 \le s_1 \le \dots \le s_n \le \dots$ is a sequence of measurable simple functions on Ω, then $f(x) := \sup_n f_n(x)$ defines a nonnegative, measurable function on Ω.*

PROOF. Part (ii) follows from Theorem 2.1. Let f be as in (i) and, for $n \ge 0$ and $1 \le j \le n2^n$, set $E_{n.j} = \{x :\ (j-1)/2^n \le f(x) < j/2^n\}$. Then,

$$s_n = \sum_{j=1}^{n2^n} \frac{j-1}{2^n} \chi_{E_{n.j}}$$

has all desired properties. □

EXERCISE 2.6. *Show that the functions defined in the proof of (ii) in fact satisfy that $\sup_n s_n(x) = f(x)$ on Ω.*

What about approximations from above?

EXERCISE 2.7. *Let f like in Theorem 2.2 (i), and suppose, more, that $f \le C$. After choosing $n(C)$ such that $n(C)2^{n(C)} \ge C$, for $n \ge n(C)$ define, with the same $E_{n.j}$ as in the proof of the theorem,*

$$S_n = \sum_{j=1}^{n2^n} \frac{j}{2^n} \chi_{E_{n.j}}.$$

Show that $S_{n+1} \ge S_n \ge f$, and $f(x) = \inf_n S_n(x)$.

We will use without mention the following fact.

PROPOSITION 2.3. *Let $s = \sum_{j=1}^{n} a_j \chi_{E_j}$ be a (measurable) simple function. We can find measurable, disjoint sets $F_1, \dots, F_N$ and numbers $b_1, \dots, b_N$ such that $s = \sum_{l=1}^{N} b_l \chi_{F_l}$. Here, $N = N_n$ only depends on n.*

PROOF. The property is clear for $n = 1$. If it holds for $n - 1$, then

$$\begin{aligned}
\sum_{j=1}^{n} a_j \chi_{E_j} &= \sum_{l=1}^{N_{n-1}} b_l \chi_{F_l} + a_n \chi_{E_n} \\
&= \sum_{l=1}^{N_{n-1}} b_l \chi_{F_l \setminus E_n} + \sum_{l=1}^{N_{n-1}} (b_l + a_n) \chi_{F_l \cap E_n} + a_n \chi_{E_n \setminus \cup_{l=1}^{N_{n-1}} F_l}.
\end{aligned}$$

Incidentally, in the worst case scenario this gives $N_1 = 1$ and $N_n = 2N_{n-1} + 1$. □

2.2.1.3. *Properties of measures.*

Proposition 2.4. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space.*

(i) *If $A_1, \dots, A_n$ are disjoint, measurable sets, then $\mu\left(\bigcup_{j=1}^n \mu(A_j)\right) = \sum_{j=1}^n \mu(A_j)$.*

(ii) *If $A, B \in \mathcal{F}$ and $A \subset B$, then $\mu(A) \leq \mu(B)$.*

(iii) *If $A_1 \subseteq \dots \subseteq A_n \subseteq A_{n+1} \subseteq \dots$ are measurable, then $\mu(\bigcup_n A_n) = \lim_n \mu(A_n) = \sup_n \mu(A_n)$.*

(iv) *If $A_1 \supseteq \dots \supseteq A_n \supseteq A_{n+1} \supseteq \dots$ are measurable and $\mu(A_1) < \infty$, then $\mu(\bigcap_n A_n) = \lim_n \mu(A_n) = \inf_n \mu(A_n)$.*

Proof. Statement (i) follows from countable additivity: set $A_{n+1} = A_{n+2} = \dots = \emptyset$. For (ii), use $A \cup (B \setminus A) = B$. About (iii),

$$\begin{aligned}\mu\left(\bigcup_n A_n\right) &= \mu(A_1) + \sum_{n=1}^{\infty} \mu(A_{n+1} \setminus A_n) \\ &= \mu(A_m) + \sum_{n=m}^{\infty} \mu(A_{n+1} \setminus A_n),\end{aligned}$$

and the two series both converge, or both diverge. In both cases, $\mu(A_m) \nearrow \mu\left(\bigcup_n A_n\right)$. (iv) Similarly, since all terms are finite (including the series):

$$\begin{aligned}\mu\left(\bigcap_n A_n\right) &= \mu(A_1) - \sum_{n=1}^{\infty} \mu(A_n \setminus A_{n+1}) \\ &= \mu(A_m) - \sum_{n=m}^{\infty} \mu(A_n \setminus A_{n+1}),\end{aligned}$$

so that $\mu(A_m) \searrow \mu\left(\bigcap_n A_n\right)$. □

Exercise 2.8. *Show that the non-degeneracy condition on the measure μ is equivalent to requiring that, for $A \in \mathcal{F}$, $\mu(A) = \sup\{\mu(B) :\ B \subseteq A,\ B \in \mathcal{F},\ \mu(B) < \infty\}$.*

2.2.2. The Lebesgue integral of a function. From now on, when we talk of simple functions, we assume them to be measurable unless otherwise stated. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let $s\colon \Omega \to \mathbb{C}$ be a measurable, simple function, $s = \sum_{i=1}^n a_i \chi_{E_i}$. Its *integral* is

$$\int_\Omega s d\mu := \sum_{i=1}^n a_i \mu(E_i).$$

and let $f \geq 0$ be a measurable function. Then,

$$\int_\Omega f d\mu := \sup \left\{ \int_\Omega s d\mu : \text{ with } 0 \leq s \leq f \text{ simple} \right\}.$$

EXERCISE 2.9. *Show that for a simple function the two definitions agree. That is, using notation $I(s) = \sum_{i=1}^n a_i \mu(E_i)$, show that*

$$I(s) = \max \left\{ I(\sigma) : \textit{ with } 0 \leq \sigma \leq s \textit{ simple} \right\}.$$

EXERCISE 2.10. *Let $f, g \geq o$ be measurable. Prove the following.*

(i) *If $0 \leq f \leq g$, then $\int_\Omega f d\mu \leq \int_\Omega g d\mu$.*

(ii) *If $\lambda > 0$, then $\int_\Omega \lambda f d\mu = \lambda \int_\Omega f d\mu$.*

We say that a measurable function $f \geq 0$ is *integrable* if $\int_\Omega f d\mu < \infty$. When f is real valued, we split it $f = f_+ - f_-$ with $f_+ = \max(f, 0)$ and $f_- = \max(-f, 0)$, and we define $\int_\Omega f d\mu = \int_\Omega f_+ d\mu - \int_\Omega f_- d\mu$, provided one of the two summands on the right is finite. We also have that $|f| = f_+ + f_-$, and we say that f is *integrable* if $|f|$ is. Similarly, for $f = u + iv$ complex valued, the integral is defined componentwise, $\int_\Omega (u + iv) d\mu = \int_\Omega u d\mu + i \int_\Omega v d\mu$. provided both summands on the right are finite. The L^1 *norm* of f is

$$\|f\|_{L^1(\mu)} := \int_\Omega |f| d\mu.$$

We have the inequality

$$\left| \int_\Omega f d\mu \right| \leq \int_\Omega |f| d\mu. \tag{2.2.1}$$

EXERCISE 2.11. *Prove* (2.2.1). **Hint.** *There is $\alpha \in \mathbb{C}$ such that $\left|\int_\Omega f d\mu\right| = \alpha \int_\Omega f d\mu$.*

With our definition of a complex integral, we have that, if $f \in L^1$ and $\lambda \in \mathbb{C}$, then

$$\lambda \int_\Omega f d\mu = \int_\Omega \lambda f d\mu. \tag{2.2.2}$$

EXERCISE 2.12. *Prove* (2.2.2).

2.3. Limit theorems for integrals

The *raison d'être* of Lebesgue's theory of integration is that, under rather general assumptions, we can pass limits under the integral sign.

2.3.1. Monotone Convergence Theorem.

THEOREM 2.3. *[Monotone convergence theorem] Let* $0 \leq f_1 \leq f_2 \leq \cdots \leq f_n \leq f_{n+1} \leq \ldots$ *be a sequence of measurable functions on* $(\Omega, \mathcal{F}, \mu)$. *Then,*

$$\lim_{n\to\infty} \int_\Omega f_n d\mu = \int_\Omega \lim_{n\to\infty} f_n d\mu.$$

PROOF. The expression $\lim_{n\to\infty} f_n(x) = f(x)$ defines a function $f \colon \Omega \to [0, +\infty]$ which is measurable. Since $f_n \leq f$, the inequality $\leq$ is obvious.

In the other direction, let $0 < \alpha < 1$, and $0 \leq s = \sum_{i=1}^m a_i \chi_{I_i}$, simple, such that $s \leq f$ and let $E_n = \{x : \ f_n(x) \geq \alpha s(x)\} \subseteq E_{n+1}$. Observe that $\Omega = \cup_{n=1}^\infty E_n$: if $0 < \alpha s(x) < s(x) \leq f(x)$, then $f_n(x) \geq \alpha s(x)$ for some n, and if $s(x) = 0$, then $f_n(x) \geq s(x)$, hence $x \in E_n$. Thus,

$$\begin{aligned} \int_\Omega f_n d\mu &\geq \int_{E_n} \alpha s(x) d\mu(x) \\ &= \alpha \sum_{i=1}^m a_i \mu(I_i \cap E_n) \\ &\nearrow \alpha \sum_{i=1}^m a_i \mu(I_i) = \alpha \int_\Omega s(x) d\mu(x) \text{ as } n \to \infty, \end{aligned}$$

i.e.

$$\lim_{n\to\infty} \int_\Omega f_n d\mu \geq \alpha \int_\Omega s(x) d\mu(x)$$

holds for all $\alpha < 1$ and $0 \leq s \leq f$, hence, $\lim_{n\to\infty} \int_\Omega f_n d\mu \geq \int_\Omega f d\mu$. □

EXERCISE 2.13. *Use Monotone Convergence Theorem and approximation by simple functions to show that* $\int_\Omega (f+g) d\mu = \int_\Omega f d\mu + \int_\Omega g d\mu$ *if* $f, g \geq 0$ *are measurable.*

EXERCISE 2.14. *Find a sequence* $\{f_n\}$ *of measurable, nonnegative functions such that* $f_n(x) \to f(x)$ *converges pointwise, but*

$$\lim_{n\to\infty} \int_\Omega f_n d\mu \neq \int_\Omega f d\mu.$$

THEOREM 2.4. *Let* $f_n \geq 0$ *be a sequence of measurable functions, and* $f = \sum_{n=1}^\infty f_n$. *Then,* f *is measurable and*

$$\int_\Omega \left(\sum_{n=1}^\infty f_n \right) d\mu = \sum_{n=1}^\infty \int_\Omega f_n d\mu.$$

PROOF. By iterating Exercise 2.13 we have that $0 \leq s_n = \sum_{j=1}^n f_j \leq s_{n+1} \nearrow f$, and

$$\sum_{j=1}^{n} \int_\Omega f_n d\mu = \int_\Omega s_n d\mu \nearrow \int_\Omega f d\mu,$$

by Monotone Convergence. □

2.3.2. Fatou's Lemma.

THEOREM 2.5. *[Fatou's Lemma] Let $\{f_n\}$ be a sequence of nonnegative measurable functions. Then,*

$$\int_\Omega \liminf_{n\to\infty} f_n(x)d\mu(x) \leq \liminf_{n\to\infty} \int_\Omega f_n(x)d\mu(x). \tag{2.3.1}$$

PROOF. Since for a real valued sequence $\{a_n : \ n \geq 1\}$ the associated sequence $\{\inf_n\{a_n : \ n \geq m\} : \ m \geq 1\}$ is increasing (use this in the first line), and $\inf\{f_n(x) : \ n \geq m\} \leq f_n(x)$ for $n \geq m$ (use this for the inequality in the second), we have:

$$\begin{aligned}
\liminf_{n\to\infty} \int_\Omega f_n d\mu &= \lim_{m\to\infty} \inf_{n\geq m} \int_\Omega f_n d\mu \\
&\geq \lim_{m\to\infty} \int_\Omega \inf_{n\geq m} f_n(x)d\mu(x) \\
&= \int_\Omega \lim_{m\to\infty} \inf_{n\geq m} f_n(x)d\mu(x) \text{ by MCT} \\
&= \int_\Omega \liminf_{n\to\infty} f_n(x)d\mu(x),
\end{aligned}$$

as wished. □

EXERCISE 2.15. *Find an example of a sequence f_n defined on Ω with $\mu(\Omega) = 1$ such that strict inequality holds in Fatou's Lemma.*

THEOREM 2.6. *Let $f : \Omega \to \mathbb{C}$ be measurable. Then,*

$$\left| \int_\Omega f d\mu \right| \leq \int_\Omega |f| d\mu.$$

PROOF. For some complex α with $|\alpha| = 1$,

$$\begin{aligned}
\left| \int_\Omega f d\mu \right| &= \alpha \int_\Omega f d\mu = \int_\Omega \alpha f d\mu \\
&= \int_\Omega Re(\alpha f) d\mu \leq \int_\Omega |\alpha f| d\mu \\
&= \int_\Omega |f| d\mu.
\end{aligned}$$

□

Given a measure μ on Ω and $f \geq 0$ measurable, we can define a new measure ν by

$$\nu(E) = \int_E f d\mu.$$

We also write

$$f = \frac{d\nu}{d\mu}. \tag{2.3.2}$$

EXERCISE 2.16. *Use Monotone Convergence to show that ν defines, in fact, a measure.*

2.3.3. Dominated Convergence Theorem.

THEOREM 2.7 (Dominated Convergence Theorem). *Let $\{f_n\}$ be a sequence of functions with values in $\mathbb{R}$, with $\lim_{n\to\infty} f_n(x) = f(x)$ pointwise, and suppose that $|f_n(x)| \leq g(x)$, with the dominating g integrable, $\int_\Omega g d\mu < \infty$. Then,*

$$\lim_{n\to\infty} \int_\Omega |f_n - f| d\mu = 0.$$

In particular,

$$\lim_{n\to\infty} \int_\Omega f_n d\mu = \int_\Omega f d\mu.$$

PROOF. By assumption, $0 \leq 2g - |f_n - f|$, hence, by Fatou's Lemma,

$$\begin{aligned} 2\int g d\mu &= \int \liminf_{n\to\infty} [2g - |f_n - f|] d\mu \\ &\leq \liminf_{n\to\infty} \int [2g - |f_n - f|] d\mu \\ &= 2\int g d\mu - \limsup_{n\to\infty} \int |f_n - f| d\mu, \end{aligned}$$

from which we deduce $\limsup_{n\to\infty} \int |f_n - f| d\mu \leq 0$, as wished. □

We say that a property $\mathcal{P}(x)$ *holds μ-almost everywhere on* Ω ($\mu - a.e.$) if

$$\mu(\{x : \mathcal{P}(x) \text{ does not hold}\}) = 0.$$

e.g. we write $f = g \ \mu - a.e.$ if $f, g \colon \Omega \to \mathbb{C}$ only differ on a set of measure zero.

EXERCISE 2.17. *Prove the series version of Dominated Convergence. If $\sum_n \int |f_n| d\mu < \infty$, then $\int (\sum_n f_n) d\mu = \sum_n \int f_n d\mu$.*

2.4. Some example of measures

2.4.1. Discrete measures. Let X be a set. The *counting measure* $\sharp$ on X is defined on 2^X,

$$\sharp(A) = (\text{the number of the elements of } A) \in \mathbb{N} \cup \{\infty\}. \tag{2.4.1}$$

Let $a \in X$. The *unit mass* at a (or, *Dirac's delta at* a), δ_a, is as well defined on 2^X,

$$\delta_a(A) = \begin{cases} 1 \text{ if } a \in A, \\ 0 \text{ if } a \notin A. \end{cases} \tag{2.4.2}$$

These measures are related,

$$\sharp(A) = \sum_{a \in X} \delta_a(X).$$

A measure μ is *discrete* if singletons are measurable and, for all measurable A,

$$\mu(A) = \sum_{a \in A} \mu(\{a\}). \tag{2.4.3}$$

Equivalently, a discrete measure on a measurable space $(X, \mathcal{F})$ is a (possibly infinite) positive, linear combination of Dirac's deltas,

$$\mu = \sum_{a \in X} \alpha(a)\delta_a,$$

where $\alpha(a) \geq 0$.

An interesting family of discrete measures can be obtained by selecting (i) a countable set $\{x_n\}_{n=1}^{\infty}$ in $\mathbb{R}$, then, (ii) a summable sequence $\{\alpha_n\}_{n=1}^{\infty}$ in $(0, +\infty)$, $\sum_n \alpha_n = \infty$. The measure

$$\mu = \sum_{n=1}^{\infty} \alpha_n \delta_{x_n}$$

is defined on all subsets of $\mathbb{R}$, in particular on its Borel σ-algebra. If $\{x_n\}$ is dense in $\mathbb{R}$, then $\mu(a, b) > 0$ for all (nonempty) open intervals (a, b).

2.4.2. Measures defined by integrals. Let μ be a measure on a measurable space $(X, \mathcal{F})$, and let $E \subseteq X$ be measurable. Then,

$$\mu_E(A) := \mu(E \cap A) \tag{2.4.4}$$

defines a measure on $(X, \mathcal{F})$, the *restriction of* μ *to* E.

More generally, if μ is a measure on $(X, \mathcal{F})$ and $f \geq 0$ belongs to $L^1(\mu)$, then

$$\nu(A) := \int_A f d\mu \tag{2.4.5}$$

defines a measure on $(X, \mathcal{F})$.

PROPOSITION 2.5. *The set function* ν *defined by* (2.4.5) *is a finite measure. Moreover, if* $\mu(A) = 0$*, then* $\nu(A) = 0$.

PROOF. Let $\{E_n\}_{n=1}^{\infty}$ be a family of disjoint, measurable sets. Let $f_n = f \sum_{j=1}^{n} \chi_{E_j} = \chi_{\cup_{j=1}^{n} E_j} f$. By monotone convergence,

$$\begin{aligned} \nu(\cup_{n=1}^{\infty} E_n) &= \int_{\cup_{n=1}^{\infty} E_n} f d\mu = \lim_{n\to\infty} \int_{\cup_{j=1}^{n} E_j} f d\mu \\ &= \lim_{n\to\infty} \sum_{j=1}^{n} \nu(E_j) = \sum_{n=1}^{\infty} \nu(E_n). \end{aligned}$$

If $\mu(A) = 0$ and s is a simple function such that $0 \leq s \leq f\chi_A$, then $\int_X s d\mu = 0$, hence, $\nu(A) = \int_X f\chi_A d\mu = 0$. □

When $\nu(A) = \int_A f d\mu$ as above, with $f \in L^1(\mu)$, we write $d\nu = f d\mu$, or

$$f = \frac{d\nu}{d\mu}. \tag{2.4.6}$$

This is a special case of the *Radon-Nikodym derivative* of a measure ν with respect to another measure μ.

If μ, ν are measures on the same measurable space $(X, \mathcal{F})$, and $\nu(A) = 0$ whenever $\mu(A) = 0$, we say that ν is *absolutely continuous* with respect to μ, and write $\nu \prec \mu$. The converse of the proposition above holds: if $\nu \prec \mu$ and $\nu(X) < \infty$, then $d\nu = f d\mu$ for a unique $f \geq 0$ in $L^1(\mu)$. In order to prove this fact, we need a theorem on the existence of the Radon-Nikodym derivative.

2.4.3. The Lebesgue measure. At some point, we will prove the following.

THEOREM 2.8. *There exists a unique measure m on $\mathcal{B}(\mathbb{R})$, the Borel σ-algebra of $\mathbb{R}$ (endowed with the Euclidean distance), such that $m((a,b)) = b-a$ for all $a<b$ in $\mathbb{R}$.*

The measure m is called the *Lebesgue measure* on $\mathbb{R}$.

There are good reasons to define measures on σ-algebras, and not on all subsets. One reason is that often, e.g. in the theory of stochastic processes, σ-algebras encode "available information". An even more stringent one is that, as we show below, the Lebesgue measure can not be defined on $2^{\mathbb{R}}$. In order to do that, we use the Axiom of Choice.

EXAMPLE 2.1. *On $[0,1]$, consider the equivalence relation $x \sim y$ if $x-y \in \mathbb{Q}$ is rational. The relation $\sim$ is clearly an equivalence one. Let $E \subseteq [0,1]$ be a set containing exactly one point from each equivalence class. Observe, next, that*

$$[0,1] \subset \bigcup_{q \in \mathbb{Q} \cap [-1,1]} (E+q) \subset [-1,2]:$$

each point x in $[0,1]$ has the form $x = x_0 + q$ for some $x_0 \sim x$, and $q \in [-1,1]$. Also, the sets $E+q$ are disjoint (because just one element for each equivalence class was selected), and there is a countable number of them. Also, $m(E) = m(E+q)$ by translation invariance.

- *If $m(E) = 0$, then $1 = m([0,1]) \le \sum_q m(E+q) = 0$, which is absurd.*
- *If $m(E) > 0$, then $3 = m([-1,2]) \ge \sum_q m(E+q) = \infty$, which is equally absurd.*

2.5. Some applications

2.5.1. Derivatives under integral sign. An extremely useful, and in fact used, fact is that we can take the derivative under the integral sign.

THEOREM 2.9. *Let $(X, \mathcal{F}, \mu)$ be a measure space, and $F\colon (a,b) \times X \to \mathbb{R}$ such that:*

(i) *for each x in X, $t \mapsto F(t,x)$ is differentiable on (a,b);*

(ii) *for each t in (a,b), $x \mapsto F(t,x)$ is measurable on X;*

(iii) *$f(x) := \sup_{t \in (a,b)} |\partial_t F(t,x)|$ defines a function in $L^1(\mu)$.*

Then,

$$\frac{d}{dt}\int_X F(t,x)d\mu(x) = \int_X \partial_t F(t,x)d\mu(x) \tag{2.5.1}$$

exists for all t in (a,b).

PROOF. Let $I(t) = \int_X F(t,x)d\mu(x)$ and $J(t) = \int_X \partial_t F(t,x)d\mu(x)$, and let $\mathbb{R} \ni h_n \to 0$. Then,

$$\begin{aligned}\frac{I(t+h_n)-I(t)}{h_n} - J(t) &= \int_X \left(\frac{F(t+h_n,x)-F(t,x)}{h_n} - \partial_t F(t,x)\right) d\mu(x)\\ &= \int_X \left(\partial_t F(t+\Theta h_n, x) - \partial_t F(t,x)\right) d\mu(x),\end{aligned}$$

with $\Theta = \Theta(x,t,h_n) \in [0,1]$, by Lagrange theorem. The integrand is dominated in absolute value by $2f(x)$, and it tends to zero as $n \to \infty$ for each x, hence we can apply Dominated Convergence:

$$\lim_{n\to\infty} \frac{I(t+h_n)-I(t)}{h_n} = \int_X \partial_t F(t,x)d\mu(x).$$

Since this holds for all sequences $h_n \to 0$, $I'(t) = J(t)$. □

This statement has many variants, which are useful for different problems, with proofs which are generally variations on the one here provided.

2.5.2. The Severini-Egorov Theorem. Almost everywhere, pointwise convergence is a weak notion. In 1910 Carlo Severini, and independently in 1911 Dmitri Egorov, showed that outside a set of small measure convergence is in fact uniform; a fact which is crucially useful in the proof of many important theorems (see e.g. Lusin's Theorem in the next chapter).

THEOREM 2.10. *[Severini-Egorov] Let $\{f_n : n \geq 1\}$ be a sequence of measurable functions defined on a finite measure space, $\mu(X) < \infty$, with values in $\mathbb{R}$ (or, in fact, in any separable metric space), and suppose that $\lim_{n\to\infty} f_n(x) = f(x)$ for a.e. x in X. Then, for all $\epsilon > 0$ there exists $B \subseteq X$ such that $\mu(X \setminus B) \leq \epsilon$ and f_n converges uniformly on B.*

PROOF. For $n, k \geq 1$, let

$$E_{n,k} = \{x \in X : |f_m(x) - f(x)| \geq 1/k \text{ for some } m \geq n\} \supseteq E_{n+1,k}.$$

Observe that $x \in \cap_n E_{n,k}$ if and only if $|f_m(x) - f(x)| \geq 1/k$ for infinitely many values of m; in particular, $\{f_n(x)\}$ does not converge. Hence, $\mu\left(\cap_n E_{n,k}\right) = 0$.

By dominated convergence applied to $\{\chi_{E_{n,k}} : n \geq 1\}$, then, $\mu(E_{n,k}) \to 0$ as $n \to \infty$ for each fixed k. Fix $\epsilon > 0$ and, for each $k \geq 1$, select n_k such that $\mu(E_{n_k,k}) \leq \epsilon/2^k$. Finally, let $B = \cup_k E_{n_k,k}$, so that $\mu(B) \leq \epsilon$.

For $x \in X \setminus B = \cap_j (X \setminus E_{n_j,j})$, and $m \geq k$, we have that

$$|f_m(x) - f(x)| \leq 1/k$$

because $x \in X \setminus E_{n_k,k}$, hence we have uniform convergence. □

EXERCISE 2.18. *Consider the real line with the Lebesgue measure. Show that the functions* $f_n = \chi_{[n,n+1]}$ *provide a counterexample to the version to the Severini-Egorov Theorem in spaces of infinite measure.*

2.5.3. L^p spaces; definition. Recall that $L^1(\mu)$ is the space of the absolutely integrable functions,

$$L^1(\mu) \ni f : \Omega \to \mathbb{C} \iff \|f\|_{L^1(\mu)} := \int_\Omega |f| d\mu < \infty. \tag{2.5.2}$$

In $L^1(\mu)$, we identify two functions which are equal *a.e.* (you might want to prove that being equal *a.e.* is an equivalence relation).

EXERCISE 2.19. *In a measure space, the relation* $f = g\ \mu - a.e.$ *defines an equivalence relation.*

To be precise, then, $L^1(\mu)$ is a space of equivalence classes of functions. Once a measure μ is fixed, viewing instead the elements of $L^1(\mu)$ as functions, with the clause that we identify two of them if they coincide *a.e.* does not cause any problem. One has to be careful, however, when working simultaneously with a family of different measures, since the equivalence relation involves a specified measure only.

EXERCISE 2.20.
(i) *Show that if* $f \geq 0$ *and* $\int_\Omega f d\mu = 0$, *then* $f = 0$ *a.e.*
(ii) *Show that, for a complex valued function, if* $\left|\int f d\mu\right| = \int |f| d\mu$, *then there exists complex* α *with* $|\alpha| = 1$ *such that* $\alpha f(x) \geq 0$ *for a.e.* x.

More generally, for $1 \leq p < \infty$ we define

$$\|f\|_{L^p} = \left(\int_\Omega |f|^p d\mu\right)^{1/p},$$

and $L^p(\mu) = \{f : \|f\|_{L^p} < \infty\}$. We will show shortly that $\|\cdot\|_{L^p}$ defines a norm on $L^p(\mu)$, which is a vector space. The only subtle point is proving *Minkovski's inequality*: $\|f + g\|_{L^p} \leq \|f\|_{L^p} + \|g\|_{L^p}$.

Define the *essential supremum* of a real valued function which is measurable on Ω as

$$\text{ess-sup} f(x) := \inf\{\lambda \in \mathbb{R} : \mu(\{x : f(x) \geq \lambda\}) = 0\}, \tag{2.5.3}$$

where as usual we set $\inf \emptyset = +\infty$.

The L^∞-norm of a measurable $f\colon \Omega \to \mathbb{C}$ is

$$\|f\|_{L^\infty(\mu)} = \text{ess-sup}_x |f(x)|.$$

As above, $L^\infty(\mu) = \{f : \ \|f\|_{L^\infty} < \infty\}$.

EXERCISE 2.21. *Show Minkovski's inequality for $p = 1, \infty$. When do we have equality?*

EXERCISE 2.22. *For $1 \leq p \leq \infty$, $\mu = 0$ if and only if $L^p(\mu)$ contains the constant function only.*

2.6. Some integral inequalities

Integral inequalities play a prominent role in analysis and its applications. Norms are often defined in terms of integrals, as in the case of the L^p spaces, and the triangle inequality in L^p is an integral inequality. Integral inequalities often appear when we consider the problem of comparing two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on the same function space, or when we study the boundedness of operators, in Partial Differential Equations, and in many other contexts. The classic and still inspiring monograph on the subject is G. H. Hardy, J. E. Littlewood, G. Pólya, Inequalities, Cambridge UP, 1934.

2.6.1. Jensen, Hölder, and Minkovski. In this subsection, we consider three classical, and, in real analysis, foundational, integral inequalities. Jensen's inequality might be seen as a reformulation of the notion of convexity, and the other two inequalities can be easily derived from it.

2.6.1.1. *Jensen inequality.* A function $\Phi\colon I \to \mathbb{R}$ defined on an interval I of the real line is *convex* if

$$\Phi(t_1x_1 + t_2x_2) \leqslant t_1\Phi(x_1) + t_2\Phi(x_2)$$

whenever $x_1, x_2 \in I$ and $t_1, t_2 \geqslant 0$, $t_1 + t_2 = 1$, and *concave* if the opposite inequality holds. By induction (exercise) we can show that this is equivalent to requiring that for $x_1, \dots, x_n \in I$ and $t_1, \dots, t_n \geqslant 0$ with $t_1 + \dots + t_n = 1$, one has

$$\Phi\left(\sum_{j=1}^n t_jx_j\right) \leqslant \sum_{j=1}^n t_j\Phi(x_j).$$

THEOREM 2.11 (Jensen's inequality). *Let $\Phi\colon [0, +\infty) \to [0, +\infty)$ be an increasing, convex function, and let (X, μ) be a probability space: μ is a positive measure on X and $\mu(X) = 1$. Then, for all measurable $f\colon X \to [0, +\infty)$,*

$$\Phi\left(\int_X f d\mu\right) \leqslant \int_X \Phi(f) d\mu.$$

PROOF. Let $0 \leqslant s = \sum_{j=1}^{n} s_j \chi_{E_j}$ be a simple function with the E_j's measurable and disjoint, and $\cup_{j=1}^{n} E_j = X$. Then,

$$\Phi\left(\int_X s d\mu\right) = \Phi\left(\sum_{j=1}^{n} s_j \mu(E_j)\right) \leqslant \sum_{j=1}^{n} \Phi(s_j)\mu(E_j) = \int_X \Phi(s) d\mu.$$

For general, measurable $f \geqslant 0$ we have, denoting by s a generic simple function, using monotonicity and continuity of Φ from first to second and from third to fourth line,

$$\begin{aligned}
\Phi\left(\int_X f d\mu\right) &= \Phi\left(\sup_{0 \leqslant s \leqslant f} \int_X s d\mu\right) \\
&= \sup_{0 \leqslant s \leqslant f} \Phi\left(\int_X s d\mu\right) \\
&\leqslant \sup_{0 \leqslant s \leqslant f} \int_X \Phi(s) d\mu. \\
&\leqslant \int_X \Phi(f) d\mu.
\end{aligned}$$

□

Actually, in third to fourth line we have equality: why?

EXERCISE 2.23. *Suppose Φ is strongly convex, $\Phi(t_1 x_1 + t_2 x_2) < t_1 \Phi(x_1) + t_2 \Phi(x_2)$ if $x_1 \neq x_2$ and $t_1, t_2 \neq 0$. Then, equality holds in Jensen's inequality if and only if f is a.e. equal to a constant function.*

COROLLARY 2.2. *Let $1 \leqslant p < q \leqslant \infty$ and let (X, μ) be a probability space. Then, for measurable $f : X \to [0, +\infty]$,*

$$\|f\|_{L^p} \leqslant \|f\|_{L^q}.$$

PROOF. The case $q = \infty$ is clear. If $q < \infty$, since $t \mapsto t^{q/p}$ is convex on $[0, \infty)$, by Jensen's:

$$\left(\int_X |f|^p d\mu\right)^{q/p} \leqslant \int_X (|f|^p)^{q/p} d\mu = \int_X |f|^q d\mu.$$

□

2.6.1.2. *Hölder's inequality.* Let $1 \leqslant p \leqslant \infty$. The *conjugate exponent* is $1 \leqslant p' \leqslant \infty$ satisfying $\frac{1}{p} + \frac{1}{p'} = 1$.

THEOREM 2.12 (Hölder's inequality). *Let $f, g\colon X \to [0, +\infty]$ be measurable, and $1 \leqslant p \leqslant p' \leqslant \infty$ be conjugate exponents. Then, if $f \in L^p$ and $g \in L^{p'}$, $fg \in L^1$ and*

$$\int_X fg d\mu \leqslant \|f\|_{L^p}\|g\|_{L^{p'}}.$$

Moreover, equality holds if and only if the functions f^p and $g^{p'}$ are a.e. linearly dependent; i.e. there are $\lambda, \mu \in [0, \infty)$, not both zero, such that $\lambda f^p = \mu g^{p'}$ a.e.

We consider g as fixed, and build a probability measure around it, so that Jensen's inequality can be applied.

PROOF. If $\|g\|_{L^{p'}} = 0$, there is nothing to prove. Otherwise, let $E = \{x : g(x) \neq 0\}$.

$$\begin{aligned}
\int_X fg d\mu &= \int_E fg^{1-p'} \frac{g^{p'} d\mu}{\|g\|^{p'}_{L^{p'}}} \|g\|^{p'}_{L^{p'}} \\
&\leqslant \left(\int_E (fg^{1-p'})^p \frac{g^{p'} d\mu}{\|g\|^{p'}_{L^{p'}}} \right)^{1/p} \|g\|^{p'}_{L^{p'}} \\
&\quad \text{because } \frac{g^{p'} d\mu}{\|g\|^{p'}_{L^{p'}}} \text{is a probability measure on } E \\
&\quad \text{and } t \mapsto t^p \text{ is convex} \\
&= \left(\int_E f^p g^{(1-p')p+p'} d\mu \right)^{1/p} \|g\|^{p' - \frac{p'}{p}}_{L^{p'}} \\
&\leq \|f\|_{L^p}\|g\|_{L^{p'}}
\end{aligned}$$

because $(1 - p')p + p' = 0$ and $p' - \frac{p'}{p} = 1$.

Since $t \mapsto t^p$ is strongly convex, we have equality when $fg^{1-p'} = (f^p/g^{p'})^{1/p}$ is a.e. equal to a constant. □

EXERCISE 2.24. *Show that Hölder' s inequality holds with more than two exponents. If $\frac{1}{p_1} + \cdots + \frac{1}{p_n} = 1$, $1 \leqslant p_j \leqslant \infty$, then, for measurable $f_1, \ldots, f_n\colon X \to [0, \infty]$,*

$$\int_X f_1(x) \ldots f_n(x) d\mu(x) \leqslant \|f_1\|_{L^{p_1}} \ldots \|f_n\|_{L^{p_n}}.$$

Write down "continuous" versions of this inequality, with $f = f(x, t)$, $t \in Y$, where (Y, λ) is a measure space.

COROLLARY 2.3 (Cauchy-Schwarz inequality for integrals). *Let $f, g \in L^2$, with values in $\mathbb{R}$ or $\mathbb{C}$. Then,*

$$\left| \int_X f\bar{g} d\mu \right| \leqslant \|f\|_{L^2}\|g\|_{L^2}.$$

EXERCISE 2.25. *Provide another proof of Cauchy-Schwarz inequality by expanding the square in*

$$\int_X \int_X |f(x)g(y) - g(x)f(y)|^2 d\mu(x)d\mu(y).$$

You will have to change order of integration (Fubini's Theorem).

Hölder's inequality can be used to give a useful characterization of the L^p norm.

THEOREM 2.13. *Let $\mu \neq 0$ and $1 \leq p \leq \infty$. For $f \in L^p(\mu)$,*

$$\|f\|_{L^p} = \sup\left\{\left|\int_X fg d\mu\right| : \ \|g\|_{L^{p'}} = 1\right\}. \tag{2.6.1}$$

PROOF. In (2.6.1) $\geq$ follows from Hölder's inequality.

For the opposite inequality, we consider the case $1 < p < \infty$, and we can assume $\|f\|_{L^p} > 0$, otherwise any $g \in L^{p'}$ with $\|g\|_{L^{p'}} = 1$ would give equality. Let $E = \{x : \ f(x) \neq 0\}$ and $g = \bar{f}|f|^{p-2}\chi_E\|f\|_{L^p}^{1-p}$, so that $\int fg d\mu = \|f\|_{L^p}$, while

$$\|g\|_{L^{p'}}^{p'} = \int_E |f|^{p'(p-1)} d\mu \cdot \|f\|_{L^p}^{(1-p)p'} \leq 1.$$

We have then $\int fg d\mu = \|f\|_{L^p}^p$. □

We deduce a useful "linearization" of the L^p norm of a function.

COROLLARY 2.4. *Let $\mu \neq 0$ and let $f \geq 0$ be measurable. Then,*

$$\|f\|_{L^p} = \sup\left\{\left|\int_X fg d\mu\right| : \ g \geq 0, \ \|g\|_{L^{p'}} = 1\right\} \in [0, \infty].$$

EXERCISE 2.26. *Prove inequality $\leq$ in (2.6.1) for $p = 1$ and $p = \infty$.*

EXERCISE 2.27. *Prove corollary 2.4.*

2.6.1.3. *Minkowski's inequality.*

THEOREM 2.14. *Let $1 \leq p \leq \infty$. Then, if $f, g \in L^p(\mu)$,*

$$\|f + g\|_{L^p(\mu)} \leq \|f\|_{L^p(\mu)} + \|g\|_{L^p(\mu)}. \tag{2.6.2}$$

PROOF. If $\mu = 0$, there is nothing to prove. Otherwise, let $\frac{1}{p} + \frac{1}{p'} = 1$. By Theorem 2.13, we can "linearize" the nonlinear inequality:

$$\begin{aligned}\|f + g\|_{L^p(\mu)} &= \sup\left\{\left|\int_X (f+g)h d\mu\right| : \ \|h\|_{L^{p'}} = 1\right\} \\ &\leq \sup\left\{\left|\int_X (|fh| + |gh|) d\mu\right| : \ \|h\|_{L^{p'}} = 1\right\}\end{aligned}$$

$$
\begin{aligned}
&\leq \sup\left\{\left|\int_X |fh|d\mu\right| :\ \|h\|_{L^{p'}} = 1\right\} \\
&\quad + \sup\left\{\left|\int_X |gh|d\mu\right| :\ \|h\|_{L^{p'}} = 1\right\} \\
&= \|f\|_{L^p(\mu)} + \|g\|_{L^p(\mu)}.
\end{aligned}
$$

□

2.7. More on L^p spaces

2.7.1. Completeness of L^p spaces. We have seen that for $1 \leq p \leq \infty$ Minkowski's inequality holds,

$$\|f + g\|_{L^p} \leq \|f\|_{L^p} + \|g\|_{L^p}. \tag{2.7.1}$$

Hence, L^p is a normed space w.r.t. linear operations on functions, and the L^p-norm.

THEOREM 2.15. *Let (Ω, μ) be a measure space. For $1 \leq p \leq \infty$, L^p is a Banach space.*

PROOF. We consider the more difficult case $1 \leq p < \infty$ first. Let $\{f_n\}$ be a Cauchy sequence in L^p, and construct a subsequence $\{f_{n_k}\}$ as follows:

- pick n_1 such that $\|f_n - f_{n+j}\|_{L^p} \leq \frac{1}{2}$ for $n \geq n_1$ and $j \geq 1$;
- iterate by choosing $n_k \geq n_{k-1}$ such that $\|f_n - f_{n+j}\|_{L^p} \leq \frac{1}{2^k}$ for $n \geq n_k$ and $j \geq 1$.

Consider then the function series

$$s_m(x) = \sum_{j=1}^{m} |f_{n_j}(x) - f_{n_{j-1}}(x)| + |f_{n_0}(x)| \nearrow s_\infty(x), \text{ where } f_{n_0}(x) = f_1(x). \tag{2.7.2}$$

By our choice of the n_j's,

$$
\begin{aligned}
\|s_\infty\|_{L^p} &= \lim_{m\to\infty} \|s_m\|_{L^p} \text{ by MCT} \\
&\leq \lim_{m\to\infty} \sum_{j=1}^{m} \|f_{n_j} - f_{n_{j-1}}\|_{L^p} + \|f_1\|_{L^p} \text{ by Minkowski's inequality} \\
&\leq \sum_{j=1}^{\infty} \frac{1}{2^j} + \|f_{n_1} - f_1\|_{L^p} + \|f_1\|_{L^p} < \infty,
\end{aligned}
$$

hence $s_\infty(x) < \infty$ a.e. A fortiori, the telescopic series

$$f(x) = \sum_{j=1}^{m} f_{n_j}(x) - f_{n_{j-1}}(x) + f_{n_0}(x) = \lim_{m\to\infty} f_{n_m}(x), \tag{2.7.3}$$

converges *a.e.*, and we have a candidate limit function f. In fact,

$$\begin{aligned} \|f - f_n\|_{L^p} &\leq \|f - f_{n_j}\|_{L^p} + \|f_{n_j} - f_n\|_{L^p} \\ &\leq \sum_{k=j}^{+\infty} \|f_{k+1} - f_k\|_{L^p} + 1/2^j \text{ if } n \geq n_j \\ &\leq \sum_{k=j}^{+\infty} 1/2^k + 1/2^j = C/2^j \to 0 \text{ as } j \to \infty. \end{aligned}$$

Case $p = \infty$. If $\{f_n : \ n \geq 1\}$ is Cauchy in L^∞, then for each $\epsilon > 0$ there is $n(\epsilon) > 0$ so that, for $n \geq n(\epsilon)$ and $j \geq 1$,

$$\text{ess-sup}_{x\in X} |f_{n+j}(x) - f_n(x)| \leq \epsilon. \tag{2.7.4}$$

Set $f_0 = 0$ and for $m, n \geq 0$ consider the sets $E_{m,n} = \{x \in X : \ |f_n(x) - f_m(x)| \geq \text{ess-sup}_{x\in X}|f_n(x) - f_m(x)|\}$. Each of them has zero measure, and (2.7.4) holds for all x in $X_1 := X \setminus \cup_{m,n} E_{m,n}$. Hence, for $x \in X_1$ we have that $\{f_n(x) : \ n \geq 1\}$ is Cauchy in $\mathbb{C}$, hence it converges to some $f(x)$ for $n \to \infty$, uniformly on X_1 because

$$|f_n(x) - f(x)| = \lim_{m\to\infty} |f_n(x) - f_m(x)| \leq \epsilon,$$

provided $n > n(\epsilon)$, with $n(\epsilon)$ as above. Hence, $\|f - f_n\|_{L^\infty} \to 0$ as $n \to \infty$.

Also, $f \in L^\infty$ because

$$|f(x)| \leq |f(x) - f_n(x)| + |f_n(x)| \leq \|f - f_n\|_{L^\infty} + \|f_n\|_{L^\infty} < \infty$$

(a priori, if n is chosen large enough; but a posteriori any f_n will do). □

As a byproduct of the proof we obtain a useful result on pointwise convergence.

THEOREM 2.16. *Let $1 \leq p < \infty$. If $\{f_n\}$ is a sequence in L^p, converging to f in L^p-norm, then there it has a subsequence $\{f_{n_j}\}$ such that* $\lim_{j\to\infty} f_{n_j}(x) = f(x)$ *a.e.*

PROOF. It follows from (2.7.3). □

We state and prove a useful fact that we will use several times.

THEOREM 2.17. *For $1 \le p < \infty$, the simple functions are dense in L^p.*

PROOF. Suppose $0 \le f \in L^p$ and let $0 \le s_n \nearrow f$ be a sequence of simple functions. Also observe that $|f - s_n|^p \le 2^p |f|^p \in L^1$. By Monotone Convergence Theorem, $\|f - s_n\|_{L^p} \to 0$. For the general case, split f into real and imaginary parts, and each of them in positive and negative part. □

2.7.2. Elementary, but useful.

2.7.2.1. *Series as integrals.* It is an interesting exercise translating the objects and theorems we have seen so far in the case where X is a set (the cases in which X is finite or countable are especially important), the σ-algebra is $\mathcal{F} = 2^X$, and the measure is the *counting measure* $\sharp$, where $\sharp(A)$ denotes the cardinality of A. We have then the following facts:

(i) For $f\colon X \to [0, \infty)$, $\int_X f d\sharp = \sum_{x \in X} f(x) = \sup_{A \subset X;\ \sharp(A) < \infty} f(x)$. Moreover, if $\sum_{x \in X} f(x) < \infty$, then $\{x \in X :\ f(x) \neq 0\}$ is at most countable.

(ii) **Monotone Convergence.** If $0 \le f_n(x) \le f_{n+1}(x)$ for all x in X and $n \ge 1$, then,

$$\lim_{n \to \infty} \sum_{x \in X} f_n(x) = \sum_{x \in X} \lim_{n \to \infty} f_n(x).$$

(iii) **Fatou's Lemma.** If $f_n(x) \ge 0$ for all x in X and $n \ge 1$, then

$$\sum_{x \in X} \liminf_{n \to \infty} f_n(x) \le \liminf_{n \to \infty} \sum_{x \in X} f_n(x),$$

and examples show that strict inequality can occur.

(iv) **Dominated Convergence.** If there is a summable function $g \ge 0$ (i.e. $\sum_x g(x) < \infty$) and $|f_n(x)| \le g(x)$ for all x in X and $n \ge 1$, then

$$\lim_{n \to \infty} \sum_{x \in X} f_n(x) = \sum_{x \in X} \lim_{n \to \infty} f_n(x).$$

(v) For $1 \le p \le \infty$, the $\ell^p = \ell^p(X)$ norm of f is defined by

$$\|f\|_{\ell^p} = \left(\sum_x |f(x)|^p \right)^{1/p}$$

if $1 \le p < \infty$, and $\|f\|_{\ell^\infty} = \sup_x |f(x)|$.

Hölder's and Minkowski's inequalities hold.

EXERCISE 2.28. *Prove the second assertion in item (i).*

2.7.2.2. *Inclusions of L^p spaces.* If $1 \le p < q \le \infty$, then:

(i) if $\mu(X) < \infty$, then $L^p \supseteq L^q$;

(ii) $\ell^p \subseteq \ell^q$.

The first inclusion follows from Jensen's inequality:

$$\int_X |f|^q d\mu = \mu(X) \int_X (|f|^p)^{q/p} \frac{d\mu}{\mu(X)} \ge \mu(X) \left(|f|^p \frac{d\mu}{\mu(X)} \right)^{q/p}.$$

The second follows from the fact that if $0 \le t \le 1$ and $r > 1$, then $t \ge t^r$:

$$\begin{aligned} 1 &= \sum_x \frac{|f(x)|^p}{\sum_y |f(y)|^p} \\ &\ge \sum_x \left(\frac{|f(x)|^p}{\sum_y |f(y)|^p} \right)^{q/p} \\ &= \frac{\sum_x |f(x)|^q}{\left(\sum_y |f(y)|^p\right)^{q/p}}, \end{aligned}$$

hence, $\|f\|_{\ell^p} \ge \|f\|_{\ell^q}$.

EXERCISE 2.29. *Let $\varphi\colon [0,\infty) \to [0,\infty)$ be convex, $\varphi(0) = 0$. Prove that, if $\{a_n\}$ is a positive sequence, $a_n \ge 0$, then*

$$\sum_n \varphi(a_n) \le \varphi\left(\sum_n a_n\right).$$

2.8. Signed measures

Additive set functions taking both positive and negative values exist in nature: think of a distribution of electric charge as an example from physics. Such *signed measures* also play an important role in mathematics *per se.* For instance, as we shall see, they provide a unified and easy to work with framework for classical objects such as functions of bounded variation, absolutely continuous functions, Lipschitz functions, and more. The extension of much of what we cover to complex valued measures, or even measures with values in a linear space, is rather straightforward. We mostly stick to the real case in order to avoid unnecessarily cumbersome notation.

2.8.1. Absolutely continuous and mutually orthogonal measures. Let μ, ν be measures on a measurable space $(X, \mathcal{F})$. We say that ν is *absolutely continuous with respect to $\nu \ll \mu$*, if $\nu(E) = 0$ whenever $\mu(E) = 0$. We say that $\nu \perp \mu$ are *mutually orthogonal* if there are disjoint sets A, B such that $\mu(B) = 0 = \nu(A)$.

Here are some basic properties of these two relations.

(i) We have $\mu \ll \mu$. Also, if $\lambda \ll \mu \ll \nu$, then $\lambda \ll \mu$.

(ii) If $\lambda, \mu \leq \nu$ and $s, t \geq 0$, then $a\lambda + b\mu \ll \nu$.

(iii) If $\lambda \ll \mu$ and $\mu \perp \nu$, then $\lambda \ll \nu$.

(iv) If $\lambda, \mu \perp \nu$ and $s, t \geq 0$, then $s\lambda + t\mu \perp \nu$.

(v) More generally, if, for $n \geq 1$, $\lambda_n \perp \mu$ and $t_n \geq 0$, then $\sum_{n=1}^{\infty} t_n \lambda_n \perp \mu$.

(vi) We have $\lambda \ll \mu \ll \lambda$ if and only if λ and μ have the same null-sets.

(vii) If $\lambda \ll \mu$ and $\lambda \perp \mu$, then $\lambda = 0$.

EXERCISE 2.30. *Show (i-vii).*

2.8.2. Definition and basic properties. Let $(X, \mathcal{F})$ be a measurable space. A *signed measure* on X (in fact, on $\mathcal{F}$) is a map $\mu : \mathcal{F} \to (-\infty, \infty]$ such that $\mu(\emptyset) = 0$ and, if $\{A_n\}_{n=1}^{\infty}$ is a family of disjoint subsets in $\mathcal{F}$, then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n), \tag{2.8.1}$$

the convergence being absolute when the left hand side is finite.

EXERCISE 2.31. *If (2.8.1) holds with both sides finite, then convergence of the series is necessarily absolute. Why?*

EXERCISE 2.32. *Look back at the proof for positive measures, and show that, if $\{a_n\}_{n=1}^{\infty}$ is a family of sets in $\mathcal{F}$, and μ is a signed measure, then*

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu\left(\bigcup_{j=1}^{n} A_j\right).$$

EXERCISE 2.33. *Show that the following are signed measures on the Borel σ-algebra of $\mathbb{R}$.*

(i) *$\mu(E) = \int_E f(x)dx$, where $f \in L^1$ is real valued.*

(ii) *$\mu(E) = m(E) - \delta_0(E)$, where m is Lebesgue measure and δ_0 is Dirac's delta (a unit mass at the origin): $\delta_0(E) = 1$ if $0 \in E$ and $\delta_0(E) = 0$ if $0 \notin E$.*

(iii) *Let $\{x_n\}$ and $\{y_m\}$ be countable, disjoint sequences, both dense in $\mathbb{R}$. Define:*

$$\mu(E) = \sum_{x_n \in E} 2^{-n} - \sum_{y_m \in E} 2^{-m}.$$

Then, μ is a signed measure (where positive and negative charges are shuffled in a rather messy way).

(iv) $\mu(E) = \int_E f(x) d\alpha(x)$, *where* $\alpha : \mathbb{R} \to \mathbb{R}$ *is increasing and bounded, and* f *is continuous.*

Examples (ii) and (iii) could be kept in mind while the theory is developed. They suggest that it is a tricky endeavour finding the locations where the charge is positive/negative just by measuring $\mu(E)$ when, say, E ranges among open intervals. The existence of such locations is our next goal.

2.8.3. The Hahn decomposition theorem. Here we follow Doss, Raouf The Hahn decomposition theorem. Proc. Amer. Math. Soc. 80 (1980), no. 2, 377. We assume throughout that μ is a signed measure on $(X, \mathcal{F})$.

LEMMA 2.1. *Suppose* $A \in \mathcal{F}$ *is such that* $\mu(A) < \infty$. *Then, there is a negative set* $N \subseteq A$ *such that* $\mu(N) \leq \mu(A)$.

PROOF. **Claim.** We prove first that for all $\epsilon > 0$ there is $A_\epsilon \in \mathcal{F}$ such that $\mu(A_\epsilon) \leq \mu(A)$ and for all measurable $B \subseteq A_\epsilon$ one has $\mu(B) \leq \epsilon$.

Suppose the claim is false. Then there is $\epsilon > 0$ such that, if $A_\epsilon \subseteq A$ is measurable and $\mu(A_\epsilon) \leq \mu(A)$, then there is $B \subseteq A_\epsilon$ with $\mu(B) > \epsilon$. Thus,

(1) With $A_\epsilon = A$, There is $B_1 \subseteq A$ such that $\mu(B_1) > \epsilon$.

(2) Set $A_\epsilon = A \setminus B_1$ and observe that $\mu(A_\epsilon) = \mu(A) - \mu(B_1) < \mu(A)$. Then, there is $B_2 \subseteq A \setminus B_1 \subset A$ with $\mu(B_2) > \epsilon$.

(n) Iterating and setting $A_\epsilon = A \setminus (B_1 \cup \cdots \cup B_{n-1})$, which satisfies $\mu(A_\epsilon) = \mu(A) - \sum_{j=1}^{n-1} \mu(B_j) < \mu(A)$, we find $B_n \subseteq A \setminus (\cup_{j=1}^{n-1} B_j)$ with $\mu(B_n) > \epsilon$.

The sets B_n (≥ 1) are disjoint in A, hence $\mu(\cup_{n=1}^{\infty} B_n) = \infty$, contrary to our assumption on μ. Hence, the claim holds.

We use the claim iteratively.

(1) There is $A_1 \subseteq A$ such that $\mu(A_1) \leq \mu(A)$ such that, for all $B \subseteq A_1$, we have $\mu(B) \leq 1$.

(2) There is $A_2 \subseteq A_1$ such that $\mu(A_2) \leq \mu(A_1) \leq \mu(A)$ such that, for all $B \subseteq A_2$, we have $\mu(B) \leq 1/2$.

(n) There is $A_n \subseteq A_{n-1}$ such that $\mu(A_n) \leq \mu(A_{n-1}) \leq \mu(A)$ such that, for all $B \subseteq A_n$, we have $\mu(B) \leq 1/n$.

Let now $N = \cap_{n=1}^{\infty} A_n$, so that for $B \subseteq N$ we have $\mu(B) \leq 1/n$ for all $n \geq 1$, hence $\mu(B) \leq 0$: N is a negative set. Also, $\mu(N) = \lim_{n\to\infty} \mu(A_n) \leq \mu(A)$, as wished. □

THEOREM 2.18 (Hahn Decomposition Theorem). *Let μ be a signed measure. Then, there exist disjoint, measurable sets E, F such that $E \cup F = X$, E is negative, and F is positive. The decomposition is unique in the sense that, for an analogous decomposition with negative E_1 and positive F_1, $E \Delta E_1 = F \Delta F_1$ is a null set. Moreover, $\mu(E) < 0$ if μ is not a positive measure, and $\mu(F) > 0$ if $-\mu$ is not a positive measure.*

EXERCISE 2.34. *Find couples E, F for the examples in Exercise 2.33.*

PROOF. Let

$$a = \inf\{\mu(A) : A \in \mathcal{F}\},$$

and let $\{A_n\}_{n=1}^{\infty}$ be a sequence of sets such that $a = \lim_{n\to\infty} \mu(A_n)$. By Lemma 2.1, there are negative $N_n \subseteq A_n$ with $\mu(n_n) \leq \mu(A_n)$, hence, $\lim_{n\to\infty} \mu(N_n) = a$. Set $N = \cup_{n=1}^{\infty} N_n$, which is a negative set with $\mu(N) \leq \mu(N_n)$ for all n, hence $\mu(N) = a$. Let $P = X \setminus N$. If P where not positive, we could find a measurable $E \subseteq P$ with $\mu(E) < 0$, but this way

$$\mu(N \cup E) = \mu(N) + \mu(E) < a,$$

which is a contradiction. Hence, P is positive and $X = N \cup P$ is the desired decomposition. □

2.8.4. The Jordan decomposition theorem.

THEOREM 2.19 (Jordan Decomposition Theorem). *Let μ be a signed measure on $(X, \mathcal{F})$. Then, there exist positive measures μ_+, μ_-, such that (i) $\mu = \mu_+ - \mu_-$; and (ii) $\mu_+ \perp \mu_-$. Moreover, such decomposition is unique.*

EXERCISE 2.35. *Consider $\mu = \delta_0 - \delta_1$ on $\{0, 1\}$. Find all couples of measures $\alpha, \beta \geq 0$ such that $\alpha - \beta = \mu$.*

PROOF. Let $X = E \cup F$ be a Hahn decomposition of X with respect to μ, and let $\mu_+ = \chi_F \mu$, $\mu_- = -\chi_E \mu$. Then, $\mu_+ \perp \mu_-$, and $\mu_+ - \mu_- = \mu$.

Suppose $\mu = \nu_+ - \nu_-$ is a different decomposition of μ satisfying properties (i-ii), with $\nu_+(G) = 0 = \nu_-(H)$, $G \cup H = X$. Then,

$$\mu_+(F) = \mu(F) = \nu_+(F) - \nu_-(F) \leq \nu_+(F),$$

with strict inequality if and only $0 < \nu_-(F) = \mu(F \cap G)$. If this were the case,

$$0 < \nu_+(E) = \mu(E \cap H),$$

contradicting the assumption that E is negative for μ. Thus, $\nu_+(F) = 0$ The same argument shows that $\nu_-(E) =$.

Finally, for any measurable A we have

$$\begin{aligned}\nu_+(A) &= \nu_+(A\cap E)+\nu_+(A\cap F)=\nu_+(A\cap E)\\ &= \nu_+(A\cap F)-\nu_-(A\cap F)=\mu(A\cap F)\\ &= \mu_+(A),\end{aligned}$$

and similarly $\nu_-(A)=\mu_-(A)$. □

The positive measure

$$|\mu| := \mu_+ + \mu_- \tag{2.8.2}$$

is called the *total variation measure* of the measure μ. We define

$$\|\mu\| = \|\mu\|_{\mathcal{M}(X)} := |\mu|(X) \in [0,\infty)] \tag{2.8.3}$$

to be the *total variation* of μ, and $\mathcal{M}(X)=\mathcal{M}(\mathcal{F})=\{\mu$ measure on $(X,\mathcal{F})$: $\|\mu\|_{\mathcal{M}(X)}\}<\infty$, the space of the *bounded signed measures.*

EXERCISE 2.36. (i) *The space $\mathcal{M}(X)$ is a linear space (with respect to which sum operation?), and the expression $\|\mu\|_{\mathcal{M}(X)}$ defines a norm on it.*

(ii) *Let $\lambda\ge 0$ be a measure on $(X,\mathcal{F})$ and, for $f\in L^1(\lambda)$ (real valued), define $d\lambda_f = f d\lambda$ (i.e.: $\lambda_f(E)=\int_E f d\lambda$). Show that $\|\lambda_f\|_{\mathcal{X}}=\|f\|_{L^1(\lambda)}$ (i.e., $f\mapsto\lambda_f$ is an isometry of $L^1(\lambda)$ into $\mathcal{M}(X)$).*

2.9. The Radon-Nikodym theorem

2.9.1. Orthogonality and absolute continuity for signed measures. Let μ,ν be signed measures on a measurable space $(X,\mathcal{F})$. We say that $\nu\ll\mu$, ν is *absolutely continuous with respect to* μ, in any of the two cases:

(i) $\mu\ge 0$ and $|\nu|\ll\mu$ (here μ can be an infinite measure);

(ii) μ and ν are finite and signed, and $|\nu|\ll|\mu|$.

We say that μ and ν are *mutually orthogonal*, $\mu\perp\nu$, if $|\mu|\perp|\nu|$.

We collect some immediate consequences of the definition.

PROPOSITION 2.6. (i) *If ν is signed and $\mu\ge 0$, then $\nu\ll\mu$ if and only if for all measurable E, $\mu(E)=0$ implies $\nu(E)=0$.*

(ii) *If ν is signed, $\nu=\nu_+-\nu_-$ is its Jordan decomposition, and $\mu\ge 0$; then $\nu\ll\mu$ if and only if $\nu_+\ll\mu$ and $\nu_-\ll\mu$.*

(iii) *$\mu\perp\nu$ if and only if $\nu_+\perp\mu_\pm$ and $\nu_-\perp\mu_\pm$.*

(iv) *If* λ, μ *are positive measures, or finite signed measures, then* $(\lambda - \mu)_+ \le \lambda_+ + \mu_-$ *and* $(\lambda - \mu)_- \le \lambda_- + \mu_+$. *If, in particular,* $\lambda, \mu \ge 0$, *then* $(\lambda - \mu)_+ \le \lambda$ *and* $(\lambda - \mu)_- \le \mu$.

Proof. (i) We prove the "if" part. Suppose $\mu(E) = 0$ and let $X = P \cup N$ be a Hahn decomposition of X with respect to ν, with P positive and N negative for ν. If $\mu(E) = 0$, then

$$|\nu|(E) = \nu_+(E) + \nu_-(E) = \nu(E \cap P) - \nu(E \cap N) = 0,$$

because $\nu(E \cap P) \le \mu(E \cap P) \le \mu(E) = 0$, and similarly $|\nu(E \cap N)| \le \mu(E \cap N) \le \mu(E) = 0$.

In the "only if direction" if E is measurable and $\mu(E) = 0$, then $\nu_+(E) = \nu(E \cap P) = 0$ because $\mu(E \cap P) = 0$, and similarly $\nu_-(E) = 0$, hence $|\nu|(E) = 0$.

(ii) By definition $\nu \ll \mu$ if and only if $|\nu| = \nu_+ + \nu_- \ll \mu$, which is equivalent to $\nu_\pm \ll \mu$.

(iii) Suppose $|\mu| \perp |\nu|$ ("only if" direction), and let $X = A \cup B$ with $|\nu|(A) = 0 = |\mu|(B)$. Then, $\nu_+(A) = 0 = \mu_+(B)$, hence $\nu_+ \perp \mu_+$. All other combinations of signs are similar. In the "if" direction, using (v) in exercise 2.30, we have

$$|\mu| = \mu_+ + \mu_- \perp \nu\pm,$$

and for the same reason $|\mu| \perp \nu_- + \nu_+ = |\mu|$.

(iv) Let $X = P \cup N$ be a Hahn decomposition for $\lambda - \mu$. For E measurable,

$$\begin{aligned}(\lambda - \mu)_+(E) &= (\lambda - \mu)(E \cap P) = \lambda(E \cap P) - \mu(E \cap P) \\ &\le \lambda_+(E \cap P) + \mu_-(E \cap P) \\ &\le \lambda_+(E) + \mu_-(E),\end{aligned}$$

and similarly one argues for $(\lambda - \mu)_-(E)$.

□

Next, we prove a sort of "uniform continuity" result when a measure is absolutely continuous with respect to another.

Theorem 2.20. *Let* $\mu \ge 0$ *be a measure on* $(X, \mathcal{F})$, *and* ν *be a finite, signed measure on the same space. The following are equivalent:*

(i) $\nu \ll \mu$

(ii) *for all* $\epsilon > 0$ *there is* $\delta > 0$ *such that for any measurable* E, *if* $\mu(E) \le \delta$, *then* $|\nu(E)| \le \epsilon$.

PROOF. Clearly (ii) implies (i). If $\mu(E) = 0$, then $\mu(E) \le \delta$ for all $\delta > 0$. This implies that $|\nu(E)| \le \epsilon$ for all $\epsilon > 0$, hence that $\nu(E) = 0$.

In the other direction, it suffices to show the property for $\nu \ge 0$, finite: we can then use the result for $\nu_\pm$, hence for $\nu = \nu_+ - \nu_-$. Suppose by contradiction that there is $\epsilon > 0$ such that, for all $n \ge 1$ there is a set E_n such that $\mu(E_n) \le \frac{1}{2^n}$ and $\nu(E_n) \ge \epsilon$. Set $F_m = \cup_{n \ge m}^m E_n \supseteq F_{m+1}$, and let $F = \cap_{m=1}^\infty F_m$. Since $\infty > \nu(F_m) \ge \nu(E_m) \ge \epsilon$, we have $\nu(F) \ge \epsilon$. On the other hand,

$$\mu(F) = \mu(F_m) \le \sum_{n \ge m} \mu(E_n) \le \frac{2}{2^m}$$

for all $m \ge 1$, hence, $\mu(F) = 0$, and we have contradicted (i). □

COROLLARY 2.5. *If* $f \in L^1(\mu)$, *with* $\mu \ge 0$, *for all* $\epsilon > 0$ *there is* $\delta > 0$ *such that, if* $\mu(E) \le \delta$, *then* $\int_E |f| d\mu \le \epsilon$.

We now show that, if μ and ν are finite and positive, and they are not mutually orthogonal, then ν ”contains” a nontrivial portion of μ (and viceversa).

LEMMA 2.2. *Let* $\mu, \nu \ge 0$ *be finite measures. Then, either* $\nu \perp \mu$, *or there exist a measurable* E *with* $\mu(E) > 0$ *and* $\epsilon > 0$ *such that, whenever* $A \subseteq E$ *is measurable,* $\nu(A) \ge \epsilon\mu(A)$.

The thesis, that is, is that $\chi_E(d\nu - \epsilon d\mu)$ is a positive measure.

PROOF. For $n \ge 1$ consider the signed measures $\nu - \frac{1}{n}\mu$, and let $X = P_n \cup N_n$ be the corresponding Hahn decomposition. Set $N = \cap_{n \ge 1} N_n$ and $P = \cup_{n \ge 1} P_n = X \setminus N$. Since $\nu \ge 0$ and N is a negative set for each $\nu - \frac{1}{n}\mu$, we have

$$0 \le \nu(N) \le \frac{\mu(N)}{n}$$

for all $n \ge 1$, hence, $\nu(N) = 0$. Now, there are two cases. Either $\mu(P) = 0$, then $\nu \perp \mu$; or $\mu(P) > 0$. In this second case, there exists $n_0 \ge 1$ such that $\mu(P_{n_0}) > 0$. Moreover, the measure $\nu - \frac{\mu}{n_0}$ is positive on P_{n_0} by definition of Hahn decomposition. Set $E = P_{n_0}$. □

2.9.2. The Radon-Nikodym theorem. A positive measure μ on $(X, \mathcal{F})$ is *σ-finite* if $X = \cup_{n=1}^\infty X_n$, where each X_n is measurable and $\mu(X_n) < \infty$.

THEOREM 2.21 (Radon-Nikodym decomposition of a measure). *Let $\mu \geq 0$ be σ-finite and ν either (i) signed and finite, or (ii) positive and σ-finite. Then, there exist measures $\rho \perp \mu$ and $\pi \ll \mu$ such that*

$$\nu = \rho + \pi. \tag{2.9.1}$$

The measures ρ, π are positive and σ-finite if $\nu \geq 0$, or signed and finite if ν is signed.

Also, there exists a measurable $f : X \to \mathbb{R}$ such that $d\pi = f d\mu$, i.e.

$$d\nu = d\rho + f d\mu. \tag{2.9.2}$$

Moreover, $f \geq 0$ if ν is positive and σ-finite, and $f \in L^1(\mu)$ if ν is signed and finite.

Finally, ρ and f (μ-a.e., the latter) are uniquely determined by μ and ν.

The main part of the proof is based on the following speculation, where μ and ν are positive and finite. If we have (2.9.2) with everything positive and $\rho \perp \mu$, how do we recover f? If $0 \leq g \leq f$ and E is measurable, then the *key inequality* $\int_E g d\mu \leq \nu(E)$ holds. Let $\mathcal{A}$ be the class of the g's satisfying the key inequality for all E. It is an educated guess that $g = f$ is maximal in some sense. Maximal for what? We introduce the functional $L(g) = \int_X g d\mu$, which measures the "size" of $g \in \mathcal{A}$. It turns out that the function f we are looking for is the one maximizing L.[1]

PROOF. **Case $\mu, \nu \geq 0$, finite.** Let

$$\mathcal{A} = \{g : X \to [0, \infty] : \ g \text{ is measurable and } \int_E g d\mu \leq \nu(E) \text{ for all measurable } E\}.$$

We have $0 \in \mathcal{A}$, and $\max(g, h) \in \mathcal{A}$ if $g, h \in \mathcal{A}$:

$$\begin{aligned} \int_E \max(g,h) d\mu &= \int_{\{x:g(x)<h(x)\}\cap E} h d\mu + \int_{\{x:g(x)\geq h(x)\}\cap E} g d\mu \\ &\leq \nu(\{x : g(x) < h(x)\} \cap E) + \nu(\{x : g(x) \geq h(x)\} \cap E) = \nu(E). \end{aligned}$$

[1] The fact that the mathematical object solving a specific problem is that maximizing a seemingly loosely related functional, is a recurrent theme in mathematical analysis.

Set

$$\begin{aligned}
a &= \sup\{\int_X g d\mu : g \in \mathcal{A}\} \\
&= \lim_{n\to\infty} \int_X g_n d\mu \\
&\quad \text{for some sequence } \{g_n\} \text{ in } \mathcal{A} \\
&\leq \lim_{n\to\infty} \int_X h_n d\mu \\
&\quad \text{where } h_n = \max\{g_1, \ldots, g_n\} \in \mathcal{A},\ h_n \leq h_{n+1} \\
&= \int_X \lim_{n\to\infty} h_n d\mu \\
&\quad \text{by monotone convergence} \\
&\leq a,
\end{aligned}$$

where the last inequality holds because $h_n \in \mathcal{A}$ for all $n \geq 1$. Let $f = \lim_{n\to\infty} h_n d\mu = \sup_{n\geq 1} h_n$. We have $f \in \mathcal{A}$ since, by monotone convergence,

$$\int_E f d\mu = \lim_{n\to\infty} \int_E h_n d\mu \leq \nu(E).$$

That is, $d\nu - f d\mu \geq 0$. We claim that $d\nu - f d\mu \perp d\mu$. If such were not the case, by lemma 2.2 there would exist measurable E with $\mu(E) > 0$ and $\epsilon > 0$ such that

$$\chi_E(d\nu - f d\mu) \geq \epsilon \chi_E d\mu,$$

i.e.

$$(d\nu - f d\mu) - \epsilon \chi_E d\mu \geq (d\nu - f d\mu - \epsilon d\mu)\chi_E \geq 0.$$

For all measurable F, then,

$$\int_F (f + \epsilon \chi_E) d\mu \leq \nu(F),$$

so that $f + \epsilon \chi_E \in \mathcal{A}$. On the other hand,

$$\int_E (f + \epsilon \chi_E) d\mu = a + \epsilon \mu(E) > a,$$

which contradicts the definition of a. Hence, $d\rho := d\nu - f d\mu \perp d\mu$, or, $d\nu = d\rho + f d\mu$, with $\rho \perp \mu$. Observe that in this case $f \in L^1(\mu)$, $f \geq 0$, and $\rho \geq 0$ is finite.

To show uniqueness, suppose that we have two decompositions

$$d\rho + f d\mu = d\nu = d\rho_0 + f_0 d\mu,$$

with $\rho \perp \mu \perp \rho_0$ and $f, f_0 \geq 0$ in $L^1(\mu)$. Then,

$$d\lambda_0 - d\lambda = (f - f_0)d\mu.$$

We have disjoint decompositions $X = A \cup B = A_0 \cup B_0$ with $\mu(A) = \mu(A_0) = 0$ and $\lambda(B) = \lambda_0(B_0) = 0$. Let $A' = A \cup A_0$ and $B' = B \cap B_0$, so that $X = A' \cup B'$ is disjoint decomposition. Moreover, $\mu(A') = 0$ and, by (iv) in proposition 2.6, $(\lambda - \lambda_0)_+(B') \leq \lambda(B') \leq \lambda(B) = 0$, and similarly $(\lambda - \lambda_0)_-(B') \leq \lambda_0(B') \leq \lambda_0(B_0) = 0$. Thus, $|\lambda - \lambda_0|(B') = 0$, and we conclude that $\lambda - \lambda_0 \perp \mu$.
Case when $\mu, \nu \geq 0$ are σ-finite, If $X = \cup_m X_m = \cup_n Y_n$ are disjoint, measurable decompositions of X with $\mu(X_m) < \infty$ and $\nu(Y_n) < \infty$ for all $m, n \geq 1$, then $X = \cup_{m,n}(X_m \cap Y_n)$ is a disjoint, measurable decomposition where each set has finite μ- and ν-measure. After renaming the sets, we might write $X = \cup_l W_l$, with $\mu(W_l) < \infty$ and $\nu(W_l) < \infty$. Applying the first case to $\mu_l = \mu|_{W_l}$ and $\nu_l = \nu|_{W_l}$, we have that

$$d\nu_l = f_l d\mu_l + d\rho_l,$$

where $f_l(x) = 0$ a.e. on $X \setminus W_l$, $\rho_l \perp \mu_l$, and $\rho_l(X \setminus W_l) = 0$. Let $f = \sum_l f_l$ and $\rho = \sum_l \rho_l$, so that

$$d\nu = \sum_l d\nu_l = \sum_l f_l d\mu_l + \sum_l \rho_l = f d\mu + d\rho.$$

We have to verify that $\rho \perp \mu$. If $W_l = A_l \cup B_l$ is a disjoint union with $\lambda_l(A_l) = 0 = \mu(B_l)$, then $X = (\cup_l A_l) \cup (\cup_l B_l)$, and after denoting $A = \cup_l A_l$ and $B = \cup_l B_l$, we have $\lambda(A) = 0 = \mu(B)$, as wished.
Case when $\mu \geq 0$ is σ-finite and ν is signed and finite. Let $X = P \cup N$ be the Hahn decomposition for ν, and use the previous case to decompose $d\nu_\pm = f_\pm d\mu + d\rho_\pm$, with $f_+|_N = f_-|_P$, $\rho_\pm \perp \mu$, and $\rho_+(N) = 0 = \rho_-(P)$. Set $\rho = \rho_+ - \rho_-$. Then, $|\rho| = \rho_+ + \rho_-$ is finite and $|\rho| \perp \mu$. Also, $f = f_+ - f_-$ belongs to $L^1(\mu)$, and

$$d\nu = f d\mu + d\rho$$

as wished.

About uniqueness, if we also had $d\nu = f_0 d\mu + d\rho_0$, then

$$d\rho_0 - d\rho = (f - f_0)d\mu.$$

Using again (iv) in proposition 2.6, we show that $\rho_0 - \rho \perp \mu$, but also $\rho_0 - \rho \ll \mu$, hence, $\rho_0 - \rho = 0$. This also implies that $f = f_0$ a.e. ◻

2.9.3. Application: the existence of the conditional expectation. Let $(X, \mathcal{F}, \mu)$ be a measure space, $\mu \geq 0$, and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra of $\mathcal{F}$. The *conditional expectation* $g = \mathbb{E}[f|\mathcal{G}]$ of a function $f \in L^1 X, \mathcal{F}, \mu)$ is a function $g : X \to [-\infty, +\infty]$ such that

(i) $g \in L^1(X, \mathcal{G}, \mu)$ is $\mathcal{G}$-measurable;

(ii) for all $G \in \mathcal{G}$, $\int_G g d\mu = \int_G g d\mu$.

This notion is foundational in probability theory, where the measure μ is typically assumed to be a probability measure: $\mu(X) = 1$.

THEOREM 2.22. *If $(X, \mathcal{F}, \mu)$ is σ-finite, $f \in L^1(X, \mathcal{F}, \mu)$, and $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$, then $\mathbb{E}[f|\mathcal{G}]$ exists, and it is uniquely a.e. defined as a function in $fL^1(X, \mathcal{G}, \mu)$.*

PROOF. For $G \in \mathcal{G}$, define $\nu(G) = \int_G f d\mu$. The set function ν defines a (finite) signed measure on $(X, \mathcal{G})$. In fact, if $G = \cup_{n=1}^{\infty} G_n$ is a countable union of disjoint sets in $\mathcal{G}$, then

$$\begin{aligned}
\nu(g) &= \int_X \sum_{n=1}^{\infty} \chi_{G_n} f d\mu \\
&= \sum_{n=1}^{\infty} \int_X f \chi_{G_n} d\mu \\
&\quad \text{by dominated convergence, since } f \in L^1, \\
&= \sum_{n=1}^{\infty} \nu(G_n).
\end{aligned}$$

The measure $\mu|_{\mathcal{G}}$, the restriction of μ to the class $\mathcal{G}$, is am positive, σ-finite measure on $\mathcal{G}$ (which we still denote by μ), and $\nu \ll \mu$. For $G \in \mathcal{G}$, in fact, if $\mu(G) = 0$, then $f\chi_G = 0$ a.e., hence,

$$\nu(G) = \int_G f d\mu = 0.$$

If we apply Radon-Nikodym theorem, we find $g \in L^1(X, \mathcal{G}, \mu)$ such that $d\nu = g d\mu$, i.e.

$$\int_G f d\mu = \nu(G) = \int_G g d\mu,$$

hence g satisfies the properties defining the conditional expectation. About uniqueness, if g' has the same properties of g, then

$$\int_G (g - g') d\mu = 0$$

for all G in $\mathcal{G}$. Considering the subsets of $\mathcal{G}$ where $g - g'$ is, respectively, positive or negative, we have that $0 = \int_X (g - g')_+ d\mu = \int_X (g - g')_- d\mu$, which implies that $(g - g')_+ = (g - g')_- = 0$ μ-a.e., hence $g = g'$ a.e. □

An alternative proof of the existence of the conditional expectation is by means of Hilbert space projections.

2.9.4. Application: the dual space of L^p for $1 \le p < \infty$. We consider here real valued functions. Complex valued functions can be dealt with in exactly the same way.

Let $(X, \mathcal{F}, \mu)$ be a measure space, and let $1 \le p, p' \le \infty$ be *conjugate*, $\frac{1}{p} + \frac{1}{p'} = 1$. Let $f \in L^{p'}$, and define the functional

$$\Lambda_f(h) = \int_X h f d\mu. \tag{2.9.3}$$

The following is a simple consequence of Hölder's inequality.

Proposition 2.7. *We have that $\Lambda_f : L^p \to \mathbb{R}$ is well defined. Moreover,*

$$\|\Lambda_f\|_{\mathcal{B}(L^p,\mathbb{R})} := \sup\left\{ \frac{|\Lambda_f(H)|}{\|h\|_{L^p}} : h \in L^p \setminus \{0\} \right\} = \|f\|_{L^{p'}}. \tag{2.9.4}$$

Proof. In one direction, by Hölder we have

$$\frac{|\Lambda_f(H)|}{\|h\|_{L^p}} = \frac{\left|\int_X hf\right|}{\|h\|_{L^p}} \le \frac{\|h\|_{L^p}\|f\|_{L^{p'}}}{\|h\|_{L^p}},$$

which shows that $\|\Lambda_f\|_{\mathcal{B}(L^p,L^{p'})} \le \|f\|_{L^{p'}}$. To obtain the opposite inequality, let $h = f|f|^{p'-2}$, so that

$$\|h\|_{L^p} = \left(\int_X |f|^{(p'-1)p} d\mu\right)^{1/p} = \left(\int_X |f|^{p'} d\mu\right)^{1/p} = \|f\|_{L^{p'}}^{p'/p} = \|f\|_{L^{p'}}^{p'-1},$$

while

$$\int_X h f d\mu = \int_X |f|^{p'} d\mu = \|f\|_{L^{p'}}^{p'}.$$

Hence,

$$\frac{|\Lambda_f(H)|}{\|h\|_{L^p}} = \|f\|_{L^{p'}},$$

so $\|\Lambda_f\|_{\mathcal{B}(L^p,L^{p'})} \ge \|f\|_{L^{p'}}$. □

When $1 \le p < \infty$, a converse to the proposition holds, at least in the σ-finite case.

THEOREM 2.23. *Let μ be a σ-finite measure on a measurable space $(X, \mathcal{F})$, $1 \le p < \infty$, and let $\Lambda : L^p(\mu) \to \mathbb{R}$ be a linear functional such that*

$$\|\Lambda\|_{\mathcal{B}(L^p,\mathbb{R})} := \sup\left\{\frac{|\Lambda(H)|}{\|h\|_{L^p}} : h \in L^p, \|h\|_{L^p} \neq 0\right\} < \infty.$$

Then, there exists $g \in L^{p'}(\mu)$ such that $\Lambda = \Lambda_g$, the functional defined in (2.9.3). Moreover, $\|\Lambda\|_{\mathcal{B}(L^p,\mathbb{R})} = \|g\|_{L^{p'}}$.

The function representing Λ is unique.

Like in all representation theorems, the problem is to extract a *concrete* object from an *abstract* one. Our abstract object here is the functional Λ. If A, B are disjoint, measurable sets with finite measure, then $\chi_A, \chi_B \in L^p(\mu)$ and $\chi_A + \chi_B = \chi_{A\cup B}$, so $A \mapsto \nu(A) := \Lambda(\chi_A)$ induces a finitely additive set function on $\mathcal{F}$. If we can prove that ν is a measure on $\mathcal{F}$, we are in business with a "concrete", measure theoretic object to work with.

PROOF. Consider first the case when $\mu(X) < \infty$. For A in $\mathcal{F}$, define

$$\nu(A) := \Lambda(\chi_A), \tag{2.9.5}$$

which is defined since $\chi_A \in L^p$ because μ is finite.

We show that the set function ν is countably additive, hence a signed measure. Suppose that $\{A_n\}$ is a family of disjoint, measurable sets in X. Then,

$$\sum_{n=1}^{\infty} \chi_{A_n} = \chi_{\cup_{n=1}^{\infty} A_n} \tag{2.9.6}$$

converges in L^p. In fact $S_n := \sum_{j=1}^{n} \chi_{A_j} \nearrow \sum_{n=1}^{\infty} \chi_{A_n}$, and it is Cauchy in L^p,

$$\left\|\sum_{i=n}^{n+j} \chi_{a_i}\right\|_{L^p}^p = \sum_{i=n}^{n+j} \mu(A_i) \searrow 0$$

as $n \to \infty$. By the boundedness of Λ on $L^p(\mu)$,

$$\nu\left(\cup_{n=1}^{\infty} A_n\right) = \Lambda\left(\sum_{n=1}^{\infty} \chi_{A_n}\right) = \sum_{n=1}^{\infty} \Lambda(L_{A_n}) = \sum_{n=1}^{\infty} \nu(A_n). \tag{2.9.7}$$

Also, if $\mu(A) = 0$, then $\chi_A = 0$ in $L^p(\mu)$, hence $\nu(A) = \Lambda\left(\chi_A\right) = 0$, showing that ν is absolutely continuous with respect to μ, $\nu \ll \mu$. By Radon-Nikodym Theorem, there is $g \in L^1(\mu)$ such that

$$\nu(A) = \int_A g d\mu = \int_X \chi_A g d\mu. \tag{2.9.8}$$

By linearity, the same happens for all simple functions s,

$$\Lambda(s) = \int_X sg d\mu. \tag{2.9.9}$$

We have more: $g \in L^{p'}$. Since the simple functions are dense in L^p, in fact,

$$\begin{aligned} \|g\|_{L^{p'}} &= \sup\left\{\left|\int_X sg d\mu\right| : \ s \text{ simple}, \|s\|_{L^p} \leq 1\right\} \\ &= \sup\{|\Lambda(s)| : \ s \text{ simple}, \|s\|_{L^p} \leq 1\} \\ &= \|\Lambda\|_{\mathcal{B}(L^p,\mathbb{R})} < \infty. \end{aligned}$$

Fix $h \in L^p$, and let $\{s_n\}$ be a sequence of simple functions converging to h in L^p. Then,

$$\left|\Lambda(h) - \int_X s_n g d\mu\right| = |\Lambda(h - s_n)| \leq \|\Lambda\|_{\mathcal{B}(L^p,\mathbb{R})} \|h - s_n\|_{L^p} \to 0, \tag{2.9.10}$$

as $n \to \infty$. On the other hand, by Hölder's inequality,

$$\begin{aligned} \left|\int_X hg d\mu - \int_X s_n g d\mu\right| &\leq \|g\|_{L^{p'}} \|h - s_n\|_{L^p} \\ &\to 0 \text{ as } n \to \infty. \end{aligned}$$

Together with (2.9.10), we have then that $\Lambda(h) = \int_X hg d\mu$, as wished.

Suppose now that μ is σ-finite on X: $X = \cup_{n=1}^{\infty} X_n$, where the summands are disjoint and $\mu(X_n) < \infty$ for each of them. For $n \geq 1$ let $\mu_n = \mu|_{X_n}$ be the restriction of μ to X_n, $\mu_n(E) = \mu(E \cap X_n)$, and for $h \in L^p(\mu)$, let $h_n = h\chi_{X_n}$. Any function in $L^p(\mu_n)$ can be isometrically identified with a function of the form h_n. Observe that $h = \sum_n h_n$ converges in $L^p(\mu)$ by dominated convergence. Apply the finite case of the theorem to the measure μ_n. The functional Λ restricted to functions vanishing outside X_n, which can be identified with a functional Λ_n on $L^p(\mu_n)$, has the form

$$\Lambda(\chi_{X_n} h) = \Lambda_n(h_n) = \int_X h_n g_n d\mu_n = \int_{X_n} h g_n d\mu,$$

where $g_n \in L^{p'}(\mu_n)$ vanishes outside X_n.

Let $g = \sum_n g_n$. We first prove that $g \in L^{p'}(\mu)$. For any h in $L^p(\mu)$, in fact,

$$\begin{aligned}
\left| \int_{\cup_{n=1}^N X_n} hg d\mu \right| &= \left| \int_X \left(\sum_{n=1}^N h_n \right) g d\mu \right| \\
&= \left| \Lambda_n \left(\sum_{n=1}^N h_n \right) \right| \\
&\leq \|\Lambda\|_{\mathcal{B}(L^p(\mu),\mathbb{R})} \left\| \sum_{n=1}^N h_n \right\|_{L^p(\mu)} \\
&\leq \|\Lambda\|_{\mathcal{B}(L^p(\mu),\mathbb{R})} \|h\|_{L^p(\mu)} .
\end{aligned}$$

Passing to sup over $\|h\|_{L^p(\mu)} \leq 1$, we obtain that $\|g\|_{L^{p'}(\mu)} \leq \|\Lambda\|_{\mathcal{B}(L^p(\mu),\mathbb{R})}$.

If $h \in L^p(\mu)$, by dominated convergence and the fact that $\left| hg\chi_{\cup_{n=1}^N X_n} \right| \leq |hg| \in L^1(\mu)$, we have:

$$\lim_{N\to\infty} \int_{\cup_{n=1}^N X_n} hg d\mu = \int_X hg d\mu. \tag{2.9.11}$$

We can finally conclude, for any $h \in L^p(\mu)$:

$$\begin{aligned}
\Lambda(h) &= \lim_{N\infty} \Lambda \left(\sum_{n=1}^N h_n \right) = \lim_{N\to\infty} \int_{\cup_{n=1}^N X_n} hg d\mu \\
&= \int_X hg d\mu,
\end{aligned}$$

as wished. □

2.10. Summary

Contents of the chapter §2.1 was purely motivational. Our first experience with integrals is *Cauchy's definition of integral*, where integrands are continuous functions on compact intervals. *Riemann's integral* allows more general integrands, but only under very special assumptions we can pass the limit under the integral sign. Both notions are based on approximations where the x-axis is partitioned into small intervals. The *Lebesgue integral* moves the partition to the y-axis: approximating functions are constant on *measurable sets*, of which we can calculate the *measure*.

In §2.2 we have developed the definition of the Lebesgue integral in an abstract setting. We have defined *σ-algebras* of measurable sets, hence *measurable functions* from measurable spaces to metric spaces, the most basic example of the latter being *simple functions*, which are piece-wise constant in a measurable sense. Finally, we have defined the Lebesgue integral of a positive, measurable function by a limiting procedure, and that of a complex valued function by splitting it into a linear combination of positive functions. An important example of σ-algebra is the *Borel σ-algebra* associated to a topology.

The promised reward for this effort comes in §2.3, where we have seen the three most used theorems about limits under the integral sign: the *Monotone Convergence Theorem* (MCT), *Fatou's Lemma*, the *Dominated Convergence Theorem* (DCT).

In §2.6 some integral inequalities related to convexity are proved: Jensen, Hölder, Minkowski. Finally, in §2.5.3 we define the L^p spaces, the Banach space of the measurable functions whose p-power is integrable (for $p < \infty$; L^∞ functions are those which are *essentially bounded*) and we prove the basic facts concerning them in §2.7.

So far, the only concrete measure we have is the counting measure. The results of the section, when particularized to this setting, provide nonetheless interesting facts about series, which are summarized in §2.7.2.1.

Spaces and operators In this chapter we have introduced the family of the Banach spaces L^p, $1 \leq p \leq \infty$. We also have defined the space of the signed measures.

CHAPTER 3

Product measures

Integrals with respect to two or more variables are an essential tool not just in higher dimensional calculus, but also in analysis on the real line. Think of the theorem on existence of weak derivatives of increasing functions, where we switched integration with respect to a Borel measure μ and Lebesgue measure. We will have an even stronger need of integrating on product structures when we introduce convolution.

3.1. σ-algebras on product spaces, product measures and Fubini's Theorem

3.1.1. Product σ-algebras. Let $(X, \mathcal{F}, \mu)$ and $(Y, \mathcal{G}, \nu)$ be measure spaces. A *measurable rectangle* is a set of the form $A \times B$, with $A \in$ $mathcalF$ and $B \in \mathcal{G}$. We consider the class $(\mathcal{F} \otimes \mathcal{G})_0$ containing disjoint unions of measurable rectangles. It is a routine exercise verifying that $(\mathcal{F} \otimes \mathcal{G})_0$ is a an algebra (draw some pictures): it is closed under union, complementation, and it contains $\emptyset$. The *product σ-algebra* is

$$\mathcal{F} \otimes \mathcal{G} := \sigma((\mathcal{F} \otimes \mathcal{G})_0),$$

the generated σ-algebra.

For a set E in $X \times Y$, $x \in X$, and $y \in Y$, consider the *x-section* E_x and the *y-section* E^y of E to be

$$E_x = \{y \in Y : \ (x, y) \in E\}, \ E^y = \{x \in X : \ (x, y) \in E\}.$$

Define analogously for $f : X \times Y \to \mathbb{C}$,

$$f_x(y) = f^y(x) = f(x, y) \text{ for all } x \in X, \ y \in Y; \ f_x : Y \to \mathbb{C}, \ f^y : X \to \mathbb{C}.$$

Lemma 3.1. (i) *If $E \in \mathcal{F} \otimes \mathcal{G}$, then $E_x \in \mathcal{G}$ and $E^y \in \mathcal{F}$ for all x in X and y in Y.*

(ii) *If $f : X \times Y \to \mathbb{C}$ is measurable with respect to $\in \mathcal{F} \otimes \mathcal{G}$, the f_x is $\mathcal{G}$-measurable and f^y is $\mathcal{F}$-measurable.*

Proof. (i) We show that the collection $\mathcal{H}$ of the sets E for which $E_x \in \mathcal{G}$ for all $x \in X$ (a) is a σ-algebra, (b) it contains the measurable rectangles; hence, it contains $E \in \mathcal{F} \otimes \mathcal{G}$.

(b) is clear: $(A \times B)_x = B$ if $x \in A$, and $(A \times B)_x = \emptyset$ if $x \in X \setminus A$, and both are elements of $\mathcal{G}$. (a) If $E \in \mathcal{H}$, then $(X \times Y \setminus E)_x = Y \setminus E_x \in \mathcal{G}$, so $X \times Y \setminus E \in \mathcal{H}$. Similarly, using the fact that $\left(\bigcup_n E_n\right)_x = \bigcup_n (E_n)_x$ we see that $\mathcal{H}$ is closed under countable unions.

(ii) For $f = \chi_E$, $E \in \mathcal{F} \otimes \mathcal{G}$, statements (ii) reduces to (i). By linearity of $f \mapsto f_x$, statement (ii) holds for simple functions. Let now f be measurable and positive on $X \times Y$, and let $f_n \nearrow f$ an approximation from below of f by means of positive, simple functions. Then $(f_n)_x \nearrow f_x$ is an approximation from below of f_x by simple functions, hence each f_x is measurable.

If f is real valued, split it $f = f_+ - f_-$ and use again linearity of $f \mapsto f_x$, and similarly is done if f is complex valued. □

3.1.2. The Monotone Class Lemma. Below, we need a general lemma on σ-algebras, which is also often used, for instance, in probability theory. The proof is not fun, but the result itself is very useful.

A family $\mathcal{M}$ of subsets of X is a **monotone class** when:

$$\text{if } \mathcal{M} \ni A_i \nearrow A, \text{ then } A \in \mathcal{M}, \text{ and if } \mathcal{M} \ni A_i \searrow A, \text{ then } A \in \mathcal{M}. \tag{3.1.1}$$

Lemma 3.2. *Let now $\mathcal{A}_0$ be a set algebra on X, $\mathcal{A} = \sigma(\mathcal{A}_0)$ be the smallest σ-algebra containing $\mathcal{A}_0$, and $\mathcal{M}(\mathcal{A}_0)$ be the smallest monotone class containing $\mathcal{A}_0$. Then, $\mathcal{A} = \mathcal{M}(\mathcal{A}_0)$.*

The way the theorem is normally used is the following.

Corollary 3.1. *If a property $\mathbf{P} = \mathbf{P}(A)$ is satisfied by the sets A belonging to a set algebra $\mathcal{A}_0$, and the $\mathbf{P}$ is preserved under monotone limits of sets ($A_i \nearrow A$, $A_i \searrow A$), then $\mathbf{P}$ holds for all sets in $\sigma(\mathcal{A}_0)$, the σ-algebra generated by $\mathcal{A}_0$.*

Proof. The family of the monotone classes is closed under intersections, hence, $\mathcal{M}(\mathcal{A}_0)$ is the intersection of all monotone classes containing $\mathcal{A}_0$. Since a σ-algebra is already a monotone class and it contains $\mathcal{A}_0$, then $\mathcal{A} \supseteq \mathcal{M}(\mathcal{A}_0)$.

Also, the class $\mathcal{M}(\mathcal{A}_0)_* = \{X \setminus E : \ E \in \mathcal{M}(\mathcal{A}_0)\}$ (i) is a monotone class (the complement, $E \mapsto X \setminus E$, switches countable unions of increasing families of sets, and countable intersections of decreasing families of sets); and (ii) contains all $E = (X \setminus E) \setminus E$, as E ranges over $\mathcal{A}_0$. Thus, $\mathcal{M}(\mathcal{A}_0)_* \supseteq \mathcal{M}(\mathcal{A}_0)$, and the opposite inclusion holds for the same reason. Hence, $\mathcal{M}(\mathcal{A}_0)$ is closed under complements.

For $E \in \mathcal{A}_0$, let $\mathcal{C}_E$ be the class of those $F \in \mathcal{M}(\mathcal{A}_0)$ such that

$$F \setminus E,\ E \setminus F,\ F \cap E,\ X \setminus (F \cup E) \in \mathcal{M}(\mathcal{A}_0).$$

Clearly $\mathcal{A}_0 \subseteq \mathcal{C}_E \subseteq \mathcal{M}(\mathcal{A}_0)$. Also, $\mathcal{C}_E$ is a monotone class. Let's verify it. If $F_n \nearrow$ in $\mathcal{C}_E$, then

$$\left(\bigcup_{n=1}^{\infty} F_n\right) \setminus E = \bigcup_{n=1}^{\infty} (F_n \setminus E) \in \mathcal{M}(\mathcal{A}_0),\ E \setminus \left(\bigcup_{n=1}^{\infty} F_n\right) = \bigcap_{n=1}^{\infty} (E \setminus F_n) \in \mathcal{M}(\mathcal{A}_0),$$

$$\left(\bigcup_{n=1}^{\infty} F_n\right) \cap E = \bigcup_{n=1}^{\infty} (F_n \cap E) \in \mathcal{M}(\mathcal{A}_0),$$

and finally

$$X \setminus \left(\left(\bigcup_{n=1}^{\infty} F_n\right) \cup E\right) = \bigcap_{n=1}^{\infty} (X \setminus (F_n \cup E)) \in \mathcal{M}(\mathcal{A}_0),$$

This shows that $\mathcal{C}_E$ is closed under unions of increasing sequences of sets, and the verification for the intersection of decreasing sequences is similar. By definition of $\mathcal{M}(\mathcal{A}_0)$, we have that $\mathcal{M}(\mathcal{A}_0) \subseteq \mathcal{C}_E$. Hence, $\mathcal{M}(\mathcal{A}_0) = \mathcal{C}_E$.

Let now $\mathcal{D}$ be the class of those sets E in $\mathcal{M}(\mathcal{A}_0)$ such that

$$F \setminus E,\ E \setminus F,\ F \cap E,\ X \setminus (F \cup E) \in \mathcal{M}(\mathcal{A}_0)$$

for all F in $\mathcal{M}(\mathcal{A}_0)$. We saw above that $\mathcal{D}$ contains $\mathcal{A}_0$. Also, $\mathcal{D}$ is a monotone class. Let's verify that it is closed under unions of increasing sequences of sets. If $E_n \nearrow$ is a sequences in $\mathcal{D}$ and $F \in \mathcal{M}(\mathcal{A}_0)$, then

$$F \setminus \left(\bigcup_n E_n\right) = \bigcap_n (F \setminus E_n) \in \mathcal{M}(\mathcal{A}_0),$$

because it is the intersection of a decreasing set sequence in $\mathcal{M}(\mathcal{A}_0$. Similarly on deals with the other properties. The punchline is that $\mathcal{D} \subseteq \mathcal{M}(\mathcal{A}_0)$ is a monotone class containing $\mathcal{A}_0$, hence $\mathcal{D} = \mathcal{M}(\mathcal{A}_0)$.

We put the pieces together. Since $\mathcal{M}(\mathcal{A}_0) = \mathcal{D}$, then $\mathcal{M}(\mathcal{A}_0)$ is closed under finite intersections, hence, being closed under complements, it is closed under finite unions; thus, it is an algebra. A set algebra which is closed under unions of increasing sequences of sets, is obviously closed under countable unions. We have proved that $\mathcal{M}(\mathcal{A}_0)$ is a σ-algebra containing $\mathcal{A}_0$, hence, it contains $\sigma(\mathcal{A}_0) = \mathcal{A}$. $\square$

3.1.3. Product measures and Cavalieri Lemma. The key step in the proof of Fubini Theorem is this special case, which is a sort of Cavalieri principle.

Lemma 3.3 (Cavalieri). *Suppose μ, ν are σ-finite. Then, for $E \in \mathcal{F} \otimes \mathcal{G}$,*

(i) *$x \mapsto \nu(E_x)$ is $\mathcal{F}$-measurable, and $y \mapsto \mu(E^y)$ is $\mathcal{G}$-measurable;*

(ii) *we have*

$$\int_X \nu(E_x) d\mu(x) = \int_Y \mu(E^y) d\nu(y). \tag{3.1.2}$$

We can write (3.1.2) as the equality of two iterated integrals:

$$\int_X \int_Y \chi_E(x, y) d\nu(y) d\mu(x) = \int_Y \int_X \chi_E(x, y) d\mu(x) d\nu(y).$$

Proof. Let's start with the case $\mu(X), \nu(Y) < \infty$. We show first that the class $\mathcal{A}$ of the subsets of $X \times Y$ for which (i)-(ii) hold, contains the algebra $(\mathcal{F} \otimes \mathcal{G})_0$. For a measurable rectangle $E = A \times B$ we have

$$\nu(E_x) = \chi_A(x)\nu(B); \;\; \mu(E^y) = \chi_B(y)\mu(A),$$

both measurable, and

$$\int_X \nu(E_x) d\mu(x) = \int_A \nu(B) d\mu(x) = \nu(B)\mu(A) = \int_B \mu(A) d\nu(y) = \int_Y \mu(E^y) d\nu(y),$$

thus $E = A \times B \in \mathcal{A}$.

If $E = \bigcup_{i=1}^n E_i$ is a finite union of disjoint measurable rectangles, and $x \in X$, then $E_x = \bigcup_{i=1}^n (E_i)_x$ is a finite union of disjoint measurable sets. Hence,

$$\nu(E_x) = \sum_{i=1}^n \nu((E_i)_x)$$

is a finite sum of $\mathcal{F}$-measurable functions of x, which is $\mathcal{F}$-measurable, so $E \in \mathcal{A}$.

Suppose now that $\mathcal{A} \ni E_n \nearrow E$. Then, $(E_n)_x \nearrow E_x$ and, by Monotone Convergence, $\nu((E_n)_x) \nearrow \nu(E^y)$. Similarly, $\mu((E_n)^y) \nearrow \mu(E^y)$, hence, again by Monotone Convergence,

$$\int_Y \mu(E^y) d\nu(y) = \lim_n \int_Y \mu((E)_n^y) d\nu(y) = \lim_n \int_X \nu((E_n)_x) d\mu(x) = \int_X \nu(E_x) d\mu(x),$$

showing that $E \in \mathcal{A}$.

If $\mathcal{A} \ni E_n \searrow E$, we repeat the reasoning *verbatim*, replacing Monotone by Dominated convergence, which we can do because $\mu(X), \nu(Y) < \infty$.

We have verified that

$$\mathcal{F} \times \mathcal{G} \supseteq \mathcal{A} \supseteq (\mathcal{F} \otimes \mathcal{G})_0$$

is a monotone class, hence, by the Monotone Class Theorem,

$$\mathcal{F} \times \mathcal{G} \supseteq \mathcal{A} \supseteq \sigma\left((\mathcal{F} \otimes \mathcal{G})_0\right) = \mathcal{F} \times \mathcal{G},$$

as promised.

Suppose now μ, ν are σ-finite:

□

We define the *product measure* $\mu \otimes \nu$ on $X \times Y$ by

$$(\mu \otimes \nu)(E) := \int_X \nu(E_x) d\mu(x) = \int_Y \mu(E^y) d\nu(y) \tag{3.1.3}$$

for E in $\mathcal{F} \otimes \mathcal{G}$. It satisfies $(\mu \otimes \nu)(A \times B) = \mu(A)\nu(B)$.

EXERCISE 3.1. *Verify that $\mu \otimes \nu$ is σ-additive.*

On the σ-finiteness condition When the measures are not σ-finite, Cavalieri's Lemma fails in a spectacular way, as the following example shows. In $\mathbb{R} \times \mathbb{R}$, consider on the first factor the counting measure $\mathcal{H}^0$ (which is not σ-finite), and the Lebesgue measure m on the second. Let $E = \{(x, x) : 0 \leq x \leq 1\}$. Then,

$$\int_{\mathbb{R}} \mathcal{H}^0(E^y) dm(y) = \int_0^1 1 dm(y) = 1,$$

while

$$\int_{\mathbb{R}} m(E_x) d\mathcal{H}^0(x) \int_{\mathbb{R}} 0 d\mathcal{H}^1(x) = 0.$$

Measurability is a delicate issue. If (i) E^y is measurable in X for all y, and $y \mapsto \mu(E^y)$ is measurable on Y, then the expression

$$\int_Y \mu(E^y) d\nu(y)$$

makes perfect sense. Still, E might not be measurable in $X \times Y$, as the following example shows. Consider $\mu = \nu = m$ to be Lebesgue measure on $\mathbb{R}$, and let L be a non-measurable set in $[0, 1]$. Let $E \subset \mathbb{R}^2$ be the union a horizontal copy of the interval $[0,1]$ for each $0 \le y \le 1$, translated differently according to the value of y:

$$E = \left(\bigcup_{y \in L} \{(x,y) : -1 \le x \le 0\} \right) \cup \left(\bigcup_{y \in [0,1] \setminus L} \{(x,y) : 0 \le x \le 1\} \right).$$

Each E^y is either empty or a segment, hence it is measurable; and $m(E^y) = \chi_{[0,1]}(y)$ is measurable, but E is not measurable in $\mathbb{R}^2$. If it were, by Cavalieri's Lemma, E_x would be measurable for all x, but $E_x = L$ for $x \in [-1,1] \setminus \{0\}$.

3.1.4. Fubini Theorem. Here, $(X, \mathcal{F}, \mu)$ and $(Y, \mathcal{G}, \nu)$ are measure spaces.

Theorem 3.1 (Fubini Theorem). *Suppose $(X, \mathcal{F}, \mu)$ and $(Y, \mathcal{G}, \nu)$ are σ-finite, and let $f : X \times Y \to \mathbb{C}$ be $\mathcal{F} \otimes \mathcal{G}$.measurable. If one of the following holds,*

(i) $f \ge 0$; *or*

(ii) $\iint_{X \times Y} |f(x,y)| d(\mu \otimes \nu)(x,y) < \infty$;

then,

(a) *for each $x \in X$ the function $f_x(y) := f(x,y)$ is $\mathcal{G}$-measurable, and for each $y \in Y$ the function $f^y(x) := f(x,y)$ is $\mathcal{F}$-measurable;*

(b) *the function $x \mapsto \int_Y f(x,y) d\nu(y)$ is $\mathcal{F}$-measurable, and the function $y \mapsto \int_Y f(x,y) d\mu(y)$ is $\mathcal{G}$-measurable;*

(c) *we have*

$$\iint_{X \times Y} f(x,y) d(\mu \otimes \nu)(x,y) = \int_X \left(\int_Y f(x,y) d\nu(y) \right) d\mu(x) \tag{3.1.4}$$

$$= \int_Y \left(\int_X f(x,y) d\mu(x) \right) d\nu(y). \tag{3.1.5}$$

Proof. (a-c)Equation reduce to the Lemma of Cavalieri when $f = \chi_E$, with E measurable. By linearity, they hold for all simple functions. By Monotone Convergence, they continue to hold for positive f's: first you apply MC to $s_n^y \nearrow f^y$ for each fixed y, then to $\left(y \mapsto \int_X s_n^y(x) d\mu(x)\right) \nearrow \left(y \mapsto \int_X f^y(x) d\mu(x)\right)$ (and repeat, switching the role of x and y). This gives Fubini for $f \ge 0$ (which is sometimes called *Tonelli Theorem*).

For real valued f, you split $f = f_+ - f_-$, and apply the previous case. ☐

There is a point we swept under the rug. It might be the case that (ii) holds and for some, say, $x \in X$, $\int_Y f_+(x,y)d\nu(y) = \int_Y f_-(x,y)d\nu(y) = \infty$. In these cases we redefine $\int_Y f(x,y)d\nu(y) = 0$. This is harmless, since the set of such points has zero μ-measure. If it did not, by the case (i) we would have in fact that the hypothesis in (ii) fails, $\iint_{X\times Y} |f(x,y)|d(\mu\otimes\nu)(x,y) = \infty$.

EXERCISE 3.2. *Write down the proof of Fubini's Theorem with all details.*

3.2. Some applications

3.2.1. Minkovski integral inequality. Just because, for $1 \leq p \leq \infty$, $\|\cdot\|_{L^p}$ is a norm, we have that

$$\begin{aligned} \left(\int_X \left|\sum_{i=1}^n a_i f_i(x)\right|^p d\mu(x)\right)^{1/p} &= \|\sum_{i=1}^n a_i f_i\|_{L^p} \\ &\leq \sum_{i=1}^n |a_i| \|f_i\|_{L^p} \\ &= \sum_{i=1}^n |a_i| \left(\int_X |f_i(x)|^p \, d\mu(x)\right)^{1/p}, \end{aligned} \tag{3.2.1}$$

where for $p = \infty$ we have the esssup instead,

$$\text{esssup}_{x\in X}\left|\sum_{i=1}^n a_i f_i(x)\right| \leq \sum_{i=1}^n |a_i| \text{esssup}_{x\in X}|f_i(x)|,$$

which is rather obvious.

By replacing the finite sum by an integral, we obtain a much used inequality.

THEOREM 3.2. *Let $(X, \mathcal{F}, \mu)$ and $(Y, \mathcal{G}, \nu)$ be measure spaces and let $f : X \times Y \to \mathbb{R}_+$ be measurable. Then,*

$$\left(\int_X \left(\int_Y f(x,y)d\nu(y)\right)^p d\mu(x)\right)^{1/p} \leq \int_Y \left(\int_X f(x,y)^p d\mu(x)\right)^{1/p} d\nu(y). \tag{3.2.2}$$

PROOF. By Fubini Theorem, $y \mapsto \int_Y f(x,y)d\nu(y)$ and $y \mapsto \int_X f(x,y)^p d\mu(x)$ are both measurable, with respect to $\mathcal{G}$ and $\mathcal{F}$, respectively; hence, both sides of the inequality makes sense. □

3.3. Convolution and Young's inequalities

3.3.1. Convolution. Let $f, g : \mathbb{R} \to \mathbb{C}$ be measurable functions. Their convolution is the function $f * g : \mathbb{R} \to \mathbb{C}$ defined as

$$(f * g)(x) = \int_{\mathbb{R}} f(x-y)g(y)dy.$$

There are issues of measurability (of the function $H(x,y) = f(x-y)g(y)$) and of convergence (of the integral defining $f * g$) which we will for the moment ignore.[1]

To illustrate the definition, consider $g_a(y) = \frac{1}{a}\chi_{[0,a]}$. Then

$$(f * g_a)(x) = \frac{1}{a}\int_{[0,a]} f(x-y)dy = \frac{1}{a}\int_{[x-a,x]} f(y)dy$$

is a *moving average* of f. Roughly speaking, all convolution operators $f \mapsto f * g$ can be viewed as linear combinations of moving averages.

Moving averages commute with translations. Let $\tau_b f(x) = f(x-b)$ be the forward shift of the function f by b units of (say) time. It is rather obvious that the moving average $g_a * f$ is shifted correspondingly.

PROPOSITION 3.1. *Let $f, g : \mathbb{R} \to \mathbb{C}$, and $b \in \mathbb{R}$. Then,*

$$\tau_b(f * g) = (\tau_b f) * g. \tag{3.3.1}$$

PROOF.

$$\begin{aligned} \tau_b(f * g)(x) &= (f * g)(x-b) = \int f(x-b-y)g(y)dy \\ &= \int \tau_b f(x-y)g(y)dy = (\tau_b f) * g(x). \end{aligned}$$

□

In other words, the operator $T_g f = f * g$ satisfies

$$\tau_b \circ T_g = T_g \circ \tau_b. \tag{3.3.2}$$

Invariance of the laws of Nature (and, hopefully, of technological items) in time and space, can be rephrased as

$$\tau_{t_0,x_0} T(f) = T(\tau_{t_0,x_0} f),$$

where $f = f(t,x)$ is an *input function* depending on time t and position x, τ is a shift in time and space, and $T(f)$ is the *output function* produced by a natural process, or by an artificial device, (a *system*) T, which still depends on time and space. A basic, heuristic principle, which has several mathematical avatars, is the following.

Principle of invariance by translations. *All* **linear** *systems T which are invariant with respect to position and time, have the form $f \mapsto f * g_T$, where g_T is a suitable mathematical object.*

[1](∗) Write the proof.

Laws of Nature and devices are invariant under spatial rotations as well. Mathematically taking this into account leads to convolutions on (some) Lie groups, which is not a topic we will touch.

Another important notion with entertaining mathematical developments, which are outside the scope of these lectures, is that of *causality*. Consider again $f = f(x)$, a function of time. If we have a *simultaneous* knowledge of the values of f (as we do when, for instance, have the record of a music piece), at each x we can compute the moving average $f * \left(\frac{1}{a}\chi_{[-a/2,a/2]}\right)$ considering both future and past values of $f(y)$ $(-a/2 + x \leq y \leq a/2 + x)$, since our record has all that information. If we want to compute an average in "real time", however (like in a transmission system, where we do not want any unnecessary delay), we are bound to use values of $f(y)$ with $y \leq x$. In the general case of a system $f \mapsto f * g$, we have then to require $g(y) = 0$ for $y < 0$, so that

$$(f * g)(x) = \int_0^\infty f(x - y)g(y)dy$$

only needs the information $f(z)$ when $z = x - y \leq x$. This kind of analysis naturally, and surprisingly, leads to holomorphic function theory.

3.3.2. Young's inequality.

THEOREM 3.3 (Young's inequality with $q = 1$). *Let* $f \in L^p(\mathbb{R})$ *and* $g \in L^1(\mathbb{R})$. *Then,*

$$\|f * g\|_{L^p} \leq \|f\|_{L^p}\|g\|_{L^1}. \tag{3.3.3}$$

PROOF. Using Minkovski's integral inequality applied to $\tau_y f(x) = f(x - y)$, $\|\tau_y f\|_{L^p} = \|f\|_{L^p}$,

$$\begin{aligned}
\|f * g\|_{L^p} &= \left(\int_{\mathbb{R}} \left|\int_{\mathbb{R}} f(x-y)g(y)dy\right|^p dx\right)^{1/p} \\
&= \left\|\int_{\mathbb{R}} \tau_y f \cdot g(y)dy\right\|_{L^p} \\
&\leqslant \int_{\mathbb{R}} \|\tau_y f\|_{L^p}|g(y)|dy \\
&= \|f\|_{L^p}\int_{\mathbb{R}} |g(y)|dy \\
&= \|f\|_{L^p}\|g\|_{L^1}.
\end{aligned}$$

□

A critical analysis of the proof shows that its main ingredient is the fact that the Lebesgue measure dx is invariant under translations, $\int_{\mathbb{R}} f(x-y)dx = \int_{\mathbb{R}} f(x)dx$. This hints at the fact that similar results hold in much greater generality (locally compact groups endowed with their left-invariant Haar measure).

3.3.3. Supplement: a more general Young's inequality. There is a family of Youngs' inequalities with an extra degree of freedom in the involved exponents.[2]

THEOREM 3.4. *Young's inequality with $q > 1$] et $f \in L^p(\mathbb{R})$ and $g \in L^q(\mathbb{R})$, and let $1 < p, q, r < \infty$ be such that $\frac{1}{p} + \frac{1}{q} - \frac{1}{r} = 1$. Then,*

$$\|f * g\|_{L^r} \leqslant \|f\|_{L^p} \|g\|_{L^q}$$

PROOF. We only use Hölder's inequality with three exponents and translation invariance of Lebesgue's measure. We can assume all functions are positive.

$$\begin{aligned}
\|f * g\|_{L^r} &= \left(\int_{\mathbb{R}} \left[\int_{\mathbb{R}} f(x-y) g(y) dy \right]^r dx \right)^{1/r} \\
&\leqslant \left(\int_{\mathbb{R}} \left[\left(\int_{\mathbb{R}} f(x-y)^{ar} g(y)^{cr} dy \right)^{1/r} \left(\int_{\mathbb{R}} f(x-y)^{bu} dy \right)^{1/u} \right. \right. \\
&\qquad \left. \left. \left(\int_{\mathbb{R}} g(y)^{dv} \right)^{1/v} \right]^r dx \right)^{1/r} \\
&\qquad \text{with exponents } r, u, v \text{ such that } \frac{1}{r} + \frac{1}{u} + \frac{1}{v} = 1 \\
&\qquad \text{and } a + b = 1,\ c + d = 1 \text{ to be chosen} \\
&\qquad \text{in such a way } ar = bu = p \text{ and } cr = dv = q \\
&= \left(\int_{\mathbb{R}} \int_{\mathbb{R}} f(x-y)^{ar} g(y)^{cr} dy \left(\int_{\mathbb{R}} f(x-y_1)^{bu} dy_1 \right)^{r/u} \right. \\
&\qquad \left. \left(\int_{\mathbb{R}} g(y_2)^{dv} dy_2 \right)^{r/v} dx \right)^{1/r} \\
&= \left(\int_{\mathbb{R}} \int_{\mathbb{R}} f(x-y)^p dx \left(\int_{\mathbb{R}} f(z_1)^p dy_1 \right)^{r/u} \left(\int_{\mathbb{R}} g(z_2)^q dy_2 \right)^{r/v} g(y)^q dy \right)^{1/r} \\
&= \|f\|_{L^p}^{p/r+p/u} \|g\|_{L^q}^{q/v+q/r}.
\end{aligned}$$

Our *desiderata* hold if $1 = a + b = p/r + p/u = p(1/r + 1/u)$ and $1 = c + d = q(1/r + 1/v)$, with the extra condition that

$$1 = \frac{1}{r} + \frac{1}{u} + \frac{1}{v} = \frac{1}{p} + \frac{1}{q} - \frac{1}{r},$$

[2]The best constant in Young's inequality was found by William Beckner in 1975. See https://arxiv.org/abs/math/9704210 for a much simplified proof. The coordinates of Beckner's article are in the references.

which is exactly the hypothesis. We can then solve and find:

$$1/u = 1/p - 1/r = 1/q', 1/v = 1/q - 1/r = 1/p',$$

and

$$a = p/r, b = p/u = 1 - p/r, c = q/r, d = q/v = 1 - q/r,$$

so that $p/r + q/u = 1 = q/v + q/r$. □

EXERCISE 3.3. *Create a Young's inequality for* $\|f * g * h\|_{L^r}$ *and prove it. Extend to the convolution of n functions.*

3.4. Some properties of convolution

3.4.1. Convolutions and continuity.

PROPOSITION 3.2. *Let* $f, g : \mathbb{R} \to \mathbb{R}$*, with* $f \in C_c$ *and* $g \in L^1_{loc}$*. Then,* $f * g$ *is continuous.*

If $g \in L^p$ *for some* $1 \leq p < \infty$*, then* $f * g \in C_0 \cap L^p$*. If* $g \in L^\infty$*, then* $f * g \in C_b$ *and it is uniformly continuous.*

PROOF. Suppose $\mathrm{supp}(f) \subset [-R, R]$, and $|h| \leq \delta \leq 1$. Then,

$$\begin{aligned} |(f*g)(x+h) - (f*g)(x)| &\leq \int_{\mathbb{R}} |f(x+h-t) - f(x-t)| \cdot |g(t)| dt \\ &\leq \int_{x-R-1}^{x+R+1} |g(t)| dt \cdot \sup_{y \in \mathbb{R},\ |h| \leq \delta} |f(y+h) - f(y)|, \end{aligned}$$

and the first factor is finite, while the second can be made smaller than any $\epsilon > 0$ by uniform continuity of f. This gives continuity of $f * g$. Membership in L^p follows from Young's inequality. If $g \in L^\infty$, the first factor is bounded by $2(R+1)\|g\|_{L^\infty}$, hence $f * g$ is uniformly continuous. We are left with showing that, for $1 \leq p < \infty$ and $g \in L^p$, $f * g \in C_0$. This follows from the estimate, where we use Hölder with respect to the (finite) measure $d\mu(t) := |f(x-t)|dt$:

$$\begin{aligned} |f * g(x)| &\leq \left(\int_{\mathbb{R}} |f(x-t)| \cdot |g(t)|^p \right)^{1/p} \left(\int_{\mathbb{R}} |f(x-t)| \right)^{1/p'} \\ &\leq \|f\|_{L^\infty}^{1/p} \|f\|_{L^1}^{1/p'} \left(\int_{x-R}^{x+R} |g(t)|^p \right)^{1/p}, \end{aligned}$$

which tends to 0 as $x \pm \infty$ by Dominated Convergence. □

Observe that in order to have $\lim_{x \to \pm\infty} f * g(x) = 0$ we just need f to be bounded, and to vanish outside some compact set.

In this, like in other similar "regularity results", the hypothesis that f have compact support can be relaxed by asking f to have some decay at $\pm\infty$, provided the growth of g at $\pm\infty$ is correspondingly kept under control. Often the function f is related to the solution of some partial differential equation, and its decay at $\pm\infty$ depends on the equation itself.

There are cases where continuity of the convolution is guaranteed even if none of the factors are continuous.

EXERCISE 3.4. *Let $f = \chi_{[-1/2,1/2]}$. Compute $f * f$, and verify that it is in fact continuous.*

More generally, we will see that, if $f, g \in L^2$, then $f * g$ is continuous. Observe that this corresponds to the case $p = q = 2$, $r = \infty$, of Young's inequality.

3.4.2. Derivative of a convolution. Derivatives can be thought of as averages on endpoints of infinitesimal intervals, so it is not surprising that they enter convolution products.

THEOREM 3.5. *Let $f, g : \mathbb{R} \to \mathbb{R}$, with $f \in C^1_c$ and $g \in L^1_{loc}$. Then, $f * g \in C^1$ and:*

$$(f * g)' = (f') * g. \tag{3.4.1}$$

Moreover,

(i) *if $g \in L^p$ with $1 \leq p < \infty$, then $f * g \in C^1_0 \cap L^p$;*

(ii) *if $g \in L^\infty$, then $f * g \in C^1_b$.*

PROOF. Suppose $\text{supp}(f) \subset [-R, R]$. The function $F(t, x) = f(t-x)g(x)$ satisfies the hypothesis of Theorem 2.9 on any interval $[a, b] \ni t$, since

$$|f(t - x)g(x)| \leq \max |f| \cdot |g(x)|\chi_{[-R+a,R+b]}(x),$$

hence, (3.4.1) holds. Since $f' \in C_c$, $f' * g$ is continuous, hence $f * g \in C^1$. Statements (i) and (ii) follow from Proposition 3.2. □

3.4.3. Approximate identities. We want here to construct smooth, compactly supported *approximate identities* (or *mollifiers*). These are families of functions $\{\varphi_\epsilon\} : \epsilon > 0$ in $C^\infty_c(\mathbb{R})$, indexed on ϵ, which should be considered as smooth approximations, in the L^1 sense, of the Dirac unit mass at the origin.

LEMMA 3.4. *There exists a family $\{\varphi_\epsilon\}_{\epsilon>0}$ in $C^\infty_c(\mathbb{R})$ such that:*

(i) *$\varphi_\epsilon \geq 0$ and $\int \varphi_\epsilon(x)dx = 1$;*

(ii) $supp(\varphi_\epsilon) \subseteq [-\epsilon/2, \epsilon/2]$.

Proof. Let $\varphi(x) = \begin{cases} ce^{\frac{1}{4x^2-1}} \text{ if } |x| < 1/2 \\ 0 \text{ if } |x| \geq 1/2. \end{cases}$ It is easy to see that (i) $\varphi > 0$; (ii) $\text{supp}(\varphi) = [-1/1, 1/2]$; (iii) $\int \varphi(x)dx = 1$, for a choice of $c > 0$. For $\epsilon > 0$, then, define

$$\varphi_\epsilon(x) := \frac{1}{\epsilon}\varphi\left(\frac{x}{\epsilon}\right). \tag{3.4.2}$$

Observe that $\|\varphi_\epsilon\|_{L^1} = 1$, $\varphi_\epsilon \in C_c^\infty(\mathbb{R})$, that $\text{supp}\varphi_\epsilon) \subseteq [-\epsilon/2, \epsilon/2]$, as wished. □

Sometimes one needs approximate identities which are supported on the positive half-axis. It suffices to start with $\psi(x) = \varphi(x - 1/2)$, then let $\psi_\epsilon(x) = \frac{1}{\epsilon}\psi(\frac{x}{\epsilon})$.

3.4.4. The smooth Urysohn lemma.

Theorem 3.6. *Let $K \subset V \subset \mathbb{R}$, K compact, V open. Then, there exists $h \in C_c^\infty$ such that $K \prec h \prec V$.*

Proof. We assume here $V \neq \mathbb{R}$, the other case being similar and easier. Let $3\epsilon = d(K, \mathbb{R} \setminus K) := \min\{|x - y| : \ x \in K, \ y \notin V\}$, and let $K_\epsilon = \{y : d(y, K) \leq \epsilon\} \subset V$. Since $y \mapsto d(y, K)$ is continuous (even 1-Lipschitz), K_ϵ is closed and bounded, hence compact. Moreover, for $y \in K_\epsilon$, $d(y, \mathbb{R} \setminus V) = 2\epsilon$.

Let $h = \varphi_\epsilon * \chi_{K_\epsilon}$. The support of h lies in $K_{2\epsilon} \subset V$, hence $h \prec V$. For $x \in K$, $h(x) = \int_{K_\epsilon} \psi_\epsilon(x - y)dy = \int \psi_\epsilon(x - y)dy = 1$, since $B(x, \epsilon/2) \subset K_\epsilon$. Hence, $K \prec h$. □

The proofs of a number of properties we saw depended on Urysohn Lemma. Critical reading shows that most of those statements, and their consequences, have then C^∞ versions, which we are now going to state. Their proof is the same, but for the fact that the smooth Urysohn Lemma is used instead. We only have to be careful and verify that in the proofs only C^∞-preserving operations are performed on the "Urysohn functions": sums, product... Such is not the case for the operation $(f, g) \mapsto \min(f, g)$, which preserves continuity, but not smoothness. Here we consider $\mathbb{R}$, but everything can be extended to $\mathbb{R}^d$ (hence, to manifolds).

3.4.4.1. *Some consequences.* Partitions of unity extend with no effort.

Theorem 3.7. *[Smooth partition of unity in $\mathbb{R}$] Let K be compact in $\mathbb{R}$, $K \subseteq V_1 \cup \dots V_n$, where each V_j is open. Then, there exist $h_i \prec V_i$, $h_i \in C_c^\infty(\mathbb{R})$, such that*

$$h_1 + \dots + h_n = 1$$

on K.

We have an important density theorem.

THEOREM 3.8. *If μ is a regular measure, then $C_c^\infty(\mathbb{R})$ is dense in L^p, if $1 \leq p < \infty$.*

The fundamental lemmas of the calculus of variations extend with no change in the proof.

LEMMA 3.5 (1^{st} Fundamental Lemma of the Calculus of Variations in $\mathbb{R}$). *Let $f \in L^1_{loc}(\mathbb{R})$ and suppose that for all $\varphi \in C_c^\infty(\mathbb{R})$ one has*

$$\int_{\mathbb{R}} f\varphi d\mu = 0. \tag{3.4.3}$$

Then, $f = 0$ a.e..

LEMMA 3.6 (2^{nd} Fundamental Lemma of the Calculus of Variations). *Let $f \in L^1_{loc}(\mathbb{R})$ and suppose that for all $\varphi \in C_c^\infty(\mathbb{R})$ with $\int_{\mathbb{R}} \varphi d\mu = 0$ one has*

$$\int_{\mathbb{R}} f\varphi d\mu = 0. \tag{3.4.4}$$

Then, f is a.e. equal to a constant.

Then, we have this consequence of the 2^{nd} lemma.

EXERCISE 3.5. *Write down the proof of some of the statements above.*

3.4.4.2. *The closure of the unit ball of $C^1[0,1]$ in the uniform norm.*

THEOREM 3.9. *The closure of B_1 in $C[0,1]$ with respect to the uniform norm is the unit ball of $Lip[0,1]$.*

PROOF. We have to show that any function f with $\|f\|_{\text{Lip}} \leq 1$ can be uniformly approximated by functions in B_1. First, extend such $f : [0,1] \to \mathbb{R}$ to the whole real line by setting $f(x) = f(0)$ if $x < 0$ and $f(x) = f(1)$ if $x > 1$. Such extension clearly preserves the Lipschitz constant of f. Let then $\varphi \in C_c^\infty(\mathbb{R})$, $\text{supp}(\varphi) \subset [-1,1]$, $\varphi \geq 0$, $\int_{\mathbb{R}} \varphi(t)dt = 1$. The existence of such a function will be proved later in the notes. We introduce an *approximation of identity* (or *Friedrich's mollifier*) by setting $\varphi_n(x) = n\varphi(nx)$, which has all the properties we listed for φ and, more, $\text{supp}(\varphi) \subset [-1/n, 1/n]$. Define

$$f_n(x) = f * \varphi_n(x) = \int_{-\infty}^{+\infty} f(x-t)\varphi_n(t)dt,$$

for $0 \leq x \leq 1$. We use the following properties of convolution.

(i) $\|f_n - f\|_u \to 0$ as $n \to \infty$ and $\|f_n\|_u \leq 1$.

(ii) $f_n' = f * \varphi_n' \in C[0,1]$.

(iii) $\|f_n'\|_u \leq 1$ for all $x \in [0,1]$, since

$$\begin{aligned}
\left|\frac{f_n(x+h) - f_n(x)}{h}\right| &= \left|\int_0^1 \frac{f(x+h-t) - f(x-t)}{h}\varphi(t)dt\right| \\
&\leq \int_0^1 \left|\frac{f(x+h-t) - f(x-t)}{h}\right| \varphi(t)dt \\
&\leq \|\varphi\|_{\text{Lip}}.
\end{aligned}$$

Properties (i-iii) say that $f_n \in B_1$ and they uniformly converge to f. $\square$

CHAPTER 4

Constructing measures

In this chapter we see some useful methods to construct measures, or to recognize when a measure looms in the shadow. A basic tool in the craft is Carathéodory's construction of a measure from an outer measure. The Extension Theorem of Carathéodory has many applications, such as the construction of the Lebesgue measure and, more generally, Lebesgue-Stieltjes measures; the construction of Hausdorff measures; et cetera. Here, we use it to prove Riesz Representation Theorem for measures, and deducing Lebesgue-Stieltjes measures from it. Riesz Representation Theorem might be seen as a statement about particular *distributions* (in the sense of the Theory of Distributions) living in rather general spaces. We will pursue further this *distributional viewpoint* trying to extract information from Lebesgue-Stieltjes measures. Distributions themselves will be the subject of a later chapter in the notes.

Here is a quick overview. Section 4.1 deals with outer measures and falls squarely into abstract measure theory. In Section 4.2 we consider measure structures associated to metric spaces: more precisely, Radon measures on locally compact metric spaces. The main result here is Riesz Representation Theorem for Measures, which is a cornerstone of real analysis, with many theoretical and practical applications. We specialize all this to the real line in the long Section 4.5, where Riesz Theorem allows us to quickly define the Lebesgue measure and, much more generally, Lebesgue-Stieltjes measures. The notion of distribution function creates a natural bijection between Borel measures on the line and right continuous, increasing functions. All operations involving increasing functions can be translated in operations involving measures, and viceversa: *the class of the Borel measures coincides with that of the Lebesgue-Stieltjes measures.* Since we believe (perhaps too optimistically) to know everything important about increasing functions, it follows that we know (more realistically) many interesting properties of Borel measures. Of particular interest for present and future developments is Theorem 4.12, showing that the Borel measures are the derivatives of the increasing functions, at least in a "weak", distributional sense. We end with the remark that the limiting process defining Lebesgue-Stieltjes measures goes through if the starting function is not increasing, but just of *bounded variation*; a fact we will be able to extract gratification from after we consider, in a later chapter, the connection between signed measures and functions of bounded variation. We conclude with Section 4.5, where we compare Lebesgue integrability and Riemann integrability. The story of Lebesgue integration started with Borel's characterization of the functions which are Riemann integrable, and here we reach the end of the loop we started when trying to motivate a new definition of integral.

There are at least two other extremely useful, and beautiful, ways to construct measures. One is the construction of Hausdorff measures, which opens the way to Geometric Measure Theory, and to a whole world of wondrous objects having "fractal dimension" and "self-similar features". A quick, but rigorous and in depth, introduction to this topic, is the short book by K.J. Falconer[1].

The other one is Kolmogorov Extension Theorem, which is a way to mathematically construct a stochastic process developing in time (where the number of "instants" to take into account is infinite), starting with the observation of its distribution at finite collections of such "instants". The theorem is part of any standard course in stochastic processes, and its proof can be found e.g. in Durrett's (advanced) introduction to probability[2].

4.1. Outer measures and Carathéodory's Extension Theorem

Outer measures are a special class of set functions. Their *raison d'être* is that they are defined on all subsets of a given set X: no *a priori* structure (algebra, σ-algebra, topology...) is assumed. We are interested in them in view of their applications to measure theory, but there are important set functions which are very different from measures, and are generated by certain outer measures. This is the case, for instance, of *set capacities* in Potential Theory, which generalize and formalize the notion of conductor capacity from electrostatics. If you want to know more about these natural, but tricky objects, you might start here[3].

4.1.1. Outer measures. Let X be a set. An *outer measure* μ^* on X is a set function $\mu^*\colon 2^X \to [0, +\infty]$ with the properties:

(i) $\mu^*(\emptyset) = 0$;

(ii) if $E \subset F$, then $\mu^*(E) \leq \mu^*(F)$;

---

[1] K.J. Falconer The Geometry of Fractal Sets: 85 (Cambridge Tracts in Mathematics) 1986

[2] Rick Durrett, Probability: Theory and Examples, Version 5 January 11, 2019 https://services.math.duke.edu/~rtd/PTE/PTE5_011119.pdf

[3] Irina Markina, Potential theory: the origin and applications, expository article

(iii) $\mu^*\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} \mu^*(E_n)$ if $E_1, \ldots, E_n, \ldots$ are subsets of X.

Outer measures often arise when we extend some positive set function from a subclass of sets to 2^X.

Proposition 4.1. *Let $\mathcal{A}$ be a family of sets in 2^X, such that X is covered by countably many elements in $\mathcal{A}$, $X = \bigcup_{n=1}^{\infty} A_n$, with each A_n in $\mathcal{A}$. Let $l\colon \mathcal{A} \to [0, +\infty]$ be a set function satisfying $l(\emptyset) = 0$.*

Define

$$\mu^*(E) := \inf\left\{\sum_{n=1}^{\infty} l(A_n) : \; E \subseteq \bigcup_{n=1}^{\infty} A_n\right\} \tag{4.1.1}$$

Then, μ^* is an outer measure.

Proof. Properties (i-ii) are obvious. If $\mu^*(E_n) = +\infty$ for some n, then (iii) holds for trivial reasons. Otherwise, for any $\epsilon > 0$ and each n we can find $\{A_m^n\}_{m=1}^{\infty}$ in $\mathcal{A}$ such that $E_n \subseteq \cup_{m=1}^{\infty} A_m^n$ and

$$\sum_{m=1}^{\infty} l(A_m^n) \leq \mu^*(E_n) + \frac{\epsilon}{2^n}.$$

Now, $\bigcup_{n=1}^{\infty} E_n \subset \bigcup_{n,m=1}^{\infty} A_m^n$, hence,

$$\mu^*\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \sum_{n,m=1}^{\infty} l(A_m^n) \leq \sum_{n=1}^{\infty}\left[\mu^*(E_n) + \frac{\epsilon}{2^n}\right] = \sum_{n=1}^{\infty} \mu^*(E_n) + \epsilon.$$

(iii) follows. □

Definition 4.1. *Let μ^* be an outer measure on X. A subset A of X is μ^*-**measurable** if for all subsets E in X:*

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \setminus A). \tag{4.1.2}$$

Observe that the inequality $\leq$ in (4.1.2) holds for all A's, so, in practice, we only have to verify $\geq$. A possible intuition of how (4.1.2) might be considered a natural guess for extracting measures from outer measures will be attempted at the end of the next section.

Theorem 4.1. *Let μ^* be an outer measure on X. Then, the class $\mathcal{F} = \mathcal{F}(\mu^*)$ of the μ^*-measurable sets is a σ-algebra, containing all sets A such that $\mu^*(A) = 0$. Moreover, μ^* is a measure on $\mathcal{F}$.*

Proof. Equation (4.1.2) becomes an identity for $A = \emptyset$, hence $\emptyset \in \mathcal{F}$. Also, (4.1.2) is symmetric with respect to A and $X \setminus A$: $E \setminus (X \setminus A) = E \cap A$, $E \cap (X \setminus A) = E \setminus A$, hence $A \in \mathcal{F}$ if and only if $X \setminus A \in \mathcal{F}$.

Let A, B two subsets in $\mathcal{F}$. Let $E \subseteq X$. Using twice the definition of $\mathcal{F}$, and twice subadditivity of μ^*, we have:

$$\begin{array}{rcl} \mu^*(E) & = & \mu^*(E \cap A) + \mu^*(E \setminus A) \\ & = & \mu^*(E \cap A \cap B) + \mu^*((E \cap A) \setminus B) \\ & + & \mu^*((E \setminus A) \cap B) + \mu^*((E \setminus A) \setminus B) \\ & = & \mu^*(E \cap (A \cap B)) + \mu^*(E \cap (A \setminus B) \\ & + & \mu^*(E \cap (B \setminus A)) + \mu^*(E \setminus (A \cup B)) \\ & \geq & \mu^*(E \cap (A \cup B)) + \mu^*(E \setminus (A \cup B)) \\ & \geq & \mu^*(E). \end{array}$$

Hence, we have equality all the way, hence, $A \cup B \in \mathcal{F}$. Since $\mathcal{F}$ is closed under complementation, $A \mapsto X \setminus A$, we also have that $A \cap B = X \setminus ((X \setminus A) \cup (X \setminus B)) \in \mathcal{F}$ if $A, B \in \mathcal{F}$. By iteration, $A_1 \cup \cdots \cup A_n \in \mathcal{F}$ if $A_1, \ldots, A_n \in \mathcal{F}$. For A, B in $\mathcal{F}$, disjoint, we have

$$\begin{array}{rcl} \mu^*(A \cup B) & = & \mu^*((A \cup B) \cap A) + \mu^*((A \cup B) \setminus A) \\ & = & \mu^*(A) + \mu^*(B), \end{array}$$

hence, μ^* is finitely additive on $\mathcal{F}$.

Consider now a countable $\{A_n\}_{n=1}^{\infty}$ of disjoint subsets of X in $\mathcal{F}$. For each $n \geq 1$ and E in X,

$$\begin{array}{rcl} \mu^*(E \cap (\bigcup_{i=1}^{n} A_i)) & = & \mu^*([E \cap (\bigcup_{i=1}^{n} A_i)] \cap A_n) + \mu^*([E \cap (\bigcup_{i=1}^{n} A_i)] \setminus A_n) \\ & = & \mu^*(E \cap A_n) + \mu^*(E \cap (\bigcup_{i=1}^{n-1} A_i)) \\ \cdots & & \\ & = & \sum_{i=1}^{n} \mu^*(E \cap A_i). \end{array}$$

Hence,

$$\begin{array}{rcl} \mu^*(E) & = & \mu^*(E \cap (\bigcup_{i=1}^{n} A_i)) + \mu^*(E \setminus (\bigcup_{i=1}^{n} A_i)) \\ & = & \sum_{i=1}^{n} \mu^*(E \cap A_i) + \mu^*(E \setminus (\bigcup_{i=1}^{n} A_i)) \\ & \geq & \sum_{i=1}^{n} \mu^*(E \cap A_i) + \mu^*(E \setminus (\bigcup_{i=1}^{\infty} A_i)). \end{array}$$

Let $n \to \infty$, and use subadditivity of μ^*:
(4.1.3)
$$\mu^*(E) \geq \sum_{i=1}^{\infty} \mu^*(E \cap A_i) + \mu^*(E \setminus (\bigcup_{i=1}^{\infty} A_i)) \geq \mu^*(E \cap (\bigcup_{i=1}^{\infty} A_i)) + \mu^*(E \setminus (\bigcup_{i=1}^{\infty} A_i)).$$

This shows that $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$, and also that, letting $E = \bigcup_{i=1}^{\infty} A_i)$ in (4.1.3), μ^* is countably additive.

If $\{B_n\}$ is a countable family in $\mathcal{F}$, then $\bigcup_{n=1}^{\infty} B_b = \bigcup_{n=1}^{\infty} [B_n \setminus (B_1 \cup \cdots \cup B_{n-1})]$ can be viewed as disjoint union of countably many sets in $\mathcal{F}$, hence it lies in $\mathcal{F}$.

Finally, if $\mu^*(A) = 0$, then, for E in X, $\mu^*(E \cap A) \leq \mu^*(A) = 0$, thus

$$\mu^*(E \cap A) + \mu^*(E \setminus A) = \mu^*(E \setminus A) \leq \mu^*(E),$$

hence, $A \in \mathcal{F}$. □

4.1.2. Carathéodory Extension Theorem. We see here a method to produce a measure from "conditionally σ-additive" set functions. Let $\mathcal{A} \ni X$ be an *algebra* of subsets of X: $\emptyset \in \mathcal{A}$; if $A, B \in \mathcal{A}$, then $A \cup B$ and $A \setminus B$ belong to $\mathcal{A}$. A *pre-measure* l on $\mathcal{A}$ is a function $l\colon \mathcal{A} \to [0, +\infty]$ such that $l(\emptyset) = 0$ and, if $\{A_n\}_{n=1}^{\infty}$ is a family of disjoint sets in $\mathcal{A}$ such that $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, then

$$l\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} l(A_n). \tag{4.1.4}$$

It follows that if $A \subset B$, with $A, B \in \mathcal{A}$, then $l(B) = l((B \setminus A) \cup A \cup \emptyset \cup \emptyset \dots) = l(B \setminus A) + l(A) \geq l(A)$.

THEOREM 4.2. *[Carathéodory Extension Theorem] Let l be a pre-measure on an algebra $\mathcal{A}$. For $E \subset X$ define*

$$\mu^*(E) = \inf \left\{ \sum_{n=1}^{\infty} l(A_n) : \ A_n \in \mathcal{A} \text{ and } \bigcup_{n=1}^{\infty} A_n \supseteq E \right\}.$$

Then,

(i) *μ^* is an outer measure;*

(ii) *$\mu^*(A) = l(A)$ if $A \in \mathcal{A}$;*

(iii) *all sets in $\mathcal{A}$ and all μ^*-null sets are μ^*-measurable.*

(iv) *Moreover, if l is σ-finite ($X = \cup_{n=1}^{\infty} A_n$ with $l(A_n) < \infty$), then the extension of l to $\sigma(\mathcal{A})$, the σ-algebra generated by $\mathcal{A}$, is unique: if ν is a measure on $\sigma(\mathcal{A})$ and $\nu(A) = l(A)$ for all A in $\mathcal{A}$, then $\nu = \mu^*$.*

Proof. (i) By Proposition 4.1, μ^* defines an outer measure.

(ii) The inequality $\mu^*(A) \le l(A)$ for $A \in \mathcal{A}$ is obvious: we test the inf in the definition of μ^* on the decomposition $A = A \cup \emptyset \cup \emptyset \cup \dots$. For the reverse inequality, suppose $A \in \mathcal{A}$ and $A \subseteq \bigcup_{n=1}^{\infty} A_n$. We recover a disjoint cover of A by setting $B_m = (A_m \setminus (A_1 \cup \cdots \cup A_{m-1})) \cap A \subseteq A_m$, so that $A = \bigcup_{m=1}^{\infty} B_m$ is a disjoint union in $\mathcal{A}$. We have

$$\begin{aligned}\sum_{n=1}^{\infty} l(A_n) &\ge \sum_{n=1}^{\infty} l(B_n) = l\left(\cup_{n=1}^{\infty} B_n\right) \\ &= l(A).\end{aligned}$$

Passing to inf over all covers, $\mu^*(A) \ge l(A)$, as wished.

(iii) We use here that $\mathcal{A}$ is an algebra of sets. If $A \in \mathcal{A}$, $E \subseteq X$, and $E \subseteq \bigcup_{n=1}^{\infty} B_n$ for $n \ge 1$, then $E \cap A \subseteq \bigcup_{n=1}^{\infty}(B_n \cap A)$ and $E \setminus A \subseteq \bigcup_{n=1}^{\infty}(B_n \setminus A)$, hence,

$$\begin{aligned}\sum_{n=1}^{\infty} l(B_n) &= \sum_{n=1}^{\infty} l(B_n \cap A) + \sum_{n=1}^{\infty} l(B_n \setminus A) \\ \ge \mu^*(E \cap A) + \mu^*(E \setminus A),\end{aligned}$$

thus, passing to inf over all covers $\bigcup_{n=1}^{\infty} B_n$ of E,

$$\mu^*(E) \ge \mu^*(E \cap A) + \mu^*(E \setminus A),$$

which is the measurability condition. Measurability of null-μ^* sets is in Proposition 4.1.

(iv) Suppose ν is another extension of the pre-measure l to $\sigma(\mathcal{A})$, the σ-algebra generated by $\mathcal{A}$. We first show that $\mu^*(E) \ge \nu(E)$ for $E \in \sigma(\mathcal{A})$. If $E \subseteq \bigcup_{n=1}^{\infty} A_n$ with $A_n \in \mathcal{A}$, then,

$$\nu(A) \le \sum_n \nu(A_n) = \sum_n l(A_n),$$

and passing to inf we have $\nu(E) \le \mu^*(E)$.

The opposite inequality is where we use σ-finiteness of l. Suppose $E \in \sigma(\mathcal{A})$, $\mu^*(E) < \infty$, and fix $\epsilon > 0$. Then, there exist $A_n \in \mathcal{A}$ ($n \ge 1$) such that

$$\mu^*(E) + \epsilon \ge \sum_n l(A_n) = \sum_n \mu^*(A_n) \ge \mu^*\left(\bigcup_n A_n\right),$$

hence $\mu^*\left(\left(\bigcup_n A_n\right) \setminus E\right) \le \epsilon$. Thus, using the fact that μ^* and ν agree on $\mathcal{A}$ from the first to the second line,

$$\mu^*(E) \le \mu^*\left(\bigcup_n A_n\right)$$

$$
\begin{aligned}
&= \lim_{m\to\infty} \mu^*\left(\bigcup_{n=1}^{m} A_n\right) \\
&= \lim_{m\to\infty} \mu^*\left(\bigcup_{n=1}^{m} A_n\right) = \nu\left(\bigcup_{n} A_n\right) \\
&= \nu\left(E\right) + \nu\left(\left(\bigcup_{n} A_n\right)\setminus E\right) \\
&\leq \nu(E) + \mu^*\left(\left(\bigcup_{n} A_n\right)\setminus E\right) \\
&\leq \nu(E) + \epsilon.
\end{aligned}
$$

Since $\epsilon > 0$ was arbitrary, $\mu^*(E) \leq \nu(E)$.

Since l is σ-finite, we can exhaust $X = \cup_n X_n$ where $\mathcal{A} \ni X_n \subseteq X_{n+1}$ and $l(X_n) < \infty$. Define $l_n(A) := l(A \cap X_n)$ for $A \in \mathcal{A}$. For $F \subseteq X_n$, we can compute $\mu^*(F)$ just by testing on covers by subsets of X_n in $\mathcal{A}$. For E in $\sigma(\mathcal{A})$ we have, then, by the finite case,

$$
\begin{aligned}
\mu^*(E) &= \lim_{n\to\infty} \mu^*(E \cap X_n) \\
&= \lim_{n\to\infty} \nu(E \cap X_n) = \nu(E).
\end{aligned}
$$

□

The *Lebesgue measure* on $\mathbb{R}$ is a particular example of a measure produced by Carathéodory Extension Theorem. Consider the family $\mathcal{A}$ populated by finite, disjoint unions of the intervals $(a, b]$ and (a, ∞), with $-\infty \leq a \leq b < \infty$. It is easy to see that $\mathcal{A}$ is an algebra. If $-\infty < a_1 < b_1 < a_2 < b_2 < \cdots < a_n < b_n < +\infty$, let

$$l\left((a_1, b_1] \cup \cdots \cup (a_n, b_n]\right) := (b_1 - a_1) + \cdots + (b_n - a_n),$$

be the length function of $(a_1, b_1] \cup \cdots \cup (a_n, b_n]$, and let $l(E) = +\infty$ if $E \in \mathcal{A}$ is unbounded. Then, l defines an additive function on $\mathcal{A}$. The corresponding measure $m = \mu^*$ is the Lebesgue measure. It is easy to see that $\sigma(\mathcal{A})$ contains the Borel algebra of $\mathbb{R}$. Also, by considering different cases it is easy to see that if I is an interval in $\mathbb{R}$, having endpoints $-\infty \leq a \leq b \leq +\infty$, then $m(I) = b - a$. However, the Lebesgue, and other, measures will be introduced in the next section as particular applications of Riesz Representation Theorem for measures. Let μ be a measure defined on a σ-algebra $\mathcal{G}$ on X.

4.1.2.1. *The outer measure associated to a measure.* Carathéodory's idea of recognizing (4.1.2) as the condition which characterizes a σ-algebra on which μ^* is a measure is a brilliant one, and it deserves some comments. By

Proposition 4.1, with $\mathcal{A} = \mathcal{G}$ and $l = \mu$, we can associate to μ the outer measure μ^* which is defined, for any E in X, by

$$\mu^*(E) = \inf\{\sum_n \mu(A_n) : \ A_n \in \mathcal{F}, \ \cup_n A_n \supseteq E\}. \tag{4.1.5}$$

By the properties of a measure, we can test the infimum over mutually disjoint A_n's. The main lemma on outer measures generated by additive set functions says that

$$\mu^*(A) = \mu(A) \tag{4.1.6}$$

whenever A is in $\mathcal{F}$. Suppose now $B \in \mathcal{F}$ and $E \subseteq X$, and $E \subseteq \cup_n A_n$, where the A_n's are disjoint element of $\mathcal{F}$. Then, $A_n \cap B$ and $A_n \setminus B$ are in $\mathcal{F}$, $\cup_n(A_n \cap B) \supseteq E \cap B$, $\cup_n(A_n \setminus B) \supseteq E \setminus B$, so

$$\begin{aligned} \mu^*(E \cap B) + \mu^*(E \setminus B) &\leq \sum_n \mu(A_n \cap B) + \sum_n \mu(A_n \setminus B) \\ &= \sum_n \mu(A_n), \end{aligned}$$

and passing to inf we have

$$\mu^*(E \cap B) + \mu^*(E \setminus B) \leq \mu^*(E), \tag{4.1.7}$$

which is Carathéodory's condition (4.1.2) for B.

4.2. Radon measures

So far, but for the definition of the Borel σ-algebra, we have considered metric structures and measurable structures as separate entities. Indeed, this state of affairs depends on the tendency towards generalization, hence, abstraction, which is part of the mathematical enterprise. Both theories stem in fact from the same base space, the real line, and more generally Euclidean spaces, which has deeply intertwined metric and (length) measure. After Riemann-Stieltjes integrals were introduced at the end of the XIX century, and immediately used in a variety of applications, it became clear the the metric structure on the real line supports many different, useful measures. In 1909, Frigyes (Frederic) Riesz[4] proved a theorem that, among other applications, shows that Riemann-Stieltjes integrals exhaust the linear, positive operators acting on continuous functions supported on a compact interval. Extensions of the result in the context of Lebesgue theory of integrals, and with more general topological spaces, were provided by Andrey Markov in 1938 and Shizuo Kakutani in 1941.

[4]The brothers Frederic and Marcel Riesz were both important mathematicians, and gave fundamental contributions to analysis. Marcel moved to Stockholm, then Lund, while Frederic remained in Hungary, mostly at Szeged.

4.2.1. Riesz Representation Theorem. Let X be a locally compact metric space. We do not specify the metric, since we will only be using Urysohn Lemma and its consequences. In fact, the theory developed in this section applies, more generally, to locally compact topological spaces. A functional $\Lambda\colon C_c(X)$ is *positive* if $\Lambda(f) \geq 0$ whenever $f \geq 0$. Examples of positive functionals are provided by Borel measures.

EXERCISE 4.1. *Let μ be a measure defined on the Borel σ-algebra of a metric space X, such that $\mu(K) < \infty$ if K is compact in X (we say that μ is a* **Borel measure***). Then, the map*

$$\Lambda_\mu\colon f \mapsto \int_X f d\mu$$

defines a positive functional on $C_c(X)$.

Riesz Theorem provides a converse of the above remark.

Let X be a locally compact metric space. A measure μ defined on the Borel σ-algebra $\mathcal{B}(X)$ is *inner regular* if, for E in $\mathcal{B}(X)$,

$$\mu(E) = \sup\{\mu(K) : \ K \text{ is compact and } E \supseteq K\}. \tag{4.2.1}$$

The measure μ is *outer regular* if

$$\mu(E) = \inf\{\mu(V) : \ V \text{ is open and } E \subseteq V\}. \tag{4.2.2}$$

The measure μ is a *Radon measure* if it is finite on compact sets, outer regular and inner regular on open sets: (4.2.1) holds when E is open.

THEOREM 4.3 (Riesz Representation Theorem for Measures). *Let Λ be a positive functional on $C_c(X)$. Then, there exists a unique Radon measure μ on X such that*

$$\Lambda(f) = \int_X f d\mu \text{ for all } f \in C_c(X). \tag{4.2.3}$$

Moreover,

(i) *if V is open in X, then*

$$\mu(V) = \sup\{\Lambda(f) : \ f \in C_c(X),\ f \prec V\}\,; \tag{4.2.4}$$

(ii) *if K is compact in X, then*

$$\mu(K) = \inf\{\Lambda(f) : \ f \in C_c(X),\ f \succ K\}\,. \tag{4.2.5}$$

The idea of the proof is natural. If (4.2.3) held for $f = \chi_E$, with E Borel measurable, we could define $\mu(E) = \Lambda(\chi_E)$. Unfortunately, characteristic functions are typically not continuous, but we can set up some approximation scheme. The F. Riesz way consists in approximating (from above) characteristic functions χ_K of compact sets by continuous functions with compact support; the Radon way, which we follow below, approximates (from below) characteristic functions χ_V of open sets instead. If Λ and μ are related as in (4.2.3) and $f \prec V$, then $\Lambda(f) \leq \mu(V)$ and in fact it can be proved that $\mu(V) = \sup\{\Lambda(f) : f \prec V\}$. This will be our starting point for defining μ. The proof that μ extends to a positive set function on a σ-algebra containing the Borel sets, as you will see, requires the verification of a number of properties (think of the Lebesgue measure on the real line).

Proof. In order to prove uniqueness, we show that a Radon measure μ satisfying (4.2.3) is determined by the values $\Lambda(f)$. For V open, we have that $\mu(V) \geq \Lambda(f)$ if $f \prec V$, so that (4.2.4) holds with $\geq$. Also, for any $K \subseteq V$ compact there exists, by Urysohn Lemma, $K \prec f \prec V$, hence, $\mu(K) \leq \Lambda(f) \leq \mu(V)$. Then,

$$\sup\{\mu(K) : \ K \subseteq V, \text{ compact}\} \leq \sup\{\Lambda(f) : \ f \prec V\}.$$

By inner regularity (4.2.1) on open sets, the sup on the left hand side is $\mu(V)$, then we also have $\leq$ in (4.2.4). Hence, $\mu(V)$ can be expressed in terms of the functional Λ.

By outer regularity (4.2.2), this determines μ on the Borel σ-algebra.

About existence, what we have seen so far suggests that, given Λ, we should first define μ on open sets. For V open, define $\mu(V)$ by (4.2.4):

$$\mu(V) = \sup\left\{\Lambda(f) : \ f \in C_c(X), \ f \prec V\right\}.$$

Let then, for any $E \subseteq X$,

$$\mu^*(E) = \inf\{\mu(V) : \ V \text{ is open and } E \subseteq V\}, \tag{4.2.6}$$

as in (4.2.2).

By definition we have monotonicity of μ on open sets and of μ^* on subsets of X: $\mu(U) \leq \mu(V)$ if $U \subset V$ and $\mu^*(E) \leq \mu^*(F)$ if $E \subset F$.

The proof of the existence statement will proceed in four steps.

(i) *We show first that μ^* is an outer measure, and that $\mu^*(V) = \mu(V)$ when V is open.* For the property of being outer, it suffices to show that, for a countable family of open sets $\{V_j\}_{j=1}^{\infty}$, we have:

$$\mu\left(\bigcup_j V_j\right) \leq \sum_j \mu(V_j). \tag{4.2.7}$$

If this holds, in fact, (4.2.6) implies that, if $E = \cup_{j=1}^{\infty} E_j$,

$$\begin{aligned} \mu^*(E) &= \inf\left\{\mu\left(\bigcup_j V_j\right) : \text{ where each } V_j \text{ is open and } E \subseteq \bigcup_j V_j\right\} \\ &\leq \inf\left\{\sum_j \mu(V_j) : \text{ where each } V_j \text{ is open and } E \subseteq \bigcup_j V_j\right\}, \end{aligned} \tag{4.2.8}$$

where we have equality in the first line because any open set can be written as countable union of open sets.

Fix $\epsilon > 0$ and, for each E_j, choose U_j open such that

$$\mu(U_j) \leq \mu^*(E_j) + \frac{\epsilon}{2^j}.$$

Since the union of such U_j's contain E, the final expression in the last chain of inequalities satisfies

$$\inf\left\{\sum_j \mu(V_j) : \text{ where each } V_j \text{ is open and } E \subseteq \bigcup_j V_j\right\} \leq \sum_j \mu^*(E_j) + \epsilon.$$

We now prove (4.2.7). Consider $f \prec \bigcup_j V_j$, and let $K = \operatorname{supp}(f) \subseteq \bigcup_j V_j$. By compactness, $K \subseteq V_1 \cup \cdots \cup V_n$ for some n. By Partition of Unity, for $1 \leq j \leq n$ there are $h_j \prec V_j$ such that $h_1 + \cdots + h_n = 1$ on K. Clearly, $h_j f \prec V_j$, hence,

$$\begin{aligned} \Lambda(f) &= \Lambda\left(\sum_{j=1}^n h_j f\right) = \sum_{j=1}^n \Lambda(h_j f) \\ &\leq \sum_{j=1}^n \mu(V_j), \end{aligned}$$

and passing to sup over all such f's, we obtain (4.2.7).

We show next that, for V open,

$$\mu^*(V) = \mu(V). \tag{4.2.9}$$

In fact,

$$\mu(V) \leq \mu^*(V) := \inf\{\mu(U) : \ U \supseteq V, \text{ open}\} \leq \mu(V) :$$

the first inequality since $\mu(U) \geq \mu(V)$ when $U \supseteq V$; the second testing the inf with $U = V$.

(ii) *We have to show that, for V open and $E \subseteq X$ with $\mu^*(E) < \infty$, one has*

$$\mu^*(E) \geq \mu^*(E \cap V) + \mu^*(E \setminus V).$$

We consider first the case when E is open. For $\epsilon > 0$ fixed, let $f \prec E \cap V$ such that $\mu(E \cap V) \leq \Lambda(f) + \epsilon$, and $g \prec E \setminus \operatorname{supp}(f) \supseteq E \setminus V$ such that $\mu(E \setminus \operatorname{supp}(f)) \leq \Lambda(g) + \epsilon$. Then, $f + g \prec E$ and

$$\begin{aligned} \mu^*(E) &= \mu(E) \geq \Lambda(f+g) \\ &= \Lambda(f) + \Lambda(g) \geq \mu(E \cap V) + \mu(E \setminus \operatorname{supp}(f)) - 2\epsilon \\ &= \mu^*(E \cap V) + \mu^*(E \setminus \operatorname{supp}(f)) - 2\epsilon \\ &\geq \mu^*(E \cap V) + \mu^*(E \setminus V) - 2\epsilon. \end{aligned}$$

Since $\epsilon > 0$ is arbitrary, Carathéodory's test is passed by any open E.

For arbitrary E, and V open,

$$\begin{aligned} \mu^*(E) &= \inf\{\mu(U) : U \supseteq E, \text{ open}\} = \inf\{\mu^*(U) : U \supseteq E, \text{ open}\} \\ &= \inf\{\mu^*(U \cap V) + \mu^*(U \setminus V) : U \supseteq E, \text{ open}\} \\ &\geq \inf\{\mu^*(U \cap V) : U \supseteq E, \text{ open}\} + \inf\{\mu^*(U \setminus V) : U \supseteq E, \text{ open}\} \\ &= \mu^*(E \cap V) + \mu^*(E \setminus V), \end{aligned}$$

where the last inequality holds because μ^* is monotone and $U \cap V \supseteq E \cap V$, and $U \setminus V \supseteq E \setminus V$, if $U \supseteq E$.

(iii) (4.2.5) *holds. In particular, $\mu(K) < \infty$ for compact K. Also, μ is inner regular on open sets.* Let K be compact and $f \succ K$. For $0 < \lambda < 1$, let $V_\lambda = \{x : f(x) > \lambda\} \supseteq K$. For any $g \prec V_\lambda$, $\lambda^{-1} f \geq g$, hence, $\lambda^{-1}\Lambda(f) \geq \Lambda(g)$. Thus,

$$\mu(K) \leq \mu(V_\lambda) = \sup\{\Lambda(g) : g \prec V_\lambda\} \leq \lambda^{-1}\Lambda(f).$$

As $\lambda \to 1$, we have $\mu(K) \leq \Lambda(f)$, and this gives the direction $\leq$ in (4.2.5). In the other direction, if $V \supseteq K$ is open, by Urysohn Lemma there is $K \prec f \prec V$, hence, $\Lambda(f) \leq \mu(V)$. Since μ is outer regular by its very definition,

$$\mu(K) = \inf\{\mu(V) : K \subset V \text{ open}\} \geq \inf\{\Lambda(f) : f \succ K\},$$

as wished.

We now prove inner regularity on open sets. Let V be open, fix $\epsilon > 0$, let $f \prec V$ such that $\Lambda(f) \geq \mu(V) - \epsilon$, and set $K = \operatorname{supp}(f) \subseteq V$. For all $g \succ K$, $g \geq f$, thus, by positivity of Λ, $\Lambda(g) \geq \Lambda(f)$. Then,

$$\Lambda(f) \leq \inf\{\Lambda(g) : g \succ K\} = \mu(K) \leq \mu(V) \leq \Lambda(f) + \epsilon,$$

implying that $\mu(V) - \mu(K) \leq \epsilon$. As $\epsilon \to 0$, we obtain inner regularity of μ for V.

(iv) *Finally, we show that* $\Lambda(f) = \int_X f d\mu$ *if* $f \in C_c(X)$. We can assume $f(X) \subseteq [0,1]$, since any other function in $C_c(X)$ can be written as a linear combination of functions of this form. We proceed by discretizing f. For $n \geq 1$ and $1 \leq j \leq 2^n$, let

$$f_j(x) = \begin{cases} 0 \text{ if } f(x) \leq \frac{j-1}{2^n}, \\ f(x) - \frac{j-1}{2^n} \text{ if } \frac{j-1}{2^n} \leq f(x) \leq \frac{j}{2^n}, \\ \frac{1}{2^n} \text{ if } f(x) \geq \frac{j}{2^n}. \end{cases}$$

After setting $K_0 = \text{supp}(f)$ and $K_j = \{x : \ f(x) \geq \frac{j}{2^n}\}$, we see that $f_j \in C_c(X)$, $\chi_{K_j} \leq 2^n f_j \leq \chi_{K_{j-1}}$. By outer regularity and (4.2.5),

$$\mu(K_j) \leq \Lambda(2^n f_j) \leq \mu(K_{j-1}).$$

The second inequality depends on the the fact that $2^n f_j \prec V$ for all open V's containing K_{j-1}, hence $\Lambda(2^n f_j) \leq \mu(V)$ for such V's, hence $\Lambda(2^n f_j) \leq \mu(K_{j-1})$.

The functions f_j decompose f:

$$\frac{1}{2^n}\sum_{j=1}^{2^n} \chi_{K_j} \leq f = \sum_{j=1}^{2^n} f_j \leq \frac{1}{2^n}\sum_{j=0}^{2^n-1} \chi_{K_j}. \tag{4.2.10}$$

Thus,

$$\frac{1}{2^n}\sum_{j=1}^{2^n} \mu(K_j) \leq \Lambda(f) = \sum_{j=1}^{2^n} \Lambda(f_j) \leq \frac{1}{2^n}\sum_{j=0}^{2^n-1} \mu(K_j).$$

Similarly, by the properties of the integral,

$$\frac{1}{2^n}\sum_{j=1}^{2^n} \mu(K_j) \leq \int_X f d\mu = \sum_{j=1}^{2^n} \int_X f_j d\mu \leq \frac{1}{2^n}\sum_{j=0}^{2^n-1} \mu(K_j).$$

We obtain the estimate

$$\begin{aligned} \left|\int_X f d\mu - \Lambda(f)\right| &\leq \frac{1}{2^n}\sum_{j=0}^{2^n-1} \mu(K_j) - \frac{1}{2^n}\sum_{j=1}^{2^n} \mu(K_j) \\ &= \frac{1}{2^n}(\mu(K_0) - \mu(K_{2^n})) \leq \frac{\mu(\text{supp}(f))}{2^n} \end{aligned}$$

which tends to 0 as $n \to \infty$. Hence, (4.2.3) holds. □

It is clear from the proof that the metric structure enters only through Urysohn's Lemma, and the Partition of Unity which follows from it. The Lemma holds, more generally, in *locally compact Hausdorff spaces* (LCH), with the attached price tag of a proof which is not two lines long as the one for metric spaces. All results of this section hold, in fact, for general LCH spaces.

4.2.2. Regularity and approximation theorems. A metric space X is σ-compact if $X = \cup_{n\geq 1} K_n$ can be exhausted as union of compact sets K_n, where we can clearly assume $K_n \subseteq K_{n+1}$. A set $A \subseteq X$ belongs to the class F_σ if it countable union of closed sets, and to the class G_δ if it is countable intersection of closed sets.

THEOREM 4.4. *Let X be a σ-compact and let Λ be linear and positive on $C_c(X)$. Then, the measure μ constructed in Riesz' Theorem has the following extra properties:*

(i) *For each $E \in \mathcal{F}$ and $\epsilon > 0$, there are $C \subseteq E \subseteq V$, C closed and V open, such that $V \setminus C \leq \epsilon$.*

(ii) *μ is regular.*

(iii) *If $E \in \mathcal{F}$, then there are $A \in F_\sigma$ and $B \in G_\delta$ such that $A \subseteq E \subseteq B$ and $\mu(B \setminus A) = 0$.*

PROOF. Let $E \subseteq X$, fix K_n in the exhaustion of X, and find $V_n \supseteq K_n \cap E$ open such that $\mu(V_n \setminus (K_n \cap E)) \leq \frac{\epsilon}{2^n}$. Set $V = \bigcup_n V_n$. Then,

$$V \setminus E \subseteq \bigcup_{n=1}^{\infty} (V_n \setminus (K_n \cap E)),$$

and so

$$\mu(V \setminus E) \leq \epsilon.$$

Apply the same reasoning to $X \setminus E$: there is open $W \supseteq (X \setminus E)$ such that $\epsilon \geq \mu(W \setminus (X \setminus E)) = \mu(E \setminus (X \setminus W))$, and $X \setminus W$ is closed. This shows (i).

Let $E \in \mathcal{F}$. By (i), for each $\epsilon > 0$ we can find $C \subseteq E$ such that $\mu(E \setminus C) \leq \epsilon$. On the other hand, C can be exhausted by compact sets $K_n \cap C$, and $\mu(K_n \cap C) \nearrow \mu(C)$ as $n \to \infty$. (ii) follows.

To obtain (iii), use (i) with $\epsilon = 1/n$ to obtain open V_n and closed C_n, then set $A = \cup_n C_n \in F_\sigma$ and $B = \cap_n V_n \in G_\delta$. □

The next theorem provides a very practical sufficient condition for a measure to be regular, which in particular holds when the metric space is σ-compact.

THEOREM 4.5. *Let X be locally compact metric space in which every open set is σ-compact, and let μ be a Radon measure on X. Then, μ is regular.*

PROOF. If $\mu(E) < \infty$ and $\epsilon > 0$, by outer regularity there open $V \supseteq E$ such that $\mu(V \setminus E) < \epsilon$, and by inner regularity on open sets there is a compact $K \subseteq V$ such that $\mu(V \setminus K) < \epsilon$. Again by outer reularity, there is open $W \supseteq U \setminus E$ such that $\mu(W) < \epsilon$. Set $C = K \setminus W$, which is compact and $C \subseteq E$. We finally estimate

$$\mu(C) = \mu(K) - \mu(K \cap W) > \mu(E) - \epsilon - \mu(W) > \mu(E) - 2\epsilon,$$

which shows inner regularity for E. If $\mu(E) = +\infty$, by σ-finiteness there exists a sequence $E_n \nearrow E$ with $\mu(E_n) < \infty$ for all n, and $\mu(E_n) \nearrow \infty$. By the previous step, there are compact sets $K_n \subseteq E_n \subseteq E$ with $\mu(K_n) \geq \mu(E_n) - 1$, hence $\mu(K_n) \to \infty = \mu(E)$, and regularity hilds in this case as well. □

In many cases, including Euclidean spaces, Borel measures which are finite on compact sets are automatically regular. The hypothesis of the following theorem, for instance, hold in Euclidean spaces.

THEOREM 4.6. *Let X be a locally compact metric space in which all open sets are σ-compact. Then, all Borel measures on X which are finite on compact sets are regular (hence, Radon).*

PROOF. If compact sets have finite measure, then $C_c(X)$ functions are integrable, hence $\Lambda : C_c(X) \to \mathbb{C}$, $\Lambda(f) = \int_X f d\mu$ defines a positive, linear functional. By the Riesz representation theorem, there exists a Radon measure ν such that $\Lambda(f) = \int_X f d\nu$. We have to show (i) that $\nu = \mu$ on the Borel σ-algebra, and (ii) that ν is also inner regular.

Let U be an open set: by assumption there are compact sets K_n such that $K_n \nearrow U$. We construct a sequence of functions $f_n \in C_c(X)$ as follows. We start with $K_1 \prec f_1 \prec U$. For each $n \geq 2$, we recursively choose $K_{n-1} \cup \operatorname{supp}(f_{n-1}) \prec f_n \prec U$. This way, $f_n \nearrow \chi_U$, hence,

$$\mu(U) = \lim_{n\to\infty} \int_X f_n d\mu = \lim_{n\to\infty} \int_X f_n d\nu = \nu(U) :$$

$\mu = \nu$ on open sets. For a Borel set E and $\epsilon > 0$, use theorem 4.4 to find $C \subseteq E \subseteq V$, with V open and C closed, such that $\mu(V \setminus C) = \nu(V \setminus C) < \epsilon$, the equality holding because $V \setminus C$ is open. Thus,

$$\mu(V) \leq \mu(C) + \epsilon \leq \mu(E) + \epsilon,$$

which shows outer regularity of μ, and

$$\mu(E) \leq \mu(V) \leq \mu(C) + \epsilon.$$

Now, C is σ-compact because X is: let $K_n \nearrow C$ be an increasing exhaustion of C by compact sets, so that $\mu(K_n) \nearrow \mu(C) \geq \mu(E) - \epsilon$. In particular, we have that $\mu(K_n) \geq \mu(C) - 2\epsilon$ for some n, and this shows inner regularity of μ.

The measures μ and ν coincide on open sets and their are outer regular, hence they coincide on the whole Borel σ-algebra. □

We have a very useful approximation theorem, which will be used several times.

THEOREM 4.7. *If μ is a regular measure, then $C_c(X)$ is dense in L^p, if $1 \le p < \infty$.*

PROOF. We start by approximating the characteristic function of a measurable set E. For $\epsilon > 0$ fixed, consider $K \subseteq E \subseteq V$ with $K, \overline{V}$ compact and $\mu(V \setminus K) \le \epsilon$. By Urysohn lemma, there is $K \prec f \prec V$. Then,

$$\begin{aligned} \|f - \chi_E\|_{L^p} &\le \|f - \chi_K\|_{L^p} + \|\chi_K - \chi_E\|_{L^p} \\ &\le 2\|\chi_V - \chi_K\|_{L^p} \le 2\epsilon^{1/p}. \end{aligned}$$

Let now $s = \sum_{j=1}^n a_j \chi_{E_j}$ be a simple function where the sets E_j are disjoint, and let $f_j \in C_c(X)$ be functions approximating the χ_{E_j}'s within errors ϵ_j's to be chosen. The function $f = \sum_{j=1}^n a_j f_j$ is continuous and

$$\begin{aligned} \|s - f\| &= \left\| \sum_{j=1}^n a_j \chi_{E_j} - \sum_{j=1}^n a_j f_j \right\|_{L^p} \\ &\le \sum_{j=1}^n |a_j| \, \|\chi_{E_j} - f_j\|_{L^p} \\ &\le \epsilon, \end{aligned}$$

if $\|\chi_{E_j} - f_j\|_{L^p} \le \frac{\epsilon}{n \sum_{j=1}^n |a_j|}$. The theorem follows because simple functions are dense in L^p. □

4.2.3. Lusin's Theorem. Lusin's Theorem says that measurable functions are continuous when restricted to large sets.

We will use Severini-Egorov's Theorem, that in the locally compact case assumes the following form.

THEOREM 4.8 (Severini-Egorov in LCH spaces). *Let μ be a regular Borel measure on a locally compact metric space, and f_n be a sequence of measurable functions converging pointwise a.e. to a function f. Then, for all $\epsilon > 0$ there exists a* **compact set** *K with $\mu(X \setminus K) \le \epsilon$, such that f_n converges to f uniformly on K.*

PROOF. Egorov-Severini proves that uniform convergence holds on a measurable E such that $\mu(X \setminus E) \le \epsilon/2$. On the other hand, by regularity we can find compact K in E with $\mu(E \setminus K) \le \epsilon/2$. □

THEOREM 4.9 (Lusin). *Let μ be a finite, regular Borel measure on a locally compact metric space, and let $f \colon X \to \mathbb{R}$ be measurable. Then, for all*

$\epsilon > 0$ there is a closed subset F in X with $\mu(X \setminus F) < \epsilon$ and a continuous function $f_\epsilon \colon X \to \mathbb{R}$ such that $f = f_\epsilon$ on F. In particular, f is continuous on F.

Proof. Suppose first that $h = \sum_{i=1}^m a_i \chi_{E_i}$ is a simple function, $\bigcup_i E_i = X$, and the E_i's are disjoint. Let $K_i \subset E_i$ be compact sets with $\mu(E_i \setminus K_i) \leq \epsilon/n$. Then, h is continuous on $F = K_1 \cup \cdots \cup K_m$, and $\mu(X \setminus F) = \mu(\bigcup_i (E_i \setminus K_i)) \leq \epsilon$.

Let now $f \colon X \to \mathbb{R}$ be measurable, and let $\{h_n : n \geq 1\}$ be a sequence of simple functions converging to f *a.e.* For fixed $\epsilon > 0$, let K_n be a compact set such that $\mu(X \setminus K_n) \leq \epsilon/2^n$ and h_n is continuous on K_n. By the Severini-Egorov Theorem, there is H compact such that h_n converges to f uniformly on H and $\mu(X \setminus H) \leq \epsilon$. Let $K = (\bigcap_n K_n) \cap H$. Then, each h_n is continuous on K, $h_n \to f$ uniformly on K (hence, f is continuous on K), and

$$\mu(X \setminus K) \leq \sum_n \mu(X \setminus K_n) + \mu(X \setminus H) \leq 2\epsilon,$$

as wished. By Tietze extension theorem, $f|_K$ extends to a continuous function f_ϵ on X, having the same minimum and maximum. □

4.2.4. The Fundamental Lemmas of the Calculus of Variations. The material in this subsection is much used in Calculus of Variations, whence the name. In particular, it is used in rigorously deriving the Euler-Lagrange equations associated to a functional. Their proofs depend on the properties of a Radon measure, not on the Riesz representation theorem.

Here, (X, d) is a locally compact metric space, and $\mu \geq 0$ is a Radon measure on X.

Lemma 4.1 (1^{st} Fundamental Lemma of the Calculus of Variations). *Let $f \in L^1_{loc}$ and suppose that for all $\varphi \in C_c(X)$ one has*

$$\int_X f \varphi d\mu = 0. \tag{4.2.11}$$

Then, $f = 0$ a.e.

Exercise 4.2. *Provide a two lines proof of Lemma 4.1, then of Lemma 4.2, under the extra assumption that f is continuous.*

Proof. The proof is by contradiction. If $f(x) \neq 0$ $\mu - a.e.$, then there are $r > 0$ and E measurable such that $\mu(E) > 0$, and $f(x) \geq r$ on E or $f(x) \leq -r$ on E. Assume the first holds. Let then $K \subseteq E$ be compact such that $\mu(E \setminus K) \geq \mu(E)/2$, and let $U \supseteq K$ be open, to be chosen later. By Urysohn Lemma, there exists $K \prec \psi \prec U$. Then,

$$\begin{aligned}\int_X f\psi d\mu &\geq \int_K f d\mu - \int_{U\setminus K} |f| d\mu \\ &\geq \mu(K) r - \int_{U\setminus K} |f| d\mu.\end{aligned}$$

Since $K = \cap_n U_n$ with $\mu(U_n) \nearrow \mu(K)$, by dominated convergence we can choose $U = U_n$ such that $\int_{U\setminus K} |f| d\mu \leq \mu(K) r/2$, in which case $\int_X f\psi d\mu > 0$. □

Lemma 4.2 (2^{nd} Fundamental Lemma of the Calculus of Variations). *Let $f \in L^1_{loc}$ and suppose that for all $\varphi \in C_c(X)$ with $\int_X \varphi d\mu = 0$ one has*

$$\int_X f\varphi d\mu = 0. \tag{4.2.12}$$

Then, f is a.e. equal to a constant.

Proof. It is a corollary of the 1^{st} Lemma's proof. The function f is not *a.e.* equal to a constant if and only if $\text{essinf} f < \text{esssup} f$, i.e. if there exist real C and $r > 0$ such that

$$\mu(\{x : f(x) \geq C + r\}) = \mu(E_+) > 0, \text{ and } \mu(\{x : f(x) \leq C - r\}) = \mu(E_-) > 0.$$

Let $K_\pm \subseteq E_\pm$ be such that $\mu(K_\pm) \geq \mu(E_\pm)/2$, and let $U_\pm \supseteq K_\pm$ be open and disjoint, to be chosen later (we first choose U_+ and U_- disjoint, and later we shrink both of them). Apply to each couple $(K_\pm, U_\pm)$ the procedure above. We find positive functions $\psi_\pm$ in $C_c(X)$ with disjoint supports such that $\int_X f\psi_+ d\mu > 0$, $\int_X f\psi_+ d\mu < 0$. The "dipole"

$$\varphi = \frac{\psi_+}{\|\psi_+\|_{L^1(\mu)}} - \frac{\psi_-}{\|\psi_-\|_{L^1(\mu)}}$$

has vanishing integral and $\int_X \varphi f d\mu > 0$. □

The name of the lemmas above comes from the fact that they are widely used in Calculus of Variations. There, the base space has a differentiable structure, and the statement often requires the testing functions φ to be smooth, which in our context does not make sense. However, a critical analysis of the proofs shows that smooth versions of the lemmas can be proved without changes, provided we have smooth versions of Urysohn Lemma, which are standard in Euclidean spaces and in manifolds.

4.3. The dual of $C_0(X)$

Let μ be a positive, *finite*, Radon measure on a locally compact metric space X. Then,

$$\Lambda_\mu : h \to \int_X h d\mu$$

defines a positive, linear functional on $C_0(X)$, not just on $C_c(X)$. Moreover, this functional is bounded: $|\Lambda_\mu(h)| \leq \mu(X)\|h\|_u$, where $\|h\|_u = \|h\|_{L^\infty}$ is the sup norm. Moreover, by inner regularity on open sets (hence, on X), we see that the constant $\mu(X)$ is best possible.

PROPOSITION 4.2. *Let μ be a positive, finite, Radon measure on a locally compact metric space X. Then,*

$$\|\Lambda_\mu\|_{C_0(X)^*} := \sup_{h\neq 0} \frac{|\Lambda_\mu(h)|}{\|h\|_u} = \mu(X). \tag{4.3.1}$$

PROOF. Let in fact $K_n \subseteq K_{n+1}$ be an increasing sequence of open sets such that $\mu(K_n) \to \mu(X)$, and consider $h_n \succ K_n$. Then, $\|h_n\|_u = 1$ and

$$\mu(K_n) \leq \Lambda_\mu(h_n) \leq \mu(X),$$

hence, $\Lambda_\mu(h_n) \to \mu(X)$. □

A variation on Riesz' representation for measures ensures that the converse statement holds.

PROPOSITION 4.3. *Let X be σ-compact, and let $\Lambda : C_0(X) \to \mathbb{C}$ be a bounded, positive, linear functional on $C_0(X)$. Then, there exists a positive, bounded, Radon measure μ on X such that $\Lambda = \Lambda_\mu$.*

PROOF. First, by Riesz representation theorem there exists a positive, Radon measure μ such that

$$\Lambda(\varphi) = \int_X \varphi d\mu$$

for all $\varphi \in C_c(X)$.

Given an increasing sequence $\{K_n\}$ of compact sets whose union is X, we produce a sequence $\{V_n\}$ of open sets with compact closure such that $\overline{V_n} \subseteq V_{n+1}$ and that $K_n \subseteq V_n$, so that the union of the V_n's is X. Consider $\eta_0 \succ K_1$ and $V_1 = \{x : \eta_0(x) > 0\} \supseteq K_1$. Next, consider $\eta_1 \succ K_2 \cup \overline{V_1}$, and let $V_2 = \{x : \eta_1(x) > 0\} \supseteq K_2$. Iterate, choosing $\eta_n \succ K_{n-1} \cup \overline{V_{n-1}}$. We have found a sequence $\{V_n\}$ with the desired properties. Moreover, we obtained functions

$$\overline{V_{n-1}} \prec \eta_n \prec V_n,$$

so that $\eta_n \nearrow 1$ on X, uniformly on compact sets. By monotone convergence,

$$\mu(X) = \lim_{n\to\infty} \int_X \eta_n d\mu = \lim_{n\to\infty} \Lambda(\eta_n) \le \|\Lambda\|_{C_0(X)^*},$$

and, in particular, the measure μ is finite.

For $h \in C_0(X) \subseteq L^1(\mu)$ (because μ is finite), $|\eta_n h| \le |h|$ and $\eta_n h \to h$ uniformly, hence, by dominated convergence and continuity of Λ,

$$\Lambda(h) = \lim_{n\to\infty} \Lambda(\eta_n h) = \lim_{n\to\infty} \int_X \eta_n h d\mu = \int_X h d\mu,$$

as desired. □

The next theorem identifies the dual of $C_0(X)$ with the space of the bounded, Radon, complex measures of X.

THEOREM 4.10 (Riesz representation theorem for signed measures). *Let X be σ-compact, and let $\Lambda : C_0(X) \to \mathbb{C}$ be a bounded, complex, linear functional on $C_0(X)$. Then, there exists a bounded, complex, Radon measure μ on X such that*

$$\Lambda(h) = \int_X h d\mu \tag{4.3.2}$$

for all h in $C_0(X)$. Moreover, $\|\Lambda\|^*_{C_0(X)} = |\mu|(X)$ is the total variation of μ.

Let $\mathcal{M}(X)$ be the linear space of the signed, finite, Radon measures on X, normed by $\|\mu\|_{\mathcal{M}} = |\mu|(X)$. Another way to state the theorem is that $\mu \mapsto \Lambda_\mu$ is an isometric isomorphism between $\mathcal{M}(X)$ and $C_0(X)^*$.

PROOF. By splitting Λ into real and imaginary part, it suffices to show the theorem for real valued functionals, and we can thgus show (4.3.2) for real valued h. We start by splittig Λ into a positive and negative part in a way which somehow mimics the proof of Jordan decomposition theorem. For $h \ge 0$ in $C_0(X)$, define

$$\Lambda_+(h) = \sup\{\Lambda(g) : 0 \le g \le h\} \ge \Lambda(h), \tag{4.3.3}$$

so that $\|g\|_u \le \|h\|_u$, and $|\Lambda(g)| \le \|\Lambda\|_{C_0(X)^*}\|g\|_u \le \|\Lambda\|_{C_0(X)^*}\|h\|$. Thus,

$$0 \le \Lambda_+(h) \le \|\Lambda\|_{C_0(X)^*}\|h\|_u, \tag{4.3.4}$$

where the first inequality follows by taking $g = 0$. By steps, we show the linearity of Λ_+ on $C_0(X)$.

(I) If $c \geq 0$ and $h \geq 0$, by homogeneity of Λ we have that $\Lambda_+(ch) = c\Lambda_+(h)$.

(II) Let $0 \leq g_i \leq h_i$ for $i = 1, 2$. Then, $0 \leq g_1 + g_2 \leq h_1 + h_2$, hence,

$$\Lambda_+(h_1) + \Lambda_+(h_2) \leq \Lambda_+(h_1 + h_2).$$

In the opposite direction, if $0 \leq g \leq h_1 + h_2$, we have that $\min\{g, h_1\} \leq h_1$ and $g - \min\{g, h_1\} \leq h_2$; hence,

$$\Lambda_+(g) = \Lambda_+(\min\{g, h_1\}) + \Lambda_+(g - \min\{g, h_1\}) \leq \Lambda_+(h_1) + \Lambda_+(h_2),$$

so $\Lambda_+(h_1 + h_2) = \Lambda_+(h_1) + \Lambda_+(h_2)$.

(III) For signed $h = h_+ - h_-$, we are forced to define $\Lambda_+(h) = \Lambda_+(h_+) - \Lambda_+(h_-)$. If h can be differently decomposed as difference of positive functions in $C_0(X)$, $h = f - g$, then $g + h_+ = f + h_-$, and by (II) we have

$$\Lambda_+(g) + \Lambda_+(h_+) = \Lambda_+(f) + \Lambda_+(h_-),$$

i.e.

$$(*)\ \Lambda_+(f) - \Lambda_+(g) = \Lambda_+(h_+) - \Lambda_+(h_-),$$

independently of the decomposition. Linearity follows easily. If $h, g \in C_0(X)$, then $h + g = (h_+ + g_+) - (h_- + g_-)$, hence, by $(*)$,

$$\begin{aligned} \Lambda_+(h + g) &= \Lambda_+(h_+ + g_+) - \Lambda_+(h_- + g_-) \\ &= \Lambda_+(h_+) + \Lambda_+(g_+) - \Lambda_+(h_-) - \Lambda_+(g_-) \\ &= \Lambda_+(h) + \Lambda_+(g). \end{aligned}$$

(IV) The linear functional Λ_+ is bounded on $C_0(X)$, since, by (4.3.4),

$$\begin{aligned} |\Lambda_+(h)| &\leq \max\{\Lambda_+(h_+), \Lambda_+(h_-)\} \\ &\leq \|\Lambda\|_{C_0(X)^*} \max\{\|h_+\|_u, \|h_-\|_u\} \\ &= \|\Lambda\|_{C_0(X)^*} \|h\|_u. \end{aligned}$$

By (IV) and proposition 4.3, there is a positive, bounded, Radon measure μ_+ on X such that

$$\Lambda_+(h) = \int_X h d\mu_+$$

for $h \in C_0(X)$, and $\mu_+(X) \leq \|\Lambda\|_{C_0(X)^*}$.

Let now $\Lambda_-(h) := \Lambda_+(h) - \Lambda(h)$, which is a positive functional since, for $h \geq 0$, $\Lambda(h) \leq \Lambda_+(h)$. If μ_- is the measure associated to Λ_- by proposition 4.3, and $\mu = \mu_+ - \mu_-$, then (4.3.2) holds.

We are left with the proof that $\|\Lambda\|_{C_0(X)^*} = |\mu|(X)$. □

COROLLARY 4.1. *The normed space $\mathcal{M}(X)$ is Banach with respect to the total variation norm.*

4.4. The Lebesgue measure and some of its variations

One of the applications of Riesz representation theorem is that any time we have a method to integrate C_c functions, we can extend that notion of integration to a much wider class, and "see" the measure which does that.

4.4.1. Lebesgue measure. For instance, we can start from Cauchy definition of integral from basic calculus, which applies to functions in $f \in C_c(\mathbb{R})$,

$$\Lambda(f) := \lim_{n\to\infty} \sum_{j=-\infty}^{+\infty} f\left(\frac{j}{2^n}\right)\frac{1}{2^n}. \tag{4.4.1}$$

A few remarks are in order.

- Cauchy integral is defined on a closed interval $[a, b]$. Since f has compact support, for each n the points $\frac{j}{2^n}$ which are considered in the sum are finitely many. They scan supp(f) at a finer and finer resolution $\frac{1}{2^n}$, hence (4.4.1) provides the usual approximations for the integral of f on an interval having, say, integer coordinates and containing supp(f).
- The existence of the limit is proved in calculus classes, and it depends on the uniform continuity of f. If you had Riemann's integral first, observe that (for $n > R$) the sum inside the limit in (4.4.1) is a Riemann sum for the integral. However, for completeness and with extensions in mind, we prove the convergence of the limit below.

PROPOSITION 4.4. *The limit in (4.4.1) exists.*

PROOF. Fix $n \ge 1$. Suppose that supp$(f) \subseteq [-R, R]$, so that the number of points of the form $\frac{j}{2^n}$ for which $f\left(\frac{j}{2^n}\right) \ne 0$ is bounded by $(2R+1)2^n$. For $\epsilon > 0$ fixed, let $n(\epsilon) > 0$ be such that, for $n \ge n(\epsilon)$ and $|x - y| \le \frac{1}{2^n}$ we have that $|f(x) - f(y)| \le \epsilon$. Set

$$\Lambda_n(f) = \sum_j f\left(\frac{j}{2^n}\right)\frac{1}{2^n}. \tag{4.4.2}$$

For $n \ge n(\epsilon)$ and $j \ge 1$,

$$|\Lambda_{n+j}(f) - \Lambda_n(f)| = \left|\sum_k f\left(\frac{k}{2^{n+j}}\right)\frac{1}{2^{n+j}} - \sum_l f\left(\frac{l}{2^n}\right)\frac{1}{2^n}\right|$$

$$\begin{aligned} &= \left| \sum_k f\left(\frac{k}{2^{n+j}}\right)\frac{1}{2^{n+j}} - \sum_l f\left(\frac{l2^j}{2^{n+j}}\right)\frac{2^j}{2^{n+j}} \right| \\ &\le \sum_l \sum_{m=1}^{2^j} \left| f\left(\frac{l2^j+m}{2^{n+j}}\right) - f\left(\frac{l2^j}{2^{n+j}}\right)\right| \frac{1}{2^{n+j}} \\ &\le \sum_{k\in\mathbb{Z}: f(k2^{-(n+j)})\neq 0} \frac{\epsilon}{2^{n+j}}, \\ &\quad \text{because } \left|\frac{l2^j+m}{2^{n+j}} - \frac{l2^j}{2^{n+j}}\right| \le \frac{1}{2^n}, \\ &\le (2R+1)\epsilon. \end{aligned}$$

This shows that $\{\Lambda_n(f):\ n\ge 1\}$ is a Cauchy sequence. □

COROLLARY 4.2. *There is a Radon measure m such that $\Lambda(f)=\int_{\mathbb{R}} f dm$ for $f\in C_c(\mathbb{R})$. Moreover, m is translation invariant, $m(E+a)=m(E)$ for $E\in\mathcal{F}$, and $m((a,b))=b-a$.*

PROOF. We can prove this in a number of ways. Here is one.

(i) For $a<b$ real, consider the interval (a,b) and for $0<\epsilon<(b-a)/4$ consider continuous functions $f_\epsilon(x)=\begin{cases}1 \text{ if } a+2\epsilon\le x\le b-2\epsilon,\\ 0 \text{ if } x\le a+\epsilon \text{ or } x\ge b-\epsilon,\end{cases}$ and that are linear in the intervals $[a+\epsilon,a+2\epsilon]$ and $[b-2\epsilon,b-\epsilon]$. Then, $f_\epsilon \prec (a,b)$ and it is easy to see that

$$b-a-4\epsilon\le\Lambda(f_\epsilon).$$

Hence, $m((a,b))\ge b-a$. In the other direction, if $f\prec(a,b)$, then $\Lambda(f)\le b-a$, thus we have $m((a,b))=b-a$.

When $a=-\infty$ or $b=+\infty$ we have $m((a,b))=+\infty$ by exhausting (a,b) by an increasing sequence of intervals.

(ii) If V is open in $\mathbb{R}$, then V is the union of at most countably many disjoint intervals (a_n,b_n), so that, by (i) and countable additivity,

$$\begin{aligned} m(V+x) &= m\left(\bigcup_{n=1}^{\infty}[(a_n,b_n)+x]\right) = \sum_{n=1}^{\infty} m((a_n,b_n)+x) \\ &= \sum_{n=1}^{\infty} m((a_n,b_n)) = m(V). \end{aligned}$$

For $E\subseteq\mathbb{R}$, $m(E)=\inf\{m(V):\ V\supseteq E \text{ open}\}=m(E+x)$ because $m(V)=m(V+x)$.

□

EXERCISE 4.3. *Complete the proof of (i).*

Let E be the non-Lebesgue measurable set in example 2.1, and define on E the σ-algebra $\mathcal{F}_E$ having as elements $A \cap E$, with $A \in \mathcal{B}(\mathbb{R})$ Borel in $\mathbb{R}$. On $\mathcal{F}_E$, define

$$m_E(A \cap E) := m^*(A \cap E),$$

where m^* is the outer measure generated by m. By our results on outer measures, we know that m_E is in fact a (Borel) measure on E. This is a notable example of a measure space, on which no concrete computation can be performed.

4.4.2. Lebesgue-Stieltjes measures. A critical analysis of the construction of the Lebesgue measure we saw above leads to an immediate generalization, with important consequences for analysis on the real line. Let $\alpha \colon \mathbb{R} \to \mathbb{R}$ be an increasing function. For $n \geq 1$ and $f \in C_c(\mathbb{R})$, let

$$\Lambda_n(f) := \sum_{j=-\infty}^{+\infty} f\left(\frac{j}{2^n}\right)\left(\alpha\left(\frac{j}{2^n}\right) - \alpha\left(\frac{j-1}{2^n}\right)\right). \tag{4.4.3}$$

Clearly, Λ_n is a well-defined, linear, positive functional on $C_c(\mathbb{R})$.

We will write

$$\Delta\alpha(n;j) = \alpha\left(\frac{j}{2^n}\right) - \alpha\left(\frac{j-1}{2^n}\right). \tag{4.4.4}$$

LEMMA 4.3. *For f in $C_c(\mathbb{R})$, the limit*

$$\Lambda(f) := \lim_{n\to\infty} \Lambda_n(f)$$

exists in $\mathbb{C}$, and it defines a linear, positive functional on $C_c(\mathbb{R})$.

PROOF. Consider $n, l \geq 1$ and $f \in C_c(\mathbb{R})$. We write

$$\begin{aligned}
\Delta\alpha(n;j) &= \alpha\left(\frac{j}{2^n}\right) - \alpha\left(\frac{j-1}{2^n}\right) \\
&= \sum_{m=1}^{2^l}\left[\alpha\left(\frac{(j-1)2^l+m}{2^{n+l}}\right) - \alpha\left(\frac{(j-1)2^l+m-1}{2^{n+l}}\right)\right] \\
&= \sum_{m=1}^{2^l} \Delta\alpha(n+l;(j-1)2^l+m)
\end{aligned}$$

as a telescopic sum. For fixed $\epsilon > 0$, choose $n(\epsilon) > 0$ so large that, for $n \geq n(\epsilon)$, $|f(x) - f(y)| \leq \epsilon$ when $|x - y| \leq \frac{1}{2^n}$. Then, using the telescopic sum in the first equality,

$$\begin{aligned}|\Lambda_{n+l}(f) - \Lambda_n(f)| &= \Bigg| \sum_{j=-\infty}^{+\infty} \sum_{m=1}^{2^l} \left[f\left(\frac{j}{2^n}\right) - f\left(\frac{(j-1)2^l + m}{2^{n+l}}\right)\right] \\ &\quad \cdot \Delta\alpha(n+l; (j-1)2^l + m) \Bigg| \\ &\leq \epsilon \sum_{k:\ k2^{-(n+l)} \in \mathrm{supp(f)}} \Delta\alpha(n+l; k) \\ &\leq \epsilon(\alpha((R+1)) - \alpha(-(R+1))),\end{aligned}$$

provided $[-R, R] \supset \mathrm{supp}(f)$. Thus, $\{\Lambda_n(f)\}$ is a Cauchy sequence, hence it converges to a value $\Lambda(f)$, which clearly satisfies the properties listed in the statement. □

We write

$$\int_{\mathbb{R}} f d\alpha = \Lambda(f), \tag{4.4.5}$$

which makes sense for $f \in C_c(\mathbb{R})$ as limit of the Λ_n's.

From Lemma 4.3 and Riesz Theorem we immediately obtain an important result.

COROLLARY 4.3. *There is a unique Borel measure μ_α on $\mathbb{R}$ such that, for all f in $C_c(\mathbb{R})$:*

$$\int_{\mathbb{R}} f d\mu_\alpha = \int_{\mathbb{R}} f d\alpha. \tag{4.4.6}$$

In practice, we will often use the notation on the right of (4.4.6) also when f is a positive Borel function, or f is integrable with respect to μ_α. Below, we will have a more complete picture of the relations between Borel measures and increasing functions. We refer to

$$\int_{\mathbb{R}} f(x) d\alpha(x)$$

as to the Lebesgue-Stieltjes integral of f with respect to the measure $d\alpha$. Of course, we would like to have a more concrete understanding of μ_α.

We pause a moment on the functionals Λ_n defined in (4.4.3). Trivially, they are positive functionals on $C_c(\mathbb{R})$, and we do not need Riesz Theorem to understand which measure they are associated to:

$$\mu_n = \sum_{j=-\infty}^{\infty} \delta_{\frac{j}{2^n}} \left(\alpha\left(\frac{j}{2^n}\right) - \alpha\left(\frac{j-1}{2^n}\right)\right) = \sum_{j=-\infty}^{\infty} \delta_{\frac{j}{2^n}} \Delta\alpha\,(n; j)\,. \tag{4.4.7}$$

Lemma 4.3 can be restated as the limit

$$\lim_{n\to\infty}\int_{\mathbb{R}} f d\mu_n = \int_{\mathbb{R}} f d\mu_\alpha \tag{4.4.8}$$

whenever $f \in C_c(\mathbb{R})$. A way to rephrase this fact is that μ_α is the *weak* limit* of the discrete measures μ_n.

Following usual notation, we let

$$\alpha(a^+) = \lim_{x\to a^+}\alpha(x);\ \alpha(a^-) = \lim_{x\to a^-}\alpha(x). \tag{4.4.9}$$

THEOREM 4.11. *Let* $\alpha\colon \mathbb{R}\to\mathbb{R}$ *be increasing. Then, for* $-\infty \le a < b <\le \infty$,

$$\mu_\alpha((a,b)) = \alpha(b^-) - \alpha(a^+). \tag{4.4.10}$$

PROOF. The general case follows from an easy limiting argument from the finite case $-\infty < a < b < \infty$, which is what we are going consider. We first show that $\mu_\alpha((a,b)) \le \alpha(b^-) - \alpha(a^+)$. Let $f \prec (a,b)$. Then, there are $N \ge 1$ and $a < \frac{l_m}{2^N} < \frac{l_m+1}{2^N} < \frac{l_M}{2^N} < b$ such that $f(x) = 0$ if $x \notin \left(\frac{l_m+1}{2^N}, \frac{l_M}{2^N}\right)$. For $n \ge N$ (draw a picture!):

$$\begin{aligned}
\Lambda_n(f) &= \sum_j f\left(\frac{j}{2^n}\right)\Delta\alpha(n;j) \\
&\le \sum_{j:\ \frac{l_m+1}{2^N}<\frac{j}{2^n}<\frac{l_M}{2^N}} \Delta\alpha(n;j) \\
&\le \alpha\left(\frac{l_M}{2^N}\right) - \alpha\left(\frac{l_m}{2^N}\right) \\
&\le \alpha(b^-) - \alpha(a^+).
\end{aligned}$$

Taking the limit as $n\to\infty$, then the sup over $f \prec (a,b)$, we have $\le$ in (4.4.10).

In the other direction, let $N \ge 1$ fixed, and l_m and l_M as above. Let $f_N \prec (a,b)$ be such that $f_N\left(\frac{l}{2^N}\right) = 1$ for $l = l_m + 1, \dots, l_{M-1}$, and $f_N(x) = 0$ when $x \ge \frac{l_M}{2^N}$ or $x \le \frac{l_m}{2^N}$. For $n \ge N$,

$$\Lambda_n(f_N) = \sum_j f_N\left(\frac{j}{2^n}\right)\Delta\alpha(n;j) \ge \alpha\left(\frac{l_M - 1}{2^N}\right) - \alpha\left(\frac{l_m}{2^N}\right).$$

Taking the limit over n, then the sup over N:

$$\mu((a,b)) \ge \sup_N \Lambda(f_N) \ge \lim_{N\to\infty}\left[\alpha\left(\frac{l_M-1}{2^N}\right) - \alpha\left(\frac{l_m}{2^N}\right)\right] = \alpha(b^-) - \alpha(a^+),$$

as wished. □

Here and in Proposition 4.4, there is nothing magic about the choice of points of the form $\frac{j}{2^n}$. We used them mostly to avoid cumbersome notation, to carry out the argument in the clearest way, and also because the points considered at step n are a subset of those considered at step $n+1$, and this nicely illustrates the fact that $\frac{1}{2^n}$ should be considered as a the "resolution" of a "sampling". As a matter of fact, for each $n \geq 1$ we might choose an increasing, two-sided sequence of points $\{x_j^n\}_{j=-\infty}^{+\infty}$ with $\lim_{j\to\pm\infty} x_j^n = \pm\infty$, and define

$$\tilde{\Lambda}_n(f) = \sum_{j=-\infty}^{+\infty} f(x_j^n)[\alpha(x_j^n) - \alpha(x_{j-1}^n)].$$

If the sequences $\{x_j^n\}_{j=-\infty}^{+\infty}$ become thinner and thinner, the positive functionals $\tilde{\Lambda}_n(f)$ converge. Also, there is nothing magic in the choice of computing f in the upper endpoint of $[x_{j-1}^n, x_j^n]$: we might replace that value by $f(\tilde{x}_j^n)$ with any $\tilde{x}_j^n$ in the interval, and the uniform continuity would do its job. It is noteworthy that in stochastic integration, which is not a topic we are touching, the function α is not increasing (it is not, in fact, of *bounded variation*), and the choice of the point in the interval becomes a crucial issue.

PROPOSITION 4.5. *Let $\alpha : \mathbb{R} \to \mathbb{R}$ be increasing. Suppose that, for all $R > 0$,*

$$\lim_{n\to\infty} \left(\sup_{j:\ x_j^n \in [-R,R]} (x_j^n - x_{j-1}^n) \right) = 0.$$

Define

$$\check{\Lambda}_n(f) = \sum_{j=-\infty}^{+\infty} f(t_j^n)[\alpha(x_j^n) - \alpha(x_{j-1}^n)],$$

with $x_{j-1}^n \leq t_j^n \leq x_j^n$. Then, for $f \in C_c(\mathbb{R})$, $\lim_{n\to 0} \check{\Lambda}_n(f) = \Lambda(f)$, the functional in Lemma 4.3.

EXERCISE 4.4. *Prove the proposition. To show that it converges to the functional Λ in Lemma 4.3, show that the measures associated to both functionals agree on open intervals (a, b).*

4.4.3. Signed Lebesgue-Stieltjes measures and function of bounded variation. Further critical analysis of the proof of Lemma 4.3, shows that the Cauchy-type estimate works the same for functions α which are not necessarily increasing, provided they do not oscillate too much.

Let $\alpha : I \to \mathbb{C}$, with $I \subseteq \mathbb{R}$ an interval. The *total variation* $V(\alpha) \in [0, +\infty]$ of α on I is a measure of its oscillation,

$$V(\alpha; I) = \sup \left\{ \sum_{j=1}^{n} |\alpha(x_j) - \alpha(x_{j-1})| : \; x_0 < x_1 < \cdots < x_n; \; x_0, x_1, \ldots, x_n \in I \right\}. \tag{4.4.11}$$

When the interval is fixed, we simply write $V(\alpha)$. Clearly, $V(\alpha) = 0$ if and only if α is constant.

EXERCISE 4.5. *Let $\alpha : I \to \mathbb{R}$ be increasing on the interval I. Find an expression for $V(\alpha; I)$ when $I = (a,b), [a,b], (a,b], [a,b)$.*

The function α has *bounded variation* if $V(\alpha) < \infty$, and $BV(I)$ is the space of such functions. We say that α is *locally of bounded variation* on an interval I, $\alpha \in BV_{loc}(I)$, if $\alpha \in BV(J)$ for all intervals J compactly contained in I.

LEMMA 4.4. *Let $\alpha : \mathbb{R} \to \mathbb{C}$ be a function in $BV_{loc}(\mathbb{R})$. For $f \in C_c(\mathbb{R})$, define $\Lambda_n(f)$ as in (4.4.3). Then, the limit*

$$\Lambda(f) := \lim_{n\to\infty} \Lambda_n(f)$$

exists in $\mathbb{C}$, and it defines a linear functional on $C_c(\mathbb{R})$.

Of course, the functional Λ is not positive anymore, in general. Consider, for instance, a function α which decreases on $(-\infty, 0]$ and it increases on $[0, +\infty)$. We use the same notation as before,

$$\Lambda(f) =: \int_{\mathbb{R}} f d\alpha. \tag{4.4.12}$$

The educated guess here is that $d\alpha$ defines a *signed Borel measure*. This is correct, but to have a proof we need more information on functions of bounded variation, which will come in a later chapter.

The proof of lemma 4.4 is similar to that of lemma 4.3.

PROOF. Let $n, l \geq 1$ and $f \in C_c(\mathbb{R})$. Observe that:

$$\begin{aligned} \left| \alpha\left(\frac{j}{2^n}\right) - \alpha\left(\frac{j-1}{2^n}\right) \right| &\leq \sum_{m=1}^{2^l} \left| \alpha\left(\frac{(j-1)2^l + m}{2^{n+l}}\right) - \alpha\left(\frac{(j-1)2^l + m - 1}{2^{n+l}}\right) \right| \\ &\leq V\left(\alpha; \left(\frac{j-1}{2^n}, \frac{j}{2^n}\right]\right). \end{aligned}$$

For fixed $\epsilon > 0$, choose $n(\epsilon) > 0$ so large that, for $n \geq n(\epsilon)$, $|f(x) - f(y)| \leq \epsilon$ when $|x - y| \leq \frac{1}{2^n}$. Then, using the above inequality,

$$\begin{aligned}
|\Lambda_{n+l}(f) - \Lambda_n(f)| &= \left| \sum_{j=-\infty}^{+\infty} \sum_{m=1}^{2^l} \left[f\left(\frac{j}{2^n}\right) - f\left(\frac{(j-1)2^l + m}{2^{n+l}}\right)\right] \right. \\
&\quad \left. \cdot \left(\alpha\left(\frac{(j-1)2^l + m}{2^{n+l}}\right) - \alpha\left(\frac{(j-1)2^l + m - 1}{2^{n+l}}\right)\right)\right| \\
&\leq \epsilon \sum_{k:\ k2^{-(n+l)} \in \mathrm{supp(f)}} \left|\alpha(k2^{-(n+l)}) - \alpha((k-1)2^{-(n+l)})\right| \\
&\leq \epsilon V(\alpha, [-R-1, R+1]),
\end{aligned}$$

provided $[-R, R] \supset \mathrm{supp}(f)$. Thus, $\{\Lambda_n(f)\}$ is a Cauchy sequence, hence it converges to a value $\Lambda(f)$, which clearly satisfies the properties listed in the statement. □

EXERCISE 4.6. *Prove Lemma 4.4.*

EXERCISE 4.7. *Let I be an interval and x_0 be a point in I.*

(i) *Verify that $BV(I)$ is a (complex) linear space.*

(ii) *Show that $\|\alpha\|_{BV} := V(\alpha) + |\alpha(x_0)|$ defines a norm on $BV(I)$, and that $\sup_{x \in I} |\alpha(x)| \leq \|\alpha\|_{BV}$.*

(iii) *Show that a monotone function $\alpha : I \to \mathbb{R}$ lies in $BV(I)$ if and only if it is bounded, and, if α is increasing, $V(\alpha) = \sup_{x\in I} \alpha(x) - \inf_{x \in I} \alpha(x)$. In particular, monotone functions are BV on any compact interval $J \subseteq I$.*

4.4.4. More on increasing functions and Borel measures on $\mathbb{R}$.

In this subsection, the reader is invited to further investigate the relationship between (positive) Lebesgue-Stieltjes measures and increasing function. A broader and more complete picture of the topic will be presented in the last chapter.

4.4.4.1. *Distribution functions of Borel measures.* Recall that a *Borel measure* μ on $\mathbb{R}$ (in general, on a locally compact space (X, d)) is one which is defined on the Borel σ-algebra of $\mathbb{R}$ (or, of (X, d)), such that $\mu(K) < \infty$ whenever K is compact.

EXERCISE 4.8. *Let $\mu \geq 0$ be a Borel measure on $\mathbb{R}$. Define its* **distribution function** *α_μ to be*

$$\alpha_\mu(t) = \begin{cases} \mu((0,t]) & \text{if } t > 0, \\ -\mu((t, 0]) & \text{if } t \leq 0, \end{cases} \tag{4.4.13}$$

so that, in particular, $\alpha_\mu(0) = 0$.

(i) *Show that α_μ is increasing: if $s \leq t$, then $\alpha_\mu(s) \leq \alpha_\mu(t)$.*

(ii) *Show that* α_μ *is right continuous,* $\lim_{t\to a^+}\alpha_\mu(t)=\alpha_\mu(a)$*, and it has left limits,* $\lim_{t\to a_-}\alpha_\mu(t)$ *exists at all* $a\in\mathbb{R}$*.*

(iii) *Show that* α_μ *has a discontinuity at* a *if and only if* $\mu(\{a\})>0$*. Relate this with*

$$\lim_{t\to a^+}\alpha_\mu(t)-\lim_{t\to a^-}\alpha_\mu(t).$$

(iv) *Find an expression for* $\mu((a,b))$*,* $\mu([a,b))$*,* $\mu((a,b])$*,* $\mu([a,b])$ *in terms of* α_μ*, with* $-\infty\le a<b\le+\infty$ *(*$a,b=\pm\infty$ *are allowed only if the interval is open at the corresponding extreme).*

This definition of distribution function is different from the one usually given in Probability Theory. The reason is that here we deal with possibly infinite measures, and $\mu((-\infty,x])$ might be infinite for all real x. The choice of having $\alpha_\mu(0)=0$ is purely conventional. The difference between the two definitions is inconsequential: for a finite measure, the more analytic definition and the probabilistic one might differ by a constant, which is "lost in differentiation", or in differences at the endpoints of an interval.
Notation. For a function $\alpha\colon\mathbb{R}\to\mathbb{R}$ we denote by $\alpha(a^+)=\lim_{t\to a^+}\alpha(t)$ (if it exists) the limit from the right, and by $\alpha(a^-)=\lim_{t\to a^-}\alpha(t)$ that from the left. If they both exist, we denote by $\Delta\alpha(a)=\alpha(a^+)-\alpha(a^-)$ the *jump* of α at a.

EXERCISE 4.9. (i) *Let* $\alpha\colon\mathbb{R}\to\mathbb{R}$ *be increasing. Show that it has at most countably many points* $\{x_n:\ n\ge1\}$ *of discontinuity, that they are all jump discontinuities, and that for each interval* $[a,b]$*,*

$$0\le\sum_{x_n\in[a,b]}\Delta\alpha(x_n)<\infty.$$

(ii) *Find an increasing function* α *with a dense set of discontinuities.*

(iii) *Let* α *be as in (i), and define its* **right continuous regularization** α_r *to be*

$$\alpha_r(a)=\lim_{t\to a^+}\alpha(t).$$

Show that

(a) $\alpha_r(t)=\alpha(t)$ *for all points* t *which are not jumps for* α*;*

(b) α *and* α_r *have the same jumps points, and the same jumps at them.*

4.4.4.2. *Cantor's function.* Recall the definition of the Cantor set $C\subseteq[0,1]$, which was constructed by inductively removing "middle thirds" from intervals. Define the **Cantor function** $V\colon[0,1]\to[0,1]$ as follows. On $[0,1]\setminus C$

$$\begin{aligned} V(x) &= \frac{1}{2}\text{ on the middle third of } I_1^0=C_0;\\ V(x) &= \frac{1}{2^2}\text{ on the middle third of } I_1^1 \text{ and } \frac{3}{2^2}\text{ on that of } I_2^1; \end{aligned} \tag{4.4.14}$$

etcetera. The function V is now defined on $[0,1] \setminus C$, which is dense in $[0,1]$. For $x \in C$, define

$$V(x) = \sup_{y \in [0,1]\setminus C,\ y<x} V(y) = \lim_{y \to x^-} V(y).$$

Also, set $V(0) = 0$.

EXERCISE 4.10. *Show that V is continuous.*

EXERCISE 4.11. *Show that the Cantor set C has zero Lebesgue measure.*

Recall from the previous exercise set that $C = \cap_{n\geq 1} C_n$, where $C_n = \cup_{l=1}^{2^n} I_l^n$ is what is left of $[0,1]$ after removing middle thirds for n times, $I_l^n = [a_l^n, b_l^n]$.

EXERCISE 4.12. *For a function $f \in C_c(\mathbb{R})$ and integer $n \geq 1$, define*

$$\Lambda_n(f) = \sum_{j=0}^{2^n} f\left(b_j^n\right) \frac{1}{2^n}. \tag{4.4.15}$$

(i) *Show that $\Lambda(f) := \lim_{n\to\infty} \Lambda_n(f)$ exists and that Λ defines a positive, linear functional on $C_c(\mathbb{R})$.*
(ii) *Let μ_C be the associated Riesz measure. Show that μ_C is supported on C: $\mu_C(\mathbb{R}\setminus C) = 0$. Also, show that $\mu_C(I_j^n) = \frac{1}{2^n}$.*
(iii) *Let $\alpha(x) = \mu_C(-\infty, x]$ be the distribution function of μ_C. Show that $\alpha = V$ is Cantor's function.*

4.4.4.3. *Generalized Cantor sets.* We construct here the *generalized Cantor set with ratios* $\{\lambda_n\}_{n=1}^\infty$, $0 < \lambda_n < 1$.

(i) Start with $C^0 = I_1^0 = [0,1]$.
(ii) At the first step, let $C^1 = I_1^1 \cup I_2^1$, the two intervals left after removing the central portion of I_1^0 having length $1 - \lambda_1$. The length of each interval is $\frac{\lambda_1}{2}$.
(iii) Iterate the construction. Assuming at step $n-1$ we have $C^{n-1} = \cup_{j=1}^{2^{n-1}} I_j^{n-1}$, let $C^n \subset C^{n-1}$ be the union of the 2^n intervals obtained by removing from each I_j^{n-1} the middle portion which is in a ratio $1 - \lambda_n$ with the whole of I_j^{n-1}.
(iv) Let $C = \cap_{n=1}^\infty C^n$.

The *(classical) Cantor set* corresponds to $\lambda_n = \frac{2}{3}$ for all $n \geq 1$.

EXERCISE 4.13. *Show that C is compact, that all points of C are accumulation points of C (i.e. C is* **perfect***), and C totally disconnected.*

EXERCISE 4.14. *Recall that m denotes Lebesgue measure. Show that*

$$m(C) = \lim_{n\to\infty} \lambda_1 \lambda_2 \dots \lambda_n =: \Pi_{n=1}^\infty \lambda_n \tag{4.4.16}$$

(the last expression is a symbol denoting the **infinite product***, which is defined by the limit on its left). In particular, if $\lambda_n = \lambda \in (0,1)$ is constant (as in the case of the classical Cantor set), then $m(C) = 0$.*

EXERCISE 4.15. *Again with $0 < \lambda_n < 1$, show that $\Pi_{n=1}^{\infty}\lambda_n > 0$ if and only if*

$$\sum_{n=1}^{\infty}(1-\lambda_n) < \infty. \tag{4.4.17}$$

Hint. *Use the logarithm to turn the product into a sum, then use an estimate for $f(\lambda) = \log(\lambda)$ as $\lambda \to 1^-$.*

Use (4.4.17) to produce a "fat" Cantor set having positive Lebesgue measure.

EXERCISE 4.16. *Use Riesz representation theorem to define a Borel measure supported on the Cantor set in which interval of generation n have measure $\frac{1}{2^n}$. Is its distribution function always continuous?*

4.4.5. Weak derivatives of increasing functions.

Although increasing functions can have a dense set of (jump) discontinuities, they are rather differentiable, in the sense that the measure of the points where their derivative does not exist has zero Lebesgue measure. (Beware! The problem does not just consist in the discontinuities, but also in sets playing the role the Cantor set plays for the Cantor function). This result is hard to prove, and it will be presented in a later chapter. It is although easy to show that any (right continuous) increasing function α satisfies an integration-by-parts formula, in which the measure μ of which α is the distribution plays the role of α's derivative. We say that $\mu = \alpha'$ *in the weak sense*. For a more general discussion of *weak derivatives*, and of their relation with the usual *"strong" derivatives*, the best mathematical universe is that of *distribution theory*, in one or another of its avatars, depending on the particular context.

THEOREM 4.12 (Weak derivatives of increasing functions). *The following are equivalent for a real valued, measurable function f on $\mathbb{R}$:*

(i) *f is a.e. equal to a right continuous increasing function $\tilde{f}$;*

(ii) *$f \in L^1_{loc}$ and there exists a Borel measure $\mu \geq 0$ such that, for all $\varphi \in C^1_c(\mathbb{R})$,*

$$\int_{\mathbb{R}} \varphi'(x)f(x)dx = -\int_{\mathbb{R}} \varphi(t)d\mu(t). \tag{4.4.18}$$

Also, $\tilde{f}(x) = \tilde{f}(0) + \mu(0,x]$ if $x > 0$, and $\tilde{f}(x) = \tilde{f}(0) - \mu(x,0]$ if $x \leq 0$.

Lebesgue-Stieltjes notation gives (4.4.18) a more suggestive form:

$$\int_{\mathbb{R}} \varphi'(x)f(x)dx = -\int_{\mathbb{R}} \varphi(t)df(t). \tag{4.4.19}$$

EXERCISE 4.17. *Specialize the statement to the case of $f \in C^1$, and find what the measure μ is in this case.*

In the proof, we exchange order of integration a couple of times, using Fubini's Theorem.

PROOF. Suppose (i) holds. Then $f - f(0)$ is *a.e.* equal to the distribution function of a positive measure μ. Using the fact that $\int_{\mathbb{R}} \varphi'(x)dx = 0$ and assuming that $\text{supp}(\varphi) \subset (-R, R)$:

$$\begin{aligned}
\int_{\mathbb{R}} \varphi'(x)f(x)dx &= \int_{\mathbb{R}} \varphi'(x) \int_{(-R,x]} d\mu(t)dx = \int_{\mathbb{R}} \varphi'(x) \int_{\mathbb{R}} \chi_{(-R,x]}(t)d\mu(t)dx \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} \chi_{(-R,x]}(t)\varphi'(x)dxd\mu(t) = \int_{(-R,R)} \int_t^R \varphi'(x)dxd\mu(t) \\
&= \int_{(-R,R)} (\varphi(R) - \varphi(t))d\mu(t) \\
&= -\int_{\mathbb{R}} \varphi(t)d\mu(t).
\end{aligned}$$

Suppose (ii) holds. Then, using the same calculations as above,

$$\begin{aligned}
\int_{\mathbb{R}} \varphi'(x)f(x)dx &= -\int_{\mathbb{R}} \varphi(t)d\mu(t) \\
&= \int_{\mathbb{R}} \varphi'(x) \int_{(-R,x]} d\mu(t)dx
\end{aligned}$$

for all $\varphi \in C^1_c(\mathbb{R})$, i.e. for all choices of $\psi = \varphi' \in C_c(\mathbb{R})$ such that $\int_{\mathbb{R}} \psi(x)dx = 0$. By the 2^{nd} Fundamental Lemma 4.2, there is real C such that $f(x) = \mu(0, x] + C$, i.e. f is *a.e.* equal to the increasing function on the right. After changing the values of $f(x)$ on a set of null measure, we can assume equality everywhere, with $C = f(0)$.

The use of Fubini theorem was justified, since $F(x, t) = \chi_{(-R,x]}(t)\varphi'(x)$ belongs to $L^1(\mathbb{R} \times \mathbb{R})$. □

4.5. Riemann integration vs. Lebesgue integration

In the chapter on measure theory, we saw some shortcomings of Riemann's definition of integral. Actually, it is not even a priori clear which functions are Riemann integrable. At the end of the XIX century, Borel characterized the Riemann integrable functions in terms of *sets having vanishing length*. This paved the way to a broader definition of length itself, and to Lebesgue's definition of integral.

In this section we state and prove Borel's Theorem. Also, when we defined Riemann's integral, for expository reasons we limited ourselves to dyadic partitions of the interval, and one might worry that this way we illicitly enlarged the class of the Riemann integrable functions. We will show below that the two definitions are, in fact, equivalent.

4.5.1. Riemann's integral and oscillations.

4.5.1.1. *Partitions and oscillations of a function.* Recall that the *oscillation* on a set I of a bounded, real valued function f defined on I is

$$\operatorname{Osc}(f, I) = \sup_{x\in I} f(x) - \inf_{x\in I} f(x) \geq 0. \tag{4.5.1}$$

It is immediate that if $J \subset I$, then $\operatorname{Osc}(f, I) \geq \operatorname{Osc}(J)$, hence, if A and B are disjoint intervals in $\mathbb{R}$, and $C = A \cup B$, we have

$$\operatorname{Osc}(f, C)m(C) \geq \operatorname{Osc}(f, A)m(A) + \operatorname{Osc}(f, B)m(B). \tag{4.5.2}$$

Here, $m((a, b)) = m([a, b]) = m((a, b]) = m([a, b)) = b - a$ is the length of the interval, which is also its Lebesgue measure.

Let $\mathcal{A} = \{I_i,\ i = 1, \dots, n\}$ be a *partition* of the interval $[0, 1]$: $0 = x_0 < x_1 < \cdots < x_{n-1} < x_n = 1$, $I_1 = [x_0, x_1]$ and $I_i = (x_{i-1}, x_i]$ for $i = 2, \dots, n$. Define the *resolution* of $\mathcal{A}$ to be $\delta(\mathcal{A}) = \max\{m(I_i) :\ i = 1, \dots, n\}$. The *average oscillation* of f with respect to the partition $\mathcal{A}$ is

$$\operatorname{AvOsc}(f, \mathcal{A}) := \sum_{I\in\mathcal{A}} \operatorname{Osc}(f, I)m(I).$$

Draw a picture to have an intuition of the quantity $\operatorname{AvOsc}(f, \mathcal{A})$. Inequality (4.5.2) implies monotonicity of $\operatorname{AvOsc}(f, \cdot)$:

$$\text{if } \mathcal{B} \subset \mathcal{A}, \text{ then } \operatorname{AvOsc}(f, \mathcal{B}) \geq \operatorname{AvOsc}(f, \mathcal{A}). \tag{4.5.3}$$

We say that $f : [0, 1] \to \mathbb{R}$ is *Riemann integrable* if

$$\inf \{\operatorname{AvOsc}(f, \mathcal{A}) :\ \mathcal{A} \text{ is a partition of } [0, 1]\} = 0. \tag{4.5.4}$$

Riemann integrability has an $\epsilon - \delta$ formulation.

PROPOSITION 4.6. *A bounded, real valued function f on $[0, 1]$ is Riemann integrable if and only if, for any $\epsilon > 0$, there is $\delta > 0$ dependent on ϵ and f alone, such that, if $\mathcal{A}$ is a partition of $[0, 1]$, then*

$$\text{if } \delta(\mathcal{A}) \leq \delta, \text{ then } AvOsc(f, \mathcal{A}) \leq \epsilon. \tag{4.5.5}$$

PROOF. The "if" part is clear since, by (4.5.3), by choosing partitions with smaller and smaller resolution, the assumption implies (4.5.4).

In the other direction, suppose that (4.5.4) does not hold. Then, there is a positive ϵ such that, for any $\delta > 0$ there is a partition $\mathcal{A}$ with $\delta(\mathcal{A}) \leq \delta$, yet $\operatorname{AvOsc}(f, \mathcal{A}) \geq \epsilon$.

Let now $\mathcal{B}$ be any partition of $[0,1]$, and let $N+1 = \sharp(\mathcal{B})$, $\delta > 0$ a number to be chosen in such a way $\delta < \delta(\mathcal{B})$, and $\mathcal{A}$ with $\delta(\mathcal{A}) < \delta$ as above. Also, let $\mathcal{A} \vee \mathcal{B} = \{I_i \cap J_j;\ I_i \in \mathcal{A},\ J_j \in \mathcal{B}\}$ be the partition having a set of endpoints the union of the endpoints of $\mathcal{A}$ and $\mathcal{B}$. We have then that $\delta(\mathcal{A} \vee \mathcal{B}) \leq \delta$ and

$$\begin{aligned} \mathrm{AvOsc}(f,\mathcal{B}) &\geq \mathrm{AvOsc}(f,\mathcal{B}\vee\mathcal{A}) \\ &= \mathrm{AvOsc}(f,\mathcal{A}) - [\mathrm{AvOsc}(f,\mathcal{A}) - \mathrm{AvOsc}(f,\mathcal{B}\vee\mathcal{A})] \\ &\geq \epsilon - [\mathrm{AvOsc}(f,\mathcal{A}) - \mathrm{AvOsc}(f,\mathcal{B}\vee\mathcal{A})]. \end{aligned}$$

Let's look at the difference of sums given by: $\mathrm{AvOsc}(f,\mathcal{A}) - \mathrm{AvOsc}(f,\mathcal{B}\vee\mathcal{A})$. If $I \in \mathcal{A}$ is an element of $\mathcal{B}\vee\mathcal{A}$ (and this happens if and only if it is included in some $J \in \mathcal{B}$), then it is cancelled out in the difference. The I's giving a contribution are then at most the $2(N-1)$ ones possibly containing some of the $N-1$ endpoints of the partition $\mathcal{B}$ which lie in $(0,1)$. Let $\mathcal{E}$ be the set of such intervals, and for $I \in \mathcal{E}$, $I = I_- \cup I_+$ is split into two parts, where $I_-, I_+ \in \mathcal{B}\vee\mathcal{A}$ are the left and right part into which I is divided by an endpoint of $\mathcal{B}$. For each I in $\mathcal{E}$, we simply estimate

$$\begin{aligned} \mathrm{Osc}(f,I)m(I) - [\mathrm{Osc}(f,I_-)m(I_-) + \mathrm{Osc}(f,I_+)m(I_+)] &\leq \mathrm{Osc}(f,[0,1])m(I) \\ &\leq \mathrm{Osc}(f,[0,1])\delta. \end{aligned}$$

Summing the contributions,

$$\begin{aligned} \mathrm{AvOsc}(f,\mathcal{A}) - \mathrm{AvOsc}(f,\mathcal{B}\vee\mathcal{A}) &\leq (N-1)\mathrm{Osc}(f,[0,1])\delta \\ &\leq \epsilon/2 \end{aligned}$$

provided $\delta > 0$ is chosen small enough. Thus, $\mathrm{AvOsc}(f,\mathcal{B}) \geq \epsilon/2$ for all partitions $\mathcal{B}$, and f is not Riemann integrable. □

4.5.1.2. *The defition of the Riemann integral.* We can now define the *Riemann integral* of a Riemann integrable function $f : [0,1] \to \mathbb{R}$. For a partition $\mathcal{A}$, define the corresponding *inferior and superior sums* to be

$$\begin{aligned} s(f,\mathcal{A}) &= \sum_{I\in\mathcal{A}} \inf_I f, \\ S(f,\mathcal{A}) &= \sum_{I\in\mathcal{A}} \sup_I f, \text{ so that} \\ \mathrm{AvOsc}(f,\mathcal{A}) &= S(f,\mathcal{A}) - s(f,\mathcal{A}). \end{aligned}$$

Since for any partitions $\mathcal{A}$, $\mathcal{B}$ we have

$$s(f,\mathcal{A}) \leq s(f,\mathcal{A}\vee\mathcal{B}) \leq S(f,\mathcal{A}\vee\mathcal{B}) \leq S(f,\mathcal{B}),$$

we have that

$$\underline{\int_0^1} f(x)dx := \sup_{\mathcal{A}} s(f,\mathcal{A}) \le \inf_{\mathcal{A}} S(f,\mathcal{A}) =: \overline{\int_0^1} f(x)dx, \tag{4.5.6}$$

and equality holds if and only if f is Riemann integrable, since it is easy to verify that

$$\overline{\int_0^1} f(x)dx - \underline{\int_0^1} f(x)dx = \inf_{\mathcal{A}} \mathrm{AvOsc}(f,\mathcal{A}).$$

If f is Riemann integrable,

$$\int_0^1 f(x)dx := \underline{\int_0^1} f(x)dx = \overline{\int_0^1} f(x)dx$$

is the *Riemann integral* of f. Proposition 4.6 can be reformulated as follows: f is Riemann integrable if and only if, for all $\epsilon > 0$, there is $\delta > 0$ such that, whenever $\delta(\mathcal{A}) \le \delta$, $S(f,\mathcal{A}) - s(f,\mathcal{A}) \le \epsilon$.

LEMMA 4.5. *If f is Riemann integrable and $\{\mathcal{A}_m : m \ge 1\}$ is a sequence of partitions with $\delta(\mathcal{A}_m) \to 0$ as $m \to \infty$, then*

$$\lim_{m\to\infty} s(f,\mathcal{A}_m) = \lim_{m\to\infty} S(f,\mathcal{A}_m) = \int_0^1 f(x)dx.$$

PROOF. We have that

$$\begin{aligned}
\int_0^1 f(x)dx &\le S(f,\mathcal{A}_m) \\
&= [S(f,\mathcal{A}_m) - s(f,\mathcal{A}_m)] + s(f,\mathcal{A}_m) \\
&\le [S(f,\mathcal{A}_m) - s(f,\mathcal{A}_m)] + \int_0^1 f(x)dx \\
&\to \int_0^1 f(x)dx \qquad (4.5.7)
\end{aligned}$$

as $m \to \infty$, hence $\lim_{m\to\infty} S(f,\mathcal{A}_m) = \int_0^1 f(x)dx$, and similarly $\lim_{m\to\infty} s(f,\mathcal{A}_m) = \int_0^1 f(x)dx$. □

4.5.1.3. *Riemann sums.* Indeed, we would like to have some more practical way to compute the integral of f. This is provided by the next result. The relevant objects are *summing frames* $(\mathcal{A}, A)$, where $\mathcal{A} = \{I_i : i = 1,\ldots,n\}$ is a partition of $[0,1]$ and $A = \{t_i : i = 1,\ldots,n\} \subset [0,1]$ has an element in each interval of the partition, $t_i \in I_i$. The associated *Riemann sum* is

$$\Sigma(f;\mathcal{A},A) := \sum_{i=1}^n f(t_i)m(I_i) \in [s(f,\mathcal{A}), S(f,\mathcal{A})]. \tag{4.5.8}$$

Riemann sums approximate Riemann integrals.

THEOREM 4.13. *Let $f : [0,1] \to \mathbb{R}$ be bounded. If f is Riemann integrable and $\{(\mathcal{A}_m, A_m) : \ m \geq 1\}$ is a sequence of summing frames, $\mathcal{A}_m = \{I_{m,i} : 1 \leq i \leq n_m\}$, $A_m = \{t_{m,i} : \ 1 \leq i \leq n_m\}$, such that*

$$\lim_{m\infty} \delta(\mathcal{A}_m) = 0,$$

then

$$\int_0^1 f(x)dx = \lim_{m\to\infty} \Sigma(f; \mathcal{A}, A) = \lim_{m\to\infty} \sum_{i=1}^{n_m} f(t_{m,i}) m(I_{m,i}) \tag{4.5.9}$$

PROOF. We have

$$s(f, \mathcal{A}_m) \leq \sum_{i=1}^{n_m} f(t_{m,i}) m(I_{m,i}) \leq S(f, \mathcal{A}_m),$$

and, since f is Riemann integrable, by Lemma 4.5, $\Sigma(\mathcal{A}_m, A_m) \to \int_0^1 f(x)dx$. □

4.5.2. A characterization of Riemann integrable functions. Here is the result where the story of Lebesgue integration started. Borel gave a definition of set of zero measure in terms of covering, that soon after Lebesgue generalized and used to define the integral that goes under his name. Borel's motivation was characterizing Riemann integrable functions.

THEOREM 4.14. *A bounded function $f : [0,1] \to \mathbb{R}$ is Riemann integrable if and only if*

$$m(\{x; \ f \text{ is discontinuous at } x\}) = 0. \tag{4.5.10}$$

PROOF. Let $D(f)$ be the set of points in $[0,1]$ where f is discontinuous.

We begin with the easy "only if direction". Fix $\epsilon \geq 0$, and let

$$D(\epsilon) = \{x : \text{ for all } \delta > 0 \text{ there is } y \in [0,1] \text{ such that } |f(x) - f(y)| > \epsilon\}.$$

Then $[0,1] \supseteq D(\eta) \supseteq D(\epsilon)$ if $\eta < \epsilon$, and $D(f) = \bigcup_{\epsilon>0} D(\epsilon)$. Consider the dyadic partition $\mathcal{A}_n$ of $[0,1]$, having as endpoints the dyadic rationals $j/2^n$ $(0 \leq j \leq 2^n)$, and let $A_n = \{j : \ \mathrm{Osc}(f, I_{n,j}) \geq \epsilon\}$. If $x \in D(\epsilon)$ and $x \notin \mathbb{Q}_2$ is not a dyadic rational, it lies in the interior of $I_{n,j}$ for some j, hence $\mathrm{Osc}(f, I_{n,j}) \geq \epsilon$, so $D(\epsilon) \setminus \mathbb{Q} \subseteq A_n$. We have, then

$$\begin{aligned} m(D(\epsilon)) &= m(D(\epsilon) \setminus \mathbb{Q}_2) \leq 2^{-n} \sharp(A_n) \\ &\leq 2^{-n} \frac{1}{\epsilon} \sum_{j=1}^{2^n} \mathrm{Osc}(f, I_{n,j}) \end{aligned}$$

$$\begin{aligned} & \text{because } \mathrm{Osc}(f, I_{n,j}) \geq \epsilon \text{ for } j \in A_n \\ = \;& \frac{1}{\epsilon}\mathrm{AvOsc}(f, \mathcal{A}_n) \\ \to \;& 0 \text{ as } n \to \infty \end{aligned}$$

because f is Riemann integrable. Hence, $D(\epsilon) > 0$ for all positive ϵ. Choose $\epsilon_k = 1/k$. The set of discontinuities is $\bigcup_k D(1/k)$, hence it has zero measure by Dominated Convergence.

In the other direction, suppose by contradiction that f is not Riemann integrable and let $\mathcal{A}_n$'s be the dyadic partitions with resolution $1/2^n$. By Proposition 4.6, there is $c > 0$ such that $\mathrm{AvOsc}(f, \mathcal{A}_n) \geq c$ for all n, and by rescaling we can assume $c = 1$. Fix $\epsilon > 0$, to be chosen later, and let

$$D(\epsilon, n) = \{x : \; |f(x) - f(y)| > \epsilon \text{ for some } y \text{ with } |x - y| \leq 1/2^n\}.$$

and observe that $D(\epsilon) = \bigcap_n D(\epsilon, n)$. Also observe that if $\mathrm{Osc}(f, I_{n.j}) \geq 2\epsilon$, then $I_{n,j} \subseteq D(\epsilon, n)$. We can now estimate

$$\begin{aligned} 1 \;\leq\;& \mathrm{AvOsc}(f, \mathcal{A}_n) = \sum_j \mathrm{Osc}(f, I_{n,j}) m(I_{n,j}) \\ =\;& \mathrm{Osc}(f, [0,1]) \sum_{j:\mathrm{Osc}(f, I_{n.j}) \geq 2\epsilon} m(I_{n,j}) + 2\epsilon \sum_{j:\mathrm{Osc}(f, I_{n.j}) \leq 2\epsilon} m(I_{n,j}) \\ \leq\;& \mathrm{Osc}(f, [0,1]) m(D(\epsilon, n)) + 2\epsilon, \end{aligned}$$

i.e.

$$m(D(\epsilon, n)) \geq \frac{1 - 2\epsilon}{\mathrm{Osc}(f, [0,1])}.$$

Choose $\epsilon = 1/3$, and let $n \to \infty$. By Dominated Convergence, $m(D(\epsilon)) \geq 1/(3\mathrm{Osc}(f, [0,1])) > 0$. ☐

Corollary 4.4. *Riemann integrable functions are Lebesgue integrable and the two notions of integral coincide.*

Proof. Let f be Riemann integrable, let $D(f)$ be the set of its points of discontinuity, and $f_1 : [0,1] \setminus D(f) \to \mathbb{R}$ be the restriction of f to $[0,1] \setminus D(f)$, which is continuous. For any real α, $f_1^{-1}(\alpha, \infty)$ is open in $[0,1] \setminus D(f)$: for any x in $f_1^{-1}(\alpha, \infty)$ there is $r_x > 0$ such that $B(x, r_x) \setminus D(f) \subseteq f_1^{-1}(\alpha, \infty)$ (here $B(x, r) = (x - r, x + r) \cap [0,1]$ is the metric unit ball for the Euclidean metric in $[0,1]$). Let $V = \cup_{x \in f_1^{-1}(\alpha,\infty)} B(x, r_x)$, so that $f_1^{-1}(\alpha, \infty) = V \setminus D(f)$. Then,

$$f^{-1}(\alpha, \infty) = [V \setminus D(f)] \cup \left[f^{-1}(\alpha, \infty) \cap D(f)\right].$$

Since $m(D(f)) = 0$ by Theorem 4.14, the set $f^{-1}(\alpha, \infty)$ is the union of two measurable, sets, hence it is measurable.

Consider now the dyadic partitions $\mathcal{A}_n$ having endpoints $j/2^n$ ($j = 0, 1, \ldots, 2^n$). The Riemann integral of f is

$$\begin{aligned}\int_0^1 f(x)dx &= \lim_{n\to\infty} \sum_j \inf_{x\in I_{n,j}} f(x)m(I_{n,j}) \\ &= \lim_{n\to\infty} \int_0^1 \sum_j \inf_{I_{n,j}} f dm \\ &= \lim_{n\to\infty} \int_0^1 f_n dm,\end{aligned}$$

where the integral in the second line is the Lebesgue integral of a step function. The functions $f_n = \int_0^1 \sum_j \inf_{I_{n,j}} f \leq f_{n+1}$ increase towards a limit function $\varphi \leq f$, and by Monotone Convergence (we can use it because f is bounded and defined on a set of finite measure)

$$\int_0^1 f(x)dx \leq \int_{[0,1]} f(x)dx.$$

where the second integral is in the Lebesgue sense.

The same reasoning with sup instead of inf shows the opposite inequality.

□

CHAPTER 5

Hilbert spaces

Among function spaces, Hilbert spaces are probably the most ubiquitous in mathematics and its applications. At its inception, "Hilbert theory" dealt with concrete L^2 spaces, but soon an abstract theory was developed. The advantage of the abstract theory is that it makes it easier to recognize when a Hilbert space structure underlies a cluster of mathematical objects and phenomena, and it helps in generating Hilbert spaces in which better state, and solve, theoretical, as well as practical problems. Although all Hilbert spaces are isomorphic to some L^2 space, the point of view that knowledge of L^2 is all that's needed is far too simplistic, and basically incorrect. Most Hilbert spaces are spaces of functions defined on some set of points, and such "points" constitute further structure, which often is at the core of the problem we have at hands. Is this chapter, however, we are mostly interested in the abstract theory.

Another useful way of thinking is that of viewing at Hilbert spaces as generalizations (often infinite-dimensional, with complex rather than real scalars) of the Euclidean space. Even if in applications each vector in the Hilbert space represents a functions, we can think of each of them as a "point" in a linear space where notions like orthogonality make perfect sense. This intuition is helpful in translating complex phenomena in simple pictures, and pictures into statements (which are sometimes true, sometimes false).

5.1. Basic geometry of Hilbert spaces and Riesz Lemma

5.1.1. Definition and basic properties. An *inner product* on a vector space V over $\mathbb{C}$ (or $\mathbb{R}$) is a map $\langle \cdot \mid \cdot \rangle : V \times V \to \mathbb{C}$ such that:

(i) $\langle x \mid x \rangle \geqslant 0$ and $\langle x \mid x \rangle = 0$ if and only if $x = 0$;

(ii) $\langle z \mid \alpha x + \beta y \rangle = \alpha \langle z \mid x \rangle + \beta \langle z \mid y \rangle$ for $x, y, z \in V$ and $\alpha, \beta \in \mathbb{C}$;

(iii) $\overline{\langle x \mid y \rangle} = \langle y \mid x \rangle$.

Two vectors x, y are orthogonal, $x \perp y$, if $\langle x \mid y \rangle = 0$. We define $\|x\| := \langle x \mid x \rangle^{1/2}$ to be the norm of $x \in V$. A simple calculation gives:

Lemma 5.1. *[Pythagorean relation] Let $x, y \in V, x \neq 0$. Then,*

$$\|y\|^2 = \left\| y - \frac{x}{\|x\|} \left\langle \frac{x}{\|x\|} \mid y \right\rangle \right\|^2 + \left| \left\langle \frac{x}{\|x\|} \mid y \right\rangle \right|^2 .$$

Proof. Expand the right hand side using the definition of norm and the properties of the inner product. □

Exercise 5.1. *Explain why the lemma is called as it is with a picture, or observing that the vectors $\frac{x}{\|x\|} \left\langle \frac{x}{\|x\|} \mid y \right\rangle$ and $y - \frac{x}{\|x\|} \left\langle \frac{x}{\|x\|} \mid y \right\rangle$ are orthogonal and their sum is y.*

Corollary 5.1. *[Cauchy-Schwarz inequality]For $x, y \in V$ we have*

$$|\langle x \mid y \rangle| \leqslant \|x\| \|y\|,$$

with equality if and only if x, y are linearly dependent.

Proof. If $x = 0$, there is nothing to prove. Otherwise, the inequality follows from dropping the first summand from the right of the Pythagorean relation. In case of equality, either $x = 0$ or the first term in the right of the Pythagorean relation vanishes, in which case x, y are linearly dependent. Conversely, if they are linearly dependent it is easy to see that equality holds in Cauchy-Schwarz. □

Proposition 5.1. *The function $x \mapsto \|x\|$ defines a norm on V.*

Proof. It is an immediate consequence of Cauchy-Schwarz:

$$\begin{aligned}
\|x + y\|^2 &= \langle x + y \mid x + y \rangle \\
&= \|x\|^2 + \langle x \mid y \rangle + \langle y \mid x \rangle + \|y\|^2 \\
&= \|x\|^2 + 2 \operatorname{Re}(\langle x \mid y \rangle) + \|y\|^2 \\
&\leqslant \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 \\
&= (\|x\| + \|y\|)^2.
\end{aligned}$$

□

An inner product space $(H, \langle \cdot \mid \cdot \rangle)$ is a *Hilbert space* if it is complete with respect to the norm induced by the inner product.

Exercise 5.2. *Show that the inner product can be written in terms of norms by means of the* **polarization identity***:*

$$\langle x \mid y \rangle = \frac{1}{4} \left[\left(\|x + y\|^2 - \|x - y\|^2 \right) + i \left(\|x + iy\|^2 - \|x - iy\|^2 \right) \right] . \tag{5.1.1}$$

Lemma 5.2. *[Parallelogram law] Let $x, y \in V$. Then,*

$$\|x + y\|^2 + \|x - y\|^2 = 2 \left(\|x\|^2 + \|y\|^2 \right) .$$

PROOF. Expand the expression on the left using the definition of $\|\cdot\|$. $\square$

EXERCISE 5.3. *Show that, if $\|\cdot\|$ is a norm on a vector space which satisfies the parallelogram law, then it comes from an inner product as defined by the polarization identity.*

EXERCISE 5.4. *Show that the (two-dimensional) Banach spaces $\ell^\infty(\{0,1\}), \ell^1(\{0,1\})$ are not inner product spaces (i.e. that their norm does not come from an inner product). A* **Hint.** *The parallelogram law fails.*

5.1.2. L^2 as a Hilbert space. Let $(X, \mathcal{F}, \mu)$ be a measure space. Observe that, if $f, g \in L^2(\mu)$, then $\overline{f}g \in L^1(\mu)$:

$$\int_x |\overline{f}g| d\mu \leq \|f\|_{L^2}\|g\|_{L^2}$$

by Hölder's inequality with $p = p' = 2$. We can then define

$$\langle f|g\rangle_{l^2} = \int_X \overline{f}g d\mu,$$

which satisfies the properties of a inner product, with associated norm $\langle f|f\rangle_{l^2} = \|f\|^2_{L^2}$. We saw that $L^2(\mu)$ is complete with respect to this norm.

5.1.3. Projections onto subspaces. Let $M \subseteq H$ be a closed, linear subspace of H. Its *orthogonal complement*, $M^\perp = H \ominus M$, is:

$$M^\perp = \{x \in H : x \perp y \text{ for all } y \in M\}.$$

LEMMA 5.3 (Projection Lemma). *Let $M \subseteq H$ be a closed, linear subspace of H. Then, for each $x \in H$ there exists a unique $u \in H$ such that*

$$\|x - u\| = \inf\{\|x - y\| : y \in M\}.$$

PROOF. Let $\{y_n\}_{n\geqslant 1}$ be a sequence in M such that $\|y_n - x\| \to \delta := \inf\{\|x - y\| : y \in M\}$. We show that it is a Cauchy sequence- By the parallelogram law,

$$\begin{aligned}
\|y_m - y_n\|^2 &= \|(y_m - x) - (y_n - x)\|^2 \\
&= 2\left(\|y_m - x\|^2 + \|y_n - x\|^2\right) - \|y_m + y_n - 2x\|^2 \\
&= 2\left(\|y_m - x\|^2 + \|y_n - x\|^2\right) - 4\left\|\frac{y_m + y_n}{2} - x\right\|^2 \\
&\leqslant 2\left(\|y_m - x\|^2 + \|y_n - x\|^2\right) - 4\delta^2 \\
&\quad \text{because } \frac{y_m + y_n}{2} \in M \\
&\to 4\delta^2 - 4\delta^2 = 0.
\end{aligned}$$

We now let $u = \lim y_n \in M$ (because M is closed).

If u' is another point with the same property, then $\frac{u+u'}{2} \in M$ and

$$\|u - u'\|^2 = 2\left(\|u - x\|^2 + \|u' - x\|^2\right) - 4\left\|\frac{u + u'}{2} - x\right\| \leqslant 4\delta^2 - 4\delta^2 = 0,$$

hence, $u = u'$. □

We call $u = \pi_M(x)$ the *orthogonal projection* of x onto M.

Proposition 5.2. *Let $E \subseteq H$ and define $E^\perp = \{x \in H : x \perp y$ for all $y \in E\}$. Then,*

(i) *$E^\perp$ is a closed, linear subspace of H, and $E \cap E^\perp = \{0\}$;*

(ii) *if M is a linear subspace of H, then $M^\perp = \overline{M}^\perp$;*

(iii) *if M is a linear subspace of H, then $\left(M^\perp\right)^\perp = \bar{M}$ is the closure of M in H.*

Proof. (i) Let $x, y \in E^\perp$, $\alpha, \beta \in \mathbb{C}$, and let $z \in E$. Then,

$$\langle z|\alpha x + \beta y\rangle = \alpha\langle z|x\rangle + \beta\langle z|y\rangle = 0,$$

hence, $\alpha x + \beta y \in E^\perp$.

Let now $E^\perp \ni x_n$ and $\lim_{n\to\infty} \|x_n - x\| = 0$. Then, for $z \in E$,

$$|\langle z|x\rangle| = |\langle z|x\rangle - \langle z|x_n\rangle| \leq \|z\| \cdot \|x - x_n\| \to 0,$$

as $n \to \infty$. So, $|\langle z|x\rangle|$ is smaller than any positive ϵ, hence $\langle z|x\rangle = 0$, i.e. $x \in E^\perp$.

Finally, if $x \in E \cap E^\perp$, then $0 = \langle v|v\rangle = \|v\|^2$, hence, $v = 0$.

(iii) $\left(M^\perp\right)^\perp$ is linear and closed by (i), and $\left(M^\perp\right)^\perp \supseteq M$: if $x \in M$, then, for $z \in M^\perp$ we have that $\langle z|x\rangle = 0$, then $x \in \left(M^\perp\right)^\perp$. We have thus proved that $\overline{M}$, the smallest closed linear subspace containing M, is a subset of $\left(M^\perp\right)^\perp$.

Consider $x \in \left(M^\perp\right)^\perp \setminus \overline{M}$. By the Projection Lemma, $x = u + v$ with $u \in \overline{M}$, and $v \in \overline{M}^\perp = M^\perp$. So, $v \in M^\perp \cap \left(M^\perp\right)^\perp$, hence, $v = 0$ and $x \in \overline{M}$, contradicting the assumption. □

Exercise 5.5. *Let E be a subset of H. Show that $(E^\perp)^\perp = \overline{span(E)}$ is the closure of the linear span of E in H.*

Theorem 5.1. *[Orthogonal Decomposition] Let $M \subseteq H$ be a closed, linear subspace of H. Then, for each $x \in H$ there is a unique decomposition $x = u + v$ with $u \in M$ and $v \in M^\perp$.*

PROOF. Let $u \in M$ be as in the Projection Lemma, and for $t \in \mathbb{R}$ and $w \in M$, so that $u + tw \in M$, define $f(t) = \|u + tw - x\|^2$. By minimality, $f'(0) = 0$, and

$$\begin{aligned} f'(0) &= \frac{d}{dt}\bigg|_{t=0} \left(\|u - x + tw\|^2\right) \\ &= \frac{d}{dt}\bigg|_{t=0} \left(\|u - x\|^2 + 2t\,\mathrm{Re}(\langle u - x \mid w\rangle) + t^2\|w\|^2\right) \\ &= \left[2\,\mathrm{Re}(\langle u - x \mid w\rangle) + 2t\|w\|^2\right]_{t=0} \\ &= 2\,\mathrm{Re}(\langle u - x \mid w\rangle). \end{aligned}$$

Hence, $\mathrm{Re}(\langle u - x \mid w\rangle) = 0$. The same reasoning with $g(t) = \|u + itw - x\|^2$ gives $\mathrm{Im}(\langle u - x \mid w\rangle) = 0$. Hence, $\langle u - x \mid w\rangle$ for all $w \in M$, i.e. $v := x - u \in M^\perp$.

If I have two decompositions $x = u + v = u' + v'$ with $u, u' \in M$ and $v, v' \in M^\perp$, then $u - u' = v' - v \in M \cap M^\perp$, hence $u - u' = v' - v = 0$. □

In the theorem on the orthogonal decomposition of a vector, we saw that, if M is a closed subspace of a Hilbert space H, and $x \in H$, then there exist unique $u \in M$ and $v \in M^\perp$ such that $x = u + v$. Define $\pi_M : H \to H$ by $\Pi_M(x) = u$, the *(orthogonal) projection* of x onto M.

EXERCISE 5.6. *Let M be a closed, linear subspace of a Hilbert space H.*

(i) *Show that $\pi_M : H \to H$ is linear and that $\|\pi_M x\| \leq \|x\|$ for all x in H.*

(ii) *Show that $\pi_M^2 = \pi_M$ and that π_M is* **self-adjoint**, *i.e. that*

$$\langle \pi_M x \mid y\rangle = \langle x \mid \pi_M y\rangle$$

for $x, y \in H$.

(iii) *Verify that $\Pi_M + \Pi_{M^\perp} = I$ is the identity operator ($I(x) = x$), and that $\Pi_M \circ \Pi_{M^\perp} = 0$.*

It is an interesting fact that the Exercise 5.6 has a converse: all operators sharing the properties of projections are in fact projections on closed subspaces. We can, that is, indifferently work with the *lattice* of the closed subspaces of H, or with the lattice of the projection operators. Since projections operators are spacial instances of linear operators on H, this identification gives much more flexibility in theory and calculations. (This is similar to the case of σ-algebras, where characteristic functions of measurable sets are special instances of measurable functions, and measurable functions have a rich structure).

PROPOSITION 5.3. *Let $\pi : H \to H$ be a linear operator satisfying:*

(i) $\pi^2 = \pi$*;*

(ii) $\pi^* = \pi$*;*

(iii) $\|\pi(x)\| \leq C\|x\|$ *for some* $C > 0$.

Then, there exists a closed, linear subspace M *of* H *such that* $\pi = \pi_M$.

Proof. Let $M = \{\pi(x) : x \in H\}$ be the range of π. For $\pi(x), \pi(y) \in M$ and $\alpha, \beta \in \mathbb{C}$ we have $\alpha\pi(x) + \beta\pi(y) = \pi(\alpha x + \beta y) \in M$, hence M is a linear space. Let $M \ni \pi(x_n) \to y$ in H. Then, $\pi(x_n) = \pi^2(x_n) \to \pi(y)$ because

$$\|\pi^2(x_n) - \pi(y)\| \leq C\|\pi(x_n) - y\| \to 0 \text{ as } n \to \infty.$$

Thus, $y = \pi(y) \in M$, showing that M is closed.

Finally, let x be an element of H, and write $x = u + v = \pi_M(x) + \pi_{M^\perp}(x)$, as in Theorem 5.1. By definition of M, $u = \pi(w)$ for some w. For $y \in H$ we have:

$$\langle \pi(v) | y \rangle = \langle v | \pi(y) \rangle = 0,$$

since $\pi(y) \in M$ and $v \in M^\perp$. Hence, $\pi(v) = 0$, and so

$$\pi(x) = \pi(u) = \pi^2(w) = \pi(w) = u = \pi_M(x),$$

showing $\pi = \pi_M$. □

5.1.4. F. Riesz representation in Hilbert spaces. F. Riesz representation theorem in Hilbert spaces, like the one we saw concerning measures, shows that any element from an abstract collection of objects can be represented as an object from a very special and concrete subclass. The first theorem we saw represented positive functionals on $C_c(X)$ in terms of measures; the present one represents bounded, linear functionals on H in terms of inner products.

Exercise 5.7. *Let* $y \in H$ *and define* $T_y(x) = \langle y \mid x \rangle$. *Then,* $T_y : H \to \mathbb{C}$ *is a linear operator (a linear functional) and* $\|T_y\| = \|y\|$.

Theorem 5.2 (F. Riesz Lemma). *Let* $T : H \to \mathbb{C}$ *be a bounded, linear functional. Then, there is a unique* y_T *in* H *such that* $T(x) = \langle y_T \mid x \rangle$.

Proof. Let $N = \operatorname{Ker} T$. If $N = H$, set $y_T = 0$. If not, by the Orthogonal Decomposition Theorem there is some $x_0 \in N^\perp \setminus \{0\}$, and we can assume that $\|x_0\| = 1$. We write

$$x = \left(x - \frac{T(x)}{T(x_0)} x_0 \right) + \frac{T(x)}{T(x_0)} x_0 = u + \lambda x_0.$$

The second summand is in $\operatorname{span}(\{x_0\}) \subseteq N^\perp$, with $\lambda = \langle x_0 \mid x \rangle$, while u clearly lies in N (incidentally, this shows that $\operatorname{Ker}(N)$ has codimension one in H).

$$T(x) = \lambda T(x_0) = \langle x_0 \mid x \rangle T(x_0) = \left\langle \overline{T(x_0)} x_0 \mid x \right\rangle,$$

hence, $y_T = \overline{T(x_0)} x_0$. □

EXERCISE 5.8. *Riesz Lemma provides an identification, in fact an isometry,* $C : T \mapsto y_T$ *of the dual space* $H^* = \{T : H \to \mathbb{C}$ *such that* $\|T\| < \infty\}$ *with* H *itself. The map* C *is conjugate linear,*

$$C(aS + bT) = \bar{a}C(S) + \bar{b}C(T).$$

Prove this and prove that, if $A : H \to K$ *is a bounded, linear map between Hilbert spaces, then* $C(TA) = A * C(T)$*, where the adjoint operator* $A^* : K \to H$ *is defined by* $\langle Ax \mid y\rangle = \langle x \mid A^* y\rangle$.

5.2. Orthonormal systems

5.2.1. Orthogonal vectors. A family $S = \{f_a : a \in I\}$ of vectors is an *orthogonal system* if any two vectors in it are orthogonal, and an *orthonormal system* if, in addition, each f_a has unit norm. An *orthonormal basis (o.n.b.)* for H is an orthonormal system which is maximal: no other vector can be added to it without breaking the orthonormality condition.

EXERCISE 5.9. . *Show that the orthonormal system* $\{e_a : a \in I\}$ *is a o.n.b. for* H *is and only if span* $\{e_a : a \in I\}$ *is dense in* H.

THEOREM 5.3. *[Bessel Inequality] Let* $\{e_i : i = 1, \dots, n\}$ *be an orthonormal system in the Hilbert space* H*, and let* $x \in H$*. Then, the vector* $x - \sum_{i=1}^{n} \langle x \mid e_i\rangle e_i$ *is orthogonal to* $\operatorname{span}\{e_1, \dots, e_n\}$*, and*

$$\|x\|^2 \geqslant \sum_{i=1}^{n} |\langle e_i \mid x\rangle|^2 .$$

If $\{e_a : a \in I\}$ *is an o.n.b., the numbers* $\langle e_a \mid x\rangle = \hat{x}(a)$ *are the Fourier coefficients of* x *w.r.t. the basis.*

PROOF. The first assertion is clear:

$$\left\langle x - \sum_{i=1}^{n} \langle x \mid e_i\rangle e_i \mid e_j \right\rangle = \langle x \mid e_j\rangle - \sum_{i=1}^{n} \langle x \mid e_i\rangle \langle e_i \mid e_j\rangle = 0,$$

by orthonormality of the system. As a consequence,

$$\begin{aligned} \|x\|^2 &= \left\| x - \sum_{i=1}^{n} \langle e_i \mid x\rangle e_i \right\|^2 + \left\| \sum_{i=1}^{n} \langle e_i \mid x\rangle e_i \right\|^2 \\ &= \left\| x - \sum_{i=1}^{n} \langle e_i \mid x\rangle e_i \right\|^2 + \sum_{i=1}^{n} |\langle e_i \mid x\rangle|^2 \end{aligned}$$

Before proceeding, we need to clarify some concepts about infinite sums. If $\{c_a : a \in I\}$ is a family of positive numbers, then

$$\sum_{a \in I} c_a := \sup \left\{ \sum_{i=1}^{n} c_{a_i} : \{a_1, \dots, a_n\} \subseteq I \right\}.$$

□

Corollary 5.2. *Let $M = span\{e_1, \dots, e_n\}$. Then,*

$$\pi_M(x) = \sum_{i=1}^{n} \langle x \mid e_i \rangle \, e_i.$$

Proof. The first assertion in Bessel's inequality says that $x = u + v$ with $u = \sum_{i=1}^{n} \langle x \mid e_i \rangle \, e_i \in M$ and $v \in M^{\perp}$, hence, $u = \pi_M(x)$. □

Exercise 5.10. *Show that, if $\sum_{a\in I} c_a < \infty$, then $\{a \in I : c_a \neq 0\}$ is at most countable. If $\{c_a : a \in I\}$ is a family of complex numbers, we say that $\sum_{a\in I} c_a$ converges absolutely if $\sum_{a\in I} |c_a| < \infty$ converges. In this case we say that $\sum_{a\in I} c_a$ converges in $\mathbb{C}$ absolutely, hence irrespective of how I is ordered (prove it if you have never done it before!).*

5.2.2. Spectral analysis and synthesis.

Theorem 5.4. *Let $\{e_a\}_{a\in I}$ be a o.n.b. of H. Then,*

(i) *For all x in H,*

$$x = \sum_{a\in I} \langle e_a | x \rangle e_a \text{ (Spectral Analysis of } x\text{)} \tag{5.2.1}$$

converges in H;

(ii) *we have*

$$\|x\|^2 = \sum_{a\in I} |\langle e_a | x \rangle|^2 \text{ (Plancherel Isometry)}; \tag{5.2.2}$$

(iii) *if $\{c_a\}_{a\in I}$ is a sequence in H such that $\sum_{a\in I} |c_a|^2$ converges, then*

$$\sum_{a\in I} c_a e_a \text{ converges in } H. \tag{5.2.3}$$

Proof. By Bessel inequality, if $A \subseteq I$ with $\sharp(A) < \infty$, then $\sum_{a\in A} |\langle e_a | x \rangle|^2 \leq \|x\|^2$. As a consequence, $I(x) := \{a \in I : \langle e_a | x \rangle\} = \{a_n\}_n$ is countable, hence, (5.2.2) holds with $\leq$. Set now $x_n = \sum_{j=1}^{n} \langle e_{a_j} | x \rangle e_{a_j}$. We show that is defines a Cauchy sequence in H:

$$\|x_{n+l} - x_n\|^2 = \sum_{j=1}^{l} |\langle e_{a_j} | x \rangle|^2 \to 0$$

as $n \to \infty$, by the convergence of the series on the right of (5.2.2). Let $y = \lim_{n\to\infty} x_n$ (the limit is taken in H-norm):

$$\begin{aligned}\langle x - y | e_{a_m}\rangle &= \left\langle x - \lim_{n\to\infty}\sum_{j=1}^{n}\langle e_{a_j}|x\rangle e_{a_j} | e_{a_m}\right\rangle \\ &= \lim_{n\to\infty}\left\langle x - \sum_{j=1}^{n}\langle e_{a_j}|x\rangle e_{a_j} | e_{a_m}\right\rangle \\ &= \lim_{n\to\infty}\left(\langle e_{a_m}|x\rangle - \langle e_{a_m}|x\rangle\right) = 0.\end{aligned}$$

The same argument shows that $\langle x - y|e_a\rangle = 0$ if $a \notin I(x)$. Hence, $x - y \perp e_a$ for all elements of the orthonormal basis, hence $x = y$, which shows (5.2.1).

We finish the proof of Plancherel's identity:

$$\begin{aligned}0 &= \lim_{n\to\infty}\left\| x - \sum_{j=1}^{n}\langle e_{a_j}|x\rangle e_{a_j}\right\|^2 \\ &= \left(\|x\|^2 - \sum_{j=1}^{n}|\langle e_{a_j}|x\rangle|^2\right) \\ &= \|x\|^2 - \sum_{n=1}^{\infty}|\langle e_{a_n}|x\rangle|^2 \\ &= \|x\|^2 - \sum_{a\in I}|\langle e_a|x\rangle|^2,\end{aligned}$$

as wished. □

5.2.3. Orthonormal basis in separable Hilbert spaces.

5.2.3.1. *Gram-Schmidt algorithm.* A Hilbert space is *separable* if it separable as a metric space, i.e. if it has a countable, dense set. Most Hilbert spaces encountered in theory and applications are separable, and in this case the existence of a orthonormal basis is constructive. We will then treat first the separable case, then we will see how things work in the general case, where Zorn's Lemma is required, and we have no way to "see" the basis.

THEOREM 5.5 (Gram-Schmidt). *Let H be a separable Hilbert space. Then, H has a countable basis.*

PROOF. We start with a countable, dense subset $G = \{g_n : n \geq 1\}$ of H. Inductively removing vectors linearly dependent from the previously chosen ones, we obtain a linearly independent subfamily $F = \{f_n : n \geq 1\}$ such that $span(F) = span(G)$ is still dense in H. We then transform F into an orthonormal system:

$$\begin{aligned} e_1 &= \frac{f_1}{\|f_1\|}, \\ e_2 &= \frac{f_2 - \langle e_1|f_2\rangle e_1}{\|f_2 - \langle e_1|f_2\rangle e_1\|}, \\ &\dots \\ e_n &= \frac{f_n - \sum_{j=1}^{n-1}\langle e_j|f_n\rangle e_j}{\|f_n - \sum_{j=1}^{n-1}\langle e_j|f_n\rangle e_j\|}, \\ &\dots \end{aligned}$$

The denominators do not vanish because F is a linearly independent family, and inductively we see that $span\{e_1, \dots, e_n\} = span\{f_1, \dots, f_n\}$. Hence, $\{e_n : n \geq 1\}$ is a orthonormal basis for H.

□

EXERCISE 5.11. *Let $H = L^2[0,1]$ with the Lebesgue measure, and consider the functions $1, x, x^2$. Apply Gram-Schmidt's algorithm to find three orthonormal vectors in H.*

5.2.3.2. *The classification of separable Hilbert spaces.*

THEOREM 5.6. *A Hilbert space H is separable if and only if it has one countable orthonormal basis $\mathcal{B}$. If it does, all o.n.b. of H have the same cardinality. Moreover,*

(i) *if $\sharp(\mathcal{B}) = d < \infty$, then H is isometrically isomorphic to $\mathbb{C}^d$;*

(ii) *if $\sharp(\mathcal{B}) = \infty$, then H is isometrically isomorphic to $\ell^2(\mathbb{N})$.*

PROOF. If H is separable, then Theorem 5.5 provides a countable o.n.b. $\mathcal{B}$. Suppose viceversa that H has a countable basis $\mathcal{B}$. If $\sharp(\mathcal{B}) = d < \infty$, $\mathcal{B} = \{e_1, \dots, e_d\}$, then each x in H can be written as $x = \sum_{j=1}^{d}\langle e_j|x\rangle e_j$ by the Analysis part of Theorem 5.4 (or by a much more elementary argument). The map

$$L_{\mathcal{B}} : x \mapsto (\langle e_j|x\rangle)_{j=1}^{d}$$

is an isometric isomorphism of H onto $\mathbb{C}^d$. Since H is finite dimensional, all its basis have the same dimension. In the countable case, $\mathcal{B} = \{e_n\}_{n=1}^{\infty}$, and we define $L_{\mathcal{B}} : H \to \ell^2(\mathbb{N})$ in the same way, $L_{\mathcal{B}} : x \mapsto \{\langle e_n|x\rangle\}_{n=1}^{\infty}$. Theorem 5.4 implies that $L_{\mathcal{B}}$ is a isometric isomorphism.

In the opposite direction, suppose H has a countable o.n.b., and consider the countable Q family of the vectors $q = \sum_{j=1}^{n} q_j e_j$, where $n \geq 1$, the q_j's are complex rationals, $q_j \in \mathbb{Q} + i\mathbb{Q}$. If $x = \sum_{j=1}^{\infty} c_j e_j$ and $\epsilon > 0$, then

$$\left\|\sum_{j=1}^{\infty} c_j e_j - \sum_{j=1}^{n} q_j e_j\right\|^2 = \sum_{j=n+1}^{\infty}|c_j|^2 + \sum_{j=1}^{n}|c_j - q_j|^2.$$

Choose first n such that the first sum is dominated by ϵ, then $q_1, \dots, q_n$ such that each summand in the second sum is dominated by ϵ/n. This shows separability.

We are left with proving that, if one o.n.b. $\mathcal{B}$ is infinite countable, then any other o.n.b. $\mathcal{B}_1$ is. Let $\mathcal{B}_1 = \{e_a\}_{a\in I}$, and observe that, for any $e_a \neq e_b$ in it, $\|e_a - e_b\|^2 = 2$. Let $\{q_n\}_{n=1}^{\infty}$ be any countable, dense set in H. Then, for each $a \in I$ there is $n = n(a)$ such that $\|e_a - q_n\| < \sqrt{2}/2$, and if $a \neq b$, then $n(a) \neq n(b)$. This provides an injective map $I \to \mathbb{N}$, hence I is countable. □

5.2.4. Supplement: orthonormal basis in general Hilbert spaces.

5.2.4.1. *Existence of o.n.b.* A widely used version of the Axiom of Choice is Zorn's Lemma, which we are stating below. We work with a partially ordered set $(A, \leq)$. A *chain* in A is a subset $B \subseteq A$ on which the partial order is in fact a total order: for any x, y in B, either $x \leq y$ or $y \leq x$. An element m in A is *maximal* for $C \subseteq A$ if $x \leq m$ for all x in C and if $n \leq m$ has the same property, then $n = m$.

THEOREM 5.7 (Zorn Lemma). *Let $(A, \leq)$ be a (nonempty) partially ordered set with the property that any chain B in A has a maximal element in A. Then, A has a maximal element.*

As a consequence, we have that an orthornormal basis exists for any Hilbert space.

THEOREM 5.8. *Every Hilbert space $H \neq \{0\}$ has a orthonormal basis.*

PROOF. Let A be the set of all orthonormal systems of H, ordered by inclusion. It is nonempty because $\{x/\|x\|\} \in A$ if $x \neq 0$ is an element of H. If B is a chain in A, then

$$\mathcal{G} = \cup_{\mathcal{F}\in B}\mathcal{F} \in A$$

is a maximal element for B: it is an orthonormal system, and all orthonormal system containing all $\mathcal{F}$ in B contain $\mathcal{G}$.

By Zorn's Lemma, A has a maximal element $\mathcal{H}$ and, unravelling definitions, $\mathcal{H}$ is a maximal orthonormal system in H, hence a basis. □

5.2.4.2. *The dimension of a Hilbert space.* By Theorem 5.4, if H be a Hilbert space and $\mathcal{B} = \{e_a : a \in I\}$ an orthonormal basis for it, then the map

$$L_{\mathcal{B}} : x = \sum_{a\in I} \langle e_a | x \rangle e_a \mapsto \{\langle e_a | x \rangle\}_{a\in I}$$

is an isometric isomorphism from H onto $\ell^2(I)$.

A natural question is whether different orthonormal basis of H have the same cardinality, as we verified the separable case. The answer is positive.

THEOREM 5.9. *Let $\{e_a : a \in I\}$ and $\{f_b : b \in J\}$ be o.n.b. of the same Hilbert space H. Then, $\sharp(I) = \sharp(J)$.*

The cardinality in question is the *dimension* of the Hilbert space.

COROLLARY 5.3. *All separable Hilbert spaces which are not finite dimensional have countable dimension, and are isomorphic to each other.*

Proof. If both I and J are finite dimensional, the statement is a well know fact from linear algebra (go back to see its proof!).

If I is finite dimensional then J is finite dimensional as well. For, suppose I is finite dimensional and let $S = \{x \in H;\ \|x\| = 1\}$ be the unit sphere in H. Being H a finite dimensional Euclidean space, and S both closed and bounded, S is compact. Suppose now J is infinite and let $\{f_{b_n} :\ n \geq 1\}$ an infinite countable subset of $\{f_b :\ b \in J\}$, which is contained in S if a subsequence of it (which we might denote the same way) converges to some x in S, by continuity of the inner product $\langle x|f_{b_n}\rangle \to \|x\|^2 = 1$ as $n \to \infty$. On the other hand, $x = \sum_{b\in J}\langle x|f_b\rangle f_b$, hence $\langle x|f_{b_n}\rangle \to 0$ as $n \to \infty$, because, for instance, $\sum_n |\langle x|f_{b_n}\rangle|^2 \leq \|x\|^2 = 1$.

The remaining case is that in which both I and J are infinite. To each $a \in I$ there corresponds an at most countable subset $J(a)$ of J such that

$$e_a = \sum_{b\in I(a)} \langle e_a|f_b\rangle f_b.$$

The number of the involved f_b's is

$$\sharp\left(\cup_{a\in I} I(a)\right) \leq \sharp(I).$$

Suppose a basis element f_{b_0} was not used. Then, f_{b_0} is orthogonal to all e_a's, hence to H: contradiction.

We have then that $\sharp(J) \leq \sharp(I)$, and the opposite inequality holds as well. □

5.3. The trigonometric system and Fourier series

In this section we consider the trigonometric system in $L^2((0, 2\pi])$, which is especially natural and important. It should be kept in mind, however, that a great amount of work goes on all over the world into finding, studying the special properties of, and applying, new orthornormal basis (*ONB*) for old and new Hilbert spaces of functions. Broadly speaking, when the functions e_n in an ONB share some common feature, so do their linear combinations; hence, using that ONB to approximate a function f means finding closer and closer approximations of f which "look like" the functions in the basis; to wit, "if all you have in your toolbox is a hammer, then everything you look at seems a hammer".

There are a number of other very good reasons to prefer one orthonormal basis to others. Sometimes they have good algebraic properties, some other times an operator of interest can be neatly expressed (maybe even diagonalized) with respect to that basis.

5.3.1. The trigonometric system. The *torus* is $\mathbb{T} = \{e^{it} :\ t \in \mathbb{R}\} = \partial\mathbb{D}$, where $\mathbb{D} = \{z \in \mathbb{C} :\ |z| < 1\}$ is the unit disc in the complex plane. We identify $e^{it} \in \mathbb{T} \longleftrightarrow t \in (-\pi, \pi]$ and functions on $\mathbb{T}$ with 2π-periodic functions on $\mathbb{R}$. In particular, the continuous function $f(e^{it})$ is identified with a continuous, 2π-periodic function $f(t) = f(t + 2\pi)$ defined on $\mathbb{R}$. The *convolution* of two functions $f, g : \mathbb{T} \to \mathbb{C}$ is here defined as

$$f * g(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t-s)g(s)ds. \tag{5.3.1}$$

Observe that $-2\pi < t-s \leq 2\pi$: this is not a problem, after we have extended f to a 2π-periodic function, $f(t+2\pi) = f(t)$, which can be done in a unique way. On $\mathbb{T}$ we consider normalized arclength measure $\frac{dt}{2\pi}$. For $1 \leq p \leq \infty$, we consider $L^p(\mathbb{T}) = L^p((-\pi, \pi], dt/2\pi)$.

The *trigonometric system* is $\{e_n(t) = e^{int} : \ n \in \mathbb{Z}\}$.

Exercise 5.12. *The functions $\{e_n(t) = e^{int} : \ n \in \mathbb{Z}\}$ form an orthonormal system in $L^2([-\pi, \pi), \frac{dt}{2\pi})$. Observe that this depends on the algebraic relation $e_n(s+t) = e_n(s)e_n(t)$.*

There are two equivalent ways to define the torus: either we choose, as we have done, the interval $(-\pi, \pi]$ and the imaginary exponentials e^{int}, or the interval $(0, 1]$ and the exponentials $e^{2\pi int}$. The latter is more algebraically elegant, and reduces the propagation of π's in the main relations; while the second is easier to use when we think of $\mathbb{T}$ as the boundary of the unit disc in the complex plane. Here we have chosen to sacrifice algebraic elegance to Euclidean geometry.

We define the n^{th} *Fourier coefficient* of $f : \mathbb{T} \to \mathbb{C}$, $f \in L^1(\mathbb{T})$,

$$\widehat{f}(n) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(x)e^{-int}dt, \ \widehat{f} : \mathbb{Z} \to \mathbb{C}. \tag{5.3.2}$$

Theorem 5.10. *The trigonometric system$\{e_n : \ n \in \mathbb{Z}\}$, is complete in $L^2 := L^2((-\pi, \pi], \frac{dt}{2\pi})$.*

The *trigonometric polynomials* $p(t) = \sum_{n=-N}^{N} c_n e^{int}$, that is, form a dense subspace of L^2. Our proof proceeds in three main steps.

5.3.2. The Poisson kernel. For $0 \leq r < 1$ and $t \in \mathbb{T}$, let

$$P_r(t) = \sum_{n=-\infty}^{+\infty} r^{|n|} e^{int}$$

be the *Poisson kernel*, which can also be thought of as a function of $z = re^{it}$, $P_r(e^{it}) = P(z)$. Integrating the series term by term, with our normalization,

$$\widehat{P_r}(n) = r^{|n|}.$$

Here are some basic properties of the Poisson kernel, which are similar to those of the approximations of the identity we have seen earlier.

Lemma 5.4. *The Poisson kernel satisfies:*

(i) $P_r(t) > 0$;

(ii) $\int_{-\pi}^{+\pi} P_r(t)\frac{dt}{2\pi} = 1$;

(iii) *for* $\delta > 0$,
$$\lim_{r\to 1}\sup_{\delta\le |t|\le\pi} P_r(t) = 0.$$

Proof. Integrating the series term by term we have (ii). An alternative expression for $P_r(t)$ can be obtained by computing the series, also proving (ii):

$$\begin{aligned} P_r(t) &= \sum_{n=0}^{\infty} z^n + \sum_{n=1}^{\infty}\overline{z}^n \\ &= \frac{1}{1-z} + \frac{1}{1-\overline{z}} - 1 \\ &= \frac{1-\overline{z}+1-z-(1-\overline{z})(1-z)}{|1-z|^2} \\ &= \frac{1-|z|^2}{|1-z|^2} \\ &= \frac{1-r^2}{1-2r\cos(t)+r^2} > 0. \end{aligned}$$

From the closed expression for P_r we have:

$$\lim_{r\to 1}\sup_{\delta\le |t|\le\pi} P_r(t) = \lim_{r\to 1}\frac{1-r^2}{1-2r\cos(\delta)+r^2} = 0.$$

□

Denote $\tau_s f(e^{it}) = f(e^{i(t-s)})$, translation by s on $\mathbb{R}$ (or, rather, rotation by t radians in $\mathbb{T}$).

Lemma 5.5 (Continuity of translations in L^p). *Let* $1 \le p < \infty$ *and* $f \in L^p(\mathbb{T})$. *Then,*
$$\lim_{s\to 0}\|\tau_s f - f\|_{L^p} = 0.$$

Proof of the lemma. We saw that any function f in L^p can be approximated in L^2 norm by a continuous function φ with any precision ϵ,
$$\|f-\varphi\|_{L^p} \le \epsilon.$$

On the other hand, it is easy to see that for a continuous function φ we have

$$\lim_{s\to 0} \|\tau_s\varphi - \varphi\|_{L^p} = 0.$$

In fact, by uniform continuity, for each $\epsilon > 0$ there is $\delta > 0$ such that for $|s| \le \delta$ we have $|\varphi(x-s) - \varphi(x)| \le \epsilon$. Integrating,

$$\frac{1}{2\pi}\int_{-\pi}^{+\pi} |\varphi(x-s) - \varphi(x)|^p dx \le \epsilon^p.$$

All together,

$$\|\tau_s f - f\|_{L^p} \le \|\tau_s f - \tau_s\varphi\|_{L^p} + \|\tau_s\varphi - \varphi\|_{L^p} + \|\varphi - f\|_{L^p} \tag{5.3.3}$$
$$= \|\tau_s\varphi - \varphi\|_{L^p} + 2\|\varphi - f\|_{L^p} \le 3\epsilon, \tag{5.3.4}$$

provided $|s| \le s(\epsilon)$. □

EXERCISE 5.13. (i) *Show by (simple) examples that Lemma 5.5 fails for $p = \infty$.*

(ii) *Show that Lemma 5.5 holds for L^∞ if we restrict to the class of the continuous functions.*

EXERCISE 5.14. *Prove that, for $1 \le p < \infty$, L^p function can be approximated in L^p norm by step functions, both in $L^p(\mathbb{T})$ and in $L^p(\mathbb{R})$.* **Hint.** *Approximate step functions by continuous functions.*

THEOREM 5.11. *Let $f \in L^1(\mathbb{T})$ and, for $0 \le r < 1$, define its* **Poisson extension**

$$P_r[f](e^{it}) = P[f](re^{it}) := P_r * f(e^{it}), \tag{5.3.5}$$

so that $P_r[f] : \mathbb{T} \to \mathbb{C}$, and $P[f] : \mathbb{D} \to \mathbb{C}$.

(i) *We have:*

$$\lim_{r\to 1} \|P_r(f) - f\|_{L^p} = 0. \tag{5.3.6}$$

(ii) *When $p = 2$,*

$$P_r[f](e^{it}) = \sum_{n=-\infty}^{\infty} \widehat{f}(n) r^{|n|} e^{int} \tag{5.3.7}$$

converges both uniformly and absolutely, and in $L^2(\mathbb{T})$.

PROOF. Observe that

$$\begin{aligned} P_r[f](e^{it}) - f(e^{it}) &= \frac{1}{2\pi}\int_{-\pi}^{+\pi} [f(e^{i(t-s)}) - f(e^{it})]P_r(e^{is})ds \\ &= \frac{1}{2\pi}\int_{-\pi}^{+\pi} [(\tau_s f)(e^{i(t)}) - f(e^{it})]P_r(e^{is})ds. \end{aligned}$$

By Minkowsky's integral inequality,

$$\begin{aligned} \|P_r[f] - f\|_{L^p} &\leq \frac{1}{2\pi}\int_{-\pi}^{+\pi} \|\tau_s f - f\|_{L^p} P_r(e^{is})ds \\ &= \frac{1}{2\pi}\int_{|s|\leq\delta} \|\tau_s f - f\|_{L^p} P_r(e^{is})ds \\ &+ \frac{1}{2\pi}\sup_{|s|>\delta} P_r(e^{is})\int_{-\pi}^{+\pi} \|\tau_s f - f\|_{L^p} ds \\ &\leq \frac{1}{2\pi}\int_{|s|\leq\delta} \|\tau_s f - f\|_{L^p} P_r(e^{is})ds \\ &+ \frac{1}{2\pi}\sup_{|s|>\delta} P_r(e^{is})\int_{-\pi}^{+\pi} (\|\tau_s f\|_{L^p}\|f\|_{L^p})\, ds \\ &= \frac{1}{2\pi}\int_{|s|\leq\delta} \|\tau_s f - f\|_{L^p} P_r(e^{is})ds \\ &+ 2\sup_{|s|>\delta} P_r(e^{is})\|f\|_{L^p}, \end{aligned}$$

because the L^p norm is translation invariant. Fix $\epsilon > 0$ and find first, by Lemma5.5, $\delta > 0$ such that $\|\tau_s f - f\|_{L^2} \leq \epsilon$ if $|s| \leq \delta$; then find $0 < r(\delta, \epsilon) < 1$ such that $P_r(e^{is}) \leq \epsilon$ if $r(\delta, \epsilon) < r < 1$. Thus,

$$\|P_r[f] - f\|_{L^p} \leq \epsilon(1 + 2\|f\|_{L^p}).$$

For the second statement, we apply Fubini's Thorem:

$$\begin{aligned} P_r[f](e^{it}) &= \frac{1}{2\pi}\int_{-\pi}^{+\pi} \sum_{n=-\infty}^{+\infty} r^{|n|}e^{in(x-y)}f(y)dy \\ &= \sum_{n=-\infty}^{+\infty} r^{|n|}\frac{1}{2\pi}\int_{-\pi}^{+\pi} e^{-iny)}f(y)dye^{-inx} \\ &= \sum_{n=-\infty}^{\infty} \widehat{f}(n)r^{|n|}e^{int}. \end{aligned}$$

Convergence is uniform since $|\widehat{f}(n)| = |\langle e_n | f\rangle| \leq \|f\|_{L^2}$. Also, by orthonormality of the imaginary exponentials and Bessel's inequality,

$$\left\|\sum_{N<n\leq N+L} \widehat{f}(n)r^{|n|}e_n\right\|_{L^2}^2 = \sum_{N<n\leq N+L} |\widehat{f}(n)|^2 r^{2|n|} \to 0 \text{ as } N \to \infty,$$

showing L^2 convergence. □

5.3.3. The trigonometric system and basic properties of Fourier series.

THEOREM 5.12. *The trigonometric system is an orthonormal basis for $L^2(\mathbb{T})$.*

PROOF. We have to show that trigonometric polynomials, those having the form

$$p(t) = \sum_{n=-N}^{N} a_n e^{int},$$

are dense in $L^2(\mathbb{T})$. Let $f \in L^2(\mathbb{T})$ and fix $\epsilon > 0$. By Theorem 5.11, there is $0 \leq r < 1$ such that

$$\|P_r[f] - f\|_{L^2}^2 \leq \epsilon.$$

On the other hand, by theorem 5.11 and the orthonormality of the trigonometric system, there exists $N(\epsilon)$ such that

$$\left\| P_r[f] - \sum_{n=-N}^{N} \widehat{f}(n) r^{|n|} e_n \right\|_{L^2}^2 = \sum_{|n|>N} \left| \widehat{f}(n) \right|^2 r^{2|n|} \leq \sum_{|n|>N} \left| \widehat{f}(n) \right|^2 \leq \epsilon.$$

The trigonometric polynomial $p = \sum_{n=-N}^{N} \widehat{f}(n) r^{|n|} e_n$ has the desired properties. □

To each function $f \in L^1()$ we associate its *Fourier series*

$$S_\infty(f)(e^{it}) = \sum_{n=-\infty}^{+\infty} \widehat{f}(n) e^{int}. \tag{5.3.8}$$

The foundational problem of Fourier theory is understanding for which f, and in which sense, the series in (5.3.8) converges. For $f \in L^2(\mathbb{T})$, a quantitative answer is provided by the fact that $\{e_n\}_{n=-\infty}^{+\infty}$ is an orthonormal basis, and the general theory of Hilbert spaces.

THEOREM 5.13. *Let $f \in L^2(\mathbb{T})$.*

(i) *The series $S_\infty(f)$ converges to f in $L^2(\mathbb{T})$. i.e. $S_\infty(f) = f$ as an L^2 function,*

(5.3.9)

$$0 = \lim_{N\to\infty} \left\| f - \sum_{n=-N}^{+N} \widehat{f}(n) e_n \right\|_{L^2}^2 = \lim_{N\to\infty} \frac{1}{2\pi} \int_{-\pi}^{+\pi} \left| f(t) - \sum_{n=-N}^{+N} \widehat{f}(n) e^{int} \right|^2 dt.$$

We write

$$f(t) = \sum_{n=-N}^{+N} \widehat{f}(n) e^{int} \text{ in } L^2(\mathbb{T}). \tag{5.3.10}$$

(ii) *We have the* **Plancherel identity**,

$$\sum_{n=-\infty}^{+\infty} \left|\widehat{f}(n)\right|^2 = \frac{1}{2\pi}\int_{-\pi}^{+\pi} |f(t)|^2 \, dt. \tag{5.3.11}$$

PROOF. Equation (5.3.10), to be read as (5.3.9), follows from the fact that $\{e_n\}$ is an orthonormal basis, and the same is true for (5.3.11). It can also be expressed by saying that

$$f \mapsto \{\widehat{f}(n)\}_{n=-\infty}^{+\infty}$$

is an isometry of $L^2(\mathbb{T})$ onto $\ell^2(\mathbb{Z})$. □

The most peculiar and important feature of the trigonometric system, and of Fourier series, is the nice behavior with respect to convolutions.

THEOREM 5.14. *Let $f, g \in L^1(\mathbb{R})$. Then,*

$$\widehat{f * g}(n) = \widehat{f}(n)\widehat{g}(n). \tag{5.3.12}$$

PROOF. We start with a formal calculation.

$$\begin{aligned}
\widehat{f * g}(n) &= \frac{1}{2\pi}\int_{-\pi}^{+\pi} (f * g)(t) e^{-int} dt \\
&= \frac{1}{2\pi}\int_{-\pi}^{+\pi} \frac{1}{2\pi}\int_{-\pi}^{+\pi} f(t-s)g(s) ds e^{-in(t-s+s)} dt \\
&= \frac{1}{2\pi}\int_{-\pi}^{+\pi} \frac{1}{2\pi}\int_{-\pi}^{+\pi} f(t-s)e^{-in(t-s)} dt g(s) e^{-ins} ds \\
&= \frac{1}{2\pi}\int_{-\pi}^{+\pi} \frac{1}{2\pi}\int_{-\pi}^{+\pi} f(u) e^{-inu} du g(s) e^{-ins} ds \\
&= \widehat{f}(n) \frac{1}{2\pi}\int_{-\pi}^{+\pi} g(s) e^{-ins} ds \\
&= \widehat{f}(n)\widehat{g}(n).
\end{aligned}$$

The first two equalities are definitions, the third an application of Fubini's Theorem, the fourth depends on the change of variables $u = u(t) = t - s$ (with s fixed), we use again the definition of Fourier coefficient. □

Observe that the proof basically depends on the the relation $e^{in(u+s)} = e^{inu}e^{ins}$, and on the fact that the Lebesgue measure is translation invariant, $d(u+s) = du$ (for fixed s).

THEOREM 5.15. (i) *If $f \in L^1(\mathbb{T})$ and $a \in \mathbb{R}$, then*

$$\widehat{\tau_a f}(n) = e^{-ina}\widehat{f}(n).$$

(ii) *If $f \in C^1(\mathbb{T})$ (i.e. f is 2π-periodic and C^1), then*

$$\widehat{f'}(n) = in\widehat{f}(n).$$

PROOF. About (i),

$$\widehat{\tau_a f}(n) = \frac{1}{2\pi}\int_{-\pi}^{+\pi} f(t-a)e^{-int}dt = e^{-ina}\frac{1}{2\pi}\int_{-\pi-a}^{+\pi-a} f(x)e^{-inx}dx = e^{-ina}\widehat{f}(n).$$

About (ii),

$$\begin{aligned}\widehat{f'}(n) &= \frac{1}{2\pi}\int_{-\pi}^{+\pi} f'(t)e^{-int}dt \\ &= \left[f(t)e^{-int}\right]_{-\pi}^{+\pi} - \frac{1}{2\pi}\int_{-\pi}^{+\pi} f(t)e^{-int}(-in)dt \\ &= in\widehat{f}(n).\end{aligned}$$

□

CHAPTER 6

Banach spaces

In this chapter we introduce some basic tools in Banach theory, *Hahn-Banach extension theorem* and *Baire's category theorem and it consequences.* The latter are, on the one hand, *(Banach-Steinhaus) uniform boundededness principle*, and on the other the sequence: *open mapping theorem*, *inverse mapping theorem*, *closed graph theorem*. We will also consider some consequences and applications of these cornerstones of functional analysis.

6.1. Zorn's lemma and some of its consequences

In the proof of Hahn-Banach Theorem we will use Zorn's lemma, which is a consequence of the axiom of choice, and it is in fact equivalent to it. In most courses Zorn's lemma is just stated as a principle, then used alongside the axiom of choice. One reason is that most instructors hope that the relationship between the two statements has been clarified in some other course. The other, probably, is that most instructors are research mathematicians with little taste for nonconstructive principles: they know them and use them, but feel that such principles are a sort of magic of last resort, and having a magic tool generating others, or having a supply of independent ones, does not really make to them much of a difference. A problem with highly nonconstructive existence theorems, in fact, is that we do not know much of the object we have proved the existence of. It is surprising, in view of this fact, that such theorems can be used to obtain more practical statements.

There is a price tag attached to it. Nonconstructive arguments allow us to prove the existence of counterexamples to very reasonable guesses. If we allow the logical basis of the nonconstructive arguments, we have to keep in mind that such monsters exist and have to be taken into account.

So much for the small talk. Before we move to Banach theory, in order to satisfy readers with a taste for logic we state Zorn's lemma and prove that it follows from the axiom of choice (but not the opposite implication).

Theorem 6.1 (Zorn's lemma). *Let $(X, \leq)$ be a partial order on a set X, and suppose that each totally ordered subset C of X has an upper bound: there is u in X such that $u \geq c$ for all $c \in C$. Then, X has a maximal element w: $w \geq x$ for all X in X.*

PROOF. While reading the proof, it is very helpful drawing pictures. A partial order is a special instance of a directed graph, and a totally ordered subset of it might be thought of, in pictorial terms, as a subset of the real line (although it is clearly false that all totally ordered sets arise this way, for instance because there is no a priori restriction on their cardinality).

Suppose by contradiction that X has no maximal element. In particular, by the axiom of choice, for each totally ordered subset C, including $C = \emptyset$, we can choose an upper bound $u(C)$ in $X \setminus C$ (if the only upper bound, which exists by hypothesis, belonged to C, then it would be maximal for X). Now, to each subset F of X and $f \in F$ we associate its *tail set* $F_{<f} <= \{x \in F : x < f\}$.

A subset A is a u-set if

(i) A is totally ordered;

(ii) A contains no infinite descending sequence $a_1 > a_2 > \dots$ (e.g. A can have the order type[1] of $\mathbb{N}$, but not that of $-\mathbb{N}$), hence, any subset A' of A has minumum;

(iii) if $a \in A$, then $u(A_{<a}) = a$.

The last property says that u-sets are linked in a very special way to the choice function u, and, as we will see below, they share many properties with the ordinals.

If A is a u-set and $a \in A$, then $A_{<a}$ has properties (i-iii), hence, it is a u-set.

If A and B are u-sets, by properties (i) and (ii) they have minimum, and $\{\min(A)\}, \{\min(B)\}$ are u-sets. By property (iii),

$$\min(A) = u(\{\min(A)\}_{<\min(A)}) = u(\emptyset) = \cdots = \min(B).$$

The proof of Zorn's lemma mostly consists in a "transfinite iteration" of this argument, to show that all u-sets are tail sets of a largest u-set E, and by further applying u to it we reach a contradiction.

Claim Let $A \neq B$ be u-sets. Then, $A = B_{<b}$ for some $b \in B$, or $B = A_{<a}$ for some $a \in A$.

Let

$$C = \{c \in A \cap B : \; A_c = B_c\}.$$

Sub-claim $C = A$ or $C = A_{<a}$ for some $a \in A$.

[1]By *order type* of a partially ordered set $(Z, <)$ we mean the equivalence class of all sets which are in a order-preserving bijection with Z.

We show the sub-claim first. Suppose $C \neq A$ and let $a \in A \setminus C$ be minimal, so that $A_{<a} \subseteq C$. In the opposite direction, if $c \in C \setminus A_{<a} \subseteq A$, then $c > a$. Hence,

$$a \in A_c = B_{<c} \subseteq B,$$

thus $a \in B$ and $A_{<a} = B_{<a}$, implying that $a \in C$, which is a contradiction. This implies that $C = A_<$.

We now prove the claim. If $A \neq C \neq B$, by the sub-claim we have that $A_{<a} = C = B_{<b}$ for some $a \in A$, $b \in B$. By property (iii) of the u-sets, $a = u(A_{<a}) = u(B_{<b}) = b$, and we have that

$$a = b \in A \cap B, \text{ and } A_{<a} = B_{<b}.$$

By definition, then, $a \in C$, but this contradicts the fact that $A_{<a} = C$. The assumption $A \neq C \neq B$ can not hold, hence we have that, again by the sub-claim, either $A = C = B_{<b}$ or, the other way around, $B = C = A_{<a}$. (Hidden in the proof is that the cases $a = \min(A)$ or $b = \min(B)$, in which $C = \emptyset$, are covered because we defined $u(\emptyset)$).

The claim allows us to glue together whatever u-sets we have into one. Let

$$E = \cup_{A \text{ a } u\text{-set}} A.$$

We first show that, if A is a u-set and $a \in E$, then $A_{<a} = E_{<a}$. The direction $A_{<a} \subseteq E_{<a}$ follows from $A \subseteq E$. In the other direction, let $x \in E_{<a}$ and let $B \ni x$ be a u-set. If $B \subseteq A$, then $x \in A$. If not, by the claim $A = B_{<b}$ for some $b \in B$, in which case $x < a < b$, and so $x \in B_{<b} = A$. In both cases, $x \in A$, hence, $E_{<a} \subseteq A$, which clearly implies $E_{<a} \subseteq A_{<a}$.

It is now easy to verify that E is a u-set.

(i) Let $a, b \in E$, and $A \ni a$, $B \ni b$ be u-sets. Then $A_{<a} = E_{<a}$ and $B_{<b} = E_{<b}$ are u-sets, hence, by the claim above, either they are equal, then $a = u(A_{<a}) = u(B_{<b}) = b$, or $A_{<a} = B_{<c}$ for some c in $B_{<b}$, and $c = u(B_{<c}) = u(A_{<a}) = a$, or the other way around. In the second case, $a = c < b$, and in the remaining case the opposite inequality holds. Hence, E is totally ordered.

(ii) Suppose $a_1 > a > 2 > \dots$ is a descending sequence in E, and that $a_1 \in A$, a u-set. For $i > 1$, $a_i \in E_{a_1} = A_{a_1}$, hence the sequence, being contained in A, can not be infinite.

(iii) If $a \in E$ and $A \ni a$ is a u-set, then $u(E_{<a}) = u(A_{<a}) = a$.

Since E is a u-set, $E \cup \{u(E)\}$ is still a u-set which is not contained in E, and we have reached a contradiction. □

In the spirit of the section, we use Zorn's lemma to prove a purely algebraic result.

THEOREM 6.2. *[Hamel's basis of a vector space] Any vector space X over a filed $\mathbb{F}$ has a basis. That is, there is a set $\{v_a\}_{a\in I}$ of linearly independent vectors of X such that $X = span(\{v_a\}_{a\in I})$.*

PROOF. Consider the set $\mathcal{V}$ of all families $V = \{v_b\}_{a\in J}$ of linearly independent vectors of X, partially ordered by inclusion. If $\{V_\alpha\}_{\alpha\in H}$ is a totally ordered subfamily in $\mathcal{V}$, then $\bigcup_{\alpha\in H} V_\alpha$ is an element of $\mathcal{V}$, and it is an upper bound for $\{V_\alpha\}_{\alpha\in H}$. By Zorn's lemma, there exists in $\mathcal{V}$ a maximal element $\{v_a\}_{a\in I}$, which is a basis for X, since a vector u which is not in $\text{span}(\{v_a\}_{a\in I})$ could be added to the family, contradicting maximality. □

Below we use basic facts about cardinal numbers. For properties of cardinal numbers, and of ordinal ones, you might look into Reed Solomon, Notes on ordinals and cardinals.

THEOREM 6.3. *Let X be a vector space over a field $\mathbb{F}$, and let $U = \{U_a\}_{a\in I}$ and $V = \{v_b\}_{a\in J}$ be two basis of X. Then, $\sharp(I) = \sharp(J)$, I and J have the same cardinality.*

PROOF. We know from linear algebra that $\sharp(I)$ is finite if and only if $\sharp(J)$ is, and they coincide. We can suppose then that both index sets are infinite.

To each $a \in I$ associate the set $B(a) := \{b_1(a), \dots, b_{n(a)}(a)\} = \{b_1, \dots, b_n\} \subset J$, where

$$U_a = \sum_{i=1}^{n} \lambda_i V_{b_i},$$

with $0 \neq \lambda_i \in \mathbb{F}$, is the unique expression of U_a with respect to the basis V. The map B is one to one from I to the family of the finite subsets of J. Since the latter class has the same cardinality as J, we have $\sharp(I) \leq \sharp(J)$, and the other inequality is proved in the same way. □

By theorem 6.3, we can attach to any vector space X the cardinality of one, hence all, of its basis. Such number is called the *algebraic dimension*, or sometimes the *Hamel dimension*, of X. Let's denote it by $\dim_{al}(X)$. The algebraic dimension is clearly a linear invariant: if $L : X \to Y$ is a linear bijection, then $\dim_{al}(X) = \dim_{al}(Y)$.

Infinite dimensional, but separable, Hilbert spaces have a countable orthonormal basis in the Hilbert sense, but their algebraic dimension is more than countable. This shows that Hilbert theory, and more general Banach theory, or even more generally topological vector spaces theory, is not reducible to linear algebra alone, even at the most basic level of dimension.

Proposition 6.1. *As a vector space, any infinite dimensional Hilbert space H does not have countable dimension.*

Proof. Let $\{v_n\}_{n=1}^{\infty}$ be any countable, linearly independent subset of H. Using Gram-Schmidt algorithm, find an orthonormal set $\{e_m\}_{m=1}^{\infty}$ such that $\mathrm{span}(\{v_n\}_{n=1}^{\infty}) = \mathrm{span}(\{e_m\}_{m=1}^{\infty})$. Consider then

$$x = \sum_{m=1}^{\infty} \frac{1}{m} e_m.$$

Then, $x \in H \setminus \mathrm{span}(\{e_m\}_{m=1}^{\infty}) = H \setminus \mathrm{span}(\{v_n\}_{n=1}^{\infty})$. This show that $\{v_n\}_{n=1}^{\infty}$ was not an algebraic basis for H. $\square$

The completeness of H was used in that x exists in H since the series converges.

Using Hamel basis, we can construct such monstrous objects as everywhere defined, unbounded linear functionals on a Hilbert space.

Proposition 6.2. *Let H be an infinite dimensional Hilbert space. Then, there exists a linear functional $l : H \to \mathbb{R}$ which is not bounded.*

Proof. Let $\{v_\alpha\}_{\alpha \in I}$ be a Hamel basis for H, let $\{v_{\alpha_n}\}_{n=1}^{\infty}$ a countable subset of it, which we might assume to be orthonormal after applying the Gram-Schmidt algorithm to it. We then define l on the basis by

$$\begin{cases} l(v_\alpha) = 0 \text{ if } \alpha \neq \alpha_n \text{ for all } n \geq 1 \\ l(v_{\alpha_n}) = n. \end{cases}$$

The operator extends to a linear operator l on H by linear algebra. On the other hand, l is not bounded,

$$\|l\|_{H^*} := \sup_{\|x\|=1} |l(x)| \geq |l(v_{\alpha_n})| = n,$$

hence, $\|l\|_{H^*} = \infty$. $\square$

6.2. The Hahn-Banach Theorem and some of its consequences

Exercise 6.1. *Let H be a Hilbert space, $M \subset H$ a closed, linear space, and $S : M \to \mathbb{C}$ a (bounded) linear functional on M. Show that there exists an extension T of S to H with $\|T\| = \|S\|$. Moreover, such an extension is unique.*

In general Banach spaces matters are more intricate.

Theorem 6.4 (Real Hahn-Banach Theorem). *Let X be a real vector space, and $p : X \to \mathbb{R}$ a real-valued, convex functional:*

$$p(ax + (1-a)y) \leq ap(x) + (1-a)p(y)$$

if $x, y \in X$ and $a \in [0,1]$. Let $l : Y \to \mathbb{R}$ be a linear functional such that for $x \in Y$

$$l(y) \leq p(y).$$

Then, there exists linear $L : X \to \mathbb{R}$ extending l to X, and such that

$$L(x) \leq p(x)$$

on X.

Proof. The main step consist in extending l on Y to λ on $span(Y, z)$, where $z \notin Y$, which we do by a sort of separation of variables. Then we use Zorn's Lemma. We must have $\lambda(y + az) = l(y) + a\lambda(z)$, and it suffices to determine $\lambda(z)$. We start with an inequality involving z and p. Let $a, b > 0$, $x, y \in Y$:

$$\begin{aligned} bl(x) + al(y) &= (b+a)l\left(\frac{b}{b+a}x + \frac{a}{b+a}y\right) \\ &\leq (b+a)p\left(\frac{b}{b+a}x + \frac{a}{b+a}y\right) \\ &= (b+a)p\left(\frac{b}{b+a}(x-az) + \frac{a}{b+a}(y+bz)\right) \\ &\leq bp(x-az) + ap(y+bz). \end{aligned}$$

That is,

$$\frac{l(x) - p(x-az)}{a} \leq \frac{p(y+bz) - l(y)}{b}$$

holds for all a, b, x, y as above. There is then real k such that $l(x) - p(x-az) \leq ka$ and $kb \leq p(y+bz) - l(y)$, or

$$l(x) - ka \leq p(x-az), \ l(y) + kb \leq p(y+bz).$$

It follows that $\lambda(z) = k$ works.

Consider now the set A having as element (Z, λ) where Z is a subspace of X containing Y and λ is an extension of l to Z in such a way $\lambda(z) \leq p(z)$ on Z. We partially order A saying that $(Z, \lambda) \leq (W, \mu)$ is Z is a subspace of W and μ extends λ.

Given a chain $\{(Z_c, \lambda_c) : \ c \in C\}$ in A, we have that $Z = \cup_{c \in C} Z_c$ and λ such that $\lambda|_{Z_c} = \lambda_c$ are a well defined element in A, and (Z, λ) is maximal for C. By Zorn's Lemma, there is a maximal (W, ν) in A. Now, if $W \neq X$, we might apply the one-dimensional extension, and (W, ν) would not be maximal. Hence, $W = X$ and the theorem is proved. □

When proving Hahn-Banach's theorem for a separable Banach space X with $p(x) = \|x\|$, as it is often the case, Zorn's lemma is not necessary. In fact, we can proceed in a way which is much similar to the Gram-Schmidt construction of a orthonormal basis for a Hilbert space.

EXERCISE 6.2. *Let $(X, \|\cdot\|)$ be a separable Banach space and let $\{x_n\}_{n=1}^{\infty}$ be a dense set. Let Z be a subspace of X and let $l : X \to \mathbb{R}$ be a linear functional such that $|l(x)| \leq K\|x\|$ for some constant $K \geq 0$.*

(i) *Let $Z_0 = \bar{Z}$ be the closure of Z in X. Show that l has a unique extension l_0 to Z_0 with $|l_1(x)| \leq K\|x\|$.*

(ii) *Let x_{i_1} be the first x_n in the sequence such that $x_n \notin Z_0$, and let $Z_1 = span(Z_0, x_{i_1})$. Verify that the inductive step in the proof of the real Hahn-Banach theorem shows that l_0 extends to a linear functional l_1 on Z_1 such that $|l_1(x)| \leq K\|x\|$.*

(iii) *Write down how to iterate the procedure in (ii). What happens if it stops afetr finitely man steps?*

(iv) *Suppose the procedure can be iterated indefinitely, exhausting $\{x_n\}$, and let $Z_\infty = \bigcup_{m=0}^{\infty} Z_m$, and $l_\infty : Z_\infty \to \mathbb{R}$ be such that $l_\infty|_{Z_m} = l_m$. Show that Z_∞ is dense in X and $|l_\infty(x)| \leq K\|x\|$. Use then again (i) to show that l_∞ has a unique extension λ to X which coincides with l_∞ on Z_∞ (hence, it coincides with l on Z), and $|\lambda(x)| \leq K\|x\|$.*

THEOREM 6.5 (Complex Hahn-Banach Theorem). *Let X be a complex vector space and $p : X \to \mathbb{R}$ be such that*

$$p(ax + by) \leq |a|p(x) + |b|p(y)$$

whenever $a, b \in \mathbb{C}$, $|a| + |b| = 1$, and $x, y \in X$. Let $l : Y \to \mathbb{C}$ be a linear functional defined on a subspace Y on X, satisfying $|l(y)| \leq p(y)$. Then, there exists a linear functional $L : X \to \mathbb{C}$ which extends l, and such that $|L(x)| \leq p(x)$ on X.

PROOF. We consider X as a real space of "twice the dimension". Let $\lambda(x) = \mathrm{Re}(l(x))$, a real functional on Y, and observe that

$$\lambda(ix) = \mathrm{Re}(l(ix)) = \mathrm{Re}(il(x)) = -\mathrm{Im}(l(x)),$$

so that $l(x) = \lambda(x) - i\lambda(ix)$ reconstructs l from λ. By the real Hahn-Banach theorem, there exists a real functional $\Lambda : X \to \mathbb{R}$ which extends λ to X, and such that $\Lambda(x) \leq p(x)$. We define $L(x) = \Lambda(x) - i\Lambda(ix)$, which defines a complex linear functional on X:

$$L(ix) = \Lambda(ix) - i\Lambda(-x) = i(\Lambda(x) - i\Lambda(ix)) = iL(x).$$

We also have, for some real t,

$$|L(x)| = e^{it}L(x) = L(e^{it}x) = \Lambda(e^{it}x) \leq p(e^{it}x) = p(x),$$

where the third equality holds because $L(e^{it}x)$ is real, and the last follows from the fact that $p(e^{it}x) = p(x)$ for all t's. □

6.3. The dual of a Banach space

Let X be a Banach space over $\mathbb{C}$ (virtually all we say holds for real Banach spaces). Its *dual space* X^* is the space of the continuous linear functionals on X. It is a Banach space in itself. In fact more is true. Let $\mathcal{B}(X, Y)$ be the linear space of the bounded, linear operators $T : X \to Y$, normed with the operator norm,

EXERCISE 6.3. *Show that the operator norm is in fact a norm on* $\mathcal{B}(X, Y)$.

THEOREM 6.6. *Let* X, Y *be Banach spaces.Then,* $\mathcal{B}(X, Y)$ *is complete.*

PROOF. Let $\{A_n : n \geq 1\}$ be Cauchy in $\mathcal{B}(X, Y)$, so that $\{A_n x : n \geq 1\}$ in Cauchy in Y for each fixed x in X, and $Ax = \lim_{n\to\infty} A_n x$ is well defined, and linear. Also, $|\|A_m\| - \|A_n\|| \leq \|A_m - A_n\|$, hence $\|A_n\| \to C$ has real limit as $n \to \infty$. We have:

$$\|Ax\| = \lim_{n\to\infty} \|A_n x\| \leq \lim_{n\to\infty} \|A_n\|\|x\| \leq C\|x\|,$$

hence, $\|A\| \leq C$.

We still have to show that $A_n \to A$ in $\mathcal{B}(X, Y)$ norm. But we have:

$$\|(A - A_m)x\| = \lim_{n\to\infty} \|(A_n - A_m)x\| \leq \lim_{n\to\infty} \|A_n - A_m\| \cdot \|x\|,$$

hence,

$$\|A - A_m\| \leq \lim_{n\to\infty} \|A_n - A_m\| \leq \epsilon$$

if $m \geq m(\epsilon)$, so $\|A - A_m\| \to 0$ as $m \to \infty$. □

EXERCISE 6.4. *Let* $S \in \mathcal{B}(X, Y)$ *and* $T \in \mathcal{B}(Y, Z)$. *Show that* $TS := T \circ S \in \mathcal{B}(X, Z)$ *satisfies* $\|TS\| \leq \|T\|\|S\|$.

The above inequality, when $X = Y = Z$, expresses the fact that $\mathcal{B}(X) := \mathcal{B}(X, X)$ is a *Banach algebra* with respect to the composition product. It is also a *unital* one, since the identity operator $Ix = x$ satisfies $IA = AI$ and $\|I\| = 1$.

We will often use without mention the following fact.

EXERCISE 6.5. *Let* $f_0 : X_0 \to Y$ *be a continuous map from a dense subspace* X_0 *of a metric space* X *and a complete metric space* Y. *Then,* f_0 *can be uniquely extended to a continuous map* $f : X \to Y$. *If* X *is a normed space,* Y *is a Banach space, and* f_0 *is linear, then* f *is linear and* $\|f\|_{\mathcal{B}(X,Y)} = \|f_0\|_{\mathcal{B}(X_0,Y)}$.

The Hahn-Banach Theorem has important consequences about X^*.

COROLLARY 6.1. *Let $i : Y \to X$ be the inclusion in the Banach space X of Y, a subspace of it. Then, the restriction map $l \mapsto l \circ i$ from X^* to Y^* is surjective: each λ in Y^* has an extension l in X^*. Moreover, there exists one extension l such that $\|l\|_{X^*} = \|\lambda\|_{Y^*}$.*

PROOF. Apply the Hanh-Banach Theorem with $p(x) = \|\lambda\|_{Y^*} \cdot \|x\|$. □

COROLLARY 6.2. *Let $y \in X$, Banach. Then, there exists $l \in X^*$, $l \neq 0$, such that $l(y) = \|l\|_{X^*}\|y\|$.*

PROOF. Set $\lambda(ay) = a\|y\|$ on $span(y)$, and extend it to $l : X \to \mathbb{C}$ by Hahn-Banach. Since $|\lambda(ay)| = \|ay\|$, we have $\|l\| = \|\lambda\| = 1$, and $l(y) = \lambda(y) = \|y\|$. □

COROLLARY 6.3. *Let Z be a subspace of a normed linear space X, and let $d(y, Z) := \inf\{\|y - z\| : \ z \in Z\}$. Then, there is $L \in X^*$ such that $L(y) = d(y, Z)$, $\|L\|_{X^*} = 1$, and $L(z) = 0$ for $z \in Z$.*

PROOF. We only have to define such L on $span(Z, y)$, then use Hahn-Banach. We are forced to define $L(z + ay) = aL(y) = ad(y, Z)$. This is a linear functional, and

$$\begin{aligned} \frac{|L(z+ay)|}{\|z+ay\|} &= |a|\frac{d(Z,y)}{\|z+ay\|} \\ &= |a|\frac{\inf\{\|w-y\| : \ w \in Z\}}{\|z+ay\|} \\ &= |a|\frac{\inf\{\|w-y\| : \ w \in Z\}}{|a| \cdot \| - \frac{z}{a} - y\|} \\ &\leq 1, \end{aligned}$$

because $-z/a \in Z$. Hence, $\|L\|_{X^*} \leq 1$.

To show the opposite inequality, for $\epsilon > 0$ find z in Z such that $\|y - z\| \leq d(y, Z) + \epsilon$, so that

$$\frac{|L(z-y)|}{\|z-y\|} \geq \frac{d(y,Z)}{d(y,Z)+\epsilon}.$$

Since $\epsilon > 0$ is arbitrary, $\|L\|_{X^*} \geq 1$. □

EXERCISE 6.6. *Give "Hilbertian" elementary proofs of the three corollaries above in the case when X is a Hilbert space.*

The *bidual* of a Banach space X, X^{**}, is the dual of X^*.

THEOREM 6.7. *Let X be a Banach space and consider the map $i : X \to X^{**}$ given by $[i(x)](\ell) := \ell(x)$ whenever $x \in X$ and $\ell \in X^*$. Then, i is an isometric imbedding of X into X^{**}.*

PROOF. We have:

$$|[i(x)](\ell)| = |\ell(x)| \le \|\ell\|_{X^*}\|x\|_X,$$

hence, $\|[i(x)]\|_{X^{**}} \le \|x\|_X$.

In the other direction, by Corollary 6.2 we can find l in X^* with $\|l\|_{X^*} = 1$, and such that $l(x) = \|x\|$. i.e.,

$$\|i(x)\|_{X^{**}} \ge \frac{|[i(x)](l)|}{\|l\|_{X^*}} = \frac{|l(x)|}{\|l\|_{X^*}} = \|x\|.$$

□

When the map $i : X \to X^{**}$ is surjective (hence, a surjective isometry), we say that X is *reflexive*. Not all Banach spaces are reflexive, as the following exercise shows.

EXERCISE 6.7. *Here $\ell^p = \ell^p(\mathbb{N})$.*

(i) *Show that for each $\varphi \in \ell^\infty$ the map $L_\varphi : \psi \mapsto \sum_{n=0}^\infty \varphi(n)\overline{\psi(n)}$, $L_\varphi : \ell^1 \to \mathbb{C}$, is a continuous linear functional in $(\ell^1)^*$ and $\|L_\varphi\|_{(\ell^1)^*} = \|\varphi\|_\ell^\infty$.*

(ii) *Given $L \in (\ell^1)^*$, show that there exists $\varphi \in \ell^1$ s.t. $L = L_\varphi$. Moreover, $\|\varphi\|_{\ell^\infty} = \|L\|_{(\ell^1)^*}$.* **Hint.** *Make use of $L(\delta_n)$ as "building blocks" to construct φ. This way, we have isometrically identified $(\ell^1)^* \equiv \ell^\infty$.*

(iii) *Consider the subspace C of ℓ^∞,*

$$C = \{\varphi \in \ell^\infty : \ \exists \lim_{n\to\infty} \varphi(n)\}.$$

Show that is is a closed subspace ℓ^∞, and that $L : \varphi \mapsto \lim_{n\to\infty} \varphi(n)$ is a closed linear functional on C.

(iv) *Show that there is no $\psi \in \ell^1$ such that $C(\varphi) = \sum_{n=0}^\infty \psi(n)\varphi(n)$. Deduce from this that ℓ^1 is not reflexive.*

EXERCISE 6.8. *Let ℓ_c be the space of the sequences $h : \mathbb{N} \to \mathbb{C}$ which vanish outside a finite set, and ℓ be the space of all sequences $h : \mathbb{N} \to \mathbb{C}$.*

(i) *For $m : \mathbb{N} \to \mathbb{C}$, define the multiplication operator with symbol m, $T_m : \ell_c \to \ell$, by $T_m(h)(n) = m(n)h(m)$. Show that $\|T_m\|_{\mathbb{B}(\ell^2,\ell^2)} = \|m\|_{\ell^\infty}$.*

(ii) *Find a similar statement with $L^2[0,1]$ instead of ℓ^2, and prove it.*

EXERCISE 6.9. *Consider $\ell^1 = \ell^1(\mathbb{N})$, its 1-dimensional, closed subspace span(δ_0), and $l(a\delta_0) = a$. Find all extensions L of l to a linear functional on ℓ^1 satisfying $\|L\|_{(\ell^1)^*} = \|l\| = 1$.*

How is this different from the Hilbert case?

Another reason why $\ell^1(\mathbb{N})$ is not reflexive is that, while $\ell^1(\mathbb{N})$ is separable, $\ell^\infty(\mathbb{N})$ is not.

THEOREM 6.8. *If X is a Banach space and X^* is separable, then X is separable.*

PROOF. We consider the case of a real Banach space, the complex case being identical. If X^* is separable, it has a dense, countable subset $\{l_n\}_{n=1}^\infty$. For each n, let x_n be such that $\|x_n\| = 1$ and $l_n(x_n) \geq \|l_n\|_{X^*}/2$. Suppose $Y_0 = \text{span}_{\mathbb{Q}}\{x_n\}_{n=1}^\infty$, the linear combinations of the x_n's with rational coefficients, is not dense in X. So, $\bar{Y_0} \neq X$, and we can find z in $X \setminus Y$. By Corollary 6.3, there exists $l \in X^*$ such that $l|_Y = 0$, and $\|l\|_{X^*} = 1$. We have that

$$\|l - l_n\|_{X^*} \geq |(l - l_n)(x_n)| = |l_n(x_n)| \geq \|l_n\|_{X^*}/2.$$

If $\{l_n\}$ were dense in X^*, we could find a subsequence $\{l_{n_k}\}$ such that

$$0 = \lim_{k\to\infty} \|l - l_{n_k}\|_{X^*} \geq \limsup_{k\to\infty} \|l_{n_k}\|_{X^*}/2,$$

then,

$$1 = \|l\|_{X^*} = \lim_{k\to\infty} \|l_{n_k}\|_{X^*} = 0,$$

which is a contradiction. □

EXERCISE 6.10. (i) *Show that $\ell^\infty(\mathbb{N})$ is not separable, while $\ell^p(\mathbb{N})$ is separable for $1 \leq p < \infty$.*

(ii) *Show that $L^\infty[0,1]$ is not separable, while $L^p[0,1]$ is separable for $1 \leq p < \infty$ (with respect to the Lebesgue measure).*

(iii) *Show that $C[0,1]$ is separable (with respect to the uniform norm).*

(iv) *Show that a Hilbert space H is separable if and only if it has a countable basis.*

(v) *Find Banach spaces X, Y with $X \subset Y$, such that the imbedding map $x \mapsto x$ is bounded from X to Y, Y is separable, and X is not separable.*

In the second part of the next exercise, we need a definition. If $A : X \to Y$ is a linear, bounded operator between Banach spaces, its *adjoint* $T' : Y^* \to X^*$ is defined on a linear functional $l \in Y^*$ by $[T'l](x) = l(Tx)$.

EXERCISE 6.11. *Let $1 \leq p \leq \infty$ and let $g : \mathbb{R} \to \mathbb{C}$ be measurable. On $L^p(\mathbb{R})$, consider the multiplication operator $M_g : f \mapsto gf$.*

(i) *Show that*

$$\|M_g\|_{\mathcal{B}(L^p)} := \sup_{0\neq f\in L^p[0,1]} \frac{\|M_g f\|_{L^p}}{\|f\|_{L^p}} = \|g\|_{L^\infty}.$$

(ii) *Let $1 \leq p < \infty$ and $g \in L^\infty(\mathbb{R})$. After identifying $[L^p(\mathbb{R})]^* = L^{p'}(\mathbb{R})$ ($1/p + 1/p' = 1$), what is the expression for $[M_g]'$, the adjoint of M_g?*

6.4. Weak and weak* topologies, and the Banach-Alaoglu theorem

Our first experience is that we fix a topology on a set Y (a notion of points "being close"), and this way we can tell which functions $f : Y \to \mathbb{R}$ are continuous (their values do not change "abruptly" from point to point). We might instead want to put in the forefront the functions (the "measurables"), rather than the points. We fix a family $\mathcal{F}$ of functions $f : Y \to \mathbb{R}$, and require that those functions are continuous (that they do not change abruptly: that we consider two points to be close, that is, if they result to be close in all our observations). The *topology* $\tau(\mathcal{F})$ *generated by* $\mathcal{F}$ is the smallest one making all functions in $\mathcal{F}$ continuous.

It is easy to show that a basis of neighborhoods for $\mathcal{F}$ is given by finite intersections of basic sets having the form

$$\mathcal{N}(f, a, r) = f^{-1}(a - r, a + r) = \{y \in Y : |f(x) - a| < r\}.$$

One advantage of such *weak topologies* is that they are most economical in terms of open sets, hence they have the largest family of compact sets, which are good when convergence of (sub)sequences is concerned. We will consider below some important instances of this construction.

6.4.1. The weak and the weak* topologies.

6.4.1.1. *The weak topology.* Let X be a Banach space, and let X^* be its dual. The *weak topology* w on X is the coarsest (weaker) which makes all functionals $l \in X^*$ continuous. By general topology, a basis of neighborhoods for the weak topology is given by the family of the subsets $N(a; l_1, \ldots, l_n; \epsilon)$ (with $a \in X$, $n \geq 1$, $l_1, \ldots, l_n \in X^*$, $\epsilon > 0$), where:

$$\begin{aligned} N(a; l_1, \ldots, l_n; \epsilon) &= \{x \in X : |l_1(x) - l_1(a)| < \epsilon, \ldots, |l_n(x) - l_n(a)| < \epsilon\} \\ &= N(a; l_1; \epsilon) \cap \cdots \cap N(a; l_n; \epsilon) \\ &= [N(0; l_1; \epsilon) + a] \cap \cdots \cap [N(0; l_n; \epsilon) + a] \\ &= [\epsilon N(0; l_1; 1) + a] \cap \cdots \cap [\epsilon N(0; l_n; 1) + a] \\ &= \epsilon N(0; l_1, \ldots, l_n; 1) + a \\ &= N(0; l_1, \ldots, l_n; \epsilon) + a \end{aligned} \tag{6.4.1}$$

are various ways to write and think of these basic neighborhoods. For instance, it is clear that the topology is invariant under translations and dilations, $x \mapsto \lambda x + a$ is a homeomorphism whenever $\lambda \neq 0$.

If (X, w) is metrizable, and Ω is a metric space, then continuity of $f : X \to \Omega$ with respect to the weak topology is equivalent to continuity by sequences: $x_n \to x$ in (X, w) implies $f(x_n) \to f(x)$ in Ω. Unfortunately $(X.w)$ might not be metrizable, and we have to use *nets* instead. But some nice features remain.

PROPOSITION 6.3. *(X, w) is a Hausdorff space.*

PROOF. By translation invariance, we just have to separate $a \neq 0$ and 0. By corollary 6.2 to the Hahn-Banach theorem, there is $l \in X^*$ with $\|l\|_{X^*} = 1$ and such that $l(a) = \|a\|$. If $N(0; l; \|a\|/2) \ni 0$ and $N(a; l; \|a\|/2) \ni a$ had a point x in common,

$$\|a\| = |l(a) - l(0)| \leq |l(a) - l(x)| + |l(x) - l(0)| < \|a\|,$$

a contradiction. □

We say that the sequence $\{x_n\}$ in X *converges weakly* to $a \in X$ if $l(x_n) \to l(a)$ for all l in X^*. Although this notion is generally weaker than "weak net convergence", it is nonetheless useful in many applications, for instance in calculus of variations. We write $x_n \underset{\mathrm{w}}{\to} a$, or $w\!-\!\lim_{n\to\infty} x_n = a$. By definition, weak convergence of x_n to a means that, for each l in X^* and each $\epsilon > 0$, there exists $n(l, \epsilon)$ such that, if $n > n(l, \epsilon)$, then $x_n \in N(l; a; \epsilon)$.

PROPOSITION 6.4. (1) *If $\|x_n - a\| \to 0$, then $x_n \underset{w}{\to} a$.*

(2) *If $x_n \underset{w}{\to} a$, then $\|a\| \leq \liminf_{n\to\infty} \|x_n\|$.*

PROOF. (1) For any l in X^*, $|l(x_n) - l(a)| \leq \|l\|_{X^*}\|x_n - a\| \to 0$.
(2) Again by corollary 6.2, there is l with $\|l\|_{X^*} = 1$ and $\|a\| = l(a)$, hence:

$$\|a\| = l(a) = \lim_{n\to\infty} |l(x_n)| \leq \|l\|_{X^*} \liminf_{n\to\infty} \|x_n\|_X = \liminf_{n\to\infty} \|x_n\|_X.$$

□

We can not improve the statement in (2). Let H be an infinite dimensional Hilbert space and let $\{e_n\}_{n=1}^{\infty}$ be a orthonormal system in it. Then,

$$e_n \underset{\mathrm{w}}{\to} 0, \tag{6.4.2}$$

but $\|e_n\| = 1$ for all n.

6.4.1.2. *The weak* topology.* Let again X be a Banach space, and let X^* be its dual. The *weak* topology* w^* on X^* is the coarsest which makes the functionals $l \mapsto l(x)$ continuous for all x in X. Unless the natural identification $X \hookrightarrow X^{**}$ is surjective, i.e. unless X is *reflexive*, the *weak* topology* on X^* is weaker than the weak topology.

A basis of neighborhoods for the origin in w^* is given by finite intersections of sets of the form

$$N(0; x; \epsilon) = \{l \in X^* : |l(x)| < \epsilon\}. \tag{6.4.3}$$

The topology is invariant under translations and a basis at any point is easy to write, as in the case of the weak topology. Its basic properties coincide with those of the weak topology.

PROPOSITION 6.5. (1) *(X^*, w^*) is Hausdorff.*

(2) *If $\|x_n - a\| \to 0$, then $x_n \underset{w^*}{\longrightarrow} a$.*

(3) *If $l_n \underset{w^*}{\longrightarrow} l$, then $\|l\|_{X^*} \leq \liminf_{n\to\infty} \|l_n\|_{X^*}$.*

PROOF. Property (2) is weaker than the analogous statement for the weak topology. Property (1) is proved like the corresponding statement for the weak topology, using the fact that if $l \neq 0$ is an element of X^*, there exists x in X with $\|x\|_X = 1$ and $|l(x)| \geq \|l\|_{X^*}/2$. If $\lambda \in N(0; x; \|l\|_{X^*}/4) \cap N(l; x; \|l\|_{X^*}/4)$, then

$$\|l\|_{X^*}/2 \leq |l(x)| \leq |(l-\lambda)(x)| + |\lambda(x)| < \|l\|_{X^*}/4 + \|l\|_{X^*}/4,$$

a contradiction.

The proof of (3) follows the same lines as that of the weak topology analog. For $\epsilon > 0$ there is x in X with $\|x\| = 1$ such that $|l(x)| \geq \|l\|_{X^*} - \epsilon$. Then,

$$\|l\|_{X^*} - \epsilon \leq |l(x)| = \lim_{n\to\infty} |l_n(x)| \leq \liminf_{n\to\infty} \|l_n\|_{X^*}.$$

Let then $\epsilon \to 0$. ☐

6.4.2. Two versions of the Banach-Alaoglu theorem. The *raison d'être* of the weak* topology is related to compactness properties. To keep the exposition self.contained, below you find a proof of the Tychonoff theorem.

6.4.2.1. *Tychonoff's theorem.* If $\{X_i : i \in I\}$ is a family of sets, $X_I = \Pi_{i\in I} X_i$, their Cartesian product, is best interpreted as the set of the functions

$$p : I \to \cup_{i\in I} X_i, \text{ with } p(i) \in X_i \text{ for each } i \in I.$$

We call I the *domain* of p.

If each X_i is a topological space, the *product topology* on $\Pi_{i\in I} X_i$ is the weakest (coarsest, minimal) making all projections $\pi_i : X_I \to X_i$, $\pi_i(p) = p(i)$, continuous. A basis of open sets for it is provided by the sets $\pi_{i_1}^{-1}(U_{i_1}) \cap \cdots \cap \pi_{i_n}^{-1}(U_{i_n})$, where $n \geq 1$, $i_1, \ldots, i_n \in I$, and U_{i_l} is open in X_{i_l}.

THEOREM 6.9. *If each X_i is compact, $i \in I$, then $\Pi_{i\in I} X_i$ is compact.*

We give the first of the three proofs surveyed in Three Proofs of Tychonoff's Theorem by É. Matheron (2020), which the author labels the *Wisconsin proof.*

Some of the usual notions associated to functions carry over and are useful. If $J \subset H$, $\Pi_{H,J}p = p|_J$ is the restriction of $p \in X_H$ to X_J, $\pi_{H,J} : X_H \to X_J$, $\pi_{H,J} \circ \pi_{J,K} = \pi_{H,K}$. We set $\pi_J = \pi_{I,J}$. If $p \in X_H$ and $q \in X_J$ with H and J disjoint, then $p \vee q \in X_{H\cup J}$ is the function such that $(p \vee q)|_H = p$, $(p \vee q)|_J = q$. The set $X_\emptyset$ reduces to the unique *empty function* from $\emptyset$ to itself, which we denote by $\emptyset$ (in agreement with the interpretation of functions as particular relations).

We fix a set I, and let $\mathbb{P} = \cup_{J\subseteq I} X_J$, which is partially ordered by the relation $p \leq q$ if $q|_J = p$, where J is the domain of p.

Proof. The proof is by contradiction. We suppose that there is a family $\mathcal{U}$ of open sets in X_I such that no finite subfamily covers X_I, and we shall show that $\mathcal{U}$ itself does not cover X_I by exhibiting an element $p \in X_I \setminus \cup_{U\mathcal{U}} U$.

We consider the set B of the *bad* elements in $\mathbb{P}$: those $p \in X_J$ such that for all open $V \ni p$ open in X_J, $\pi_J^{-1}(V)$ can not be covered by finitely many sets in $\mathcal{U}$. The empty function lies in B, which is then nonempty. The only open set in $X_\emptyset$ containing $\emptyset$, in fact, is $X_\emptyset = \{\emptyset\}$ itself, and $\pi_\emptyset^{-1}(X_\emptyset) = X_I$, which can not be covered by a finite subfamily of $\mathcal{U}$ by assumption. We will prove three facts about B.

(i) *B is downward close*: if $p \leq q \in B$, then $p \in B$. In fact, if H is the domain of p and J that of q, and $V \ni p$ is open in X_H, then $\pi_H^{-1}(V) = \pi_J^{-1}(\pi_{H,J}^{-1}(V))$ which can not be covered by a a finite subfamily of $\mathcal{U}$ because $\pi_{H,J}^{-1}(V) \ni q$ is open and q is bad. We can assume $V_a = O_a \times W_a$

(ii) *If $p \in B$ has domain $J \subset I$ and $i_0 \in I \setminus J$, then there is $a \in X_{i_0}$ such that $p \vee a \in B$.* Suppose by contradiction that for all a in X_{i_0} we have $p \vee a \notin B$; i.e. there exists $V_a \ni p \vee a$ open in $X_{J\cup\{i_0\}}$ such that $\pi_{J\cup\{i_0\}}^{-1}(V_a)$ can be covered by a finite subfamily of $\mathcal{U}$. We can assume that $V_a = O_a \times W_a$ with O_a open in X_J and $W_a \ni a$ open in X_{i_0}. Since X_{i_0} is compact, it can be covered by finitely many $W_{a_1}, \ldots, W_{a_n}$. Let $O = O_{a_1} \cap \cdots \cap O_{a_n} \ni p$, so that $O \times W_{a_j} \ni p \vee a_j$ and $\pi_{J\cup\{i_0\}}^{-1}(O \times W_{a_j})$ can be covered by a finite subfamily of $\mathcal{U}$. Now,

$$\pi_J^{-1}(O) = \bigcup_{j=1}^{n} \pi_{J\cup\{i_0\}}^{-1}\left(O \times W_{a_j}\right),$$

and each set in the union can be covered by a finite subfamily of $\mathcal{U}$. Hence, $p \notin B$.

(iii) *$(B, \leq)$ has a maximal element.* If any chain C in B has an upper bound in B, by Zorn's lemma B has a maximal element. Let C be a chain in B, let H be the union of the domains of all q's in C, and define $p \in X_H$ to be such that $p|_J = q$ if $q \in C$ has domain J. The function p is well defined because C is totally ordered. Thus p is an upper bound of C in $\mathbb{P}$. It suffices to show that $p \in B$. Let $V \ni p$ be an open subset of X_H. By definition of the product topology, we can choose $V \supseteq \pi_{H,F}^{-1}(W) \ni p$ where $F \subset H$ is finite and $W \subset X_F$ is open with $W \ni p|_F$. Now, C is a chain and each $i_l \in F$ belongs to some domain J_l in the chain, so $F \subseteq J_1 \cup \ldots J_m = J_0 \subseteq H$, where J_0 is a domain in the chain. Let $q_0 \in X_{J_0} \cap C$. Since q_0 is bad, $p|_F = q_0|_F$ is bad as well by (i), thus $\pi_F^{-1}(W))$ can not be covered by a finite subfamily of $\mathcal{U}$. A fortiori, $\pi_H^{-1}(V) \supseteq \pi_F^{-1}(W)$ can not be covered by a finite subfamily of $\mathcal{U}$, hence p is bad.

Summarizing, B has a maximal element p by (iii), having domain I by (ii). In particular $X_I \ni p$ can not be covered by a finite subfamily of $\mathcal{U}$, but this is possible only if p does not belong to any subset of $\mathcal{U}$, which henceforth does not cover the whole of X_I. $\square$

6.4.2.2. *Banach-Alaoglu theorem: the topological form.*

THEOREM 6.10 (Banach-Alaoglu). *Let $\overline{B}_1$ be the closed unit ball of X^*. Then, $\overline{B}_1$ is compact in the weak* topology.*

PROOF. For each $x \in X$, consider $\overline{D(0, \|x\|)}$, the closed unit disc in the complex plane, and let

$$K := \Pi_{x \in X} \overline{D(0, \|x\|)}, \tag{6.4.4}$$

endowed with the product topology. Such is the topology having as basis at $\{f(x) : x \in X\}$ with $|f(x)| \leq \|x\|$, finite intersections of sets of the form

$$M(f; a; \epsilon) = \{g : X \to \mathbb{C} \ \text{ s.t.} |g(x)| \leq \|x\| \text{ and } |g(a) - f(a)| < \epsilon\}.$$

Since each factor $\overline{D(0, \|x\|)}$ is compact, by Tikhonoff theorem K is compact, too.

We imbed $\overline{B}_1$ into K, $\Phi : \overline{B}_1 \to K$,

$$\Phi(l) = \{\{l(x)\}_{x \in X}\}.$$

Unraveling definitions, we have that

$$\Phi(N(l; a; \epsilon)) = M(l; a; \epsilon) \cap \Phi(\overline{B}_1),$$

i.e. Φ is a homeomorphism of $\overline{B}_1$ onto its image into K. Since the latter is compact, if we show that $\Phi(\overline{B}_1)$ is closed in K, then it is compact, hence $\overline{B}_1$ is compact as well.

Suppose f lies in the closure of $\Phi(\overline{B}_1)$, and consider $x, y \in X$. For any $\epsilon > 0$ there is $l \in \overline{B}_1$ such that $|l(x) - f(x)| < \epsilon/3$, $|l(y) - f(y)| < \epsilon/3$, and $|l(x + y) - f(x + y)| < \epsilon/3$, so that

$$|f(x) + f(y) - f(x + y)| < \epsilon.$$

Hence, $f(x)+f(y) = f(x+y)$. The same way, one shows that $f(\lambda x) = \lambda f(x)$ if λ is a complex number. This shows that $f : X \to \mathbb{C}$ is a linear functional, which also lies in $\overline{B}_1$, since $|f(x)| \leq \|x\|$. We have proved that $\Phi(\overline{B}_1)$ is closed, hence K is compact. $\square$

6.4.2.3. *Banach-Alaoglu theorem: the sequential form.* A practical problem with this version of the Banach-Alaoglu theorem is that in general compactness is not in general equivalent to sequential compactness, which is very useful in many applications (measure theory and probability, claculus of variations, PDEs...). The version involving sequential compactness below could be deduced by the previous one, but we provide a direct proof which is similar to the original one for the Ascoli-Arzelà theorem, which is essentially constructive and does not require the Tikhonoff theorem.

THEOREM 6.11 (Banach-Alaoglu, separable pre-dual). *Let X be a separable Banach space, and let $\{l_n\}$ be a sequence in $\overline{B_1}$, the unit ball in X^*. then, there exists a subsequence $\{l_{n_k}\}$ and an element l in $\overline{B_1}$ such that $l_{n_k} \underset{w^*}{\longrightarrow} l$.*

PROOF. Let $\{z_j\}_{j=1}^\infty$ be a dense sequence in X, let X_0 be the dense space it generates in X, and let $\{x_m\}_{m=1}^\infty$ be a maximal family of linearly independent vectors in $\{z_j\}_{j=1}^\infty$, so that $X_0 = \operatorname{span}(\{x_m\}_{m=1}^\infty)$. We will first proceed to define l on X_0.

Consider a subsequence $n^{(1)} = \{n_k^{(1)}\}_{k=1}^\infty$ such that

$$\lim_{k\to\infty} l_{n_k^{(1)}}(x_1) = a_1$$

exists in $\mathbb{C}$. Such sequence exists because $|l_n(x_1)| \leq \|l_n\|_{X^*}\|x_1\|_X \leq \|x_1\|_X$ is bounded in $\mathbb{C}$. Suppose that subsequences $\{n\}_{n=1}^\infty \supset \{n_k^{(1)}\}_{k=1}^\infty \supset \cdots \supset \{n_k^{(m)}\}_{k=1}^\infty$ have been chosen in such a way (i) $\lim_{k\to\infty} l_{n_k^{(j)}}(x_j) = a_j \in \mathbb{C}$ when $1 \leq j \leq m$, and (ii) $n_1^{(1)} < n_1^{(2)} < \cdots < n_1^{(m)}$. As above, we can choose a subsequence $\{n_k^{(m+1)}\}_{k=1}^\infty$ of $\{n_k^{(m)}\}_{k=1}^\infty$ such that $\lim_{k\to\infty} l_{n_k^{(m+1)}}(x_{m+1}) = a_{m+1} \in \mathbb{C}$, and $n_1^{(m+1)} > n_1^{(m)}$. The sequence $\{x_1^{(m)}\}_{m=1}^\infty$ is a subsequence of all subsequences $\{n_k^{(m)}\}_{k=1}^\infty$, hence,

$$\lim_{m\to\infty} l_{n_1^{(m)}}(x_j) = a_j$$

for all x_j's. Define $l(x_j) = a_j$.

The linear extension of l to X_0 is forced by its values on the basis elements,

$$l\left(\sum_{j=1}^n c_j x_j\right) := \sum_{j=1}^n c_j l(x_j) = \sum_{j=1}^n c_j \lim_{k\to\infty} l_{n_k}(x_j) = \lim_{k\to\infty} l_{n_k}\left(\sum_{j=1}^n c_j x_j\right).$$

As a consequence,

$$\left| l\left(\sum_{j=1}^{n} c_j x_j\right)\right| = \lim_{k\to\infty}\left| l_{n_k}\left(\sum_{j=1}^{n} c_j x_j\right)\right| \tag{6.4.5}$$

$$\leq \liminf_{k\to\infty}\|l_{n_k}\|_{X^*}\left\|\sum_{j=1}^{n} c_j x_j\right\|_X \tag{6.4.6}$$

$$\leq \left\|\sum_{j=1}^{n} c_j x_j\right\|_X . \tag{6.4.7}$$

The unique continuous extension of l to $X = \overline{X_0}$ (see Exercise 6.5) satisfies $\|l\|_{X^*} \leq 1$.

We have to verify that $\lim_{k\to\infty} l_{n_k}(x) = l(x)$ for all x in X. This is a simple 3ϵ argument. For $x \in X$ and $x_0 \in X_0$,

$$\begin{aligned} |l(x) - l_{n_k}(x)| &\leq |l(x) - l(x_0)| + |l(x_0) - l_{n_k}(x_0)| + |l_{n_k}(x) - l(x)| \\ &\leq 2\|x - x_0\|_X + |l(x_0) - l_{n_k}(x_0)|. \end{aligned}$$

For given $\epsilon > 0$, choose x_0 with $\|x - x_0\|_X \leq \epsilon$, then $k(\epsilon) > 0$ such that for $k > k(\epsilon)$ one has $|l(x_0) - l_{n_k}(x_0)| < \epsilon$. □

6.5. Baire's Theorem and the uniform boundedness principle

Let (X, d) be a metric space. A subset A of X is *nowhere dense* in X if $\overline{A}$ has empty interior (it does not contain nonempty open subsets).

THEOREM 6.12. *[Baire's Theorem] If (X, d) is a complete metric space, and $A_n : n \geq 1$ is a countable union of nowhere dense sets, then $\cup_{n=1}^{\infty} A_n \neq X$.*

PROOF. Since $\overline{A_1}$ does not contain a nonempty open set, we can find $\overline{B(x_1, r_1)} \subset X\setminus\overline{A_1}$ with $r_1 < 1/2$. Since $\overline{A_2}$ does not contain a nonempty open set, $B(x_1, r_1) \setminus \overline{A_2}$ is a nonempty open set, hence, it contains $\overline{B(x_2, r_2)}$ with $r_2 < 1/2^2$. By iteration, we find $\overline{B(x_n, r_n)} \supseteq \overline{B(x_{n-1}, r_{n-1})}$ with $r_n < 1/2^n$, and $\overline{B(x_n, r_n)} \cap \left(\overline{A_1} \cup \ldots \overline{A_n}\right) = \emptyset$.

Since X is complete, the intersection of the balls $\overline{B(x_n, r_n)}$ contains (a unique) point $z \in X \setminus (\cup_{n=1}^{\infty} A_n)$. □

The next, important theorem is due to Banach and Steinhaus.

THEOREM 6.13. *[Uniform Boundedness Principle] Let $\mathcal{F} = \{T\}$ be a family of bounded, linear operators $T : X \to Y$ from a Banach space X*

to a normed, linear space Y. Suppose that, for each x in X, $\|Tx\| \le C(x)$ independent of $T \in \mathcal{F}$, although possibly dependent on x. Then, $\mathcal{F}$ is bounded,

$$\|T\| \le C$$

for some $C > 0$ independent of $T \in \mathcal{F}$.

Proof. For $m \ge 1$, let $A_m = \{x \in X : \sup\{\|Tx\| : T \in \mathcal{F}\} \le m\}$. By hypothesis, $\cup_{m=1}^{\infty} A_m = X$, hence, some A_m contains a closed open ball $\overline{B(z,r)}$. Since $z \in A_m$,

$$\overline{B(0,r)} \subseteq A_m - z \subseteq A_m + A_m \subseteq A_{2m}.$$

This is what we need, since, for $\|x\| \le 1$ and $T \in \mathcal{F}$,

$$\|Tx\| = \frac{1}{r}\|Trx\| \le \frac{2m}{r}.$$

□

Here is an application of Banach-Steinhaus.

Corollary 6.4. *Let $T_n \in \mathcal{B}(X,Y)$ be a sequence of bounded, linear maps from X to Y, Banach spaces. If for all x in X there exists*

$$Tx := \lim_{n\to\infty} T_n x,$$

then $T \in \mathcal{B}(X,Y)$ is bounded.

Proof. By hypothesis, $\{\|T_n x\|,\ n \ge 1\}$ is bounded for each x, hence $\{\|T_n\|,\ n \ge 1\}$ is bounded by some finite $C > 0$. We have, then, if $\|x\| \le 1$,

$$\|Tx\| = \lim_{n\to\infty} \|T_n x\| \le \limsup_{n\to\infty} \|T_n\| \le C.$$

Thus, $\|T\| \le C$. □

Banach-Steinhaus allows us to transform (without control of the constants) weak *quantitative* information, into strong quantitative information. The following exercise provides an example.

Exercise 6.12. *Let $f : [a,b] \to X$ a function with values in a Banach space X, and suppose f is* **weakly Lipschitz**,

$$|l(f(t+h)) - l(f(t))| \le C(l)|h|$$

for all $t, t+h \in [a,b]$ and $l \in X^*$. Show that f is Lipschitz: there is a constant $C > 0$ such that

$$\|f(t+h) - f(t)\| \le C|h|,$$

We will see another important example in the next subsection, where we treat Banach space valued holomorphic functions.

Qualitative information, however, behaves differently. For instance, it is possible to exhibit functions $f : [0,1] \to \ell^2$ which are *weakly continuous* (for all h in ℓ^2, $x \mapsto \langle h, f(x)\rangle_{\ell^2}$ is continuous), but which are not continuous. The following example is modeled on the weak, but not strongly convergent sequence in (6.4.2). Let

$$[f(x)](n) = \begin{cases} \frac{1}{1+|n-1/x|^2} \text{ if } 0 < x \leq 1, \\ 0 \text{ if } x = 0. \end{cases} \tag{6.5.1}$$

Since for $x \neq 0$ we have $\|f(x)\|_{\ell^2}^2 \geq \sum_{n=0}^{\infty} \frac{1}{(1+n^2)^2}$, f is not continuous at $x = 0$. On the other hand,

$$\lim_{x\to 0} \langle h, f(x)\rangle_{\ell^2} = 0. \tag{6.5.2}$$

After changing variables to $y = 1/x \to \infty$, and considering WLOG $h \geq 0$, we can estimate (using an elementary estimate and Cauchy-Schwarz)

$$\begin{aligned} \langle h, f(1/y)\rangle_{\ell^2} &= \sum_{0\leq n\leq y/2} \frac{h(n)}{1+(n-y)^2} + \sum_{n>y/2} \frac{h(n)}{1+(n-y)^2} \\ &\leq \frac{C}{y} \sum_{0\leq n\leq y/2} \frac{h(n)}{(1+(n-y)^2)^{1/2}} \\ &+ 2\left(\sum_{n>y/2} h(n)^2\right)^{1/2} \left(\sum_{n\geq 0} \frac{1}{(1+n^2)^2}\right)^{1/2} \\ &\leq \frac{C}{y} \|h\|_{\ell^2} \left(\sum_{n\geq 0} \frac{1}{(1+n^2)^2}\right)^{1/2} \\ &+ 2\left(\sum_{n>y/2} h(n)^2\right)^{1/2} \left(\sum_{n\geq 0} \frac{1}{(1+n^2)^2}\right)^{1/2}. \end{aligned}$$

The last expression tends to 0 as $y \to \infty$ by monotone convergence. It is easy to see that f is continuous on $(0,1]$ (for instance, by dominated convergence).

6.5.1. Banach space-valued holomorphic functions. Complex valued power series of a complex variable define holomorphic functions, which are at the heart of an elegant and powerful theory. At the root of it, is the fact that holomorphic functions can be defined in several, equivalent ways:

through power series, through complex integrals, via the Cauchy-Riemann equations, and others. These different viewpoints can be translated to the world of functions with values in Banach spaces X; or even better in Banach spaces of the form $\mathcal{B}(X)$, where a product is part of the structure; or in Banach algebras, which is the most general Banach structure with products. This way, a number of results, tools, and techniques from holomorphic function theory become available to Functional Analysis, a fact of the greatest importance.

A function $f : \mathbb{C} \supseteq \Omega \to X$ defined from a region of the complex plane with values in a Banach space is *holomorphic* if

$$f'(z) := \lim_{h \to 0 \text{ in } \mathbb{C}} \frac{f(z+h) - f(z)}{h} \in X$$

exists for all z in Ω. The function f is *weakly holomorphic* if for all l in X^* the function

$$z \mapsto l(f(z)) \in \mathbb{C}$$

is holomorphic in the usual sense.

EXERCISE 6.13. *Show that a holomorphic function $f : \Omega \to X$ is weakly holomorphic.*

THEOREM 6.14. *Let $f : \Omega \to X$ be a map defined from an open subset of $\mathbb{C}$ with values in a Banach space X. The following are equivalent.*

(i) *The function f is holomorphic in Ω.*

(ii) *The function f is weakly holomorphic in Ω.*

(iii) *For each z_0 in Ω, there is $r > 0$ such that, for $|z - z_0| < r$,*

$$f(z) = \sum_{n=0} a_n (z - z_0)^n,$$

where $a_n \in X$ and the series converges absolutely uniformly for $|z - z_0| \leq \rho < r$.

The value of r can be taken to be that of the larger disc centered at z_0 and contained in Ω.

In fact, the series converges to an X-valued holomorphic function in $B(z_0, R)$ where

$$R = \frac{1}{\limsup_{n \to \infty} \|a_n\|^{1/n}} \tag{6.5.3}$$

is the radius of the largest disc centered at z_0 and contained in Ω.

The computational part of the proof is contained in the following.

LEMMA 6.1. *Consider the power series*

$$g(z)=\sum_{n=0}^{\infty}a_n(z-z_0)^n \tag{6.5.4}$$

with coefficients $a_n \in X$, a Banach space, and with radius of convergence as in (6.5.3).

Then, $g : B(0,R) \to X$ is strongly holomorphic, and

$$g'(z)=\sum_{n=1}^{\infty}na_n(z-z_0)^{n-1}, \tag{6.5.5}$$

which has the same radius of convergence as g.

In particular, g is infinitely differentiable, and

$$g^{(n)}(z_0)=n!a_n. \tag{6.5.6}$$

PROOF OF THE LEMMA. The usual proof from holomorphic theory makes use of tools we do not have, and we do not want to develop. We provide instead an XVIII century style proof which does not require them. We can suppose $z_0=0$. We start with the estimate

$$\left|\frac{(z+h)^n-z^n}{h}-nz^{n-1}\right|\le|h|\frac{n(n-1)}{2}(|z|+|h|)^{n-2}, \tag{6.5.7}$$

which holds for $n\ge 1$ and $z,h\in\mathbb{C}$. The proof is just a calculation,

$$\begin{aligned}\left|\frac{(z+h)^n-z^n}{h}-nz^{n-1}\right| &= \left|\sum_{j=2}^{n}\binom{n}{j}z^{n-j}h^{j-1}\right|\\ &\le |h|\sum_{l=0}^{n-2}\binom{n}{l+2}|z|^{n-2-l}|h|^l\\ &= |h|\sum_{l=0}^{n-2}\frac{n(n-1)}{(l+2)(l+1)}\binom{n-2}{l}|z|^{n-2-l}|h|^l\\ &\le |h|\frac{n(n-1)}{2}\sum_{l=0}^{n-2}\binom{n-2}{l}|z|^{n-2-l}|h|^l\\ &= |h|\frac{n(n-1)}{2}(|z|+|h|)^{n-2}.\end{aligned}$$

Let now

$$\psi(z)=\sum_{n=1}^{\infty}na_nz^{n-1},$$

which has the same radius of convergence. For $|z| < R$ and $|h| < R - |z|$, we have:

$$\begin{aligned}\left\|\frac{g(z+h)-g(z)}{h} - \psi(z)\right\| &= \left\|\sum_{n=1}^{\infty} a_n\left(\frac{(z+h)^n - z^n}{h} - nz^{n-1}\right)\right\| \\ &\leq |h|\sum_{n=2}^{\infty}\frac{n(n-1)}{2}\|a_n\|(|z|+|h|)^{n-2} \\ &\to 0\end{aligned}$$

as $h \to 0$ in $\mathbb{C}$, since the series in the line before the last converges. Hence, $\psi = g'$. Since g' has the same radius of convergence as g, we can iterate the calculation,

$$g^{(m)}(z) = \sum_{n=m}^{\infty}\frac{n!}{(n-m)!}a_n(z-z_0)^{n-m},\ g^{(m)}(z_0) = m!a_m. \tag{6.5.8}$$

□

Proof of the theorem. (i) implies (ii) by exercise 6.13, and (iii) implies (i) is the lemma above.

We show that (ii) implies (iii). We fix some notation. If $x \in X$, $\widehat{x} \in X^{**}$ is the functional on X^* for which $\widehat{x}(l) = l(x)$. We set $\widehat{X} \subseteq X^{**}$ the set of such functionals.

We start by showing that if $f : \Omega \to X$ a weakly holomorphic function, and γ is a closed curve in Ω, then $z \mapsto f(z)$ is bounded on γ. In fact, and for z on γ:

$$\begin{aligned}|\widehat{f(z)}(l)| &= |l(f(z))| \\ &\leq \sup_{z\in\gamma}|l(f(z))| \\ &= C(l),\end{aligned}$$

which is finite because $z \mapsto l(f(z))$ is continuous on γ. By Banach-Steinhaus theorem,

$$\sup_{z\in\gamma}\|f(z)\|_X = \sup_{z\in\gamma}\|\widehat{f(z)}\|_{X^{**}} < \infty.$$

Let $B(z_0, r)$ be a disc contained in Ω, and let γ be a circle of radius $\rho < r$ centered at z_0 and contained in Ω. Let again l be in X^*. Then, $z \mapsto l(f(z))$ can be expanded as a power series,

$$l(f(z)) = \sum_{n=0}^{\infty} a_n(l)(z-z_0)^n,$$

where

$$a_n(l) = \frac{1}{2\pi i}\int_\gamma \frac{l(f(z))}{(z-z_0)^{n+1}}dz.$$

The functional $l \mapsto a_n(l)$ is linear and

$$|a_n(l)| \leq \sup_{z\in\gamma} \|f(z)\|_X \frac{\|l\|_{X^*}}{\rho^n}.$$

That is, $a_n \in X^{**}$ with

$$\|a_n\|_{X^{**}} \leq \frac{\sup_{z\in\gamma}\|f(z)\|_X}{\rho^n}. \tag{6.5.9}$$

Thus,

$$g(z) = \sum_{n=0}^{\infty} a_n(z-z_0)^n \tag{6.5.10}$$

converges for $|z - z_0| < \rho$ to a function g with values in X^{**}. You just have to check that the usual proof in holomorphic function theory works when you have power series with coefficients in a Banach space (geometric series only are involved). The Hadamard formula for the radius of convergence of the series is proved like in holomorphic theory. By the lemma, g is an X^{**} valued holomorphic function. We next prove that $g(z) = \widehat{f(z)}$ is the image in X^{**} of $\widehat{f(z)}$, i.e. that $[g(z)](l) = l(f(z))$ for all $l \in X^*$.

In fact, all l in X^* we have

$$[g(z)](l) = \sum_{n=0}^{\infty} a_n(l)(z-z_0)^n = l(f(z)).$$

By the lemma, g is infinitely differentiable and

$$a_n = \frac{g^{(n)}(z_0)}{n!} = \frac{\widehat{f^{(n)}(z_0)}}{n!} \in X,$$

because X is closed in X^{**}, and the derivatives, which are limits of X valued functions, belong to X. Thus, $a_n = \widehat{\alpha_n}$, with $\alpha_n \in X$, and

$$f(z) = \sum_{n=0}^{\infty} \alpha_n(z-z_0)^n,$$

as wished. □

EXERCISE 6.14. *Let $\Omega \subseteq \mathbb{C}$ be open, and let $\lambda : \Omega \to \mathbb{C}$ be holomorphic, and $f : \Omega \to X$ be a Banach space-valued holomorphic function. Show that their product $\lambda \cdot f : \Omega \to X$ is holomorphic.*

6.6. The Open Mapping Theorem and the Closed Graph Theorem

A map $F : M \to N$ between metric spaces is open if the image of an open set in M is open in N.

EXERCISE 6.15. *If $T : X \to Y$ is a linear map between normed, linear spaces, then the following are equivalent:*

(i) *T is open;*

(ii) *there is a ball $B_X(0, r)$ in X such that $T(B_X(0, r))$ contains a ball in Y;*

(iii) *there is a ball $B_X(0, r)$ in X such that $T(B_X(0, r))$ contains a ball centered at 0 in Y.*

Moreover, if T is open, then it is onto.

Hint. *(i) $\implies$ (ii) is clear. For (ii) $\implies$ (iii), you can show that if $T(B(0, r)) \supset B(y, R)$, then $T(B(0, 2r)) \supset B(0, R)$. The proof that (iii) $\implies$ (i) is easy. Also, (iii) implies that T is onto by the homogeneity of T.*

THEOREM 6.15. *[Open Mapping Theorem] Let X, Y be Banach spaces, and let $T : X \to Y$ be a linear, bounded map from X onto Y. Then, T is open.*

PROOF. Let $B_n := B(0, n) \subset X$. Since $\cup_n T(B_n) = Y$, there is n such that $\overline{T(B_n)} \supset B(y_0, \epsilon)$. If $y \in Y$, then $y = (y + y_0) - y_0 \in \overline{T(B_n)} - \overline{T(B_n)} \subset \overline{T(B_{2n})}$ provided $\|y\|\epsilon$. i.e. $B(0, \epsilon/(2n)) = B(0, \eta) \subseteq \overline{T(B_1)}$.

It suffices then to show that $\overline{T(B_1)} \subset T(B_2)$.

Let $y \in \overline{T(B_1)}$, and pick $x_1 \in B_1$ such that $\|y - Tx_1\| < \eta/2$, so that $Tx_1 - y \in B(0, \eta/2) \subseteq \overline{T(B_{1/2})}$. We can then find pick $x_2 \in B_{1/2}$ such that $\|y - Tx_1 - Tx_2\| < \eta/2^2$.

Iterating, we have $x_n \in B_{1/2^{n-1}}$ such that $\|y - T(x_1 + \cdots + x_n)\| < \eta/2^n$. We have that $x_1 + 1dots + x_n \to x \in B_2$ as $n \to \infty$, and $T(x) = \lim_{n\to\infty} T(x_1 + \cdots + x_n) = y$, as wished. □

EXERCISE 6.16. *Let $X = \mathbb{R}^2$ with the Euclidean metric. Show that for all $\epsilon > 0$ there is a linear bijection $T : X \to X$ such that $\|T\| = 1$, yet $T(B(0, 1))$ does not contain $B(0, \epsilon)$. That is, the Open Mapping Theorem is not quantitative.*

THEOREM 6.16. *[Inverse Mapping Theorem] Let $T : X \to Y$ be a bounded bijection of Banach spaces. Then, $T^{-1} : Y \to X$ is bounded.*

PROOF. The fact that T is open means that for some $\epsilon > 0$, if $\|y\| < \epsilon$ in Y, then there is $x \in X$ with $\|x\| < 1$ and $Tx = y$, i.e. $x = T^{-1}y$. Said it differently, T^{-1} maps $B(o, \epsilon)$ into $B(0, 1)$, so that $\|T^{-1}\| \leq 1/\epsilon$. □

EXERCISE 6.17. *Show that if $\|\cdot\|_1$ and $\|\cdot\|_2$ are two norms on X, if $\|x\|_2 \leq \|x\|_1$ on X and X is Banach with respect to $\|\cdot\|_1$, then there is $C > 0$ such that $\|x\|_1 \leq C\|x\|_2$ on X.*

Exercise 6.18. *Show that (even!) in* $\mathbb{R}^2$ *one can find norms* $\|\cdot\|_1$ *and* $\|\cdot\|_2$ *with* $\|x\|_2 \le \|x\|_1$, *yet* $\|x\|_1 \le C\|x\|_2$ *only holds if* C *is (arbitrarily) large. That is, the result in the previous exercise is not quantitative.*

The *graph* of a function $f : M \to N$ is the set $\Gamma(f) = \{(x, y \in M \times N : y = f(x))\}$. If $T : X \to Y$ is a linear operator between linear spaces, then $\Gamma(T)$ is a linear subspace of $X \times Y$. If X and Y are normed, then $\|(x, y)\| := \|x\| + \|y\|$ defines a norm on $X \times Y$, hence on $\Gamma(T)$.

Exercise 6.19. *If* X *and* Y *are normed linear space and* $T \in \mathcal{B}(X, Y)$, *then* $\Gamma(T)$ *is closed.*

Theorem 6.17. *[Closed Graph Theorem] Let* $T : X \to Y$ *be a linear operator defined from a Banach space* X *to a Banach space* Y. *If* $\Gamma(T)$ *is closed, then* T *is bounded.*

Proof. By assumption, $\Gamma(T)$ is closed in a Banach space, hence Banach itself: if $(x_n, Tx_n) \to (x, y)$ is Cauchy, then $y = Tx$, so $(x, y) \in \Gamma(T)$.

Consider the projections $\pi_X : (x, Tx) \mapsto x$ and $\pi_Y : (x, Tx) \mapsto Tx$. Both projections are bounded, Π_X is invertible (hence, its inverse is continuous), and $T = \pi_Y \circ \Pi_X^{-1}$. Hence, T is bounded. □

Theorem 6.18. *[Hellinger-Toeplitz] Let* $A : H \to H$ *be an everywhere defined self-adjoint operator on a Hilbert space* H*:*

$$\langle Ax|y\rangle = \langle x|Ay\rangle \text{ if } x, y \in H.$$

Then, A *is bounded.*

Proof. Let (x_n, Ax_n) be a sequence in $\Gamma(A)$ with $x_n \to x$ and $Ax_n \to y$. For all $z \in H$:

$$\begin{aligned} \langle z|y\rangle &= \lim_{n\to\infty} \langle z|Ax_n\rangle = \lim_{n\to\infty} \langle Az|x_n\rangle \\ &= \langle Az|x\rangle = \lim_{n\to\infty} \langle z|Ax\rangle, \end{aligned}$$

hence, $Ax = y$, so $\Gamma(A)$ is closed. □

6.7. Integrals of continuous, Banach space valued functions

The integral of functions with values on a Banach space can be defined in several ways, choosing which depends on the application we have in mind[2].

[2] There are different definitions for the integral of a Banach space valued function. A reasonably general, Lebesgue style one, is provided by *Bochner integrals*. See e.g. The Bochner integral by Wenjing Wu. The weak version of the Bochner integral is the *Pettis integral*. Here, however, we are integrating continuous functions, and more elementary definitions of integral can be used.

Think of Banach space valued holomorphic functions: extending to them the notion of Cauchy integrals requires integrating vector valued functions. Having in mind holomorphic theory, we sketch here a construction of the integral for a continuous function $f : [a, b] \to X$, where X is Banach. The definition of integral we discuss here does not require any foundational result concerning Banach spaces. We only make use of linearity and completeness.

Let $f : [a, b] \to X$ be a continuous function with values in the Banach space X. By theorem 1.7, f is uniformly continuous. For fixed $n \geq 1$, let

$$S_n(f) = \sum_{j=1}^{2^n} f\left(\frac{j}{2^n}\right)\frac{b-a}{2^n} \in X. \tag{6.7.1}$$

The sequence $\{S_n(f)\}_{n=1}^{\infty}$ is Cauchy in X. The calculation is similar to the one we met when defining the Lebesgue measure. For any fixed $\epsilon > 0$, there exists $\delta > 0$ such that $\|f(s) - f(t)\| \leq \epsilon$ if $|s - t| \leq \delta$. If $1/2^n \leq \delta$, then

$$\begin{aligned}
\|S_{n+m}(f) - S_n(f)\| &= \left\|\sum_{j=1}^{2^n}\frac{b-a}{2^n}\frac{1}{2^m}\sum_{l=1}^{2^m}\left[f\left(\frac{j}{2^n}\right) - f\left(\frac{j-1}{2^n} + \frac{l}{2^{n+m}}\right)\right]\right\| \\
&\leq \sum_{j=1}^{2^n}\frac{b-a}{2^n}\frac{1}{2^m}\sum_{l=1}^{2^m}\left\|f\left(\frac{j}{2^n}\right) - f\left(\frac{j-1}{2^n} + \frac{l}{2^{n+m}}\right)\right\| \\
&\leq (b-a)\epsilon.
\end{aligned}$$

By definition,

$$\int_a^b f(t)dt = \lim_{n\to\infty} S_n(f). \tag{6.7.2}$$

You might complain that this definition depends on the choice of very special partitions of $[a, b]$ (this makes additivity of the integral problematic, for instance). Fortunately, the dependence is apparent. Given a partition $a = t_0 < t_1 < \cdots < t_{m-1} < t_m = b$ of $[a, b]$ and sampling points $t_{j-1} \leq t_j^* \leq t_j$, let

$$S(\{t_j\}_{j=0}^m, \{t_j^*\}_{j=1}^m) = \sum_{j=1}^{m} f(t_j^*)(t_j - t_{j-1}).$$

The *resolution* of the partition is $\min_{j=1,\ldots,m}(t_j - t_{j-1}) = \delta\left(\{t_j\}_{j=0}^m\right)$.

PROPOSITION 6.6. *Let X be Banach and $f : [a, b] \to X$ continuous. Then, for all $\epsilon > 0$ there is $\delta > 0$ such that if $\delta\left(\{t_j\}_{j=0}^m\right) \leq \delta$, then $\left\|\int_a^b f(t)dt - S(\{t_j\}_{j=0}^m, \{t_j^*\}_{j=1}^m)\right\| \leq \epsilon$.*

Proof. Let $n \geq 1$ be such that $\left\| \int_a^b f(t)dt - S_n(f) \right\| \leq \epsilon$ and that $\|f(s) - f(t)\| \leq \epsilon$ if $|s-t| \leq 1/2^n$. Consider a partition $\{t_j\}_{j=0}^m$ with resolution less than $\delta = 1/2^n$, and a sampling set $\{t_j^*\}_{j=1}^m$ as above. Below, we denote by $|I|$ the Lebesgue measure of an interval I.

$$\begin{aligned}
& \left\| S_n(f) - S(\{t_j\}_{j=0}^m, \{t_j^*\}_{j=1}^m) \right\| \\
= & \left\| \sum_{l=1}^{2^n} \sum_{j:[t_{j-1},t_j]\cap[(l-1)/2^n,l/2^n]\neq\emptyset} (f(l/2^n) - f(t_j^*)) |[t_{j-1},t_j] \cap [(l-1)/2^n, l/2^n]| \right\| \\
\leq & \sum_{l=1}^{2^n} \sum_{j:[t_{j-1},t_j]\cap[(l-1)/2^n,l/2^n]\neq\emptyset} \left\| f(l/2^n) - f(t_j^*) \right\| \cdot |[t_{j-1},t_j] \cap [(l-1)/2^n, l/2^n]| \\
\leq & (b-a)\epsilon.
\end{aligned}$$

□

Exercise 6.20. *Let $\alpha : [a,b] \to \mathbb{R}$ be increasing (or, more generally, let $\alpha : [a,b] \to \mathbb{C}$ be of bounded variation). For $f : [a,b] \to X$ continuous (where X is Banach) provide a definition of the vector valued Stjelties integral*

$$\int_a^b f(t) d\alpha(t),$$

and show that it is well defined.

Corollary 6.5. *Let $f : [a,b] \to X$ be a continuous function with values in a Banach space X, and let $T : X \to Y$ be a bounded operator between X and another Banach space Y. Then,*

$$T\left(\int_a^b f(t)dt \right) = \int_a^b T(f(t))dt.$$

Proof. It suffices to pass in the limit for $n \to \infty$ the two sides of the equality

$$T\left(\sum_{j=1}^m f(t_j^*)(t_j - t_{j-1}) \right) = \sum_{j=1}^m T\left[f(t_j^*) \right] (t_j - t_{j-1}).$$

□

We give a sample application to Banach space valued holomorphic functions. The reader might find it interesting and rewarding to translate in the Banach valued world the chapter of a book of complex analysis concerning complex integrals.

Corollary 6.6. *[Morera's theorem for Banach space valued functions] Let $f : \Omega \to X$ be a continuous, Banach space valued function. Then f is holomorphic if and only if $\int_{\partial T} f(z)dz = 0$ for all triangles T contained in Ω.*

Proof. The function f is holomorphic if and only if it is weakly holomorphic, and by Morera's theorem this holds if and only if

$$0 = \int_{\partial T} l(f(z))dz = l\left(\int_{\partial T} f(z)dz\right)$$

for all triangles and $l \in X^*$, the second equality following from corollary 6.5. Thus, $\int_{\partial T} f(z)dz = 0$. $\square$

CHAPTER 7

Tempered distributions and Fourier transforms

The basic idea underlying distributions is well represented by Riesz representation theorem for measures. If we want to simultaneously manipulate a set of (Borel, regular) measures, it is convenient to consider their joint action on nice (continuous, compactly supported) functions. Each such measure defines a positive, linear functional on $C_c(\mathbb{R})$ and Riesz theorem ensures that all such functionals arise in this way. In the 1940's Laurent Schwartz greatly extended the scope of such idea. He much restricted the space of the "test functions" (not just continuous, but also infinitely differentiable), and correspondingly much enlarged the space of the linear functionals (the "distributions") defined on them. What is more important, operations on test functions have analogs in the space of distributions. We saw an example of this when showing that increasing functions have "weak derivatives", and such derivatives are Borel measures.

This allows a great freedom in taking derivatives, applying integral operators, and so on. More precisely, distributions are the dual of a function space where many manipulations of functions are allowed. The adjoints of such operations are defined in the much extended universe of distributions.

The resulting body of knowledge has several striking applications. Of course, after "distributional" objects performing certain tasks are found, the problem remains to see if they correspond to more terrestrial mathematical objects (functions, measures...).

Here we give a basic introduction to distributions, with a special emphasis on tempered distributions, which are well suited to deal with the Fourier transform. We work on $\mathbb{R}$, but everything we prove can be generalized with no effort to $\mathbb{R}^d$.

More than a theory, distributions are a method, and different theories can be developed in different contexts, depending on the problems one is considering, and the structures which are offered by the environment (smoothness, algebraic structures, geometric structures, etcetera). For instance, on a locally compact metric space the natural class of distributions is that of the (signed) Borel measures.

The best place to start learning in depth the theory of distributions is still Laurent Schwartz, Théorie Des Distributions (1950/51, revised edition

1966). A nice and easy reading which can usefully supplement these notes is the 16 pages Tempered Distributions by Joel Feldman.

7.1. Tempered distributions

The Fourier transform with its different *avatars* is, together with derivatives, integrals, and holomorphic functions, the most natural, important, and useful single object of basic mathematical analysis. Tempered distributions were developed as a version of distribution theory which could accomodate Fourier transforms.

7.1.1. The Schwartz class: definition, topology, and basic operations. The *Schwartz class* $\mathcal{S} = \mathcal{S}(\mathbb{R})$ is the linear space of the functions $\varphi\colon \mathbb{R} \to \mathbb{C}$ for which the seminorms

$$[\varphi]_{m,n} = \sup_{x\in\mathbb{R}} (1+x^2)^{n/2} \left|\varphi^{(m)}(x)\right| < \infty \tag{7.1.1}$$

for all $m, n \geq 0$. Schwartz functions are C^∞ and decay at infinity faster than any polynomial, together with their derivatives of all orders.

Exercise 7.1. *Show that each $[\cdot]_{m,n}$ defines a norm on $\mathcal{S}$: if $\varphi \in \mathcal{S}$ and $[\varphi]_{m,n} = 0$, then $\varphi = 0$.*

An example of a function in $\mathcal{S}(\mathbb{R})$ which does not belong to C_c^∞ is $\varphi(x) = e^{-x^2}$. Schwartz functions come in great supply.

Proposition 7.1. (i) *$C_c^\infty(\mathbb{R})$ is contained in $\mathcal{S}$.*

(ii) *$\mathcal{S}$ is dense in $C_0(\mathbb{R})$ with respect to the uniform norm.*

(iii) *For $1 \leq p < \infty$, $\mathcal{S}$ is dense in $L^p(\mathbb{R})$.*

Proof. Item (i) is obvious, and (ii) holds because $C_c^\infty(\mathbb{R})$ is dense in $C_c(\mathbb{R})$ in the uniform norm, and the closure of the latter with respect to the uniform norm is $C_0(\mathbb{R})$. About (iii), we first verify that $\mathcal{S} \subset L^p$:

$$\int_{\mathbb{R}} |\varphi(x)|^p dx = \int_{\mathbb{R}} |(1+x^2)\varphi(x)|^p \frac{dx}{(1+x^2)^p} \leq C(p)[\varphi]_{0,2}^p,$$

with $C(p) = \int_{\mathbb{R}} \frac{dx}{(1+x^2)^p}$. We have density of $\mathcal{S}$ in L^p because C_c^∞ is already dense in L^p. □

Indeed, $\mathcal{S}$ is dense in many other relevant function spaces, and this is one of the reasons why the Schwartz class is so useful.

Given a sequence of functions $\{\varphi_j\}$ and a function φ in $\mathcal{S}(\mathbb{R})$, we say that $\lim_{j\to\infty}\varphi_j = \varphi$ *in* $\mathcal{S}$ if

$$\lim_{j\to\infty}[\varphi_j-\varphi]_{m,n}=0 \text{ for all } m,n\geq 0. \tag{7.1.2}$$

This notion of convergence is rather strong. For instance, it implies L^p-convergence.

LEMMA 7.1. *Let* $1\leq p\leq\infty$. *If* $\varphi_j\to\varphi$ *in* $\mathcal{S}$, *then* $\varphi_j\to\varphi$ *in* L^p.

PROOF. For $1\leq p<\infty$,

$$\int_{\mathbb{R}}|\varphi(x)-\varphi_j(x)|^p dx=\int_{\mathbb{R}}|(1+x^2)[\varphi(x)-\varphi_j(x)]|^p\frac{dx}{(1+x^2)^p}\leq C(p)[\varphi-\varphi]_{0,2}^p.$$

Also, $\|\varphi-\varphi_j\|_{L^\infty}=[\varphi-\varphi_j]_{0,0}$. □

Convergence in $\mathcal{S}$ can be captured by a distance on $\mathcal{S}(\mathbb{R})$:

$$d(\varphi,\psi):=\sum_{m,n\geq 0}\frac{1}{2^{m+n+2}}\frac{[\varphi-\psi]_{m,n}}{[\varphi-\psi]_{m,n}+1}<1. \tag{7.1.3}$$

By definition, d is translation invariant: $d(\varphi,\psi)=d(\varphi-\psi,0)$.

PROPOSITION 7.2. (i) $\varphi_j\to\varphi$ *in* $\mathcal{S}(\mathbb{R})$ *if and only if* $d(\varphi_j,\varphi)\to 0$.

(ii) *A basis of neighborhoods for* 0 *in* $\mathcal{S}(\mathbb{R})$ *is given by the cylinder sets:*

$$\mathcal{N}(M,\epsilon)=\{\varphi\in\mathcal{S}(\mathbb{R}):\ [\varphi]_{m,n}<\epsilon \text{ for all } m,n\leq M\}. \tag{7.1.4}$$

(iii) $(\mathcal{S},d)$ *is complete.*

PROOF. We can assume $\varphi=0$ by translation invariance.

(i) If $d(\varphi_j,0)\to 0$, then each sequence $[\varphi_j]_{m,n}$ has to converge to zero. Viceversa, suppose that $[\varphi_j]_{m,n}\to 0$ for all $m,n\geq 0$ and fix $\epsilon>0$. Find $M>0$ such that $\sum_{\max(m,n)>M}2^{-m-n-2}<\epsilon$, and find $J>0$ such that if $j>J$ then $[\varphi_j]_{m,n}<\epsilon$ for all $\max(m,n)\leq M$. Then, $d(\varphi_j,0)<2\epsilon$.

(ii) If $\varphi\in\mathcal{N}(M,\delta)$, then

$$d(\varphi,0)<M^2\delta+\frac{1}{2^M},$$

hence, the ball $B(0,\epsilon)$ in $(\mathcal{S},d)$ contains $\mathcal{N}(M,\delta)$ provided M, then δ, are appropriately chosen. In the other direction, if $d(\varphi,0)<\epsilon$, then

$$\max_{m,n\leq M}\frac{[\varphi]_{m,n}}{1+[\varphi]_{m,n}}\leq 2^{2M+2}\epsilon.$$

Choose $\epsilon > 0$ such that $2^{2M+2}\epsilon < \delta \leq 1/2$, and observe that for such δ and positive x, if $\frac{x}{1+x}$, then $x < 2\delta$. Then, $\mathcal{N}(M, 2\delta)$ contains $B(0, \epsilon)$.

(iii) If $\{\varphi_j\}$ is a Cauchy sequence, then

$$\sup_{x\in\mathbb{R}}(1+x^2)^{n/2}\left|\varphi_{l+j}^{(m)}(x) - \varphi_l^{(m)}(x)\right| \to 0 \text{ as } l \to \infty,$$

and this implies uniform convergence of φ_j to some infinitely differentiable φ, together with all the derivatives. Moreover, for $m, n \geq 0$,

$$\begin{aligned}
(1+x^2)^{n/2}\left|\varphi^{(m)}(x)\right| &\leq (1+x^2)^n\left|\varphi^{(m)}(x) - \varphi_j^{(m)}(x)\right| + (1+x^2)^{n/2}\left|\varphi_j^{(m)}(x)\right| \\
&\leq \lim_{l\to\infty}(1+x^2)^n\left|\varphi_{l+j}^{(m)}(x) - \varphi_j^{(m)}(x)\right| + [\varphi_j]_{m,n} \\
&\leq \limsup_{l\to\infty}[\varphi_{j+l} - \varphi_j]_{m,n} + [\varphi_j]_{m,n} \\
&< \infty,
\end{aligned}$$

hence $\varphi \in \mathcal{S}(\mathbb{R})$, and in particular

$$\begin{aligned}
(1+x^2)^n\left|\varphi^{(m)}(x) - \varphi_j^{(m)}(x)\right| &= \lim_{l\to\infty}(1+x^2)^n\left|\varphi_{l+j}^{(m)}(x) - \varphi_j^{(m)}(x)\right| \\
&\leq \limsup_{l\to\infty}[\varphi_{j+l} - \varphi_j]_{m,n},
\end{aligned}$$

which can be made as small as we wish by choosing $j \geq j(m, n, \epsilon)$ large enough. □

The proof of the completeness contains a useful facts.

Corollary 7.1. *A sequence $\{\varphi_n\}$ in $\mathcal{S}$ is Cauchy with respect to d if and only if it is Cauchy separately with respect to all the norms $[\cdot]_{l,m}$, $l, m \geq 0$.*

Exercise 7.2. *Prove corollary 7.1.*

Proposition 7.3. *$C_c^\infty(\mathbb{R})$ is dense in $\mathcal{S}(\mathbb{R})$.*

Proof. Pick $\eta \in C_c^\infty$, $\eta \succ [-1, 1]$, and for $R > 1$ let $\eta_R(x) = \eta(x/R)$. For $\varphi \in \mathcal{S}$, we have that $\varphi\eta_R$ is in C_c^∞, and

$$\begin{aligned}
|(\varphi(1-\eta_R))^{(m)}(x)| &\leq |\varphi^{(m)}(x)|1-\eta_R(x)| + C\sum_{j=1}^m R^{-j}|\eta^{(j)}(x/R)|\cdot|\varphi^{(m-j)}(x)| \\
&\leq |\varphi^{(m)}(x)|1-\eta_R(x)| + CR^{-1}\sum_{j=1}^m[\eta]_{j,0}\cdot|\varphi^{(m-j)}(x)|.
\end{aligned}$$

Then, for fixed $\epsilon > 0$,

$$\begin{aligned}
(1+x^2)^{n/2}|(\varphi(1-\eta_R))^{(m)}(x)| &\leq (1+x^2)^{n/2}|\varphi^{(m)}(x)|1-\eta_R(x)| \\
&\quad CR^{-1}\sum_{j=1}^m[\eta]_{j,0}[\varphi]_{m-j,n}
\end{aligned}$$

$$
\begin{aligned}
&\le (1+x^2)^{n/2+1}|\varphi^{(m)}(x)|\frac{1-\eta_R(x)}{1+x^2}+\epsilon \\
&\quad \text{provided } R \ge R(\epsilon) \text{ is large enough} \\
&\le (1+x^2)^{n/2+1}|\varphi^{(m)}(x)|\frac{1}{1+R^2}+\epsilon \\
&\le \epsilon(1+x^2)^{n/2+1}|\varphi^{(m)}(x)|+\epsilon,
\end{aligned}
$$

possibly choosing a larger value of R. Then,

$$[\varphi-\varphi\eta_R]_{m,n} \le \epsilon[\varphi]_{m,n+2}+\epsilon,$$

which implies density. □

Many important operations on functions are continuous on $\mathcal{S}(\mathbb{R})$.

PROPOSITION 7.4. (i) *The* **derivative** $\varphi \mapsto D\varphi = \varphi'$ *is continuous on* $\mathcal{S}(\mathbb{R})$.

(ii) *The* **multiplication times** x $M_x\colon \varphi \mapsto x\varphi$ *is continuous on* $\mathcal{S}(\mathbb{R})$.

(iii) **Dilations** $\varphi \mapsto \delta_t\varphi(x) = \frac{1}{t}\varphi\left(\frac{x}{t}\right)$ *and* **translations** $\varphi \mapsto \tau_a\varphi(x) = \varphi(x-a)$ *are continuous on* $\mathcal{S}(\mathbb{R})$.

(iv) *The* **multiplication** $\cdot\colon \mathcal{S}(\mathbb{R}) \times \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ *is continuous.*

(v) *The* **inversion map** $\varphi \mapsto U\varphi(x) = \varphi(-x)$ *is an isometry of* $\mathcal{S}(\mathbb{R})$.

(vi) *The* **convolution** $(\varphi,\psi) \mapsto \varphi * \psi$ *is a continuous map* $*\colon \mathcal{S}(\mathbb{R}) \times \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$.

PROOF. (i) It follows from $[D\varphi]_{m,n} = [\varphi]_{m+1,n}$.

(ii) After expanding $(x\varphi)^{(m)} = m\varphi^{(m-1)} + \varphi^{(m)}$, we have $[M_x\varphi]_{m,n} \le m[\varphi]_{n-1,m} + [\varphi]_{m,n+1}$.

(iii) The easy verification is left to the reader.

(iv) For $\varphi, \psi \in \mathcal{S}(\mathbb{R})$, we have:

$$
\begin{aligned}
[\varphi\psi]_{m,n} \quad & \sup_x \left|(1+x^2)^{n/2}(\varphi\psi)^{(m)}(x)\right| \\
\le\; & \sum_{l=0}^{m} C(m,l)\sup_x \left|(1+x^2)^{n/2}\varphi^{(l)}(x)\psi^{(m-l)}(x)\right| \\
\le\; & \sum_{l=0}^{m} [\varphi]_{l,0}[\psi]_{m-l,n}. \qquad (7.1.5)
\end{aligned}
$$

If $\varphi_j \to \varphi$ and $\psi_j \to \psi$ in $\mathcal{S}$, then

$$\begin{aligned}
[\varphi\psi - \varphi_j\psi_j]_{m,n} &\leq [\varphi(\psi - \psi_j)]_{m,n} + [(\varphi - \varphi_j)\psi_j]_{m,n} \\
&\leq C\sum_{l=0}^{m}[\varphi]_{l,0}[\psi - \psi_j]_{m-l,n} + C\sum_{l=0}^{m}[\varphi - \varphi_j]_{l,0}[\psi_j]_{m-l,n} \\
&\leq C\sum_{l=0}^{m}[\varphi]_{l,0}[\psi - \psi_j]_{m-l,n} + C_1\sum_{l=0}^{m}[\varphi - \varphi_j]_{l,0}[\psi]_{m-l,n}, \qquad (7.1.6)
\end{aligned}$$

where the third inequality holds because $[\psi_j]_{m-l,n} \to [\psi]_{m-l,n}$ for $l = 0, \ldots, m$. The last expression vanishes as $j \to \infty$ by hypothesis.

(v) Obvious.

(vi) Estimating $1 + (a + b)^2 \leq 2(1 + a^2)(1 + b^2)$, and differentiating the convolution m times, we have

$$\begin{aligned}
(1 + x^2)^{n/2}|(\varphi * \psi)^{(m)}(x)| &= \left|\int (1 + x^2)^{n/2}\varphi(x - y)\psi^{(m)}(y)dy\right| \\
&\leq C\int (1 + (x - y)^2)^{n/2}|\varphi(x - y)| \cdot \qquad (7.1.7) \\
&\quad \cdot(1 + y^2)^{n/2}|\psi^{(m)}(y)|dy \\
&\leq C[\varphi]_{0,n}\int (1 + y^2)^{n/2+1}|\psi^{(m)}(y)|\frac{dy}{1 + y^2} \\
&\leq C_1[\varphi]_{0,n}[\psi]_{m,n+2}. \qquad (7.1.8)
\end{aligned}$$

We deduce (vi) from (7.1.7) the same way we deduced (7.1.6) from (7.1.5).

□

7.1.2. Tempered distributions and the basic operations on them. A *tempered distribution* is a continuous, linear functional $T : \mathcal{S} \to \mathbb{C}$. Since the latter is a metric space, continuity is equivalent to sequential continuity,

$$\text{if } \varphi_j \to \varphi \text{ in } \mathcal{S}, \text{ then } T(\varphi_j) \to T(\varphi) \text{ in } \mathbb{C}. \qquad (7.1.9)$$

We denote by $\mathcal{S}' = \mathcal{S}'(\mathbb{R})$ the vector space of the tempered distributions.

The defining condition (7.1.9) involves all norms in $\mathcal{S}$, since $\varphi_j \to \varphi$ in $\mathcal{S}$ if and only if $[\varphi_j - \varphi]_{m,n} \to 0$ for all $m, n \geq 0$. We will see below that, for a given distribution T, only finitely many norms are involved.

The basic examples of distributions are based on functions. Let $f \in L^1_{loc}$, and define

$$T_f(\varphi) := \int_{\mathbb{R}} \varphi(x)f(x)dx. \qquad (7.1.10)$$

The integral converges, provided that $f(x)$ does not grow too fast as $|x| \to \infty$.

PROPOSITION 7.5. *If $f \in L^1_{loc}$, and there is $n \geq 0$ such that $\int_{\mathbb{R}} \frac{|f(x)|}{(1+x^2)^{n/2}} dx < \infty$, then $T_f \in \mathcal{S}'$.*

This condition holds, in particular, if $f \in L^p$ for some $1 \leq p \leq \infty$.

PROOF. It suffices to observe that

$$|T_f(\varphi)| \leq [\varphi]_{0,n} \int_{\mathbb{R}} \frac{|f(x)|}{(1+x^2)^{n/2}} dx.$$

□

Indeed, we can act on Schwartz functions by means of positive, Borel measures μ:

$$T_\mu(\varphi) = \int_{\mathbb{R}} \varphi(x) d\mu(x). \tag{7.1.11}$$

The growth condition on μ is that, for some $n \geq 0$,

$$\int_{\mathbb{R}} \frac{d\mu(x)}{(1+x^2)^{n/2}} < \infty. \tag{7.1.12}$$

This statement has a converse, which we state without proof.

THEOREM 7.1. *A distribution on $\mathcal{S}$ is positive if and only if there exists a regular Borel measure with moderate growth $\mu \geq 0$ on $\mathbb{R}$ such that:*

$$T(\varphi) = \int_{\mathbb{R}} \varphi d\mu \text{ for all } \varphi \in \mathcal{S}. \tag{7.1.13}$$

See A note on tempered measures, by Michael Baake and Nicolae Strungaru for a detailed argument. We will prove a weaker statement in §7.3.

7.1.2.1. *The order of a distribution.* Although the distance between Schwartz functions involves all derivatives, for any specific distribution T, only finitely many of them are involved in testing its continuity.

THEOREM 7.2. *Let $T : \mathcal{S} \to \mathbb{C}$ be a linear functional. Then, T is a distribution (it is continuous) if and only if there are $C > 0$, $L, M \in \mathbb{N}$, such that*

$$|T(\varphi)| \leq C \sum_{0 \leq l,m \leq N} [\varphi]_{m,l} \tag{7.1.14}$$

for all φ in $\mathcal{S}$.

The fact that for each tempered distribution T we can find finitely many norms such that (7.1.14) holds, expresses the fact that tempered distributions have *finite order*. Since (7.1.14) implies continuity of T at 0, hence everywhere, we have that (7.1.14) characterizes tempered distributions among linear functionals on $\mathcal{S}$.

PROOF. By assumption, there is $\eta > 0$ such that $|T(\varphi)| \leq 1$ if $d(\varphi, 0) \leq 2\eta$. Suppose that

$$\sum_{0\leq l,m\leq N} [\varphi]_{m,l} \leq \eta,$$

with $N > 0$ to be chosen. Then,

$$\begin{aligned} d(\varphi,0) &= \sum_{m.l\geq 0} \frac{1}{2^{m+l+2}} \frac{[\varphi]_{m.l}}{1+[\varphi]_{m,l}} \\ &\leq \sum_{0\leq m.l\leq N} [\varphi]_{m.l} + \sum_{m>N \text{ or } l>N} \frac{1}{2^{m+l+2}} \\ &\leq \eta + \frac{1}{2^N} \\ &\leq 2\eta \end{aligned}$$

if N is chosen large enough so that $\frac{1}{2^N} \leq \eta$. With this value of N, if $\sum_{0\leq l,m\leq N} [\varphi]_{m,l} \leq \eta$, then $|T(\varphi)| \leq 1$. By homogeneity, (7.1.14) holds with $C = 1/\eta$. □

COROLLARY 7.2. *Let T be a tempered distribution. Then there exists $N \geq 0$ and, for all $R > 0$ there exists $C(R)$ such that, if $\varphi \in C_c^\infty$ is supported in $[-R, R]$, then*

$$|T(\varphi)| \leq C \sum_{m=1}^{N} [\varphi]_{m,0}. \tag{7.1.15}$$

PROOF. Under the hypothesis, $[\varphi]_{m.l} \leq (1 + R^2)^{l/2} [\varphi]_{m,0}$. □

We will now define some basic operations with distributions. The heuristics consist in assuming that the distribution we start with has the form T_f with an especially nice f, to download the difficulties on the testing function φ, and to finally write down a rigorous definition. Let's see first how this works with the derivative.

7.1.2.2. *Derivative of a tempered distribution.* If f is differentiable and not too large at infinity,

$$\begin{aligned} T_{f'}(\varphi) &= \int_{-\infty}^{+\infty} f'(x)\varphi(x)dx \\ &= [f(x)\varphi(x)]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} f(x)\varphi'(x)dx. \end{aligned}$$

DEFINITION 7.1. *Let $T \in \mathcal{S}$ be a distribution. Then, its* **distributional derivative** *T' is defined as*

$$T'(\varphi) := -T(\varphi').$$

Linearity is obvious, and continuity follows from the fact that $\varphi \mapsto \varphi'$ is continuous, by Proposition 7.4. We already had experience of distributional derivatives.

THEOREM 7.3. *Let $\alpha\colon \mathbb{R} \to \mathbb{R}$ be an increasing function such that, for some $n \geq 0$,*

$$\int_{\mathbb{R}} \frac{\alpha(x)dx}{(1+x^2)^{n/2}} < \infty. \tag{7.1.16}$$

Then, $T'_\alpha = T_\mu$, where $\mu \geq 0$ is a Borel measure such that, for some $n \geq 0$,

$$\int_{\mathbb{R}} \frac{d\mu(x)}{(1+x^2)^{n/2}} < \infty. \tag{7.1.17}$$

Also, α differs a.e. by a constant c from the distribution function α_m of μ:

$$T\alpha = T_{\alpha_\mu} + cId.$$

PROOF. We can assume that (7.1.16) and (7.1.17) hold for the same n. Let $\eta \in C_c^\infty$, $\eta \succ [-1,1]$, and let $\eta_R(x) = \eta(x/R)$, so that $\eta_R \succ [-R,R]$. By the Theorem on the weak derivatives of an increasing function, we know that there exists a measure μ such that

$$-\int_{\mathbb{R}} \eta_R(x)\varphi(x)d\mu(x) = \int_{\mathbb{R}} (\varphi\eta_R)'(x)\alpha(x)dx. \tag{7.1.18}$$

for all $\varphi \in \mathcal{S}$. We wish to apply Dominated Convergence on both sides as $R \to \infty$. On the left,

$$|\eta_R(x)\varphi(x)| \leq (1+x^2)^{n/2}|\varphi(x)|/(1+x^2)^{n/2} \leq [\varphi]_{0,n}/(1+x^2)^{n/2},$$

which is integrable with respect to μ.

On the right, for $R \geq 1$ we can estimate

$$\begin{aligned} |(\varphi\eta_R)'(x)| &\leq |\varphi'(x)| + |\varphi(x)| \cdot |\eta'(x/R)| \\ &\leq ([\varphi]_{1,n} + [\eta]_{1,0}[\varphi]_{0,n})/(1+x^2)^{n/2}, \end{aligned}$$

which is integrable with respect to $|\alpha(x)|dx$. We let $R \to \infty$ in (7.1.18) and find

$$-\int_{\mathbb{R}} \varphi(x)d\mu(x) = \int_{\mathbb{R}} \varphi'(x)\alpha(x)dx,$$

as wished. □

Here are some special cases of Theorem 7.3.

(i) If $H_a(x) = \begin{cases} 1 \text{ if } x \geq a \\ 0 \text{ if } x < a \end{cases}$ is the Heaviside function with jump at a, then $H_a' = \delta_a$ is the Dirac mass at a.

(ii) If V is Vitali's function, that $V' = \mu_C$ is the probability measure uniformly distributed on the Cantor set.

It is instructive working out (i) by direct calculation, with $a = 0$. Often, in fact, derivatives of distributions are computed "by hands", without using sophisticated machinery.

$$\begin{aligned} -T'_{H_0}(\varphi) &= T_{H_0}(\varphi') = \int_{-\infty}^{+\infty} \varphi'(x)H_0(x)dx \\ &= \int_0^{\infty} \varphi'(x)dx = \lim_{R\to+\infty}(\varphi(R) - \varphi(0)) = \varphi(0) \\ &= -\delta_0(\varphi). \end{aligned}$$

Exercise 7.3. *Show that $\delta_0'(\varphi) := T'_{\delta_0}(\varphi) = -\varphi'(0)$ is the "dipole" distribution.*

7.1.2.3. *Some more operations. Translation* The calculation

$$\int \tau_a f(x)\varphi(x)dx = \int f(x-a)\varphi(x)dx = \int f(y)\varphi(y+a)dy = \int f(y)\tau_{-a}\varphi(y)dy,$$

suggests to define, for $T \in \mathcal{S}'$,

$$(\tau_a T)(\varphi) := T(\tau_{-a}\varphi). \tag{7.1.19}$$

Exercise 7.4. *Show that $\tau_a T \in \mathcal{S}'$.*

Dilation We consider here the dilation without L^1 normalization, $\tilde{\delta}_t(f)(x) = f(x/t)$

$$\int \tilde{\delta}_t f(x)\varphi(x)dx = \int f(x/t)\varphi(x)dx = \int f(y)t\varphi(ty)dy = \int f(y)\delta_{1/t}\varphi(y)dy,$$

we are led to define, for $T \in \mathcal{S}'$.

$$(\tilde{\delta}_t T)(\varphi) := T(\delta_{1/t}\varphi). \tag{7.1.20}$$

EXERCISE 7.5. *Show that $\tilde{\delta}_t T \in \mathcal{S}'$.*

Multiplication times x Here it is clear that we want to define

$$(M_x T)(\varphi) := T(M_x\varphi),$$

for $T \in \mathcal{S}'$. By iterating and taking linear combinations, we can multiply T times any polynomial.

Multiplication of distributions is a delicate operation, and it is not always defined. For instance, $f(x) = 1/|x|^{1/2}$ defines a distribution T_f, but $f(x)^2 = 1/|x|$ does not. Or think of the Borel measure $T = \delta_0$. If we want to multiply it times some function f, such f can not be defined a.e.: all the action takes place at 0! However, something we can surely do is multiplying a tempered distribution times a Schwartz function.
Multiplication times $\psi \in \mathcal{S}$. It is defined by

$$(M_\psi T)(\varphi) := T(\psi\varphi).$$

Convolution times $\psi \in \mathcal{S}$. If $T = T_f$, the formal calculation is

$$\begin{aligned} (T_f * \psi)(\varphi) &= \int f * \psi(x)\varphi(x)dx = \iint \psi(x-y)f(y)\varphi(x)ddx \\ &= \int f(y)(\int \psi(x-y)\varphi(x)dx)dy = \int f(y)[(U\psi) * \varphi](y)dy, \end{aligned}$$

so we define

$$(T * \psi)(\varphi) := T((U\psi) * \varphi). \tag{7.1.21}$$

We have to check that (i) $(U\psi) * \varphi \in \mathcal{S}$; (ii) $\varphi \mapsto (U\psi) * \varphi$ is continuous on $\mathcal{S}$. Both these properties were proved in Proposition 7.4.

We will see in theorem 7.10 that the convolution of a distribution and a test function is a concrete and smooth object.

7.2. The Fourier transform in $\mathcal{S}(\mathbb{R})$ and in $\mathcal{S}'(\mathbb{R})$

7.2.1. The Fourier transform in $\mathcal{S}(\mathbb{R})$. Let φ be a Schwartz function. Its *Fourier transform* $\mathcal{F}(\varphi) = \widehat{\varphi}\colon \mathbb{R} \to \mathbb{C}$ is defined as

$$\widehat{\varphi}(\omega) := \int_{-\infty}^{+\infty} \varphi(x) e^{-2\pi i \omega x} dx. \tag{7.2.1}$$

LEMMA 7.2. *Suppose $\varphi \in \mathcal{S}$.*

(i) *$\widehat{\varphi}$ is differentiable, and*

$$D\widehat{\varphi}(\omega) = -2\pi i (M_x \varphi)^\wedge(\omega). \tag{7.2.2}$$

(ii) *One has*

$$(D\varphi)^\wedge(\omega) = 2\pi i \omega \widehat{\varphi}(\omega). \tag{7.2.3}$$

PROOF. (i) We can take the derivative inside the integral:

$$\frac{d}{d\omega}\int \varphi(x)e^{-2\pi i\omega x}dx = \int \varphi(x)\frac{\partial}{\partial\omega}\left(e^{-2\pi i\omega x}\right)dx = -2\pi i\int x\varphi(x)e^{-2\pi i\omega x}dx.$$

The hypothesis of the theorem on differentiation under integral sign are satisfied because $|x\varphi(x)|$ is integrable.

(ii) We integrate by parts,

$$\begin{aligned} \int_{-R}^{R} \varphi'(x)e^{-2\pi i\omega x}dx &= \left[\varphi(x)e^{-2\pi i\omega x}\right]_{-R}^{R} - \int_{-R}^{R}(-2\pi i\omega)\varphi(x)e^{-2\pi i\omega x}dx \\ &\to 2\pi i\omega\widehat{\varphi}(\omega) \text{ as } R\to\infty, \end{aligned}$$

since the decay of φ kills the boundary terms, and Dominated Convergence takes care of the integral. □

PROPOSITION 7.6. *The Fourier transform $\mathcal{F}\colon \varphi \mapsto \widehat{\varphi}$ is a linear, continuous map of $\mathcal{S}(\mathbb{R})$ into itself.*

PROOF. Linearity is obvious.

The basic estimate is

$$\begin{aligned} \left|(2\pi i\omega)^n\widehat{\varphi}^{(m)}(\omega)\right| &= \left|\mathcal{F}\left\{\frac{d^n}{dx^n}\left[(-2\pi i x)^m\varphi\right]\right\}(\omega)\right| \\ &\le \int_{\mathbb{R}}\left|\frac{d^n}{dx^n}\left[(-2\pi i x)^m\varphi(x)\right]\right|dx \end{aligned}$$

$$
\begin{aligned}
&\leq C \sum_{l\leq m, j\leq n} \int_{\mathbb{R}} \left| x^l \varphi^{(j)}(x) \right| dx \\
&\leq C \sum_{l\leq m, j\leq n} \int_{\mathbb{R}} \frac{dx}{1+x^2} [\varphi]_{j,l+2}.
\end{aligned}
$$

This implies both that $\widehat{\varphi}$ belongs to $\mathcal{S}(\mathbb{R})$, if φ does, and that $\varphi \mapsto \widehat{\varphi}$ is continuous. □

The proof of the following proposition is conceptually the same as that of the analogous properties of the Fourier coefficients.

PROPOSITION 7.7. *Let* $\varphi, \psi \in \mathcal{S}$.

(i) $\mathcal{F}(\varphi * \psi) = \mathcal{F}(\varphi)\mathcal{F}(\psi)$.

(ii) $\mathcal{F}(\delta_t \varphi)(\omega) = \mathcal{F}(\varphi)(t\omega)$, *and* $\mathcal{F}(\tau_a \varphi)(\omega) = e^{-2\pi i a\omega}\mathcal{F}(\varphi)(\omega)$.

(iii) $\mathcal{F}(U\varphi) = \mathcal{F}(\varphi)(-\omega) = (U\mathcal{F}(\varphi))(\omega)$.

(iv) $\mathcal{F}(\tau_a \varphi)(\omega) = e^{-2\pi i a\omega}\mathcal{F}\varphi(\omega)$, *where* $\tau_a\varphi(x) = \varphi(x-a)$.

(v) *Let* $e_a(x) = e^{2\pi i a x}$: $\mathcal{F}(e_a\varphi) = \tau_a(\mathcal{F}\varphi)$.

Items (iv), (v) say that $\mathcal{F}$ intertwines *translation* and *modulation*.

PROOF. (i) The basic ingredients are the sum-to-product (group homomorphism) property of the exponential (from first to second line), Fubini Theorem (from second to third), and invariance of Lebesgue measure under translations (from third to fourth):

$$
\begin{aligned}
\mathcal{F}(\varphi * \psi)(\omega) &:= \int \left(\int \varphi(x-y)\psi(y)dy \right) e^{-2\pi i x\omega} dx \\
&= \int \left(\int \varphi(x-y)\psi(y)dy \right) e^{-2\pi i (x-y)\omega} e^{-2\pi i y\omega} dx \\
&= \int \left(\int \varphi(x-y) e^{-2\pi i (x-y)\omega} dx \right) e^{-2\pi i y\omega} \psi(y) dy \\
&= \int \left(\int \varphi(z) e^{-2\pi i z\omega} dz \right) e^{-2\pi i y\omega} \psi(y) dy \\
&= \mathcal{F}(\varphi)\mathcal{F}(\psi).
\end{aligned}
$$

(ii) For the first equality,

$$
\begin{aligned}
\mathcal{F}(\delta_t\varphi)(\omega) &= \int \varphi\left(\frac{x}{t}\right) \frac{1}{t} e^{-2\pi i x\omega} dx \\
&= \int \varphi(z) e^{-2\pi i t z\omega} dz = \mathcal{F}(\varphi)(t\omega).
\end{aligned}
$$

The second equality and (iii-v) are easy and they are left to the reader. □

For $t > 0$, define $N_t(x) = (\delta_{\sqrt{t}} N_1)(x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}$, the *heat kernel*.[1]

LEMMA 7.3. *We have* $\widehat{N_t}(\omega) = e^{-2\pi^2 t\omega^2}$. *Also,* $\mathcal{F}(\widehat{N_t})(x) = N_t(x)$.

PROOF. By Proposition 7.7 (ii), it suffices to show the equality for $t = 1$. Observe that $N_1 \in \mathcal{S}$ and it satisfies the differential equation

$$N_1' + xN_1 = 0.$$

Taking Fourier transforms of both sides, and using Lemma 7.2, we have

$$0 = 2\pi i\omega \widehat{N}_1(\omega) - \frac{1}{2\pi i}\widehat{N}_1'(\omega) = \frac{i}{2\pi}(4\pi^2\omega\widehat{N}_1(\omega) + \widehat{N}_1'(\omega)). \tag{7.2.4}$$

The solution of (7.2.4) with the initial condition $\widehat{N}_1(0) = \int \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1$ is $\widehat{N}_1(\omega) = e^{-2\pi^2\omega^2}$.

The second relation follows from the same argument: $\widehat{N_t}$ satisfies (7.2.4) with $\widehat{N_t}(0) = 1$, hence, $\psi = \mathcal{F}(\widehat{N_t})$ satisfies $\psi' + x\psi = 0$, with $\int_{\mathbb{R}} \psi(x)dx = 0$. $\square$

The following lemma is a variation on the theme of approximations of identity, and so is its proof.

LEMMA 7.4. *Let* $f \in C_b(\mathbb{R})$ *be uniformly continuous. Then,*

$$\lim_{t\to 0} N_t * f(x) = f(x) \tag{7.2.5}$$

uniformly on $\mathbb{R}$.

PROOF. Fix $\epsilon > 0$.

$$\begin{aligned} |N_t * f(x) - f(x)| &= \left| \int_{-\infty}^{+\infty} [f(x-y) - f(x)] N_t(y) \right| \\ &\leq \sup_{|y|\leq\delta} |f(x-y) - f(x)| + 2\|f\|_{L^\infty} \int_{|y|\geq\delta} N_t(y)dy \\ &= \sup_{|y|\leq\delta} |f(x-y) - f(x)| + 2\|f\|_{L^\infty} \int_{|z|\geq\frac{\delta}{\sqrt{t}}} N_1(z)dz. \end{aligned}$$

[1]We dilate by $\sqrt{t}$ because (i) this is what we need when studying the *heat equation*; (ii) this way we obtain the densities of a *Brownian motion*. Wait! What does heat diffusion have to do with Brownian motion? The connection was first made by Einstein in 1905, and it has since become a common theme in the theory of stochastic processes and in PDEs. See e.g. Einstein and Brownian Motion for a short historical account. A more mathematical reason to use $\sqrt{t}$ is that in this way (iii) $N_s * N_t = N_{s+t}$, i.e. that $\{N_t : t > 0\}$ is a *convolution semigroup*. Indeed, (iii) is strictly related with (i), (ii), and with physics, as well as with probability. Probabilistically, the semigroup property is closely linked with the *Markov property* of Brownian motion.

Choose first $\delta > 0$ such that the first summand is less that ϵ, then use Dominated Convergence on the second (as $t \to \infty$). $\square$

THEOREM 7.4 (Fourier inversion formula). *Let $\varphi \in \mathcal{S}$. Then,*

$$\varphi(x) = \int_{-\infty}^{+\infty} \widehat{\varphi}(\omega) e^{2\pi i \omega x} d\omega = (U \circ \mathcal{F} \circ \mathcal{F})(\varphi)(x). \tag{7.2.6}$$

In particular, $\mathcal{F} : \mathcal{S} \to \mathcal{S}$ is a homeomorphism.

The Holy Graal of Fourier theory consists in couples of topological function spaces X, Y such that $\mathcal{F}$ maps X homeomorphically onto Y. We have met one such instance when discussing Fourier series. For the Fourier transform $\mathcal{F}$ we have the following:

- $\mathcal{F}$ maps $\mathcal{S}$ homeomorphically onto itself and, dually, the same does with $\mathcal{S}'$ (this is the content of theorem 7.4);
- $\mathcal{F}$ maps $L^2(\mathbb{R})$ isometrically onto itself (Plancherel formula, which will appear shortly);
- $\mathcal{F}$ bijectively maps the "cone" of the positive, finite Borel measures on $\mathbb{R}$ onto the "cone" of the "continuous, positive definite, bounded functions" on $\mathbb{R}$. This last statement is *Bochner's theorem*, which plays a prominent role in Fourier analysis, probability, and operator theory (especially operators' semigroups). If you are interested in the precise statement, and references for the proof, you might start here Yitao Lei, Bochner's Theorem on the Fourier Transform on $\mathbb{R}$.
- Building on L^2, you can construct other nice spaces having norms defined by derivatives (Sobolev spaces).

PROOF. Once we have proved (7.2.6), we have that $\mathcal{F}$ has a two sided inverse, $\mathcal{F}^{-1} = U \circ \mathcal{F}$ (see also (iii) in Proposition 7.7), which is continuous.
We would like to compute:

$$\begin{aligned}\int_{\mathbb{R}} \widehat{\varphi}(\omega) e^{2\pi i \omega x} d\omega &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \varphi(y) e^{-2\pi i \omega y} dy \right) e^{2\pi i \omega x} d\omega \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} e^{2\pi i \omega (x-y)} d\omega \right) \varphi(y) dy,\end{aligned}$$

but we reach a dead end since the inner integral diverges; we were not careful in verifying the hypothesis of Fubini theorem, ω being the troubling coordinate.

But we can downsize the integrand by inserting the heat kernel's Fourier transform. For $t > 0$, use Lemma 7.3 in the second equality:

$$\begin{aligned} I_t(x) &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} e^{2\pi i\omega(x-y)} e^{-2\pi^2 t\omega^2} d\omega \right) \varphi(y) dy \\ &= \int_{\mathbb{R}} N_t(x-y)\varphi(y)dy \\ &\to f(x) \text{ as } t \to 0, \end{aligned}$$

by Lemma 7.4.

In the other direction,

$$\begin{aligned} I_t(x) &= \int_{\mathbb{R}} \widehat{\varphi}(\omega) e^{2\pi i\omega x} e^{-2\pi^2 t\omega^2} d\omega \\ &\to \int_{\mathbb{R}} \widehat{\varphi}(\omega) e^{2\pi i\omega x} d\omega \end{aligned}$$

as $t \to 0$ by Dominated Convergence.

The hypothesis of Fubini theorem were satisfied, since

$$F(y,\omega) = e^{2\pi i\omega(x-y)} e^{-2\pi^2 t\omega^2} \varphi(y)$$

is integrable with respect to $dy d\omega$. □

In the following calculation, Fubini can be applied. If $\varphi, \psi \in \mathcal{S}$:

$$\int \varphi(x)\overline{(\mathcal{F}\psi)(x)}dx = \iint \varphi(x)\overline{\psi(y)} e^{2\pi i xy} dx dy = \int (\mathcal{F}^{-1}\varphi)(y)\overline{\psi(y)}dy,$$

i.e.

$$\langle \mathcal{F}\psi | \varphi \rangle_{L^2} = \langle \psi | F^{-1}\varphi \rangle_{L^2} : \tag{7.2.7}$$

the inverse Fourier transform is the *Hilbert space adjoint* of the Fourier transform.

THEOREM 7.5 (Plancherel formula in $\mathcal{S}$). *For φ in $\mathcal{S}$, we have*

$$\int_{\mathbb{R}} |\varphi(x)|^2 dx = \int_{\mathbb{R}} |\widehat{\varphi}(\omega)|^2 d\omega. \tag{7.2.8}$$

PROOF. Applying (7.2.7) to $\varphi = \mathcal{F}\eta$ we obtain

$$\langle \mathcal{F}\psi | \mathcal{F}\eta \rangle_{L^2} = \langle \psi | (\mathcal{F}^{-1} \circ \mathcal{F})\eta \rangle_{L^2} = \langle \psi | \eta \rangle_{L^2}, \tag{7.2.9}$$

which is Plancherel's formula when $\psi = \eta$. □

Formal manipulation The inversion and the Plancherel formulas could have deduced with no approximation of the identity if had postulated that:

$$\int_{\mathbb{R}} e^{2\pi i\omega(x-y)} d\omega = \delta_x(y). \tag{7.2.10}$$

In fact, (i) on $\mathbb{T}$, the analogous formula expresses orthogonality of imaginary exponentials; (ii) it is what we need to carry the formal calculations to conclusion. Let's see how this works for the inversion formula:

$$\begin{aligned}
\int_{\mathbb{R}} \widehat{\varphi}(\omega) e^{2\pi i\omega x} d\omega &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \varphi(y) e^{-2\pi i\omega y} dy \right) e^{2\pi i\omega x} d\omega \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} e^{2\pi i\omega(x-y)} d\omega \right) \varphi(y) dy \\
&= \int_{\mathbb{R}} \delta_x(y) \varphi(y) dy \\
&= \varphi(x).
\end{aligned}$$

Now, it is not in our power to "postulate" pieces of a calculation. Nonetheless, physicists and electrical engineers have trusted for a long time that expressions like (7.2.10), if used to work on mathematical objects whose meaning is clear (a particle, a signal), will lead to the correct result.

As mathematicians, we have a choice: using formal formulas as a magic to speedily arrive at the end of a calculation, knowing that later we have to put the same calculation on solid ground by less magic means; or we can extend the theory to include the formula. In the case of (7.2.10), we will see shortly that it can be given a solid meaning in terms of tempered distributions.

7.2.2. Extension of the Fourier transform to L^1 and L^2.

7.2.2.1. *Fourier transform in L^1.* For $f \in L^1(\mathbb{R})$, the Fourier transform is well defined pointwise,

$$\widehat{f}(\omega) = \int_{\mathbb{R}} f(x) e^{-2\pi i\omega x} dx, \ |\widehat{f}(\omega)| \le \|f\|_{L^1}. \tag{7.2.11}$$

THEOREM 7.6. (i) *If $f \in L^1$, then $\widehat{f}$ is continuous.*

(ii) *[Riemann-Lebesgue lemma] If $f \in L^1$, then* $\lim_{\omega \to \pm\infty} \widehat{f}(\omega) = 0$.

(iii) *If $f \in L^1$ and $f \ge 0$, then $\|\widehat{f}\|_{L^\infty} = \|f\|_{L^1}$.*

PROOF. (i) Fix $\epsilon > 0$. If $\varphi \in \mathcal{S}$, then,

$$|\widehat{f}(\omega)| \le |\widehat{f}(\omega) - \widehat{\varphi}(\omega)| + |\widehat{\varphi}(\omega)|$$

$$\leq \|f - \varphi\|_{L^1} + |\widehat{\varphi}(\omega)|.$$

Choose first φ sch that the first summand is less than ϵ, then $R > 0$ large enough so that $|\widehat{\varphi}(\omega)| \leq \epsilon$ if $|\omega| > R$.

(ii) With $\epsilon > 0$, h real, and $\varphi \in \mathcal{S}$, estimate

$$\begin{aligned} |\widehat{f}(\omega+h) - \widehat{f}(\omega)| &\leq |\widehat{f}(\omega+h) - \widehat{\varphi}(\omega+h)| + |\widehat{\varphi}(\omega+h) - \widehat{\varphi}(\omega)| \\ &\quad + |\widehat{\varphi}(\omega) - \widehat{f}(\omega)| \\ &\leq 2\|f-\varphi\|_{L^1} + |\widehat{\varphi}(\omega+h) - \widehat{\varphi}(\omega)|. \end{aligned}$$

Choose first such that the first summand is less that ϵ, then $\delta > 0$ such that for $|h| < \delta$ we have that the second summand is less that ϵ, too.

(iii) This is obvious, but useful.

Having proven that $\mathcal{F} : L^1(\mathbb{R}) \to C_0(\mathbb{R})$ is bounded, it is natural to ask whether it is surjective. The answer is negative.

Proposition 7.8. *There exists functions g in $C_0(\mathbb{R})$ which are not the Fourier transform of any f in $L^1(\mathbb{R})$.*

Proof. If $\mathcal{F}\colon L^1(\mathbb{R}) \to C_0(\mathbb{R})$ were onto, since it is injective and bounded, by the inverse mapping property for operators it would have bounded inverse $\mathcal{F}^{-1}\colon C_0(\mathbb{R}) \to L^1(\mathbb{R})$. To contradict it, we exhibit $\{g_n\}_{n\geq 1}$ in $C_0(\mathbb{R}) \cap L^1(\mathbb{R})$ with norm bounded by 1, but with $\|\mathcal{F}^{-1} g_n\|_{L^1} \to \infty$ as $n \to \infty$.

Define

$$g_n = \chi_{[-n/2,n/2]} * \chi_{[-1/2,1/2]}.$$

Using the definition of convolution, we see that $\|g_n\|_{L^\infty} = 1$, and $\|g_n\|_{L^1} \leq n$ by Young's inequality. On the anti-Fourier side,

$$\begin{aligned} \mathcal{F}^{-1} g_n(x) &= \widehat{\chi}_{[-n/2,n/2]}(x)\widehat{\chi}_{[-1/2,1/2]}(x) \\ &= \frac{\sin(\pi n x)}{\pi x} \cdot \frac{\sin(\pi x)}{\pi x}, \end{aligned}$$

hence,

$$\begin{aligned} \int_{\mathbb{R}} |\mathcal{F}^{-1} g_n(x)| dx &\geq \frac{2}{\pi} \int_{|x|\leq 1/2} \left| \frac{\sin(\pi n x)}{\pi x} \right| dx \\ &= \frac{2}{\pi} \int_{|y|\leq n/2} \left| \frac{\sin(\pi y)}{\pi y} \right| dy \\ &\to \frac{2}{\pi} \int_{\mathbb{R}} \left| \frac{\sin(\pi y)}{\pi y} \right| dy \\ &= \infty \end{aligned}$$

as $n \to \infty$. □

EXERCISE 7.6. *Show that, as stated in the proof,* $\int_{\mathbb{R}} \left|\frac{\sin(\pi y)}{\pi y}\right| dy = \infty$.

□

7.2.2.2. *Fourier transform in* L^2. For $f \in L^2(\mathbb{R})$, we need to make sense of the Fourier transform of f, since the integral (7.2.11) might diverge. We can use density of $\mathcal{S}$ in L^2 and Plancherel identity (7.2.8). More explicitly, consider an approximation of f in L^2 norm by means of Schwartz functions φ_j, $\|\varphi_j - f\|_{L^2} \to 0$. Then, for all $\epsilon > 0$,

$$\|\widehat{\varphi}_{n+j} - \widehat{\varphi}_n\|_{L^2} = \|\varphi_{n+j} - \varphi_n\|_{L^2} \leq \epsilon$$

if $n > n(\epsilon)$ and $j \geq 1$. Let h be the limit of $\{\widehat{\varphi}_j\}$ in L^2. We define $\widehat{f} = h$ to be the *Fourier transform* of f.

THEOREM 7.7. *The Fourier transform on* $\mathcal{S}$ *extends uniquely to a linear isometry* $\mathcal{F} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$.

Indeed, it is nice having a uniform, linear procedure to approximate $f \in L^2$ by functions whose Fourier transform can be computed explicitly. Here is a way for doing so. For $f \in L^2$, $f_t(x) = e^{-2\pi^2 x^2 t} f(x)$ defines, as $t \to 0$, an approximation of f in L^2 by L^1 function, whose Fourier transform can be computed,

$$\widehat{f_t} = N_t * \widehat{f} \in C_0 \cap L^2.$$

We have then that $\|\widehat{f} - \widehat{f_t}\|_{L^2} \to 0$ as $t \to 0$.

EXERCISE 7.7. *Show that* $\lim_{t\to 0} \|f - f_t\|_{L^2} = 0$.

7.2.3. Fourier transforms of tempered distributions. The *raison d'être* of the tempered distributions is that on them we can perform (i) derivatives; (ii) Fourier transforms; and that (iii) they contain Borel measures (with controlled growth: some price has to be paid somewhere). We see now how the Fourier transform of a distribution is defined. We start from a formal calculation.

$$\begin{aligned} \int \widehat{f}(\omega)\varphi(\omega)d\omega &= \iint f(x)e^{-2\pi i\omega x}\varphi(\omega)dxd\omega \\ &= \int f(x)\widehat{\varphi}(x)dx. \end{aligned}$$

We make it into the definition of the Fourier transform of $T \in \mathcal{S}'$:

$$\widehat{T}(\varphi) := T(\widehat{\varphi}). \tag{7.2.12}$$

Since $\widehat{T} = T \circ \mathcal{F}$ is the composition of continuous operators, it defines a tempered distribution.

In order to be of some use, the definition must be computable in simple cases, and it has to retain some properties of the *bona fide* Fourier transform. Let's compute the Fourier transform of a Dirac delta.

$$\begin{aligned}\widehat{\delta_a}(\varphi) &= \delta_a(\widehat{\varphi}) = \widehat{\varphi}(a)\\ &= \int_{-\infty}^{+\infty} e^{-2\pi i a x}\varphi(x)dx.\end{aligned}$$

In this distributional sense,

$$\widehat{\delta_a} = M_{e^{-2\pi i a x}} \equiv e^{-2\pi i a x}, \tag{7.2.13}$$

which provides a graphic interpretation of the imaginary exponentials in $\mathbb{R}$.

PROPOSITION 7.9. *Let $T \in \mathcal{S}'$ and $\psi \in \mathcal{S}$.*

(i) $\widehat{T'} = 2\pi i \omega \widehat{T} = 2\pi i M_\omega \widehat{T}$.

(ii) $\widehat{T'} = -2\pi i (M_x T)^\wedge$.

(iii) $\widehat{\psi * T} = \widehat{T} \circ M_{\widehat{\psi}}$.

PROOF. (i) For $\varphi \in \mathcal{S}$, we have:

$$\begin{aligned}\widehat{T'}(\varphi) &= T'(\widehat{\varphi}) = -T(D\widehat{\varphi})\\ &= -T(-2\pi i M_\omega \widehat{\omega}) = 2\pi i (M_\omega T)(\widehat{\omega})\\ &= 2\pi i M_\omega (T(\widehat{\omega})) = 2\pi i M_\omega \widehat{T}(\varphi).\end{aligned}$$

(ii) Exercise.

(iii) For $\varphi \in \mathcal{S}$, we have:

$$\begin{aligned}\widehat{\psi * T}(\varphi) &= (\psi * T)(\widehat{\varphi})\\ &= T((U\psi) * \mathcal{F}(\varphi))\\ &= T(\mathcal{F}(\mathcal{F}(\psi)\varphi))\\ &= \widehat{T}((\mathcal{F}(\psi)\varphi))\\ &= \widehat{T} \circ M_{\widehat{\psi}}(\varphi).\end{aligned}$$

□

Again on formal calculations with (7.2.10) We have seen in (7.2.13) that the Fourier transform of δ_a is the imaginary exponential $\omega \mapsto e^{-2\pi i a \omega} =: e_{-a}(\omega)$. The Fourier inversion theorem can be read in a similar, distributional sense:

$$\delta_x(\varphi) = \varphi(x)$$

$$\begin{aligned} &= \int e^{2\pi i\omega x}\widehat{\varphi}(\omega)d\omega \\ &= e_x(\widehat{\varphi}) \\ &= \widehat{e_x}(\varphi), \end{aligned}$$

i.e.

$$\widehat{e_x} = \delta_x. \tag{7.2.14}$$

All equalities are definitions, but for the one from first to second line. This is *not* a proof! The inverse Fourier transform theorem has to be proved in some way. It is a translation of the theorem in the language of distributions.

We end with an *alternative proof of Plancherel formula*,

$$\int_{\mathbb{R}} |\varphi(y)|^2 dy = \int_{\mathbb{R}} |\widehat{\varphi}(\omega)|^2 d\omega,$$

which is based on the heat kernel.

Proof. Here, too, we make a false start and try to learn from it.

$$\begin{aligned} \|\widehat{\varphi}\|_{L^2}^2 &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \varphi(x)e^{-2\pi i\omega x}dx\right)\left(\int_{\mathbb{R}} \overline{\varphi(y)}e^{2\pi i\omega y}dy\right)d\omega \\ &= \iint_{\mathbb{R}^2}\left(\int_{\mathbb{R}} e^{2\pi i\omega(y-x)}d\omega\right)\varphi(x)\overline{\varphi(y)}dxdy, \end{aligned}$$

and the inner integral diverges. For $t > 0$, set

$$\begin{aligned} J_t &= \iint_{\mathbb{R}^2}\left(\int_{\mathbb{R}} e^{2\pi i\omega(y-x)}e^{-2\pi^2 t\omega^2}d\omega\right)\varphi(x)\overline{\varphi(y)}dxdy \\ &= \iint_{\mathbb{R}^2} N_t(x-y)\varphi(x)\overline{\varphi(y)}dxdy \\ &= \int_{\mathbb{R}} N_t * \varphi(y)\overline{\varphi(y)}dy \\ &\to \int_{\mathbb{R}} |\varphi(y)|^2 dy. \end{aligned}$$

The limit is justified because $N_t*\varphi$ converges uniformly to φ, and $|N_t*\varphi(x)| \leq \|N_t\|_{L^1}\|\varphi\|_{L^1} = \|\varphi\|_{L^1}$, then we can apply Dominated Convergence.

On the other hand,

$$\begin{aligned} J_t &= \int_{\mathbb{R}} |\widehat{\varphi}(\omega)|^2 e^{-2\pi^2 t\omega^2}d\omega \\ &\to \int_{\mathbb{R}} |\widehat{\varphi}(\omega)|^2 d\omega \end{aligned}$$

as $t \to 0$, by Dominated Convergence. □

Having the Fourier inversion theorem, we can deduce Plancherel's formula. We start with a calculation in which Fubini can be applied, if $\varphi, \psi \in \mathcal{S}$:

$$\int \varphi(x)\overline{(\mathcal{F}\psi)(x)}dx = \iint \varphi(x)\overline{\psi(y)}e^{2\pi ixy}dxdy = \int (\mathcal{F}^{-1}\varphi)(y)\overline{\psi(y)}dy,$$

i.e.

$$\langle \mathcal{F}\psi|\varphi\rangle_{L^2} = \langle \psi|\mathcal{F}^{-1}\varphi\rangle_{L^2} \tag{7.2.15}$$

(the inverse Fourier transform is the *Hilbert space adjoint* of the Fourier transform). Applying (7.2.15) to $\varphi = \mathcal{F}\eta$ we obtain

$$\langle \mathcal{F}\psi|\mathcal{F}\eta\rangle_{L^2} = \langle \psi|F^{-1}\circ\mathcal{F}\eta\rangle_{L^2} = \langle \psi|\eta\rangle_{L^2}, \tag{7.2.16}$$

which is Plancherel's formula.

7.3. The support of a distribution

To provide motivation for the definition to follow, consider the distribution $T(\varphi) = \int_0^1 \varphi(x)dx - \varphi(0)$ (which is in fact a signed measure). Intuition suggests that T is supported on $[0,1]$. Evidence for the intuition is provided by the fact that, if φ is a Schwartz function (more generally, a continuous function) with support on $\mathbb{R}\setminus[0,1]$, then $T(\varphi) = 0$. The set $\mathbb{R}\setminus[0,1]$ is largest with this property.

Let T be a tempered distribution. An open subset A of the real line is an *annihilation set* for T if $T(\varphi) = 0$ whenever φ has support in A. The *support* $\mathrm{supp}(T)$ of T is the complement of the union of all its annihilation sets. It is not a priori clear that the union of annihilation sets is itself an annihilation set, so we prove it here below.

Lemma 7.5. *If $\{A_i\}_{i\in I}$ is a family of annihilation sets for the distribution T, then $A = \bigcup_{i\in I} A_i$ is an annihilation set for T. In particular, $\mathbb{R}\setminus \mathrm{supp}(T)$ is an annihilation set for T.*

Proof. It suffices to show that for any $\varphi \in C_c^\infty(\mathbb{R})$ with $\mathrm{supp}(\varphi) \subset A$ we have $T(\varphi) = 0$. In fact, we saw in proposition 7.3 that any ψ in $\mathcal{S}$ can be approximated in $\mathbb{S}$ by a sequence $\{\varphi_n\}$ in $C_c^\infty(\mathbb{R})$, and the proof shows that we can take $\mathrm{supp}(\varphi_n) \subseteq \mathrm{supp}(\psi)$. So, if $\mathrm{supp}(\psi) \subseteq A$, by the continuity of T,

$$T(\psi) = \lim_{n\to\infty} T(\varphi_n) = 0.$$

Since the support of φ is compact, $\mathrm{supp}(\varphi) \subset \cup_{l=1}^n A_{i_l}$, we can apply the smooth partition of unity and find $h_l \prec A_l$ with $h_1 + \cdots + h_n = 1$ on $\mathrm{supp}(\varphi)$.

We have then,

$$\begin{aligned} T(\varphi) &= T\left(\sum_{l=1}^{n} h_l\varphi\right) = \sum_{l=1}^{n} T\left(h_l\varphi\right) \\ &= 0 \quad , \end{aligned}$$

because $h_l\varphi$ is supported in A_{i_l}, which is an annihilation set for T. □

As a consequence, the complement of the support of T is maximal with the property of being an annihilation set.

Corollary 7.3. *If T is a tempered distribution and A is an annihilation set for T, then $A \subseteq \mathbb{R} \setminus supp(T)$.*

Proof. If such were not the case, then $A \cup (\mathbb{R} \setminus \text{supp}(T))$ would be an annihilation set for T which is not contained in $\mathbb{R}\setminus\text{supp}(T)$, which contradicts the definition of support. □

For compactly supported, tempered distributions T, we have a simpler bound for $|T(\varphi)|$.

Lemma 7.6. *Suppose $T \in \mathcal{S}'$ has compact support, contained in $[-P, P]$. Then, there exist $\Gamma > 0$ and $N \geq 0$ such that*

$$|T(\varphi)| \leq \Gamma \sum_{0 \leq l \leq N} [\varphi]_{l,0} \tag{7.3.1}$$

for all $\varphi \in \mathcal{S}$.

The minimum value of N for which an estimate like (7.3.1) holds is called the *order* of the compactly supported distribution T.

Proof. Let $\eta \succ [-P-1, P+1]$, so that $\text{supp}(\varphi(1-\eta)) \subset \mathbb{R} \setminus [-P, P]$, and let $N \geq 0$ and $C > 0$ such that

$$|T(\varphi)| \leq C \sum_{0 \leq l,m \leq N} [\varphi]_{l,m}.$$

Then,

$$T(\varphi) = T(\varphi\eta) + T(\varphi(1-\eta)) = T(\varphi\eta) \leq C \sum_{0 \leq l,m \leq N} [\varphi\eta]_{l,m}.$$

We compute

$$\begin{aligned}
[\varphi\eta]_{l,m} &= \sup_{x\in\mathbb{R}}(1+x^2)^{m/2}|(\varphi\eta)^{(l)}(x)| \\
&\leq \sum_{j=0^l} C(l,j)\sup_{x\in\mathbb{R}}|\varphi^{(j)}(x)|\cdot\sup_{x\in\mathbb{R}}\sup_{x\in\mathbb{R}}(1+x^2)^{m/2}|\eta^{(l-j)}(x)| \\
&\leq \sum_{j=0^l} C(l,j)[\varphi]_{j,0}[\eta]_{l-j,m}.
\end{aligned}$$

Summing over l, m and taking the maximum of the constants, we obtain (7.3.1). □

7.3.1. Distributions supported at the origin. An especially interesting class is that of the tempered distributions having compact support. We consider here a rather extreme case.

THEOREM 7.8. *Let T be a tempered distribution which is supported on $\{0\}$. Then, there exist an integer $N \geq 0$ and scalars $a_0, \dots, a_N$ such that*

$$T(\varphi) = \sum_{l=0}^{N} a_l \varphi^{(l)}(0). \tag{7.3.2}$$

PROOF. Suppose T satisfies the assumptions in the theorem, and observe that, by density of $C_c^\infty(\mathbb{R})$ in $\mathcal{S}$, it suffices to show that (7.3.2) holds for compactly supported φ's. Let N be a positive integer such that (7.3.1) holds. Suppose that φ is supported on $[-R, R]$, .

Let $\eta_R \succ [-R-1, R+1]$. Then, using Taylor's formula,

$$\begin{aligned}
\varphi(x) &= \eta_R(x)\varphi(x) = \eta_R(x)\left(\sum_{j=0}^{N}\frac{\varphi^{(j)}(0)}{j!}x^j + R_N(x)\right) \\
&= \sum_{j=0}^{N}\frac{\varphi^{(j)}(0)}{j!}x^j\eta_R(x) + R_N(x)\eta_R(x) \\
&= \sum_{j=0}^{N}\frac{\varphi^{(j)}(0)}{j!}x^j\eta_R(x) + h(x),
\end{aligned}$$

where the remainder R_N is C^∞, thus $h \in C_c^\infty$ is supported in $[-R, R]$ and $h(0) = \cdots = h^{(N)}(0) = 0$.

The terms in the sum are easily handled,

$$
\begin{aligned}
T\left(\sum_{j=0}^{N}\frac{\varphi^{(j)}(0)}{j!}x^j\eta_R(x)\right) &= \sum_{j=0}^{N}\frac{\varphi^{(j)}(0)}{j!}T\left(x^j\eta_R(x)\right)\\
&= \sum_{j=0}^{N}\frac{\varphi^{(j)}(0)}{j!}a_{R,j},
\end{aligned}
$$

where $A_{R,j} = T\left(x^j\eta_R(x)\right)$. Actually, these scalars are independent of R. Since $x^j(\eta_{R+1}(x)-\eta_R(x))$ vanishes on $[-R-1,R+1]$ and T is supported at $\{0\}$, $T\left(x^j\eta_{R+1}(x)\right) = T\left(x^j\eta_R(x)\right)$.

The theorem is proved, then, if we show that $T(h) = 0$, which is what we do next. Pick $\psi \in C_c^\infty$, $\psi \succ [-1,1]$, and let $\psi_n(x) = \psi(nx)$. We have $T(h) = T(h\psi_n) + T(h(1-\psi_n)) = T(h\psi_n)$ because T is supported at the origin and $h(1-\psi_n)$ is supported on $\mathbb{R}\setminus\{0\}$. We want to estimate

$$
|T(h)| = |T(h\psi_n)| \leq C\sum_{m=0}^{N}[h\psi_n]_{m,0} \tag{7.3.3}
$$

$$
= C\sum_{m=0}^{N}\sup_{|x|\leq\frac{1}{n}}\left|(h\psi_n)^{(m)}(x)\right|, \tag{7.3.4}
$$

where we used that ψ_n is supported on $[-1/n,1/n]$. We estimate each summand,

$$
\left|(h\psi_n)^{(m)}(x)\right| = \left|\sum_{j=0}^{m}C(m,j)h^{(j)}(x)\psi_n^{(m-j)}(x)\right| \tag{7.3.5}
$$

$$
= \left|\sum_{j=0}^{m}C(m,j)h^{(j)}(x)n^{m-j}\psi^{(m-j)}(nx)\right|. \tag{7.3.6}
$$

Since $h(0) = \cdots = h^{(N)}(0) = 0$, for $|x| \leq \frac{1}{n}$,

$$
|h^{(j)}(x)| \leq A_j|x|^{N+1-j} \leq A_j\frac{1}{n^{N+1-j}},
$$

where the positive numbers A_j depend on h. Putting this in (7.3.5), we have

$$
\left|(h\psi_n)^{(m)}(x)\right| \leq \frac{B_m}{n}
$$

for some $B_m > 0$ which depends on h, and by (7.3.3),

$$
|T(h)| \leq \frac{C}{n}.
$$

As $n\to\infty$, we obtain $T(h) = 0$ as wished. □

7.3.2. Positive distributions having compact support. A distribution $T \in \mathcal{S}'$ is *positive* if and only if $T(\varphi) \geq 0$ for all $\varphi \geq 0$ in $\mathcal{S}$.

THEOREM 7.9. *Let T be a positive, compactly supported tempered distribution on $\mathcal{S}$. Then, there exists a positive, compactly supported Borel measure $\mu \geq 0$ such that:*

$$T(\varphi) = \int_{\mathbb{R}} \varphi d\mu \text{ for all } \varphi \in \mathcal{S}. \tag{7.3.7}$$

Moreover, μ and T have the same support.

Here, $\mathbb{R} \setminus \operatorname{supp}(\mu)$ is the largest *open* set E such that $\mu(E) = 0$.

EXERCISE 7.8. *State and prove the converse statement of theorem 7.9.*

PROOF. If T has support in $[-R, R]$, its action can be studied on functions φ supported in $[-R-1, R+1]$. The idea is one we have used before: if we fix $[-R-1/2, R+1/2] \prec \eta \prec (-R-1, R+1)$ and if $\varphi \in \mathcal{S}$, then $\varphi\eta$ is supported on $[-R-1, R+1]$ and

$$T(\varphi) = T(\varphi\eta) + T(\varphi(1-\eta)) = T(\varphi\eta)$$

because $\varphi(1-\eta)$ vanishes on $(-R-1/2, R+1/2) \supset \operatorname{supp}(T)$.
Step I. We show that there is C such that $|T(\varphi)| \leq C\|\varphi\|_{L^\infty}$ if $\operatorname{supp}(\varphi) \subseteq [-R-1, R+1]$. It suffices to show it for real valued φ, because $|T(\alpha+i\beta)| \leq \sqrt{|T(\alpha)|^2 + |T(\beta)|^2}$. Let $[-R-1, R+1] \prec \psi$, so that $|\varphi| \leq \|\varphi\|_{L^\infty}\psi$, i.e. $-\|\varphi\|_{L^\infty}\psi \leq \varphi \leq \|\varphi\|_{L^\infty}\psi$. By positivity of T, $-\|\varphi\|_{L^\infty}T(\psi) \leq T(\varphi) \leq \|\varphi\|_{L^\infty}T(\psi)$, and we can let $C = T(\psi)$ (in the real case).

The space $C_c^\infty(\mathbb{R})$ is dense in $C_c(\mathbb{R})$ with respect to the uniform norm $\|\cdot\|_{L^\infty}$, hence T extends to $\tilde{T}\colon C_c(\mathbb{R}) \to \mathbb{C}$:

$$\tilde{T}(f) = \lim_{n\to\infty} T(\varphi_n), \tag{7.3.8}$$

if $C_c^\infty \ni \varphi_n \to f \in C_c$ in the uniform norm.
Step II. By Riesz representation theorem, there is a positive Borel measure μ on $(-R-1, R+1)$ such that

$$\tilde{T}(h) = \int_{\mathbb{R}} h d\mu \tag{7.3.9}$$

for all $h \in C_c(-R-1, R+1)$. In particular,

$$T(\varphi) = \tilde{T}(\varphi) = \int_{\mathbb{R}} \varphi d\mu \tag{7.3.10}$$

holds for all φ in $C_c^\infty(-R-1, R+1)$. Actually the representation extends to all $\varphi \in \mathcal{S}$ by the reasoning above.

We finally have to prove that the supports of μ and T coincide. Suppose φ has support in $\mathbb{R} \setminus \operatorname{supp}(T)$. Then,

$$\int_{\mathbb{R}} \varphi d\mu = T(\varphi) = 0.$$

Hence,

$$\mu(\mathbb{R} \setminus \operatorname{supp}(T)) = \sup\{\int_{\mathbb{R}} \varphi d\mu : \operatorname{supp}(\varphi) \subset \mathbb{R} \setminus \operatorname{supp}(T)\} = 0,$$

which shows that $\operatorname{supp}(\mu) \subseteq \operatorname{supp}(T)$. In the other direction, if φ has support in $\mathbb{R} \setminus \operatorname{supp}(\mu)$, then

$$T(\varphi) = \int_{\mathbb{R}} \varphi d\mu = 0,$$

hence, $\mathbb{R} \setminus \operatorname{supp}(\mu)$ is an annihilation set for T, which shows $\operatorname{supp}(T) \subseteq \operatorname{supp}(\mu)$. $\square$

7.4. Convergence of tempered distributions

Let $\{T_n\}_{n=1}^\infty$ be a sequence in $\mathcal{S}'$ and $T \in \mathcal{S}'$ be a tempered distribution. We say that $\lim_{n\to\infty} T_n = T$ *in* $\mathcal{S}'$ if

$$\lim_{n\to\infty} T_n(\varphi) = T(\varphi) \tag{7.4.1}$$

for all φ in $\mathcal{S}$. This is just pointwise convergence, since tempered distributions are functions on $\mathcal{S}$. In the language of functional analysis, it is *weak* (sequential) convergence*. In §7.4.4 we will see that this notion of convergence can be captured by a suitable topology, which is in fact the *weak* topology* on $\mathcal{S}'$. However, below we use the simple sequential definition.

7.4.1. More on the convolution in $\mathcal{S}$. Proposition 7.4 (vi) has the following interesting consequence.

LEMMA 7.7. *Let $\varphi, \psi \in \mathcal{S}$. Then, the $\mathcal{S}$-valued integral*

$$\int_{\mathbb{R}} (\tau_y\psi)\varphi(y)dy := \lim_{n\to\infty} \sum_{j=-n2^n}^{n2^n} \left(\tau_{\frac{j}{2^n}}\psi\right) \cdot \varphi\left(\frac{j}{2^n}\right) \frac{1}{2^n} \in \mathcal{S} \tag{7.4.2}$$

exists as a limit in $\mathcal{S}$. Also,

$$\int_{\mathbb{R}} (\tau_y\psi)\varphi(y)dy = \psi * \varphi. \tag{7.4.3}$$

PROOF. Let $h_n \in \mathcal{S}$ be the sequence element of which we take the limit on the right hand side of (7.4.2). For $n, m, s \geq 0$,

$$\begin{aligned}
&\left| h_n^{(s)}(x) - h_{n+m}^{(s)} \right| \\
= &\left| \frac{1}{2^n} \sum_j \frac{1}{2^m} \sum_{i=1}^{2^m} \left[\tau_{\frac{j}{2^n}} \psi^{(s)}(x) \varphi\left(\frac{j}{2^n}\right) - \tau_{\frac{j}{2^n}+\frac{i}{2^{n+m}}} \psi^{(s)}(x) \varphi\left(\frac{j}{2^n} + \frac{i}{2^{n+m}}\right)\right]\right| \\
\leq &\frac{1}{2^n} \sum_j \frac{1}{2^m} \sum_{i=1}^{2^m} \Delta_{j,i}(x),
\end{aligned}$$

where

$$\begin{aligned}
\Delta_{j,i}(x) \;=\; &\left|\psi^{(s)}\left(x - \frac{j}{2^n}\right)\right| \cdot \left|\varphi\left(\frac{j}{2^n}\right) - \varphi\left(\frac{j}{2^n} + \frac{i}{2^{n+m}}\right)\right| \\
&+ \left|\psi^{(s)}\left(x - \frac{j}{2^n}\right) - \psi^{(s)}\left(x - \frac{j}{2^n} - \frac{i}{2^{n+m}}\right)\right| \cdot \left|\varphi\left(\frac{j}{2^n} + \frac{i}{2^{n+m}}\right)\right|.
\end{aligned}$$

Making use of the estimate for $\Delta_{j,i}(x)$, we estimate the sum over j separately for small and for large j's. Recall the elementary estimate

$$1 + (a+b)^2 \leq 2(1+a^2)(1+b^2) \text{ if } a, b \geq 0.$$

Let $t \geq 0$. For small j's we use Lagrange theorem:

$$\begin{aligned}
(1+x^2)^{t/2}\text{small} \;&=\; \frac{(1+x^2)^{t/2}}{2^n} \sum_{\frac{|j|}{2^n} \leq R} \frac{1}{2^m} \sum_{i=1}^{2^m} \Delta_{j,i}(x) \\
&\leq\; \frac{C(1+R^2)^{t/2}}{2^n} \sum_{\frac{|j|}{2^n} \leq R} \left([\psi]_{s,t} [\varphi]_{1,0} \frac{1}{2^n} + [\psi]_{s+1,t} [\varphi]_{0,0} \frac{1}{2^n} \right) \\
&=\; \frac{\sharp\{j : |j| \leq R2^n\}}{2^{2n}} \left([\psi]_{s,t} [\varphi]_{1,0} + [\psi]_{s+1,t} [\varphi]_{0,0} \right) \\
&\leq\; \frac{C_1 R (1+R^2)^{t/2}}{2^n} \left([\psi]_{s,t} [\varphi]_{1,0} + [\psi]_{s+1,t} [\varphi]_{0,0} \right).
\end{aligned}$$

For large j's,

$$\begin{aligned}(1+x^2)^{t/2}\text{large} &= \frac{(1+x^2)^{t/2}}{2^n}\sum_{\frac{|j|}{2^n}>R}\frac{1}{2^m}\sum_{i=1}^{2^m}\Delta_{j,i}(x)\\ &\leq \frac{C}{2^n}\sum_{\frac{|j|}{2^n}>R}\left([\psi]_{0,t}\frac{[\varphi]_{1,t+k}}{(1+(j/2^n)^2)^{k/2}}+[\psi]_{s+1,t}\frac{[\varphi]_{0,t+k}}{(1+(j/2^n)^2)^{k/2}}\right)\\ &= C(\varphi,\psi)\sum_{\frac{|j|}{2^n}>R}\frac{1}{(1+(j/2^n)^2)^{k/2}}\frac{1}{2^n}.\end{aligned}$$

The last term in the estimate for $(1+x^2)^{t/2}$large is a sequence of Riemann sums for $\int_{|z|>R}\frac{dz}{(1+z^2)^{k/2}}$, and it can be made smaller than any given $\epsilon>0$ by choosing $R=R(\epsilon)$ large enough, independently of x, n, and m. For that value of R, the estimate for $(1+x^2)^{t/2}$small can be made smaller than ϵ by choosing $n>n(\epsilon)$ large enough, independently of x and m.

We have proved that

$$[h_n-h_{n+m}]_{s,t}\leq C\epsilon$$

provided $n\geq n(\epsilon,s,t)$. Hence, $\{h_n\}_{n=1}^{\infty}$ is a Cauchy sequence, which converges to a function in $\mathcal{S}$ by completeness.

The verification of (7.4.3) is elementary. With $h_n\in\mathcal{S}$ defined as above,

$$\lim_{n\to\infty}h_n(x)=\int_{\mathbb{R}}\psi(x-y)\varphi(y)dy=(\psi*\varphi)(x).$$

On the other hand, if $h_n\to h$ as $n\to\infty$ in $\mathcal{S}$, then $h_n(x)\to h(x)$. Hence,

$$(\psi*\varphi)(x)=\left(\lim_{n\to\infty}h_n\right)(x)=\left(\int_{\mathbb{R}}\tau_y\psi\cdot\varphi(y)dy\right)(x),$$

as wished. □

7.4.2. The convolution of a distribution in $\mathcal{S}'$ and a function in $\mathcal{S}$. A distribution might be a strange object, but, after convolving it with a function in $\mathcal{S}$, it becomes the functional "integral with a smooth weight". That it becomes a functional of this kind is not a surprise: tempered distributions are modeled on functionals of this form. The proof, however, requires some care.

THEOREM 7.10. *Let T be a tempered distribution and ψ a function in $\mathcal{S}$. Define*

$$g(x) = T(\tau_x(U\psi)). \tag{7.4.4}$$

Then, $g \in C^\infty$, all of its derivatives have polynomial growth, and $T * \psi = T_g$. Moreover,

$$g^{(m)}(x) = T(\tau_x(U\psi^{(m)})). \tag{7.4.5}$$

PROOF. We prove first that $g \in C^\infty$. In order to prove the existence and compute the derivative

$$g'(x) = \frac{d}{dx} T(\tau_x U\psi),$$

we have to pass the derivative, which is a pointwise limit, inside the functional T, which is defined on functions, not on points. Using the continuity of $T\colon \mathcal{S} \to \mathbb{C}$ and $\tau_x\colon \mathcal{S} \to \mathcal{S}$,

$$\begin{aligned} g'(x) &= \frac{d}{dx} T(\tau_x U\psi) \qquad (7.4.6)\\ &= \lim_{h\to 0} \frac{T(\tau_{x+h}U\psi) - T(\tau_x U\psi)}{h} \\ &= T\left(\lim_{h\to 0} \frac{\tau_{x+h}U\psi - \tau_x U\psi}{h}\right) \\ &= T \circ \tau_x \left(\lim_{h\to 0} \frac{\tau_h U\psi - U\psi}{h}\right), \qquad (7.4.7) \end{aligned}$$

provided $\lim_{h\to 0} \frac{\tau_h U\psi - U\psi}{h}$ exists in $\mathcal{S}$.

Since the pointwise limit is

$$\lim_{h\to 0} \frac{\tau_h U\psi(y) - U\psi(y)}{h} = \lim_{h\to 0} \frac{\psi(h-y) - \psi(-y)}{h} = \psi'(-y) = (U\psi')(y),$$

what we have to prove is that $\lim_{h\to 0} \frac{\tau_h U\psi - U\psi}{h} = (U\psi)'$ holds in $\mathcal{S}$ as well. Let $m \geq 0$. Using Lagrange Theorem twice,

$$\begin{aligned} \left|\frac{d^{(m)}}{dy^m}\left[\frac{\tau_h U\psi(y) - U\psi(y)}{h} - (U\psi')(y)\right]\right| &= \left|\frac{\tau_h U\psi^{(m)}(y) - U\psi^{(m)}(y)}{h} - U\psi^{(m+1)}(y)\right| \\ &= \left|\frac{\psi^{(m)}(h-y) - \psi^{(m)}(-y)}{h} - \psi^{(m+1)}(-y)\right| \\ &= \left|\psi^{(m+1)}(\Theta_1 h - y) - \psi^{(m+1)}(-y)\right| \\ &= \left|\psi^{(m+2)}(\Theta_2\Theta_1 h - y)\right| \cdot |h|. \end{aligned}$$

where $0 \leq \Theta_1, \Theta_2 \leq 1$. Then, if $n \geq 0$ and $|h| \leq 1$,

$$\begin{aligned}
\left[\frac{\tau_h U\psi - U\psi}{h} - U\psi'\right]_{m,n} &= \sup_y \left|(1+y^2)^{n/2}\frac{d^{(m)}}{dy^m}\left[\frac{\tau_h U\psi(y) - U\psi(y)}{h} - (U\psi')(y)\right]\right| \\
&\leq |h| \sup_y \left|(1+y^2)^{n/2}\psi^{(m+2)}(\Theta_2\Theta_1 h - y)\right| \\
&\leq |h| \sup_z \left|(1+(|z|+1)^2)^{n/2}\psi^{(m+2)}(z)\right| \\
&\leq C|h|[\psi]_{m+2,n},
\end{aligned}$$

which tends to 0 as $h \to 0$. Returning to (7.4.6), we have proved that

$$\frac{d}{dx}T(\tau_x U\psi) = T(\tau_x(U\psi')), \tag{7.4.8}$$

hence (7.4.5). Also, using theorem 7.2, (7.4.8), and the elementary estimate (see the beginning of the proof of (vi) in proposition 7.4)

$$[\tau_x\psi]_{m,n} = \sup_{y\in\mathbb{R}}(1+y^2)^{n/2}|\psi^{(m)}(y-x)| \leq C_n(1+x^2)^{n/2}[\psi]_{m,n},$$

we have, for $k \geq 0$,

$$\begin{aligned}
|g^{(k)}(x)| &= |T(\tau_x(U\psi^{(k)}))| \leq C \sum_{0\leq m,n\leq N} [\tau_x(U\psi^{(k)})]_{m,n} \\
&\leq C_1 \sum_{0\leq m,n\leq N} [U\psi]_{m+k,n}(1+x^2)^{n/2} \\
&= C_1 \sum_{0\leq m,n\leq N} [\psi]_{m+k,n}(1+x^2)^{n/2} \\
&\leq C_2(1+x^2)^{N/2}.
\end{aligned}$$

Hence, $g^{(k)}$ has polynomial growth.

To conclude the proof, we have to show that $T * \psi = T_g$. For $\varphi \in \mathcal{S}$, the map

$$\psi \mapsto (U\psi) * \varphi = \int_{\mathbb{R}} \tau_y(U\psi)\varphi(y)dy$$

is continuous on $\mathcal{S}$, by lemma 7.7. Thus,

$$\begin{aligned}
(T * \psi)(\varphi) &:= T((U\psi) * \varphi) \\
&= T\left(\int_{\mathbb{R}} \tau_y(U\psi)\varphi(y)dy\right) \\
&= \int_{\mathbb{R}} T(\tau_y(U\psi))\varphi(y)dy \\
&= T_g(\varphi),
\end{aligned}$$

as wished. □

It would be nice if g were always in $\mathcal{S}$, but this is not the case. Consider the case of $T_m(\varphi) = \int_{\mathbb{R}} \varphi(x)dx$ (i.e. T_m is the Lebesgue measure), and let $\psi \in \mathcal{S}$ with $\int_{\mathbb{S}} \psi(x)dx = \alpha$. Then,

$$\begin{aligned} g(x) &= T_m(\tau_x U\psi) = \int_{\mathbb{R}} \psi(x-y)dy \\ &= \alpha \end{aligned}$$

is a constant function.

7.4.3. The density of $\mathcal{S}$ in $\mathcal{S}'$. In this subsection we show that any tempered distribution T is the limit of distribution $T_{\varphi_n}(\psi) = \int_{\mathbb{R}} \varphi_n(x)\psi(x)dx$, with φ_n in $\mathcal{S}$. Before we do that, we consider some special cases and related facts, having short proofs. They will be used in the proof of the main theorem.

Recall that $\delta_t\varphi(x) = \frac{1}{t}\varphi\left(\frac{x}{t}\right)$ if $t > 0$. We write $\varphi_t = \delta_{1/t}\varphi$.

PROPOSITION 7.10. *Let $\varphi \in \mathcal{S}$ with $\int_{\mathbb{R}} \varphi(x)dx = 1$.*

(i) *We have*
$$\lim_{t\to 0} T_{\varphi_t} = \delta_0$$
in $\mathcal{S}'$.

(ii) *We have*
$$\lim_{t\to 0} \varphi_t * \psi = \psi$$
in $\mathcal{S}$.

(iii) *If $T \in \mathcal{S}'$, then*
$$\lim_{t\to 0} T * \varphi_t = T$$
in $\mathcal{S}'$.

(iv) *Let $\eta \succ [-1,1]$, $\eta \in C_c^\infty(\mathbb{R})$, and let $\eta_R(x) = \eta(x/R)$ for $R > 0$. Also let $g \in C^\infty(\mathbb{R})$ be a function with polynomial growth, $|g(x)| \leq C(1+x^2)^{N/2}$ for some N. Then,*
$$\lim_{R\to\infty} T_{\eta_R g} = T_g.$$

Statements (i), (ii) and (iii) are avatars of the same basic principle, which we already met with approximations of identity. Since $\eta_R g \in C_c^\infty(\mathbb{R})$, in view of theorem 7.10, (iii) and (iv) suggest that $C_c^\infty(\mathbb{R})$ is dense in $\mathcal{S}'$. However, proving the latter involves a double limit, and I do not know how to deduce the result from these simpler statements.

PROOF. (i) For $\psi \in \mathcal{S}$,

$$\begin{aligned} T_{\delta_{1/t}\varphi}(\psi) &= \int_{\mathbb{R}} \frac{1}{t}\varphi\left(\frac{x}{t}\right)\psi(x)dx \\ &= \int_{\mathbb{R}} \varphi(y)\,\psi(ty)dy \\ &\to \int_{\mathbb{R}} \varphi(y)\,dy \cdot \psi(0) \\ &= \delta_0(\psi) \end{aligned}$$

by dominated convergence.

(ii) For $\psi \in \mathcal{S}$, we want to estimate $\left[(\delta_{1/t}\varphi) * \psi - \psi\right]_{m,n}$, which a variation on the theme of approximations of identity.

$$\begin{aligned} & (1+x^2)^{n/2}\left|\left[(\delta_{1/t}\varphi) * \psi\right]^{(m)}(x) - \psi^{(m)}(x)\right| \\ = \; & (1+x^2)^{n/2}\left|\left[(\delta_{1/t}\varphi) * \psi^{(m)}\right](x) - \psi^{(m)}(x)\right| \\ \leq \; & (1+x^2)^{n/2}\int_{\mathbb{R}}\left|\frac{1}{t}\varphi\left(\frac{y}{t}\right)\right| \cdot |\psi^{(m)}(x-y) - \psi^{(m)}(x)|dy \\ = \; & A + B, \end{aligned}$$

where A is the integral over $|y| \leq \delta$ and B that over $|y| > \delta$. Using Lagrange mean value theorem, there is $\Theta = \Theta(x,y)\imath[-1,1]$ such that

$$\begin{aligned} A &= (1+x^2)^{n/2}\int_{|y|\leq\delta}\left|\frac{1}{t}\varphi\left(\frac{y}{t}\right)\right| \cdot |\psi^{(m+1)}(x+\Theta\delta)| \cdot |y|dy \\ &\leq C[\psi]_{m+1,n}\delta\|\varphi\|_{L^1}, \end{aligned}$$

which vanishes ad $\delta \to 0$. Choose δ such that $A \leq \epsilon$ (with $\epsilon > 0$ fixed).

About B, assuming $0 < t < 1$ in the last few lines

$$\begin{aligned} B &\leq C\int_{|y|\geq\delta}\left|\frac{1}{t}\varphi\left(\frac{y}{t}\right)\right|(1+(x-y)^2)^{n/2}\left|\psi^{(m)}(x-y)\right|(1+y^2)^{n/2}dy \\ &+ \int_{|y|\geq\delta}\left|\frac{1}{t}\varphi\left(\frac{y}{t}\right)\right|(1+x^2)^{n/2}\left|\psi^{(m)}(x)\right|dy \\ &\leq C[\psi]_{m,n}\left(\int_{|z|\geq\delta/t}|\varphi(z)|\,(1+tz^2)^{n/2}dz + \int_{|z|\geq\delta/t}|\varphi(z)|\,dz\right) \\ &\leq C[\psi]_{m,n}\int_{|z|\geq\delta/t}|\varphi(z)|\,[1+(1+z^2)^{n/2}]dz \\ &\to 0 \end{aligned}$$

as $t \to 0$, by dominated convergence. By choosing t small enough, then, $B \leq \epsilon$.

(iii) Fix $\psi \in \mathcal{S}$, and recall that $(T * \varphi_t)(\psi) = T((U\varphi_t) * \psi) = T((U\varphi)_t * \psi)$. Using (ii) and the continuity of T,

$$\lim_{t \to 0} T((U\varphi)_t * \psi) = T\left(\lim_{t \to 0}(U\varphi)_t * \psi)\right) = T(\psi),$$

where the limit in the argument of T is a limit in $\mathcal{S}$.

(iv) For $\psi \in \mathcal{S}$,

$$\begin{aligned} |T_g(\psi) - T_{\eta_R} g(\psi)| &\leq \int_{\mathbb{R}} |1 - \eta_R(x)| \cdot |g(x)| \cdot |\varphi(x)| dx \\ &\leq \int_{|x| \geq R} |g(x)| \cdot |\varphi(x)| dx \\ &\leq C \int_{|x| \geq R} (1 + x^2)^{N/2} \cdot |\varphi(x)| dx \\ &\to 0 \end{aligned}$$

as $R \to \infty$, by dominated convergence.

□

We now come to the promised density theorem.

THEOREM 7.11. *Let $T \in \mathcal{S}'$. Then, there exists a sequence $\{h_n\}$ in $\mathcal{S}$ such that $\lim_{n \to \infty} T_{h_n} = T$ in $\mathcal{S}'$. In fact, we can choose $h_n \in C_c^\infty$.*

PROOF. Fix $\varphi \in \mathcal{S}$ such that $\int_{\mathbb{R}} \varphi(x) dx = 1$, and let $\varphi_n = \delta_{1/n}\varphi$. Let then g_n be the function in C^∞ whose derivatives have polynomial growth such that $T * \varphi_n = T_{g_n}$ (see theorem 7.10). As before, let $\eta \succ [-1, 1]$ and set $\eta_R(x) = \eta(x/R) \succ [-R, R]$. We can ask $\text{supp}(\eta) \subseteq [-2, 2]$. The function $h_n = g_n \eta_{R_n}$ belongs to C_c^∞.

For $\psi \in \mathcal{S}$ fixed, we estimate:

$$\begin{aligned} |T(\varphi) - T_{\eta_{R_n} g_n}(\psi)| &\leq |T(\psi) - T_{g_n}(\psi)| + |T_{g_n}(\psi) - T_{\eta_{R_n} g_n}(\psi)| \\ &= A + B. \end{aligned}$$

Fix $\epsilon > 0$. We have

$$A = |T(\psi) - (T * \varphi_n)(\psi)| \leq \epsilon \tag{7.4.9}$$

if $n \geq n(\epsilon)$ by proposition 7.10 (iii). We also have

$$B \leq \int_{\mathbb{R}} |g_n(x)| \cdot |\psi(x)| \cdot |1 - \eta_{R_n}(x)| dx \leq \int_{|x| \geq R_n} |g_n(x)| \cdot |\psi(x)| dx. \tag{7.4.10}$$

We need an estimate for $|g_n(x)|$. For some $N > 0$ and $C > 0$,

$$|g_n(x)| = |T(\tau_x(U\varphi_n))| \le C \sum_{0\le l,m\le N} [\tau_x(U\varphi_n)]_{l,m}, \tag{7.4.11}$$

and we proceed to estimate the summands.

$$\begin{aligned}
& (1+y^2)^{m/2}\left|[\tau_x(U\varphi_n)]^{(l)}(y)\right| \\
= \;& n^{1+l}(1+y^2)^{m/2}\varphi^{(l)}\left(\frac{x-y}{n}\right) \\
\le \;& Cn^{l+1}(1+x^2)^{m/2}(1+(x-y)^2)^{m/2}\cdot\varphi^{(l)}\left(\frac{x-y}{n}\right) \\
\le \;& C[\varphi]_{l,m}n^{l+1}(1+x^2)^{m/2}\frac{(1+(x-y)^2)^{m/2}}{(1+(|x-y|/n)^2)^{m/2}} \\
= \;& C[\varphi]_{l,m}n^{l+m+1}(1+x^2)^{m/2}\frac{(1+(x-y)^2)^{m/2}}{(n^2+(x-y)^2)^{m/2}} \\
\le \;& C[\varphi]_{l,m}n^{2N+1}(1+x^2)^{N/2}.
\end{aligned}$$

The latter is an estimate for $[\tau_x(U\varphi_n)]_{l,m}$, thus (7.4.11) gives

$$|g(x)| \le C\sum_{0\le l,m\le N}[\varphi]_{l,m}n^{2N+1}(1+x^2)^{N/2}.$$

Inserting in (7.4.10), we obtain:

$$\begin{aligned}
B &\le \int_{|x|\ge R_n}|g_n(x)|\cdot|\psi(x)|dx \\
&\le C\sum_{0\le l,m\le N}[\varphi]_{l,m}n^{2N+1}\int_{|x|\ge R_n}|\psi(x)|(1+x^2)^{\frac{N}{2}}dx \\
&\le C\sum_{0\le l,m\le N}[\varphi]_{l,m}\frac{n^{2N+1}}{(1+R_n^2)^{1/2}}\int_{|x|\ge R_n}|\psi(x)|(1+x^2)^{\frac{N+1}{2}}dx \\
&= C\sum_{0\le l,m\le N}[\varphi]_{l,m}\frac{n^{2N+1}}{(1+R_n^2)^{1/2}}\int_{|x|\ge R_n}|\psi(x)|(1+x^2)^{\frac{N+3}{2}}\frac{dx}{1+x^2} \\
&\le C[\psi]_{0,N+3}\sum_{0\le l,m\le N}[\varphi]_{l,m}\frac{n^{2N+1}}{(1+R_n^2)^{1/2}},
\end{aligned}$$

and the latter goes to 0 as $n\to\infty$ if, for instance, $R_n = n^{2N+2}$.

In total, with this choice of R_n, there is $n_1(\epsilon) > 0$ such that, for $n > n(\epsilon)$,

$$|T(\varphi) - T_{\eta_{R_n}g_n}(\psi)| \le A + B \le 2\epsilon.$$

□

7.4.4. The topology on $\mathcal{S}'$ by means of cylinder sets. The sequential definition of convergence in $\mathcal{S}$ can be interpreted in terms of a topology on $\mathcal{S}'$. Define cylinder sets at the origin,

$$\mathcal{N}(\varphi,\epsilon;0)=\{T\in\mathcal{S}' :\ |T(\varphi)|<\epsilon\}, \tag{7.4.12}$$

where $\varphi\in\mathcal{S}$ and $\epsilon>0$, and

$$\mathcal{N}(\varphi_1,\ldots,\varphi_N,\epsilon;0)=\mathcal{N}(\varphi_1,\epsilon;0)\cap\cdots\cap\mathcal{N}(\varphi_n,\epsilon;0).$$

Cylinder neighborhoods centered at $S\in\mathcal{S}'$ are defined by translation,

$$\mathcal{N}(\varphi_1,\ldots,\varphi_N,\epsilon;S)=S+\mathcal{N}(\varphi_1,\ldots,\varphi_N,\epsilon;S).$$

A subset U of $\mathcal{S}'$ is open if and only if U contains a cylinder neighborhood. Let τ be the corresponding topology.

THEOREM 7.12. *Let $\{T_n\}_{n=1}^{\infty}$ be a sequence in $\mathcal{S}'$ and $T\in\mathcal{S}'$. Then,* $\lim_{n\to\infty}T_n(\varphi)=T(\varphi)$ *for all φ in $\mathcal{S}$ if and only if* $\lim_{n\to\infty}T_n=T$ *in the topological space $(\mathcal{S}',\tau)$.*

PROOF. By translation invariance, we can assume $T=0$. Suppose T_n converges to 0 sequentially, and consider a cylinder neighborhood $\mathcal{N}(\varphi_1,\ldots,\varphi_N,\epsilon;0)$ (any open neighborhood of the origin contains one). By hypothesis, there are $n_j(\epsilon)$ such that $|T_n(\varphi_j)|<\epsilon$ if $n>n_j(\epsilon)$. For $n>\max(n_j(\epsilon):j=1,\ldots,n)$ we have that $T\in\mathcal{N}(\varphi_1,\ldots,\varphi_N,\epsilon;0)$. This shows that $\lim_{n\to\infty}T_n=0$ in the topological space $(\mathcal{S}',\tau)$.

Viceversa, suppose that $\lim_{n\to\infty}T_n=0$ in the topological space $(\mathcal{S}',\tau)$, and fix $\varphi\in\mathcal{S}$. Then, for all $k\geq 1$ there is n_k such that $|T_n(\varphi)|<1/k$ if $n>n_k$. Then, $\lim_{n\to\infty}T_n(\varphi)=0$, as wished. □

Critical analysis of the proof shows that the line of argument above has little to do with tempered distributions. This way, in fact, one constructs *weak** topologies on the dual of topological vector spaces.