Lecture Note on
Measure Theory and
Functional Analysis

Anthony Hong[1]

March 1, 2024

[1]Thanks to Professor Martinkainen for his course Math 5051-5052 Measure Theory and Functional Analysis I-II; thanks for his permission to edit his tex file of 5051 lecture note.

Contents

Chapter 1

Review of fundamentals

Make sure to review these fundamentals carefully before continuing further. We will not need a lot of background theory in this course – however, we will be performing various, sometimes tricky, manipulations with these basic concepts.

1.1 Basic set manipulations

Given a collection $\mathcal{F}$ of sets $A \subset X$, where X is some fixed ambient space, their union is

$$\bigcup_{A \in \mathcal{F}} A := \{x \in X : x \in A \text{ for some } A \in \mathcal{F}\}$$

and their intersection is

$$\bigcap_{A \in \mathcal{F}} A := \{x \in X : x \in A \text{ for all } A \in \mathcal{F}\}.$$

Often the family of sets is indexed – $\mathcal{F} = \{A_\alpha : \alpha \in \mathcal{A}\}$ for some index set $\mathcal{A}$ – and then the above unions and intersections are also denoted by

$$\bigcup_{\alpha \in \mathcal{A}} A_\alpha \qquad \text{and} \qquad \bigcap_{\alpha \in \mathcal{A}} A_\alpha.$$

Finally, if the index set $\mathcal{A}$ is obvious from the context, we may also simply write $\bigcup_\alpha A_\alpha$ and $\bigcap_\alpha A_\alpha$. Indeed, often the index set is $\mathbb{N} = \{1, 2, 3, \ldots\}$ and it can be convenient to simply write $\bigcup_n A_n$.

In general, the index set $\mathcal{A}$ need not be countable, but we are mostly interested in countable unions and intersections. Countability is an extremely important concept in measure theory. Recall that $\mathcal{F}$ is countable if $\mathcal{F} = \emptyset$ or there is an injection $\varphi \colon \mathcal{F} \to \mathbb{N}$. In practice, if $\mathcal{F}$ is countable we may always index the set $\mathcal{F}$ as $\mathcal{F} = \{A_n : n \in \mathbb{N}\}$. Remember that a countable union of countable sets is countable.

For two sets $A, B \subset X$ we define their difference as

$$A \setminus B := \{x \in X : x \in A \text{ and } x \notin B\}.$$

The complement of a set $A \subset X$ is

$$A^c := X \setminus A.$$

We almost always use the explicit notation $X \setminus A$, since the notation A^c hides how the definition depends on the choice of the ambient set X. It is often useful that

$$A \setminus B = A \cap (X \setminus B) = A \cap B^c.$$

Remember the laws of de Morgan:

$$X \setminus \bigcup_{A\in\mathcal{F}} A = \bigcap_{A\in\mathcal{F}} (X \setminus A),$$
$$X \setminus \bigcap_{A\in\mathcal{F}} A = \bigcup_{A\in\mathcal{F}} (X \setminus A).$$

It is also useful that

$$B \cap \bigcup_{A\in\mathcal{F}} A = \bigcup_{A\in\mathcal{F}} (A \cap B)$$

and

$$B \cup \bigcap_{A\in\mathcal{F}} A = \bigcap_{A\in\mathcal{F}} (A \cup B).$$

Let $f\colon X \to Y$ be a function. Recall that for $A \subset X$ we define that the image of A under f is

$$f(A) = fA := \{f(x)\colon x \in A\} \subset Y$$

and for $B \subset Y$ that the preimage of B under f is

$$f^{-1}(B) = f^{-1}B := \{x \in X\colon f(x) \in B\}.$$

Despite the notational similarities, recall that this does not require the existence of the inverse mapping f^{-1} – the function f need not be a bijection and nevertheless the preimage always makes sense. Remember that we have

$$f \bigcup_{\alpha\in\mathcal{A}} A_\alpha = \bigcup_{\alpha\in\mathcal{A}} fA_\alpha,$$
$$f^{-1} \bigcup_{\beta\in\mathcal{B}} B_\beta = \bigcup_{\beta\in\mathcal{B}} f^{-1}B_\beta,$$
$$f^{-1} \bigcap_{\beta\in\mathcal{B}} B_\beta = \bigcap_{\beta\in\mathcal{B}} f^{-1}B_\beta.$$

Notice that we always have

$$f \bigcap_{\alpha\in\mathcal{A}} A_\alpha \subset \bigcap_{\alpha\in\mathcal{A}} fA_\alpha$$

but that the inclusion may be strict (equality holds e.g. if f is an injection).

1.2 Normed spaces, metric spaces and topological spaces

We first recall normed spaces.

1.2.1 Definition. Let Y be a vector space and let $y \mapsto \|y\|_Y$ be a mapping from Y to non-negative real numbers. This mapping is a norm on Y if the following holds.

1. We have the triangle inequality $\|y_1 + y_2\|_Y \le \|y_1\|_Y + \|y_2\|_Y$ for all $y_1, y_2 \in Y$.
2. We have $\|ay\|_Y = |a|\|y\|_Y$ for all scalars a and $y \in Y$.
3. We have $\|y\|_Y = 0$ if and only if $y = 0$.

Normed spaces are fundamental for functional analysis, and they appear frequently throughout the course. For example, many spaces Y consisting of functions $f\colon X \to \mathbb{R}$ are equipped with some norm that measures the size of the function f. An example is given by the so-called L^p spaces $L^p(X)$ consisting of the p-integrable functions (with respect to some measure μ):

$$\Big(\int_X |f|^p \,\mathrm{d}\mu\Big)^{\frac{1}{p}} < \infty.$$

However, the underlying ambient space X may often be more complicated – a metric space or a topological space, or even just a set with a measure. In particular, it need not be a vector space like in the case of a normed space.

1.2.2 Definition. A metric space (X, d) is a set X equipped with a distance function $d\colon X \times X \to [0, \infty)$ with the properties

1. (triangle inequality) $d(x, z) \le d(x, y) + d(y, z)$, $x, y, z \in X$;
2. $d(x, y) = d(y, x)$, $x, y \in X$;
3. $d(x, y) = 0$ if and only if $x = y$.

Remember that a norm $\|\cdot\|_Y$ on a vector-space Y defines a metric $d_Y(y_1, y_2) = \|y_1 - y_2\|_Y$ on Y. Associated with a metric comes the notion of open sets: a set $V \subset X$ on a metric space (X, d) is open if for all $x \in V$ there is $r = r_x > 0$ so that

$$B_d(x, r) := \{y \in X : d(x, y) < r\} \subset V.$$

If the metric is clear from the context, an open ball is denoted just by $B(x, r)$ or also by $B_X(x, r)$. We can denote the open sets in X with τ_d – they depend on the choice of the metric d. It is not hard to prove that arbitrary (even uncountable) unions of open sets are open and finite intersection of open sets are open.

For many things, we do not actually even need a metric at all. This is when we only care about the open sets of X and topological notions that can be defined using them – such as, continuity, compactness etc.

1.2.3 Definition. we say that $\tau \subset \mathcal{P}(X)$ is a topology of X if

1. τ is closed under arbitrary unions,
2. τ is closed under finite intersections,
3. $X \in \tau$ and $\emptyset \in \tau$.

A pair (X, τ) is called a topological space and the sets $V \in \tau$ are called open sets of X.

Again, if (X, d) is a metric space, then (X, τ_d) is a topological space, but, in general, a topology τ may not be induced by a metric d. If you feel more comfortable, when we work on a topological space you can first think that it is actually a metric space and the topology is induced by the metric. This limits the generality somewhat. However, it is a key goal of us to learn to work in abstract settings and carefully identify what properties and concepts we really need and when.

1.2.4 Example. The power set $\mathcal{P}(X)$ is always a topology on X – then all subsets of X are open. The mini topology $\{\emptyset, X\}$ is also always a topology on X. For example, the mini topology is not given by any metric: if d is any metric on X, $a, b \in X$ and $a \neq b$, then the open ball $B(a, d(a, b))$ is neither X nor $\emptyset$.

One needs to be a bit careful when working in topological spaces – not everything that is true in metric spaces continues to hold. Also, not all definitions that make sense in metric spaces necessarily make sense in topological spaces. Many concepts make sense, though, but you have to use the correct definition – one that involves open sets only. We now go over some of the critical definitions in a topological space (X, τ) – but we will not venture deep.

1.2.5 Definition. If $U \in \tau$ and $x \in U$, we say that U is a neighbourhood of x. If $A \subset X$ and $U \in \tau$ is such that $A \subset U$, we say that U is a neighbourhood of A.

The following is a key definition – something that we often need to assume.

1.2.6 Definition. A topological space (X, τ) is Hausdorff (or T_2) if for every $x, y \in X$ with $x \neq y$ there exists $V, U \in \tau$ with $x \in V$, $y \in U$ and $V \cap U = \emptyset$.

In other words, distinct points have distinct neighbourhoods. This is trivially true in metric spaces (consider balls with small enough radius) but has to be often assumed otherwise.

1.2.7 Definition. A set F is closed if $X \setminus F$ is open.

An arbitrary intersection of closed sets is closed and a finite union of closed sets is closed using the definition and the laws of de Morgan. Moreover, X and $\emptyset$ are closed.

1.2.8 Definition. The closure of $A \subset X$ is

$$\overline{A} = \{x \in X : U \cap A \neq \emptyset \text{ for all neighbourhoods } U \text{ of } x\}.$$

1.2.9 Remark. In metric spaces an equivalent definition is given by

$$\overline{A} = \{x \in X : \ \operatorname{dist}(x, A) = 0\},$$

where

$$\operatorname{dist}(A, B) = \inf\{d(x, y) : \ x \in A, \ y \in B\}, \ \operatorname{dist}(x, A) := \operatorname{dist}(\{x\}, A).$$

The closure satisfies the following properties.

1. $A \subset \overline{A}$ and $\overline{A}$ is closed.
2. If $A \subset B$, then $\overline{A} \subset \overline{B}$.
3. A is closed if and only if $A = \overline{A}$.
4. $\overline{\overline{A}} = A$.
5. $\overline{A \cup B} = \overline{A} \cup \overline{B}$, $\overline{A \cap B} \subset \overline{A} \cap \overline{B}$.

A set A is called dense in X if $\overline{A} = X$. For example, $\mathbb{Q}$ is dense on $\mathbb{R}$.

The boundary ∂A of A is defined with

$$\partial A = \{x \in X : U \cap A \neq \emptyset \text{ and } U \cap A^c \neq \emptyset \text{ for all neighbourhoods } U \text{ of } x\}.$$

We have that ∂A is always closed,

$$\partial A = \partial(X \setminus A)$$

and

$$\overline{A} = A \cup \partial A.$$

1.2.10 Definition. A function $f\colon X \to Y$, where X and Y are topological spaces, is called continuous if $f^{-1}V$ is open in X for all open $V \subset Y$.

1.2.11 Remark. The corresponding pointwise definition is the following. A function f is continuous at $x \in X$ if for every neighbourhood V of $f(x)$ there exists a neighbourhood U of x so that $fU \subset V$.

The most relevant notion for us is compactness.

1.2.12 Definition. A set $K \subset X$ is compact if every open covering of K has a finite subcovering. That is, if $K \subset \bigcup_\alpha V_\alpha$, where each V_α is open, then there are $V_{\alpha_1}, \ldots, V_{\alpha_n}$ so that $K \subset \bigcup_{i=1}^n V_{\alpha_i}$.

1.2.13 Remark. In $\mathbb{R}^d$ we know that a set is compact if and only if it is bounded and closed (Heine–Borel).

The following are key properties of compactness. If you do not know these results in this generality, you can either take these as facts, look at Rudin's book or prove them yourself. This is not a course on topology and it is more important to know how to apply these topological facts and understand what they mean, than be extremely fluent with their proofs.

1. A closed subset of a compact set is compact.
2. The image of a compact set under a continuous mapping is compact.
3. If $K_1, K_2 \subset X$ are compact, then so is $K_1 \cup K_2$.
4. Let X be Hausdorff and $K_1, K_2 \subset X$ be compact and disjoint $K_1 \cap K_2 = \emptyset$. Then there are neighbourhoods U_1 and U_2 of K_1 and K_2, respectively, which are disjoint $U_1 \cap U_2 = \emptyset$. Notice that this can, and often is, applied with $x \notin K_1$ and $K_2 := \{x\}$.
5. If X is Hausdorff and $K \subset X$ is compact, then K is closed. This and the previous property are some of the key reasons why the Hausdorff property is often key when dealing with compactness.
6. Let X be Hausdorff and $K_1 \supset K_2 \supset \cdots$ be compact and non-empty. Then $\bigcap_i K_i \neq \emptyset$.

We move on to local compactness, our final key topological concept.

1.2.14 Definition. A topological space (X, τ) is locally compact if for every $x \in X$ there is a neighbourhood U of x with $\overline{U}$ compact.

Many common metric spaces are NOT locally compact (such as the infinite dimensional Hilbert spaces L^2 that we see later). Trivially $\mathbb{R}^d$ is locally compact.

We call a locally compact Hausdorff space X LCH.

1.2.15 Theorem. *Let X be a LCH space. If $K \subset X$ is compact and U is a neighbourhood of K, then there is a neighbourhood V of K so that $\overline{V}$ is compact and*

$$\overline{V} \subset U.$$

1.3 Euclidean spaces

We will sometimes work in the Euclidean space $\mathbb{R}^d$, which has a lot of structure missing from abstract spaces. The points x of $\mathbb{R}^d$ can be denoted by $x = (x_1, \ldots, x_d)$, $x_i \in \mathbb{R}$, and the Euclidean norm is

$$|x| = \Big(\sum_{i=1}^d |x_i|^2 \Big)^{\frac{1}{2}}.$$

This, of course, induces a metric – which in turn gives rise to open sets, etc.

1.4 $\limsup$ and $\liminf$

Finally, we recall the definitions of $\limsup$ and $\liminf$ of sequences, since these appear frequently in measure theory but not necessarily so frequently in other places. Given a sequence $a_1, a_2, \ldots$ of scalars we define

$$\limsup_{i\to\infty} a_i = \lim_{k\to\infty} \sup_{i\geq k} a_i = \inf_{k\in\mathbb{N}} \sup_{i\geq k} a_i$$

and

$$\liminf_{i\to\infty} a_i = \lim_{k\to\infty} \inf_{i\geq k} a_i = \sup_{k\in\mathbb{N}} \inf_{i\geq k} a_i$$

These always exist and the sequence has a limit if and only if

$$\liminf_{i\to\infty} a_i = \limsup_{i\to\infty} a_i.$$

We will often have a sequence of real valued functions f_i and we consider e.g. the function

$$(\liminf_{i\to\infty} f_i)(x) := \liminf_{i\to\infty} f_i(x).$$

1.5 Additional notation

We denote $A \lesssim B$ if $A \leq CB$ for some unimportant constant C that we need not track. This means that the *implicit constant* C cannot depend on anything important, such as, some key parameter ϵ appearing in the proof at question. For instance, C can be a uniform constant (e.g. $C = 10$), or some constant depending on the dimension d of the underlying Euclidean space $\mathbb{R}^d$ (e.g. $C = 100 \cdot 2^d$). We can write $A \lesssim_\epsilon B$ to mean that $A \leq C(\epsilon)B$ for some constant $C(\epsilon)$ that is now allowed to depend on some given parameter ϵ. We will also write $A \sim B$ if $A \lesssim B \lesssim A$.

We denote closed intervals by $[a, b]$ and open intervals by (a, b) (and half-open intervals obviously by $[a, b)$ and $(a, b]$).

Chapter 2

Abstract measure theory

Let X be a set. A measure μ is a mapping assigning a value on the interval $[0, \infty]$ to subsets $E \subset X$ (and satisfying some additional assumptions to be specified soon). Often, a measure μ cannot act on every single set on the power set

$$\mathcal{P}(X) = \{E \colon E \subset X\},$$

but rather on some subcollection. The subcollection has to be a σ-algebra.

2.0.1 Definition. Let X be a set. A collection $\mathcal{F} \subset \mathcal{P}(X)$ is called a σ-algebra on X if the following holds.

1. $\emptyset \in \mathcal{F}$.
2. If $E \in \mathcal{F}$ then $X \setminus E \in \mathcal{F}$.
3. If $E_1, E_2, \ldots \in \mathcal{F}$ then $\bigcup_{i=1}^{\infty} E_i \in \mathcal{F}$.

2.0.2 Remark. If $\mathcal{F}$ is a σ-algebra on X and $E_1, E_2, \ldots \in \mathcal{F}$ then also

$$\bigcap_i E_i = X \setminus \bigcup_i (X \setminus E_i) \in \mathcal{F}.$$

Of course, also $X \in \mathcal{F}$ as it is the complement of $\emptyset$.

2.1 Measure spaces

2.1.1 Definition. Let $\mathcal{F}$ be a sigma-algebra on a space X. A function $\mu \colon \mathcal{F} \to [0, \infty]$ is a measure if the following holds.

1. We have $\mu(\emptyset) = 0$.
2. (countable additivity) If $A = \bigcup_{i=1}^{\infty} A_i$, $A_i \in \mathcal{F}$, and $A_i \cap A_j = \emptyset$ for $i \neq j$, then

$$\mu(A) = \sum_i \mu(A_i).$$

The triple $(X, \mathcal{F}, \mu)$ is called a measure space. A pair $(X, \mathcal{F})$ is sometimes referred to as a measurable space – it can be equipped with any measure $\mu \colon \mathcal{F} \to [0, \infty]$ to get a measure space $(X, \mathcal{F}, \mu)$.

We have to specify both the collection of measurable sets – the σ-algebra $\mathcal{F}$ – and the measure μ, to lock down a measure space. When the context is clear, we often say that A is measurable or μ-measurable if $A \in \mathcal{F}$.

A source of general measures is e.g. probability theory. A measure μ with $\mu(X) = 1$ is called a probability measure. On the other hand, the so-called Lebesgue measure on $\mathbb{R}^d$ makes rigorous the notion of area, volume, etc. It is extremely useful to be able study abstract measure spaces, since there are many interesting measures, and it is important that our general theory is as flexible as possible.

2.1.2 Remark. A measure μ is automatically *monotonic*. If $A \subset B$, $A, B \in \mathcal{F}$, then by additivity

$$\mu(B) = \mu(A) + \mu(B \setminus A) \geq \mu(A).$$

A measure μ is also automatically *subadditive*. If $A = \bigcup_{i=1}^{\infty} A_i$, $A_i \in \mathcal{F}$, we write $A = \bigcup_{i=1}^{\infty} B_i$, where $B_i \subset A_i$, $B_i \in \mathcal{F}$ and the sets B_i are disjoint. By additivity and monotonicity we have

$$\mu(A) = \sum_i \mu(B_i) \leq \sum_i \mu(A_i).$$

Let $A \subset B$, $A, B \in \mathcal{F}$ and $\mu(A) < \infty$. Then the formula (used already above)

$$\mu(B) = \mu(A) + \mu(B \setminus A)$$

implies that

$$\mu(B \setminus A) = \mu(B) - \mu(A).$$

Notice that this makes sense written like this as $\mu(A) < \infty$.

2.1.3 Example. We will see interesting concrete measures later. For instance, look at Appendix A, which constructs the most famous measure, the Lebesgue measure in an elementary (but not necessarily easy) way.

Easy examples come from probability. A measure space $(X, \mathcal{F}, \mu)$ is called a probability space if $\mu(X) = 1$, and then $\mu\colon \mathcal{F} \to [0, 1]$ is called a probability measure (notice that monotonicity implies $\mu(A) \leq \mu(X) = 1$ for all $A \in \mathcal{F}$). Suppose we e.g. throw two dices. Then the natural event space is $X = \{(i, j) : i, j \in \{1, \ldots, 6\}\}$. We can define the associated probability measure by setting $\mathcal{F} := \mathcal{P}(X)$ and

$$\mu(A) := \frac{\#A}{36}, \qquad A \subset X,$$

where $\#A$ denotes the number of elements in A. You will check in the exercises that the counting measure $A \mapsto \#A$ is a measure, and so then is the constant multiple μ. In some simple instances like this, we can choose the biggest possible σ-algebra $\mathcal{P}(X)$ and still have additivity. We will see later that in more complicated situations it becomes necessary to choose $\mathcal{F}$ more carefully to obtain countable additivity in $\mathcal{F}$. For instance, that is the case for the Lebesgue measure (see Appendix B).

2.1.4 Definition. Let $(X, \mathcal{F}, \mu)$ be a measure space. We say that a property $P = P(x)$, $x \in X$, holds almost everywhere (with respect to the measure μ) in a set F, if there is $N \in \mathcal{F}$ with $\mu(N) = 0$ so that the property P holds on $F \setminus N$. We often write this by saying that P holds for μ-a.e. $x \in F$ – if the measure is clear from the context, we might only say that P holds almost everywhere in F. This says that the property holds everywhere in the given set F except possibly in a small set (from the point of view of the measure μ).

There is a technical detail. The above definition does not insist that

$$\{x \in F\colon P(x) \text{ does not hold}\} \in \mathcal{F}.$$

Indeed, in a general measure space it might be possible that

$$\{x \in F \colon P(x) \text{ does not hold}\} \subset N$$

for $N \in \mathcal{F}$ with $\mu(N) = 0$, but that $\{x \in F \colon P(x) \text{ does not hold}\} \notin \mathcal{F}$.

Ideally, we would like that subsets of measurable sets of measure zero would also belong to $\mathcal{F}$ (and then by monotonicity they would automatically have measure zero). Motivated by this, we make the following definition.

2.1.5 Definition. A measure space $(X, \mathcal{F}, \mu)$ is complete if $A \in \mathcal{F}$, $\mu(A) = 0$ and $B \subset A$ implies $B \in \mathcal{F}$.

Luckily, studying only complete measure spaces is not a significant restriction – it turns out that we can always add more stuff to $\mathcal{F}$ so that it becomes complete (we can complete the σ-algebra), and we can extend our measure to act on this bigger σ-algebra. We do this now.

Let $(X, \mathcal{F}, \mu)$ be a measure space and denote

$$\mathcal{N} = \{N \in \mathcal{F} \colon \mu(N) = 0\}$$

and

$$\mathcal{N}' = \{N' \subset X \colon \exists N \in \mathcal{N} \text{ with } N' \subset N\}.$$

Notice that completeness means precisely that $\mathcal{N}' = \mathcal{N}$.

2.1.6 Lemma. *Suppose $(X, \mathcal{F}, \mu)$ is a measure space. Then the collection*

$$\overline{\mathcal{F}} = \{F \cup N' \colon F \in \mathcal{F}, N' \in \mathcal{N}'\}$$

is a σ-algebra and there exists a unique measure $\overline{\mu}$ on $\overline{\mathcal{F}}$ so that $(X, \overline{\mathcal{F}}, \overline{\mu})$ is a complete measure space and

$$\overline{\mu}(F) = \mu(F)$$

whenever $F \in \mathcal{F}$.

Proof. We begin by proving that $\overline{\mathcal{F}}$ is a σ-algebra. It is clear that $\emptyset \in \overline{\mathcal{F}}$ and that $\overline{\mathcal{F}}$ is closed under countable unions (since both $\mathcal{F}$ and $\mathcal{N}'$ are). Notice then that if $F \cup N' \in \overline{\mathcal{F}}$, we have

$$X \setminus (F \cup N') = [X \setminus (F \cup N)] \cup [N \setminus (N' \cup F)],$$

where $X \setminus (F \cup N) \in \mathcal{F}$ as $F \cup N \in \mathcal{F}$ and $N \setminus (N' \cup F) \subset N$, $N \in \mathcal{F}$, $\mu(N) = 0$. This proves that $X \setminus (F \cup N') \in \overline{\mathcal{F}}$ and so $\overline{\mathcal{F}}$ is a σ-algebra.

We now define

$$\overline{\mu}(F \cup N') := \mu(F), \qquad F \cup N' \in \overline{\mathcal{F}}.$$

We have to check that this is well-defined. To this end, suppose that $F_1 \cup N_1' = F_2 \cup N_2'$, where $F_i \in \mathcal{F}$ and $N_i' \subset N_i$, $N_i \in \mathcal{F}$, $\mu(N_i) = 0$. We have to show that $\mu(F_1) = \mu(F_2)$. Notice that

$$F_1 \subset F_1 \cup N_1' = F_2 \cup N_2' \subset F_2 \cup N_2,$$

and so

$$\mu(F_1) \le \mu(F_2 \cup N_2) = \mu(F_2),$$

since

$$\mu(F_2) \le \mu(F_2 \cup N_2) \le \mu(F_2) + \mu(N_2) = \mu(F_2).$$

Similarly, we also have $\mu(F_2) \le \mu(F_1)$, and so $\mu(F_1) = \mu(F_2)$.

It is clear that $\overline{\mu}$ extends μ on $\mathcal{F}$. It is also easy to verify that $\overline{\mu}$ is a measure and we omit this. We now show that $\overline{\mu}$ (or rather the corresponding measure space) is complete. Let $A \subset F \cup N'$, where $F \cup N' \in \overline{\mathcal{F}}$ satisfies $0 = \overline{\mu}(F \cup N') = \mu(F)$. Notice that $A = A \cap (F \cup N') = (A \cap F) \cup (A \cap N')$, where $A \cap F \subset F$, $F \in \mathcal{F}$ with $\mu(F) = 0$, and $A \cap N' \subset N$, where $N \in \mathcal{F}$ with $\mu(N) = 0$. It follows that $A \in \mathcal{N}' \subset \overline{\mathcal{F}}$. This proves the completeness.

Finally, we show the uniqueness. Let ν be another complete measure on $\overline{\mathcal{F}}$ extending μ on $\mathcal{F}$. Let $F \cup N' \in \overline{\mathcal{F}}$ be arbitrary – then we have

$$\nu(F \cup N') \leq \nu(F) + \nu(N) = \mu(F) = \overline{\mu}(F \cup N').$$

The reverse inequality follows by symmetry and so $\overline{\mu} = \nu$. □

2.2 Outer measures

Sometimes, it is natural to define a set function on the whole power set $\mathcal{P}(X)$. But such a function may easily fail to satisfy countable additivity. However, we might then be able to identify exactly the sets where countable additivity holds, and build an actual measure by restricting to these sets. This is the idea of outer measures.

2.2.1 Definition. We say that $\mu \colon \mathcal{P}(X) \to [0, \infty]$ is an outer measure on X if the following holds.

1. We have $\mu(\emptyset) = 0$.
2. (monotonicity) If $A \subset B \subset X$, then $\mu(A) \leq \mu(B)$.
3. (countable subadditivity) If $A = \bigcup_{i=1}^{\infty} A_i$, $A_i \subset X$, then

$$\mu(A) \leq \sum_i \mu(A_i).$$

2.2.2 Definition. Given an outer measure μ on X we say that $E \subset X$ is μ-measurable if for all $A \subset X$ we have

$$\mu(A) = \mu(A \cap E) + \mu(A \setminus E).$$

We define the collection of μ-measurable sets

$$\mathcal{M}_\mu(X) := \{E \subset X \colon E \text{ is } \mu\text{-measurable}\}. \tag{2.2.3}$$

2.2.4 Remark. As we always have $A = (A \cap E) \cup (A \setminus E)$ it follows that for all $A, E \subset X$ we have by the subadditivity of the outer measure μ that

$$\mu(A) \leq \mu(A \cap E) + \mu(A \setminus E).$$

We thus have $E \in \mathcal{M}_\mu(X)$ if for all A with $\mu(A) < \infty$ we have

$$\mu(A) \geq \mu(A \cap E) + \mu(A \setminus E).$$

It is often useful to only aim to prove the above inequality, instead of the original equality.

2.2.5 Example. Let X be a nonempty set. Define $\mu : \mathcal{P}(X) \to \{0, 1\}$ by setting $\mu(\emptyset) = 0$ and $\mu(A) = 1$ if $A \subset X$, $A \neq \emptyset$. So we assign 1 to every non-trivial set – this seems like a very rough measure of the size of a set. We will prove that μ is an outer measure and study the μ-measurable sets $\mathcal{M}_\mu$.

By definition $\mu(\emptyset) = 0$. Let then $A \subset B$. If $A = \emptyset$ we have $\mu(A) = 0 \leq \mu(B)$ and otherwise $\mu(A) = 1 = \mu(B)$ (as also $B \neq \emptyset$ due to $A \subset B$). Thus, we have $\mu(A) \leq \mu(B)$. Let then $A = \bigcup_{i=1}^{\infty} A_i$. If $A_i = \emptyset$ for every i, we have $\mu(A) = \mu(\emptyset) = 0 = \sum_i 0 = \sum_i \mu(A_i)$. Otherwise, there is (at least one) i_0 with $A_{i_0} \neq \emptyset$ and now $\mu(A) = 1 = \mu(A_{i_0}) \leq \sum_i \mu(A_i)$. We have shown that μ is an outer measure.

Next, we study $\mathcal{M}_\mu$. We have that $\emptyset$ and X are μ-measurable: $\emptyset, X \in \mathcal{M}_\mu$. This follows e.g. from the general fact that we prove below: $\mathcal{M}_\mu$ is always a σ-algebra if μ is an outer measure. And we just proved that μ is an outer measure. However, we next prove that there are no other μ-measurable sets.

To this end, suppose $\emptyset \neq A \neq X$. Then we have

$$\mu(X) = 1 < 2 = \mu(A) + \mu(X \setminus A) = \mu(X \cap A) + \mu(X \setminus A),$$

and so A is not μ-measurable (the measurability condition fails with the test set X). Thus, $\mathcal{M}_\mu = \{\emptyset, X\}$ – this shows that if your outer measure is too silly, almost no sets need to be measurable.

What is important now, as we will next prove, is that $\mathcal{M}_\mu(X)$ is a σ-algebra and the restriction $\mu\colon \mathcal{M}_\mu(X) \to [0, \infty]$ is a measure.

2.2.6 Remark. This is a general way to construct measures from outer measures. However, not every measure is necessarily constructed from some outer measure with this process. In fact, notice that this process always yields a complete measure space $(X, \mathcal{M}_\mu, \mu|\mathcal{M}_\mu)$, where $\mu|\mathcal{M}_\mu$ is the restriction of μ to $\mathcal{M}_\mu$ (often denoted just by μ). We show the completeness claim next.

If $E \subset X$ satisfies $\mu(E) = 0$, then $E \in \mathcal{M}_\mu(X)$. This follows by noticing that for an arbitrary $A \subset X$ we have

$$\mu(A) \geq \mu(A \setminus E) = \mu(A \cap E) + \mu(A \setminus E),$$

where the inequality follows from monotonicity as $A \setminus E \subset A$ and the equality from the fact that $\mu(A \cap E) = 0$ since $0 \leq \mu(A \cap E) \leq \mu(E) = 0$ by monotonicity.

2.2.7 Theorem. *Let μ be an outer measure on X. The collection of μ-measurable sets $\mathcal{M}_\mu(X)$ is a σ-algebra on X. Moreover, if the sets $E_1, E_2, \ldots \in \mathcal{M}_\mu(X)$ are disjoint ($E_i \cap E_j = \emptyset$ for $i \neq j$), then we have countable additivity*

$$\mu\Big(\bigcup_{i=1}^{\infty} E_i\Big) = \sum_{i=1}^{\infty} \mu(E_i).$$

Proof. We will move in small steps.

Claim 1. In the remark above we showed that if $E \subset X$ satisfies $\mu(E) = 0$, then $E \in \mathcal{M}_\mu(X)$. In particular, as $\mu(\emptyset) = 0$ we have $\emptyset \in \mathcal{M}_\mu(X)$.

Claim 2. A set $E \subset X$ is measurable if and only if $X \setminus E$ is measurable. This follows directly from the definition as e.g. $A \setminus E = A \cap (X \setminus E)$.

Claim 3. If $E_1, E_2, \ldots, E_N$ are measurable, then $\bigcup_{i=1}^{N} E_i$ is measurable. Clearly, it is enough to prove that $E_1 \cup E_2$ is measurable (as then e.g. $E_1 \cup E_2 \cup E_3 = (E_1 \cup E_2) \cup E_3$ is measurable and so on). Let $A \subset X$ be a test set for measurability. By the measurability of E_2 applied with the test set $A \cap (X \setminus E_1)$ we have

$$\mu(A \setminus E_1) = \mu(A \cap (X \setminus E_1)) = \mu(A \cap (X \setminus E_1) \cap E_2) + \mu((A \cap (X \setminus E_1)) \setminus E_2).$$

Using the measurability of E_1 with the test set A first, then the above formula and then subadditivity we have

$$\begin{aligned}
\mu(A) &= \mu(A \cap E_1) + \mu(A \setminus E_1) \\
&= \mu(A \cap E_1) + \mu(A \cap (X \setminus E_1) \cap E_2) + \mu((A \cap (X \setminus E_1)) \setminus E_2) \\
&\geq \mu((A \cap E_1) \cup (A \cap (X \setminus E_1) \cap E_2)) + \mu(A \setminus (E_1 \cup E_2)).
\end{aligned}$$

It remains to notice that

$$(A \cap E_1) \cup (A \cap (X \setminus E_1) \cap E_2) = A \cap (E_1 \cup E_2).$$

Thus, $E_1 \cup E_2$ is measurable.

Claim 4. If $E_1, E_2, \ldots, E_N$ are measurable, then $\bigcap_{i=1}^N E_i$ is measurable. Follows from Claim 2 and Claim 3 as in Remark 2.0.2.

Claim 5. If E and F are measurable, then $E \setminus F$ is measurable. This follows using Claim 2 and Claim 4 as $E \setminus F = E \cap (X \setminus F)$.

Claim 6. If $E_1, E_2, \ldots, E_N$ are measurable and disjoint and $A \subset X$ is arbitrary, then

$$\mu\Big(A \cap \bigcup_{i=1}^N E_i\Big) = \sum_{i=1}^N \mu(A \cap E_i).$$

Notice that we want this in the generality that A does not need to be measurable (we need this later). Apply the measurability of E_1 with the test set $A \cap (E_1 \cup E_2)$ to get

$$\mu(A \cap (E_1 \cup E_2)) = \mu((A \cap (E_1 \cup E_2)) \cap E_1) + \mu((A \cap (E_1 \cup E_2)) \setminus E_1) = \mu(A \cap E_1) + \mu(A \cap E_2),$$

where we used the disjointness to conclude that $(A \cap (E_1 \cup E_2)) \setminus E_1 = A \cap E_2$. This is the claim for $N = 2$. The general case follows, since e.g.

$$\mu(A \cap (E_1 \cup E_2 \cup E_3)) = \mu(A \cap (E_1 \cup E_2)) + \mu(A \cap E_3) = \sum_{i=1}^3 \mu(A \cap E_i),$$

where we used the case $N = 2$ twice.

Claim 7. If $E_1, E_2, \ldots$ are measurable, then $\bigcup_{i=1}^\infty E_i$ is measurable. If the sets are, in addition, disjoint, then we have

$$\mu\Big(\bigcup_{i=1}^\infty E_i\Big) = \sum_{i=1}^\infty \mu(E_i).$$

We start with the measurability of the union (which is the last fact that we need to show that $\mathcal{M}_\mu(X)$ is a σ-algebra on X). By defining $F_1 = E_1$, $F_2 = E_2 \setminus E_1$, $F_3 = E_3 \setminus (E_1 \cup E_2)$, and so on, we get measurable sets (by previous claims) $F_i \subset E_i$ that are disjoint and satisfy

$$E := \bigcup_{i=1}^\infty E_i = \bigcup_{i=1}^\infty F_i.$$

Define also the measurable sets

$$S_k = \bigcup_{i=1}^k F_i \subset E.$$

Let A be a test set for measurability. Then we have by the measurability of S_k that

$$\mu(A) = \mu(A \cap S_k) + \mu(A \setminus S_k).$$

By monotonicity $\mu(A \setminus S_k) \geq \mu(A \setminus E)$ and by Claim 6 $\mu(A \cap S_k) = \sum_{i=1}^k \mu(A \cap F_i)$. Thus, we have

$$\mu(A) \geq \sum_{i=1}^k \mu(A \cap F_i) + \mu(A \setminus E).$$

Letting $k \to \infty$ we get

$$\mu(A) \geq \sum_{i=1}^{\infty} \mu(A \cap F_i) + \mu(A \setminus E). \tag{2.2.8}$$

Using now subadditivity we get

$$\mu(A) \geq \mu\Big(\bigcup_{i=1}^{\infty}(A \cap F_i)\Big) + \mu(A \setminus E) = \mu(A \cap E) + \mu(A \setminus E).$$

This shows the measurability of E.

Finally, we move to the countable additivity. Notice that (2.2.8) applied with $A = E$ gives

$$\mu(E) \geq \sum_{i=1}^{\infty} \mu(E \cap F_i) + \mu(E \setminus E) = \sum_{i=1}^{\infty} \mu(F_i).$$

The reverse inequality holds by subadditivity. Thus, we have

$$\mu(E) = \sum_{i=1}^{\infty} \mu(F_i).$$

If the sets E_i are already disjoint we have $F_i = E_i$, so we are done. □

Next, we note the following easy to way to construct outer measures.

2.2.9 Lemma. *Suppose $\mathcal{S} \subset \mathcal{P}(X)$ is an arbitrary collection of subsets of X with $\emptyset \in \mathcal{S}$ and $h\colon \mathcal{S} \to [0, \infty]$ is a function with $h(\emptyset) = 0$. For $A \subset X$ define*

$$\mu(A) := \inf\Big\{\sum_{i=1}^{\infty} h(S_i)\colon\ A \subset \bigcup_{i=1}^{\infty} S_i,\ S_i \in \mathcal{S}\Big\}.$$

By convention $\inf \emptyset = \infty$ *– meaning $\mu(A) = \infty$ if there is no covering of A by sets $S_i \in \mathcal{S}$. Then μ is an outer measure.*

Proof. Exercise. □

One way to construct the famous Lebesgue outer measure on $\mathbb{R}^d$ is to use the above construction as follows. A rectangle on $\mathbb{R}^d$ (with sides parallel to the coordinate axes) is a set R of the form

$$R = I_1 \times \cdots \times I_d = \prod_{i=1}^{d} I_i,$$

where I_i is an interval with endpoints $-\infty < a_i < b_i < \infty$. Define

$$\mathrm{vol}(R) := \prod_{i=1}^{d}(b_i - a_i).$$

Now use the above construction with $\mathcal{S}$ consisting of all rectangles and $h(R) := \mathrm{vol}(R)$. The resulting measure μ is the Lebesgue outer measure and the associated measurable sets $\mathcal{M}_\mu(\mathbb{R}^d) =: \mathrm{Leb}(\mathbb{R}^d)$ are called the Lebesgue measurable sets. This is the natural measure on $\mathbb{R}^d$ that gives the notion of d-dimensional volume. In this course we will postpone the careful study of the Lebesgue measure – in fact, we will later get the Lebesgue outer measure from a different, deeper, construction which has the big benefit that it directly gives that $\mathrm{Leb}(\mathbb{R}^d)$ is a big collection of sets (e.g. contains all open sets) and that the Lebesgue measure satisfies various natural regularity properties. If you are interested to see an elementary construction of the Lebesgue measure, together with all of its key properties, it is included in Appendix A. The appendix is readable (does not require other background material) after you finish the current chapter.

2.3 Convergence results for measures

Many results hold in the abstract setting of a (complete) measure space $(X, \mathcal{F}, \mu)$. For example, we will later develop integration theory in this abstract setting. However, some finer properties of measures may hold only for some particular measures, like the Lebesgue measure. The following are key convergence properties that hold for all measures.

2.3.1 Theorem. *Let $(X, \mathcal{F}, \mu)$ be a measure space.*

1. *Let $A_i \in \mathcal{F}$ and $A_1 \subset A_2 \subset \cdots$. Then we have*

$$\mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \lim_{i\to\infty} \mu(A_i).$$

2. *Let $A_i \in \mathcal{F}$, $A_1 \supset A_2 \supset \cdots$ and $\mu(A_1) < \infty$. Then we have*

$$\mu\Big(\bigcap_{i=1}^{\infty} A_i\Big) = \lim_{i\to\infty} \mu(A_i).$$

Proof. For (1) notice that setting $A_0 = \emptyset$ we have by additivity that

$$\begin{aligned}
\mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) &= \mu\Big(\bigcup_{i=1}^{\infty}[A_i \setminus A_{i-1}]\Big) \\
&= \sum_{i=1}^{\infty} \mu(A_i \setminus A_{i-1}) \\
&= \lim_{N\to\infty} \sum_{i=1}^{N} \mu(A_i \setminus A_{i-1}) = \lim_{N\to\infty} \mu\Big(\bigcup_{i=1}^{N}[A_i \setminus A_{i-1}]\Big) = \lim_{N\to\infty} \mu(A_N).
\end{aligned}$$

For (2) we first define $B_i = A_1 \setminus A_i$ to obtain an increasing sequence of sets $B_1 \subset B_2 \subset \cdots$. Notice now that

$$\begin{aligned}
\mu\Big(\bigcap_{i=1}^{\infty} A_i\Big) &= \mu(A_1) - \mu\Big(A_1 \setminus \bigcap_{i=1}^{\infty} A_i\Big) \\
&= \mu(A_1) - \mu\Big(\bigcup_{i=1}^{\infty} B_i\Big) \\
&= \mu(A_1) - \lim_{j\to\infty} \mu(B_j) = \mu(A_1) - \lim_{i\to\infty}[\mu(A_1) - \mu(A_i)] = \lim_{j\to\infty} \mu(A_i).
\end{aligned}$$

□

2.3.2 Remark. The assumption $\mu(A_1) < \infty$ is necessary in (2) above. Think carefully where we used $\mu(A_1) < \infty$ in the above proof.

2.4 Borel sets

When the space X is such that the concept of open sets make sense – X is a metric space or more generally a topological space – then we often want that open sets are measurable. This leads to the concept of a Borel set.

2.4.1 Definition. Given a collection of sets Γ on a space X the σ-algebra generated by Γ is

$$\sigma(\Gamma) := \bigcap \mathcal{F},$$

where the intersection is taken over all σ-algebras $\mathcal{F}$ satisfying $\Gamma \subset \mathcal{F}$.

2.4.2 Remark. It is easy to check that $\sigma(\Gamma)$ is a σ-algebra on X, and clearly it is the smallest σ-algebra containing Γ.

2.4.3 Definition. If X is a topological space we denote by $\mathrm{Bor}(X)$ the σ-algebra generated by open sets on X. These are called Borel sets.

The idea of $\mathrm{Bor}(X)$ is that if we want that open sets are measurable, then already a lot more sets have to be measurable – namely all of the Borel sets. Indeed, if $\mathcal{F}$ is a σ-algebra with

$$\{V \subset X : V \text{open}\} \subset \mathcal{F},$$

then already

$$\mathrm{Bor}(X) \subset \mathcal{F}$$

by the definition of $\mathrm{Bor}(X)$.

1. Countable unions of closed sets $\bigcup_i F_i$ are Borel sets. These are called $\mathcal{F}_\sigma$-sets. For example, a half-open interval in $\mathbb{R}$ is an $\mathcal{F}_\sigma$-set:

$$[a, b) = \bigcup_{i=1}^{\infty} \Big[a, b - \frac{1}{i}\Big].$$

2. Countable intersections of open sets $\bigcap_i G_i$ are Borel sets. These are called $\mathcal{G}_\delta$-sets. For example, a half-open interval is also a $\mathcal{G}_\delta$-set:

$$[a, b) = \bigcap_{i=1}^{\infty} \Big(a - \frac{1}{i}, b\Big).$$

3. The $\mathcal{F}_{\sigma\delta}$-sets are of the form $\bigcap_j A_j$, where $A_j \in \mathcal{F}_\sigma$.
4. The $\mathcal{G}_{\delta\sigma}$-sets are of the form $\bigcup_j B_j$, where $B_j \in \mathcal{G}_\delta$.

All of these more complicated sets are Borel sets. And if you want that open sets belong to some σ-algebra, all of these sets have to belong there as well.

Chapter 3

Measurable functions

We want to integrate a function against a measure. But we cannot quite integrate all functions – we need them to be measurable. That is what we study in this section.

We denote the extended real numbers by $\dot{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$. Let $(X, \mathcal{F}, \mu)$ be a measure space througout this section.

3.0.1 Definition. A function $f\colon X \to \dot{\mathbb{R}}$ is measurable if

$$f^{-1}[-\infty, a) = \{x \in X\colon f(x) < a\} = \{f < a\} \in \mathcal{F}$$

for every $a \in \mathbb{R}$.

3.0.2 Remark. In fact, we should talk about $\mathcal{F}$-measurability to be strict. The definition does not depend on the whole ambient measure space $(X, \mathcal{F}, \mu)$ – it only depends on $\mathcal{F}$.

If $E \in \mathcal{F}$ and $f\colon E \to \dot{\mathbb{R}}$, then measurability ($\mathcal{F}$-measurability) means that $\{f < a\} = \{x \in E\colon f(x) < a\} \in \mathcal{F}$ for every $a \in \mathbb{R}$. If $f\colon X \to \dot{\mathbb{R}}$ is measurable, then so is the restriction $f|E$ of f to E, since

$$\{x \in E\colon f(x) < t\} = E \cap \{x \in X\colon f(x) < t\} \in \mathcal{F}.$$

If f is measurable, the preimages of many more sets than just these intervals belong to $\mathcal{F}$. We can establish this using the following.

3.0.3 Lemma. *Let $f\colon X \to \dot{\mathbb{R}}$ and*

$$\Gamma_f := \{M \subset \dot{\mathbb{R}}\colon f^{-1}M \in \mathcal{F}\}.$$

Then Γ_f is a σ-algebra on $\dot{\mathbb{R}}$.

Proof. This is immediate as $f^{-1}(\dot{\mathbb{R}} \setminus M) = X \setminus f^{-1}M$ and $f^{-1}\bigcup_i M_i = \bigcup_i f^{-1}M_i$. □

3.0.4 Corollary. *If $f\colon X \to \dot{\mathbb{R}}$ is measurable then $f^{-1}B \in \mathcal{F}$ for every Borel set $B \subset \mathbb{R}$. In addition, $f^{-1}(\infty) \in \mathcal{F}$ and $f^{-1}(-\infty) \in \mathcal{F}$.*

Proof. Let $a < b$. Notice then that

$$[a, b) = [a, \infty] \cap [-\infty, b) = (\dot{\mathbb{R}} \setminus [-\infty, a)) \cap [-\infty, b) \in \Gamma_f,$$

since Γ_f is a σ-algebra. We can write an open set in $\mathbb{R}$ as a (disjoint) union of half-open intervals – this elementary fact is proved later in Lemma 6.0.1. As Γ_f is a σ-algebra, it contains these countable unions – hence it contains all open sets. As $\mathrm{Bor}(\mathbb{R})$ is the smallest σ-algebra containing all open sets, we must have $\mathrm{Bor}(\mathbb{R}) \subset \Gamma_f$. We also e.g. have

$$\{-\infty\} = \bigcap_{i=1}^{\infty} [-\infty, -i) \in \Gamma_f.$$

□

3.0.5 Remark. In the definition of measurability it makes no difference whether we use $\{f < a\}$, $\{f \leq a\}$, $\{f > a\}$ or $\{f \geq a\}$ in the sense that no matter which definition we choose, we can prove the above much stronger statement – preimages of all Borel sets of $\mathbb{R}$ are measurable and not just preimages of certain intervals. Corollary 3.0.4 is what is used in practice when you know that a function is measurable. However, to prove measurability it is often easier to prove the more restricted condition of the definition.

3.0.6 Remark. This formulation is very convenient. Soon we show that if e.g. $f, g \colon X \to \mathbb{R}$ are measurable, then $f - g$ is measurable. It follows that e.g. $\{x \colon f(x) = g(x)\}$ is measurable as it can be written in the form $(f-g)^{-1}\{0\}$, where $\{0\} \in \mathrm{Bor}(\mathbb{R})$.

We have to be a little bit careful when talking about the measurability of compositions of functions. Suppose e.g. $f \colon \mathbb{R} \to \mathbb{R}$ is Borel measurable and $g \colon X \to \mathbb{R}$ is $\mathcal{F}$-measurable. Then $f \circ g \colon X \to \mathbb{R}$ is also $\mathcal{F}$-measurable as for every Borel $B \subset \mathbb{R}$ we have

$$(f \circ g)^{-1} B = g^{-1}(f^{-1} B) \in \mathcal{F}, \tag{3.0.7}$$

since $f^{-1} B \in \mathrm{Bor}(\mathbb{R})$.

3.0.8 Example. Given a set E the characteristic function 1_E, which equals 1 for $x \in E$ and 0 otherwise, is measurable if and only if the set E is measurable.

3.0.9 Proposition. *Let $f, g \colon X \to \mathbb{R}$ be real-valued measurable functions and $\alpha \in \mathbb{R}$. Then the functions*

$$\alpha f, f + g, fg, f/g$$

are measurable, where we assume $g \neq 0$ in the case of f/g.

Proof. We omit the easy case of αf.

For $f + g$ we can write

$$\{f + g < a\} = \bigcup_{q,r \in \mathbb{Q} \colon \; q + r < a} [\{f < q\} \cap \{g < r\}],$$

which is a countable union of measurable sets, hence measurable.

We notice that f^2 is measurable as for $a \geq 0$ we have

$$\{f^2 < a\} = \{-\sqrt{a} < f < \sqrt{a}\}.$$

This, in turn, implies that

$$fg = \frac{1}{2}[(f+g)^2 - f^2 - g^2]$$

is measurable. It is straightforward from the definition that $1/g$ is measurable (exercise), and so the product f/g is measurable. □

3.0.10 Remark. The conclusion remains the same even if the functions would take values in the extended real numbers. The only requirement is that the said functions need to be well-defined – e.g. for $f + g$ this requires that an undefined case like $\infty - \infty$ cannot occur.

3.0.11 Proposition. *If $f, g\colon X \to \dot{\mathbb{R}}$ are measurable, then*

$$\min(f,g),\ \max(f,g) \qquad \text{and} \qquad |f|$$

are measurable.

Proof. Simply notice that

$$\{\min(f,g) < a\} = \{f < a\} \cup \{g < a\}$$

and

$$\{\max(f,g) < a\} = \{f < a\} \cap \{g < a\}$$

are measurable. It then follows that

$$|f| = \max(f,0) - \min(f,0)$$

is measurable. □

3.0.12 Proposition. *If $f\colon X \to \dot{\mathbb{R}}$ is measurable, then $|f|^p$ is measurable for all $p > 0$.*

Proof. Define $\varphi(x) = |x|^p$ – this is a continuous, especially a Borel measurable, function (preimages of open sets are even open by continuity). Hence the composition $\varphi \circ f$ is measurable. □

Finally, we will need that various limit operations of measurable functions remain measurable. Below, the functions are, of course, defined pointwise – for instance $(\sup_{j\in\mathbb{N}} f_j)(x) = \sup_{j\in\mathbb{N}} f_j(x)$.

3.0.13 Lemma. *If $f_j\colon X \to \dot{\mathbb{R}}$ are measurable, then the functions*

$$\sup_{j\in\mathbb{N}} f_j,\ \inf_{j\in\mathbb{N}} f_j,\ \limsup_{j\to\infty} f_j, \liminf_{j\to\infty} f_j$$

are measurable.

Proof. First, notice that

$$\{\sup_j f_j \le a\} = \bigcap_j \{f_j \le a\}$$

is measurable as a countable intersection of measurable sets. Notice that we used $\le$ here instead of $<$ (as in the definition), since it was important for the identity to be true, but that this still implies measurability by Remark 3.0.5. It follows that $\inf_j f_j = -\sup_j(-f_j)$ is also measurable.

Now it follows that also

$$\limsup_j f_j = \inf_J \sup_{j\ge J} f_j$$

and

$$\liminf_j f_j = \sup_J \inf_{j\ge J} f_j$$

are measurable. □

3.0.14 Corollary. *If $f_j \to f$ pointwise everywhere and the functions f_j are measurable, then f is measurable.*

Proof. Follows from above as $\lim_j f_j = \limsup_j f_j = \liminf_j f_j$. □

Limiting operations are crucial to analysis, and the above shows that the class of measurable functions is large enough for the purposes of modern analysis. Of course, this fails miserably e.g. for continuous functions.

The following is a good reason to work in a complete measure space.

3.0.15 Lemma. *Suppose $(X, \mathcal{F}, \mu)$ is a complete measure space, $f\colon X \to \dot{\mathbb{R}}$ and $f = g$ μ-a.e., where $g\colon X \to \mathbb{R}$ is measurable. Then f is measurable.*

Proof. Let $N \in \mathcal{F}$ be such that $f = g$ on $X \setminus N$ and $\mu(N) = 0$. Now, we have

$$\{f < a\} = \{x \in X \setminus N\colon g(x) < a\} \cup \{x \in N\colon f(x) < a\}.$$

Since g is measurable and $X \setminus N$ is measurable, the set

$$\{x \in X \setminus N\colon g(x) < a\} = (X \setminus N) \cap \{g < a\}$$

is measurable. Next, we have

$$\{x \in N\colon f(x) < a\} \subset N,$$

where $N \in \mathcal{F}$ and $\mu(N) = 0$, and so by completeness $\{x \in N\colon f(x) < a\} \in \mathcal{F}$. □

Often we essentially identify functions that agree almost everywhere – this can rigorously be done by studying the equivalence classes of functions that agree almost everywhere. In practice, we work with actual functions (instead of equivalence classes), but think that f and g are for all practical purposes the same function if $f = g$ almost everywhere (for example, then their integerals will agree).

3.1 Simple functions

Later, it will be obvious how to define the integral of a simple function. On the other hand, it turns out that arbitrary non-negative measurable functions can be obtained as increasing limits of simple functions. These two facts will be the key to defining the integral of a general measurable function.

3.1.1 Definition. A simple function is a measurable function $s\colon X \to \mathbb{R}$ that takes only finitely many distinct values.

Obviously, sums and products of simple functions are simple. Moreover, a function of the form

$$\sum_{i=1}^{N} c_i 1_{E_i}$$

for some scalars $c_1, \ldots, c_N$ and measurable sets $E_1, \ldots, E_N$ is simple.

A complication is that a simple function can be written in many ways, e.g. $0 = 1_E - 1_E$ for any measurable set E. However, there is a natural canonical form. Let $c_1, \ldots, c_N$ be the distinct values of a given simple function s and define

$$E_i := s^{-1}\{c_i\} = \{x\colon s(x) = c_i\}.$$

Then the sets E_i are pairwise disjoint, $X = \bigcup_i E_i$ and we can write

$$s = \sum_{i=1}^{N} c_i 1_{E_i}.$$

We call the above form the standard/canonical representation of s. For example, the function $s := 1_{[0,10]} + 1_{[0,100]}$ is simple and its standard representation is

$$s = 2 \cdot 1_{[0,10]} + 1 \cdot 1_{(10,100]} + 0 \cdot 1_{\mathbb{R}\setminus[0,100]}.$$

We now prove the key approximation result.

3.1.2 Lemma. *Let $f\colon X \to \dot{\mathbb{R}}$ be measurable and non-negative, $f \geq 0$. Then there are simple functions s_j satisfying*

$$0 \leq s_j(x) \leq s_{j+1}(x) \leq f(x)$$

and

$$f(x) = \lim_{j\to\infty} s_j(x)$$

for every $x \in X$.

Proof. The proof has a simple idea, the technical details can obscure it first. With a fixed j we define s_j by rounding f to the nearest integer multiple of 2^{-j} – however, as a simple function may take only finitely many values we stop this rounding at the height j.

We now execute this idea. Define

$$s_1(x) = \begin{cases} 0, & 0 \leq f(x) < 1, \\ 1, & f(x) \geq 1, \end{cases}$$

and

$$s_2(x) = \begin{cases} 0, & 0 \leq f(x) < 1/2, \\ 1/2, & 1/2 \leq f(x) < 1, \\ 1, & 1 \leq f(x) < 3/2, \\ 3/2, & 3/2 \leq f(x) < 2, \\ 2, & f(x) \geq 2. \end{cases}$$

Continue this in the obvious way obtaining the general definition

$$s_j(x) = \begin{cases} \frac{i-1}{2^j}, & \frac{i-1}{2^j} \leq f(x) < \frac{i}{2^j},\ i = 1, \ldots, j2^j, \\ j, & f(x) \geq j. \end{cases}$$

By the measurability of the function f the appearing sets, such as $f^{-1}[j, \infty]$, are measurable. Thus, s_j is a non-negative simple function.

By construction it is clear that

$$0 \leq s_j(x) \leq s_{j+1}(x) \leq f(x).$$

The convergence is also clear. Indeed, notice first that if $f(x) = \infty$ then $s_j(x) = j \to \infty = f(x)$. If $f(x) < \infty$, then choose $j_0 > f(x)$. For $j \geq j_0$ we have that

$$f(x) - 2^{-j} < s_j(x) \leq f(x)$$

and so $s_j(x) \to f(x)$. We are done. □

The above lemma, despite its simple proof, is very powerful. It means that we can essentially reduce to simple functions. It is extremely common in modern analysis to deal with general (possibly very rough) functions by approximating them by nicer functions – in this case by simple functions.

Chapter 4

Integration againts a general measure

The abstract integration theory that we now present can be developed in high generality. Indeed, here we work in an arbitrary fixed (complete) measure space $(X, \mathcal{F}, \mu)$, and the theory in this generality is no harder than it would be just in the case of the so-called Lebesgue measure (which is the case generalizing the Riemann integral).

4.1 The μ-integral of a non-negative simple function

4.1.1 Definition. Suppose s is simple and $s \geq 0$. Write s in its standard form

$$s = \sum_{i=1}^{N} c_i 1_{E_i}.$$

Then we define its μ-integral as follows:

$$\int_X s \, \mathrm{d}\mu = \int s \, \mathrm{d}\mu := \sum_{i=1}^{N} c_i \mu(E_i). \tag{4.1.2}$$

We use the standard representation of s to have an unambiguous definition of the integral – however, we will soon see that it does not matter (the same formula holds for all representations).

4.1.3 Remark. Notice that it is possible that $\mu(E_i) = \infty$. In this case, we use the convention

$$c_i \cdot \infty = \begin{cases} 0, & c_i = 0, \\ \infty, & c_i > 0. \end{cases}$$

With this convention, the right hand side of (4.1.2) equals ∞ if there exists $i = 1, \ldots, N$ with $c_i > 0$ and $\mu(E_i) = \infty$. Otherwise, the integral is a real number in the interval $[0, \infty)$.

4.1.4 Definition. For a simple non-negative function s and a measurable set E we define

$$\int_E s \, \mathrm{d}\mu := \int 1_E s \, \mathrm{d}\mu.$$

Notice that the above definition makes sense as $1_E s$ is simple. We will now prove important properties of the μ-integral of a simple function. Things will eventually be much smoother when we have proved all the natural properties of integrals and no longer have to work with the definition.

4.1.5 Lemma. *Let s and u be non-negative simple functions with $\int s\,\mathrm{d}\mu < \infty$ and $\int u\,\mathrm{d}\mu < \infty$. Let $\alpha \in \mathbb{R}$ be such that the simple function $s + \alpha u$ satisfies $s + \alpha u \geq 0$. Then we have the linearity of the integral:*

$$\int (s + \alpha u)\,\mathrm{d}\mu = \int s\,\mathrm{d}\mu + \alpha \int u\,\mathrm{d}\mu.$$

Proof. Write the standard representations of s and u:

$$s = \sum_{i=1}^{I} b_i 1_{B_i},$$
$$u = \sum_{j=1}^{J} c_j 1_{C_j}.$$

Let $d_1, \ldots, d_K$ be an enumeration of the set

$$\{b_i + \alpha c_j : i = 1, \ldots, I,\ j = 1, \ldots, J\}.$$

Given $k \in \{1, \ldots, K\}$ let $\mathcal{F}_k$ consist of the pairs (i, j) for which

$$d_k = b_i + \alpha c_j.$$

We obtain the standard representation

$$s + \alpha u = \sum_{k=1}^{K} d_k 1_{D_k}, \qquad D_k = \bigcup_{(i,j)\in\mathcal{F}_k} (B_i \cap C_j).$$

Having obtained the standard representation we have by definition that

$$\int (s + \alpha u)\,\mathrm{d}\mu = \sum_{k=1}^{K} d_k \mu(D_k) = \sum_{k=1}^{K} \sum_{(i,j)\in\mathcal{F}_k} (b_i + \alpha c_j)\mu(B_i \cap C_j),$$

where we used the additivity of the measure μ and the fact that the union in the definition of D_k consists of pairwise disjoint sets ($B_{i_1} \cap B_{i_2} \cap C_{j_1} \cap C_{j_2} = \emptyset$ if $i_1 \neq i_2$ or $j_1 \neq j_2$). Continuing, we have

$$\begin{aligned}
\int (s + \alpha u)\,\mathrm{d}\mu &= \sum_{i=1}^{I} \sum_{j=1}^{J} (b_i + \alpha c_j)\mu(B_i \cap C_j) \\
&= \sum_{i=1}^{I} b_i \sum_{j=1}^{J} \mu(B_i \cap C_j) + \alpha \sum_{j=1}^{J} c_j \sum_{i=1}^{I} \mu(B_i \cap C_j) \\
&= \sum_{i=1}^{I} b_i \mu(B_i) + \alpha \sum_{j=1}^{J} c_j \mu(C_j) = \int s\,\mathrm{d}\mu + \alpha \int u\,\mathrm{d}\mu,
\end{aligned}$$

where we used that $X = \bigcup_{i=1}^{I} B_i = \bigcup_{j=1}^{J} C_j$ and the unions are disjoint. □

4.1.6 Remark. If $\alpha \geq 0$ in Lemma 4.1.5, we do not need to assume that $\int s\,\mathrm{d}\mu < \infty$ and $\int u\,\mathrm{d}\mu < \infty$. This is an easy exercise that we omit here.

We have the following monotonicity, which will be key going forward.

4.1.7 Lemma. *If s, u are simple and $0 \le s \le u$, then*

$$\int s\,\mathrm{d}\mu \le \int u\,\mathrm{d}\mu.$$

Proof. If $\int u\,\mathrm{d}\mu = \infty$ the claim is trivial. We may thus assume $\int u\,\mathrm{d}\mu < \infty$ – then it also easily follows that $\int s\,\mathrm{d}\mu < \infty$ (inspect the definition). Applying Lemma 4.1.5 to $u - s \ge 0$ we have

$$0 \le \int (u-s)\,\mathrm{d}\mu = \int u\,\mathrm{d}\mu - \int s\,\mathrm{d}\mu$$

and the claim follows. The first inequality above follows simply from the definition (the integral of a non-negative simple function is non-negative). □

4.1.8 Lemma. *We have the following independence of the representation. If $s = \sum_{i=1}^N c_i 1_{E_i}$ for some scalars $c_1, \dots, c_N \in [0,\infty)$ and measurable sets $E_1, \dots, E_N$ (so this is not necessarily the standard representation of s), then*

$$\int s\,\mathrm{d}\mu = \sum_{i=1}^N c_i \mu(E_i).$$

Proof. Follows right away from the proved linearity of the integral and the fact that by definition $\int c_i 1_{E_i}\,\mathrm{d}\mu = c_i\mu(E_i)$. □

Notice that if A is measurable and $0 \le s = \sum_{i=1}^N c_i 1_{E_i}$ is simple, then

$$\int_A s\,\mathrm{d}\mu = \int \sum_{i=1}^N c_i 1_{E_i \cap A}\,\mathrm{d}\mu = \sum_{i=1}^N c_i \mu(E_i \cap A). \tag{4.1.9}$$

In particular, if $\mu(A) = 0$, then (as $\mu(E_i \cap A) \le \mu(A) = 0$) we have

$$\int_A s\,\mathrm{d}\mu = 0.$$

4.1.10 Theorem. *Fix a simple function $s \ge 0$. The mapping*

$$A \mapsto \int_A s\,\mathrm{d}\mu$$

from the μ-measurable sets to $[0,\infty]$ is a measure.

Proof. As $\mu(\emptyset) = 0$ we have (as proved above)

$$\int_\emptyset s\,\mathrm{d}\mu = 0.$$

In fact, this is obvious also simply because $1_\emptyset s = 0$. To prove the countable additivity, let $A_1, A_2, \dots$ be measurable and pairwise disjoint, and set $A = \bigcup_j A_j$. Write $s = \sum_{i=1}^N c_i 1_{E_i}$. We have by (4.1.9) and the countable additivity of the measure μ that

$$\int_A s\,\mathrm{d}\mu = \sum_{i=1}^N c_i \mu(E_i \cap A)$$

$$
\begin{aligned}
&= \sum_{i=1}^{N} c_i \sum_j \mu(E_i \cap A_j) \\
&= \sum_j \sum_{i=1}^{N} c_i \mu(E_i \cap A_j) = \sum_j \int_{A_j} s \, \mathrm{d}\mu.
\end{aligned}
$$

We are done. □

The following is a special case of Theorem 2.3.1.

4.1.11 Corollary. *Let $E_1 \subset E_2 \subset \cdots$ be measurable and $s \geq 0$ be simple. Then we have*

$$\int_{\bigcup_j E_j} s \, \mathrm{d}\mu = \lim_{j \to \infty} \int_{E_j} s \, \mathrm{d}\mu.$$

We are ready to move forward to defining the integral of a general non-negative measurable function.

4.2 The μ-integral of a non-negative measurable function

4.2.1 Definition. Let $f \geq 0$ be measurable. We define

$$\int f \, \mathrm{d}\mu := \sup \Big\{ \int s \, \mathrm{d}\mu \colon 0 \leq s \leq f, \ s \text{ simple} \Big\}.$$

Notice that the set over which we take the supremum is always non-empty as it contains the zero function. Moreover, whenever f is a simple function, the above coincides with the previous definition by Lemma 4.1.7 and so we can use the same notation.

Notice that if $0 \leq f \leq g$, then

$$\{s \colon 0 \leq s \leq f, \ s \text{ simple}\} \subset \{s \colon 0 \leq s \leq g, \ s \text{ simple}\},$$

and so we have the desired monotonicity property

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

We will prove more key properties of the integral (like linearity) after we have proved the monotone convergence theorem, since it is a useful tool even for this.

A key property of the Lebesgue integral is that it behaves well under limits. The following very important result is the first indication of this.

4.2.2 Theorem (Monotone convergence theorem (MCT))**.** *Suppose we have measurable functions*

$$0 \leq f_1 \leq f_2 \leq \ldots$$

and $f(x) := \lim_{j\to\infty} f_j(x) = \sup_j f_j(x)$. Then we have

$$\int f \, \mathrm{d}\mu = \lim_{j \to \infty} \int f_j \, \mathrm{d}\mu.$$

Proof. By the monotonicty property the sequence $\left(\int f_j \,\mathrm{d}\mu\right)_j$ is increasing and thus has a limit, and we have

$$B := \lim_{j\to\infty} \int f_j \,\mathrm{d}\mu \leq \int f \,\mathrm{d}\mu =: A.$$

We need to prove that

$$B \geq A.$$

It suffices to prove that for every $0 < \delta < 1$ we have

$$B \geq \delta A. \tag{4.2.3}$$

Pay attention how this trick with the arbitrary δ helps with the proof.

With a fixed δ and a simple function $0 \leq s \leq f$ define

$$E_j = \{f_j - \delta s \geq 0\}.$$

Notice that $f_j \leq f_{j+1}$ implies $E_j \subset E_{j+1}$. We also have $X = \bigcup_j E_j$. To see this, notice first that if $s(x) = 0$, then $x \in E_1$. Suppose then $s(x) > 0$. In this case $\delta s(x) < f(x)$ so for a large enough j we must have $f_j(x) \geq \delta s(x)$, and then $x \in E_j$. Since

$$\int_{E_j} \delta s \,\mathrm{d}\mu \leq \int_{E_j} f_j \,\mathrm{d}\mu \leq \int f_j \,\mathrm{d}\mu$$

we obtain by Corollary 4.1.11 that

$$\delta \int s \,\mathrm{d}\mu = \lim_{j\to\infty} \int_{E_j} \delta s \,\mathrm{d}\mu \leq \lim_{j\to\infty} \int f_j \,\mathrm{d}\mu = B.$$

Taking supremum over all simple $0 \leq s \leq f$ yields (4.2.3) and so we are done. □

4.2.4 Corollary. *Suppose we have measurable functions $f_i \geq 0$ satisfying*

$$f_1 \geq f_2 \geq \ldots$$

and $f(x) := \lim_{j\to\infty} f_j(x) = \inf_j f_j(x)$. If $\int f_1 \,\mathrm{d}\mu < \infty$, then we have

$$\int f \,\mathrm{d}\mu = \lim_{j\to\infty} \int f_j \,\mathrm{d}\mu.$$

Proof. Exercise. This corollary also directly follows from the dominated convergence theorem proved later. □

Next, we will prove more general properties of the integral of positive functions. The combination of the following ingredients is powerful for this:

- Approximation by simple functions, Lemma 3.1.2.
- Monotone convergence theorem.
- The corresponding property for the integral of simple functions.

4.2.5 Lemma. *Let $f, g \geq 0$ be measurable and $\alpha \geq 0$. Then we have*

$$\int (f + \alpha g) \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu + \alpha \int g \,\mathrm{d}\mu.$$

Proof. Using Lemma 3.1.2 we choose two increasing sequences of simple non-negative functions so that $s_j(x) \to f(x)$ and $u_j(x) \to g(x)$. Using the linearity of the integral of simple functions we have

$$\int (s_j + \alpha u_j)\,\mathrm{d}\mu = \int s_j\,\mathrm{d}\mu + \alpha \int u_j\,\mathrm{d}\mu.$$

It remains to take the limit on both sides and use MCT to conclude that we have

$$\int (s_j + \alpha u_j)\,\mathrm{d}\mu \to \int (f + \alpha g)\,\mathrm{d}\mu$$

and

$$\int s_j\,\mathrm{d}\mu + \alpha \int u_j\,\mathrm{d}\mu \to \int f\,\mathrm{d}\mu + \alpha \int g\,\mathrm{d}\mu.$$

□

4.2.6 Remark. Let $f \geq g \geq 0$ be measurable with $\int f\,\mathrm{d}\mu < \infty$ (notice that by monotonicity also $\int g\,\mathrm{d}\mu < \infty$). Notice that by applying the above with $f - g \geq 0$ and $g \geq 0$ we have

$$\int f\,\mathrm{d}\mu = \int ((f-g)+g)\,\mathrm{d}\mu = \int (f-g)\,\mathrm{d}\mu + \int g\,\mathrm{d}\mu$$

and so

$$\int (f-g)\,\mathrm{d}\mu = \int f\,\mathrm{d}\mu - \int g\,\mathrm{d}\mu.$$

Things will be most natural when we soon can integrate general functions (non-negative or not).

In fact, we can integrate a positive series term by term. This amounts to one more application of MCT.

4.2.7 Lemma. *Let $f_j \geq 0$ be measurable. Then we have*

$$\int \sum_{j=1}^{\infty} f_j\,\mathrm{d}\mu = \sum_{j=1}^{\infty} \int f_j\,\mathrm{d}\mu.$$

Proof. Simply notice that by the monotone convergence theorem we have

$$\int \sum_{j=1}^{\infty} f_j\,\mathrm{d}\mu = \int \lim_{N\to\infty} \sum_{j=1}^{N} f_j\,\mathrm{d}\mu = \lim_{N\to\infty} \int \sum_{j=1}^{N} f_j\,\mathrm{d}\mu.$$

Now we just apply the linearity we just proved to get

$$\int \sum_{j=1}^{\infty} f_j\,\mathrm{d}\mu = \lim_{N\to\infty} \sum_{j=1}^{N} \int f_j\,\mathrm{d}\mu = \sum_{j=1}^{\infty} \int f_j\,\mathrm{d}\mu.$$

□

4.2.8 Lemma. *Let $f \geq 0$ be measurable. Then the function*

$$A \mapsto \int_A f\,\mathrm{d}\mu$$

defines a measure.

Proof. Exercise. □

Next, we ask about the interchange of integration and limits when the convergence is not monotone. A partial answer is given by Fatou's lemma below – this result is often very useful as well.

4.2.9 Theorem (Fatou's lemma)**.** *Let $f_j \geq 0$ be measurable. Then we have*

$$\int \liminf_{j\to\infty} f_j \, \mathrm{d}\mu \leq \liminf_{j\to\infty} \int f_j \, \mathrm{d}\mu.$$

Proof. By definition we have

$$\liminf_{j\to\infty} f_j = \lim_{k\to\infty} \inf_{j\geq k} f_j =: \lim_{k\to\infty} g_k.$$

Notice that we can apply MCT to the increasing sequence g_k to have

$$\int \liminf_{j\to\infty} f_j \, \mathrm{d}\mu = \lim_{k\to\infty} \int g_k \, \mathrm{d}\mu = \liminf_{k\to\infty} \int g_k \, \mathrm{d}\mu.$$

Since clearly $g_k \leq f_k$ we have by the monotonicity of the integral that

$$\int g_k \, \mathrm{d}\mu \leq \int f_k \, \mathrm{d}\mu.$$

Combining with the above we are done. □

We end this subsection with some easy, but very useful, observations. Suppose $f \geq 0$ is measurable and

$$\int f \, \mathrm{d}\mu = 0,$$

then $f = 0$ almost everywhere. Indeed, notice that

$$\{f > 0\} = \bigcup_{j=1}^{\infty} \{f > 1/j\},$$

and so if $\mu(\{f > 0\}) > 0$ then for some j we must also have $\mu(\{f > 1/j\}) > 0$. But in this case

$$\int f \, \mathrm{d}\mu \geq \int_{\{f>1/j\}} f \, \mathrm{d}\mu \geq \frac{\mu(\{f > 1/j\})}{j} > 0.$$

Also notice that if $f = g$ almost everywhere, then

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

This follows from the fact that the integral over sets of measure zero vanish. Motivated by this, we should think that functions that agree pointwise almost everywhere are the same function. For example, we identify the function $x \mapsto 0$ and any function f satisfying $f(x) = 0$ almost everywhere. Very precisely, instead of functions, we could work with the equivalence classes

$$[f] := \{g \colon f = g \ \mu\text{-a.e.}\}$$

of functions that agree almost everywhere. We will not obsess about doing that – just remember that we identify functions that agree almost everywhere. This makes a technical difference, if we want to e.g. view

$$f \mapsto \int |f| \, \mathrm{d}\mu$$

as a norm. For instance, if $\int |f| \,\mathrm{d}\mu = 0$ it follows that $|f| = 0$ almost everywhere, but we would like to say that f is the unique zero element – this is possible if we do the above mentioned identification.

Lastly, notice that if $f \geq 0$ satisfies

$$\int f \,\mathrm{d}\mu < \infty,$$

then $f(x) < \infty$ almost everywhere. Indeed, we have for every $j \geq 1$ that

$$\begin{aligned}\mu(\{f = \infty\}) &\leq \mu(\{f \geq j\}) \\ &= \int 1_{\{f \geq j\}} \,\mathrm{d}\mu = \frac{1}{j} \int_{\{f \geq j\}} j \,\mathrm{d}\mu \leq \frac{1}{j} \int_{\{f \geq j\}} f \,\mathrm{d}\mu \leq \frac{1}{j} \int f \,\mathrm{d}\mu.\end{aligned}$$

Letting $j \to \infty$ gives the claim $\mu(\{f = \infty\}) = 0$.

4.3 The μ-integral of a general function

Let $f\colon X \to \dot{\mathbb{R}}$ be measurable. It makes sense to write

$$f = f^+ - f^-, \qquad f^+ = \max(f(x), 0) = \frac{1}{2}(|f| + f),\ f^- = -\min(f(x), 0) = \frac{1}{2}(|f| - f).$$

In this pointwise identity there is no problematic case like $\infty - \infty$ as always either $f^+(x) = 0$ or $f^-(x) = 0$. Now, these are non-negative and measurable functions and the integral is supposed to be linear, so it makes sense to define

$$\int f \,\mathrm{d}\mu := \int f^+ \,\mathrm{d}\mu - \int f^- \,\mathrm{d}\mu.$$

In general, here we can have the problem of $\infty - \infty$. However, we shall make this exact definition whenever possible – in other words, exactly when $\int f^+ \,\mathrm{d}\mu < \infty$ and $\int f^- \,\mathrm{d}\mu < \infty$ (in principle it can be made if only one of them is finite, but this is our definition for now). Notice that this is easily seen to be equivalent with

$$\int |f| \,\mathrm{d}\mu < \infty,$$

which is more convenient to state. This is because $|f| = f^+ + f^-$.

4.3.1 Definition. We say that a measurable function $f\colon X \to \dot{\mathbb{R}}$ is integrable if

$$\int |f| \,\mathrm{d}\mu < \infty$$

and in this case we define

$$\int f \,\mathrm{d}\mu := \int f^+ \,\mathrm{d}\mu - \int f^- \,\mathrm{d}\mu.$$

As usual, for a measurable set E we define integrability of f over the set E by applying the above to $1_E f$.

4.3.2 Remark. These are the L^1 functions that we will also study as part of the family of L^p spaces later.

Notice that if $f \geq 0$ this again coincides with the previous definition. Notice also that if $f\colon X \to \dot{\mathbb{R}}$ is integrable then by the definition and triangle inequality

$$\Big|\int f \,\mathrm{d}\mu\Big| \leq \int f^+ \,\mathrm{d}\mu + \int f^- \,\mathrm{d}\mu = \int |f| \,\mathrm{d}\mu,$$

since $|f| = f^+ + f^-$.

4.3.3 Proposition. *If $f, g\colon X \to \dot{\mathbb{R}}$ are integrable and $\alpha \in \mathbb{R}$, then $f + \alpha g$ is integrable*

$$\int (f + \alpha g) \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu + \alpha \int g \,\mathrm{d}\mu.$$

If $f \leq g$ we have

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

Proof. Notice that by triangle inequality and the results concerning positive functions we have

$$\int |f + \alpha g| \,\mathrm{d}\mu \leq \int |f| \,\mathrm{d}\mu + |\alpha| \int |g| \,\mathrm{d}\mu < \infty.$$

Thus, $f + \alpha g$ is integrable. For the following let first $\alpha \geq 0$. Let $h := f + \alpha g$ and notice that

$$h^+ - h^- = h = f + \alpha g = f^+ - f^- + \alpha g^+ - \alpha g^-,$$

and so

$$h^+ + f^- + \alpha g^- = h^- + f^+ + \alpha g^+.$$

Integrate both sides of the equality and use the linearity of the integral of non-negative functions, then rearrange to get the claimed linearity of the integral. The case $\alpha < 0$ requires only straightforward modifications that we omit.

For the latter claim, notice simply that if $f \leq g$, then

$$\int g \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu + \int (g - f) \,\mathrm{d}\mu \geq \int f \,\mathrm{d}\mu,$$

since $g - f \geq 0$ implies $\int (g - f) \,\mathrm{d}\mu \geq 0$. ☐

The following is an easy (since we have the corresponding result for non-negative functions, Lemma 4.2.8) result concerning integrability over disjoint sets.

4.3.4 Lemma. *Let E_j be measurable and disjoint and set $E = \bigcup_{j \in \mathbb{N}} E_j$.*

1. *If f is integrable over E, then f is integrable over each E_j and*

$$\int_E f \,\mathrm{d}\mu = \sum_{j \in \mathbb{N}} \int_{E_j} f \,\mathrm{d}\mu.$$

2. *If f is integrable over each E_j and*

$$\sum_{j \in \mathbb{N}} \int_{E_j} |f| \,\mathrm{d}\mu < \infty,$$

then f is integrable over E and

$$\int_E f \,\mathrm{d}\mu = \sum_{j \in \mathbb{N}} \int_{E_j} f \,\mathrm{d}\mu.$$

Proof. Easy optional exercise. □

The next theorem is the most important convergence theorem in Lebesgue integration.

4.3.5 Theorem (Dominated Convergence Theorem (DCT))**.** *Let $f_j \colon X \to \dot{\mathbb{R}}$ be a sequence of measurable functions so that the following holds.*

1. *There is an* ***integrable*** *function g so that for every j we have*
$$|f_j| \le g$$
almost everywhere.
2. *We have*
$$f_j(x) \to f(x)$$
as $j \to \infty$ almost everywhere.

Then f_j, f are integrable and we have
$$\int |f - f_j| \,\mathrm{d}\mu \to 0$$
as $j \to \infty$. In particular, we have
$$\lim_{j\to\infty} \int f_j \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu.$$

Proof. Since $|f_j|, |f| \le g$ almost everywhere, all of the functions are integrable. If we are careful about this, notice that outside the set of measure zero
$$\{f_j \not\to f\} \cup \bigcup_j \{|f_j| > g\}$$
we have $|f_j| \le g$ for all j and $|f| = \lim_j |f_j| \le g$. By modifying the functions in a set of measure zero, we can, for convenience, assume that these properties hold everywhere.

We aim to apply Fatou's lemma and seek for a suitable non-negative auxiliary function. To this end, notice that
$$|f - f_j| \le 2g$$
and so
$$h_j := 2g - |f - f_j| \ge 0, \qquad h_j \to 2g.$$
By Fatou's lemma we have
$$\begin{aligned}\int 2g \,\mathrm{d}\mu = \int \lim_{j\to\infty} h_j \,\mathrm{d}\mu &\le \liminf_{j\to\infty} \int h_j \,\mathrm{d}\mu \\ &= \liminf_{j\to\infty} \Big(\int 2g \,\mathrm{d}\mu - \int |f - f_j| \,\mathrm{d}\mu \Big) \\ &= \int 2g \,\mathrm{d}\mu - \limsup_{j\to\infty} \int |f - f_j| \,\mathrm{d}\mu.\end{aligned}$$
By cancelling out $\int 2g \,\mathrm{d}\mu < \infty$ we conclude that
$$\limsup_{j\to\infty} \int |f - f_j| \,\mathrm{d}\mu \le 0.$$

Since

$$0 \leq \liminf_{j\to\infty} \int |f - f_j| \,\mathrm{d}\mu \leq \limsup_{j\to\infty} \int |f - f_j| \,\mathrm{d}\mu \leq 0$$

we have

$$\lim_{j\to\infty} \int |f - f_j| \,\mathrm{d}\mu = 0.$$

The last claim also follows, since

$$\Big| \int f_j \,\mathrm{d}\mu - \int f \,\mathrm{d}\mu \Big| = \Big| \int (f_j - f) \,\mathrm{d}\mu \Big| \leq \int |f - f_j| \,\mathrm{d}\mu \to 0.$$

□

4.3.6 Example. First, recall Lemma 4.2.7. Suppose we now have measurable functions φ_j that satisfy

$$\sum_{j\in\mathbb{N}} \int |\varphi_j| \,\mathrm{d}\mu < \infty.$$

Then we know that pointwise

$$\sum_{j\in\mathbb{N}} |\varphi_j| < \infty$$

almost everywhere, and so the series $\sum_j \varphi_j$ converges absolutely outside a set of measure zero. To define the limit function, define

$$E := \Big\{ \sum_{j\in\mathbb{N}} |\varphi_j| = \infty \Big\}, \qquad \mu(E) = 0,$$

and define $f(x) = \sum_j \varphi_j(x)$ if $x \notin E$. In the set E that has measure zero we e.g. set $f = 0$ (this does not really matter). Now, we have that

$$f_j(x) := \sum_{i=1}^{j} \varphi_i(x) \to f(x)$$

almost everywhere and

$$|f_j(x)| \leq \sum_{i\in\mathbb{N}} |\varphi_i(x)| := g(x),$$

and that g is integrable (Lemma 4.2.7). Thus, by the dominated convergence theorem we have

$$\int f \,\mathrm{d}\mu = \int \lim_{j\to\infty} f_j \,\mathrm{d}\mu = \lim_{j\to\infty} \int f_j \,\mathrm{d}\mu = \lim_{j\to\infty} \sum_{i=1}^{j} \int \varphi_i \,\mathrm{d}\mu = \sum_{i\in\mathbb{N}} \int \varphi_i \,\mathrm{d}\mu.$$

4.3.7 Example. We show a simple way to see summing as integration, after which we e.g. have the monotone convergence theorem and the dominated convergence theorem available also for series. This is often useful for limit arguments involving series.

Equip $\mathbb{N}$ with the counting measure $\nu(B) = \#B$, $B \subset \mathbb{N}$. Consider a non-negative function $g\colon \mathbb{N} \to \mathbb{R}$ – we want to show that

$$\int_{\mathbb{N}} g(j) \,\mathrm{d}\nu(j) = \sum_{j=1}^{\infty} g(j).$$

We have

$$\int_{\mathbb{N}} g(j)\,\mathrm{d}\nu(j) = \sum_{i=1}^{\infty} \int_{\{i\}} g(j)\,\mathrm{d}\nu(j)$$
$$= \sum_{i=1}^{\infty} g(i) \int_{\{i\}} 1\,\mathrm{d}\nu(j) = \sum_{i=1}^{\infty} g(i)\nu(\{i\}) = \sum_{i=1}^{\infty} g(i).$$

In general, if $g\colon \mathbb{N} \to \mathbb{R}$ is integrable – $\sum_j |g(j)| < \infty$ – we have the above identity.

Again, this trick can sometimes be used to e.g. convert a series into an integral and then apply the general convergence theorems we know for integration against a measure.

4.4 Absolute continuity of measures

Let μ and ν be two measures on the same measurable space (X, Γ) – so here Γ is a σ-algebra on X and $\mu, \nu\colon \Gamma \to [0, \infty]$ are both measures defined on Γ. We say that ν is absolutely continuous with respect to μ – and denote this by $\nu \ll \mu$ – if for all $A \in \Gamma$ with $\mu(A) = 0$ we also have $\nu(A) = 0$.

4.4.1 Lemma. *Let μ and ν be two measures on the same measurable space (X, Γ) and let $\nu(X) < \infty$. Then $\nu \ll \mu$ if and only if for every $\epsilon > 0$ there is $\delta > 0$ so that $\mu(A) < \delta$ implies $\nu(A) < \epsilon$.*

Proof. It is trivial that if the condition of the lemma is satisfied, then $\nu \ll \mu$. Assume then that $\nu \ll \mu$. Aiming for a contradiction, assume that there exists $\epsilon > 0$ and a sequence $E_1, E_2, \ldots \in \Gamma$ so that $\mu(E_i) < 2^{-i}$ and $\nu(E_i) \geq \epsilon$. Define

$$A_k = \bigcup_{i \geq k} E_i.$$

Now $A_1 \supset A_2 \supset \cdots$ and $\nu(A_1) \leq \nu(X) < \infty$ so that by Theorem 2.3.1 we have

$$\nu\Big(\bigcap_{k=1}^{\infty} A_k\Big) = \lim_{k\to\infty} \nu(A_k) \geq \epsilon,$$

since by monotonicity $\nu(A_k) \geq \nu(E_k) \geq \epsilon$ for every k. To get a contradiction with $\nu \ll \mu$ it will be enough to show that $A := \bigcap_{k=1}^{\infty} A_k$ satisfies $\mu(A) = 0$. But this follows, since for **every** k we have

$$\mu(A) \leq \mu(A_k) \leq \sum_{i=k}^{\infty} \mu(E_i) \leq \sum_{i=k}^{\infty} 2^{-i} \sim 2^{-k}.$$

□

The following is an interesting, non-obvious corollary.

4.4.2 Corollary. *Suppose $f\colon X \to \dot{\mathbb{R}}$ is μ-integrable. Then for every $\epsilon > 0$ there is $\delta > 0$ so that $\mu(A) < \delta$ implies*

$$\int_A |f|\,\mathrm{d}\mu < \epsilon.$$

Proof. By Lemma 4.2.8 the function

$$A \mapsto \int_A |f|\,\mathrm{d}\mu$$

is a measure. As f is integrable, it is a finite measure. We also known that $\mu(A) = 0$ implies

$$\int_A |f| \, \mathrm{d}\mu = 0$$

so that this measure is absolutely continuous with respect to μ. Now the claim follows from Lemma 4.4.1. □

We will return to absolute continuity later when we prove the fundamental Radon–Nikodym theorem stating that all absolutely continuous measures are given by integration against some function.

Chapter 5

Riesz representation theorem

Preliminaries on topological spaces

The Riesz representation theorem will be stated on locally compact Hausdorff (LCH) spaces. Before we can prove it, we need to prove a few key results: Urysohn's lemma and partitions of unity.

Let X be a LCH space. We write $C_c(X)$ for the collection of all functions $f\colon X \to \mathbb{R}$ that are continuous and also satisfy that their support

$$\operatorname{spt} f := \overline{\{x \in X : f(x) \neq 0\}}$$

is compact. We introduce the following notation. Given sets $K, U \subset X$ and a function $f\colon X \to \mathbb{R}$ the notation $K \prec f$ means that $f \in C_c(X)$, K is compact and $1_K \leq f \leq 1$, and the notation $f \prec U$ means that $f \in C_c(X)$, U is open, $0 \leq f \leq 1$ and $\operatorname{spt} f \subset U$. Sometimes the notation $C_c(U)$ is used for the functions $f \in C_c(X)$ with $\operatorname{spt} f \subset U$, so in the definition $f \prec U$ we could equivalently say that this means $f \in C_c(U)$, U is open and $0 \leq f \leq 1$.

Urysohn's lemma precisely gives a function f with $K \prec f \prec U$ **whenever** $K \subset U$, K compact and U open. Without the continuity assumption we could trivially use 1_K – but the essence is to get a continuous function which behaves like 1_K.

5.0.1 Theorem (Urysohn's lemma)**.** *Let X be a LCH space, $K \subset X$ be compact, $U \subset X$ be open and $K \subset U$. Then there is a function $f \in C_c(X)$ with*

$$K \prec f \prec U.$$

Proof. We define $q_1 = 0$, $q_2 = 1$ and then write

$$(0,1) \cap \mathbb{Q} = \{q_3, q_4, q_5, \ldots\}.$$

By Theorem 1.2.15 we find open sets V_0 and V_1 so that their closures are compact and

$$K \subset V_1 \subset \overline{V_1} \subset V_0 \subset \overline{V_0} \subset U.$$

Suppose that $n \geq 2$ and that we have already chosen $V_{q_1}, V_{q_2}, \ldots, V_{q_n}$ so that $q_i < q_j$ implies $\overline{V_{q_j}} \subset V_{q_i}$. Consider $q_{n+1} \in (0,1)$ and let q_i, q_j, $i, j \in \{1, \ldots, n\}$, be the largest and smallest number so that

$$q_i < q_{n+1} < q_j.$$

Using Theorem 1.2.15 again we find $V_{q_{n+1}}$ so that

$$\overline{V_{q_j}} \subset V_{q_{n+1}} \subset \overline{V_{q_{n+1}}} \subset V_{q_i}.$$

Continuing we obtain (by induction) a collection of open sets $\{V_q : q \in \mathbb{Q} \cap [0,1]\}$ so that $K \subset V_1$, $\overline{V_0} \subset U$, each $\overline{V_q}$ is compact and $q > r$ implies

$$\overline{V_q} \subset V_r.$$

Given $q \in \mathbb{Q} \cap [0,1]$ define

$$f_q = q 1_{V_q}.$$

As V_q is open, it follows that for all $\alpha \in \mathbb{R}$ the set

$$\{f_q > \alpha\}$$

is open and we say that f_q is lower semicontinuous. It is immediate that the function $f := \sup_q f_q$ is also lower semicontinuous as a supremum of lower semicontinuous functions. Moreover, it is clear that $0 \leq f(x) \leq 1$, $1 \geq f(x) \geq f_1(x) = 1$ for $x \in V_1 \supset K$ and $f(x) = 0$ if $x \notin V_0$ so that $\operatorname{spt} f \subset \overline{V_0} \subset U$. So f will be our desired function if it is continuous – currently, we only know that it is lower semicontinuous.

To aid with the proof of continuity we define another function g. To this end, define now

$$g_q = 1_{\overline{V_q}} + q 1_{X \setminus \overline{V_q}}.$$

As $\overline{V_q}$ is closed, it follows that for all $\alpha \in \mathbb{R}$ the set

$$\{g_q < \alpha\}$$

is open, and we say that g_q is upper semicontinuous. It follows that the function $g := \inf_q g_q$ is upper semicontinuous as an infimum of upper semicontinuous functions. We will show that $f = g$ – then f is both upper and lower semicontinuous, hence continuous, and we are done.

The inequality $f_q(x) > g_r(x)$ is only possible if $q > r$, $x \in V_q$ and $x \notin \overline{V_r}$ – but if $q > r$ we know that $\overline{V_q} \subset V_r$, so this is never possible. Hence $f_q(x) \leq g_r(x)$ for all q, r and so $f(x) \leq g(x)$. Suppose, aiming for a contradiction, that $f(x) < g(x)$ for some x. Pick rationals q, r so that $f(x) < r < q < g(x)$. Since $f(x) < r$, we have $x \notin V_r$. Since $g(x) > q$, we have $x \notin X \setminus \overline{V_q}$ so $x \in \overline{V_q}$. But as $q > r$ we have $\overline{V_q} \subset V_r$, and so $x \in V_r$ – a contradiction. We have proved $f(x) = g(x)$ and are done. □

A partition of unity is some way of writing 1 – the unity – as a sum of other, usually continuous, functions. These other functions have some desired localization properties as well. This is a widely used technique, we will see its usefullness it the course of proving the Riesz representation theorem. In the following theorem the functions $h_1, \ldots, h_n$ are called a partition of unity on K, subordinate to the cover $V_1, \ldots, V_n$.

5.0.2 Theorem. *Let X be a LCH space, $V_1, \ldots, V_n \subset X$ be open and*

$$K \subset V_1 \cup \ldots \cup V_n$$

be compact. Then there exists functions $h_i \prec V_i$, $i = 1, \ldots, n$, such that

$$1_K \leq h_1 + \cdots + h_n \leq 1.$$

Proof. If $x \in K$, there is some $i = i(x)$ so that $x \in V_i$. If there are multiple i so that $x \in V_i$, we just pick one to be $i(x)$. By applying Theorem 1.2.15 to $\{x\}$ and V_i we find a neighbourhood W_x of x so that $\overline{W_x} \subset V_i$ is compact.

As $K \subset \bigcup_{x \in K} W_x$ is an open cover of the compact set K, we find finitely many $x_1, \dots, x_m \in K$ so that $K \subset \bigcup_{j=1}^m W_{x_j}$. Given $i \in \{1, \dots, n\}$ define

$$H_i = \bigcup_{j\colon i(x_j)=i} \overline{W_{x_j}}.$$

As a finite union of compact sets $H_i \subset V_i$ is compact, and by Urysohn's lemma we find g_i with $H_i \prec g_i \prec V_i$.

We now define

$$h_k = \Big(\prod_{i=1}^{k-1} (1-g_i) \Big) g_k, \qquad k \in \{1, \dots, n\}.$$

Clearly $h_i \prec V_i$. The idea behind the definition is the algebraic identity

$$h_1 + \cdots + h_n = 1 - \prod_{i=1}^{n} (1-g_i).$$

As $K \subset H_1 \cup \cdots \cup H_n$, for $x \in K$ we have that $1 - g_i(x) = 0$ for at least one i, and thus $h_1(x) + \cdots + h_n(x) = 1$. It is also clear from the identity that $h_1 + \cdots + h_n \le 1$ always. $\square$

Standard version of the Riesz representation theorem

Let $(X, \mathcal{F}, \mu)$ be a measure space, where X is a locally compact Hausdorff (LCH) space. Assume that μ is a Borel measure – this simply means by definition that $\mathrm{Bor}(X) \subset \mathcal{F}$ (i.e. all Borel sets are measurable). Assume that μ is locally finite – i.e. $\mu(K) < \infty$ for all compact sets $K \subset X$. If $f \in C_c(X)$, f is μ-measurable as $f^{-1}V$ is open, in particular a Borel set, for all open $V \subset \mathbb{R}$, and we assumed that $\mathrm{Bor}(X) \subset \mathcal{F}$. Given $f \in C_c(X)$ let $K = \mathrm{spt}\ f$ be the compact support of f. There exists $M < \infty$ such that $|f| \le M$ as $fK \subset \mathbb{R}$ is compact (hence bounded). Then we have

$$\int |f| \, \mathrm{d}\mu = \int_K |f| \, \mathrm{d}\mu \le M\mu(K) < \infty,$$

that is f is integrable. With these preparations done, the following linear functional makes sense

$$\Lambda\colon C_c(X) \to \mathbb{R}, \qquad \Lambda f := \int f \, \mathrm{d}\mu.$$

Notice that Λ is linear ($\Lambda(\alpha_1 f_1 + \alpha_2 f_2) = \alpha_1 \Lambda f_1 + \alpha_2 \Lambda f_2$) and positive in the following sense:

$$\Lambda f \ge 0 \qquad \textbf{if} \ \ f \ge 0.$$

Is this the only special property of these functionals – do all linear functionals $C_c(X) \to \mathbb{R}$ with this positivity property arise as integration against some measure?

In general, we call any mapping like above a positive linear functional on $C_c(X)$ (but notice that positivity means the above property not that $\Lambda f \ge 0$ always). The content of the Riesz representation theorem is that the striking converse also holds: all positive linear functionals on $C_c(X)$ are given by integration against some nice measure. This gives a way to construct measures as well – just define some positive linear functional and the Riesz representation theorem gives you an associated measure. This e.g. leads to a very clever definition of the Lebesgue measure as we will see later.

5.0.3 Theorem (Riesz representation theorem). *Suppose X is a is a locally compact Hausdorff (LCH) space and $\Lambda\colon C_c(X) \to \mathbb{R}$ is an arbitrary positive linear functional:*

1. $\Lambda(\alpha_1 f_1 + \alpha_2 f_2) = \alpha_1 \Lambda f_1 + \alpha_2 \Lambda f_2$ *if* $f_1, f_2 \in C_c(X)$ *and* $\alpha_1, \alpha_2 \in \mathbb{R}$,
2. $\Lambda f \geq 0$ *if* $f \in C_c(X)$ *and* $f \geq 0$.

Then there exists a σ-algebra $\mathcal{F} \supset \operatorname{Bor}(X)$ and a unique measure $\mu\colon \mathcal{F} \to [0, \infty]$ so that $(X, \mathcal{F}, \mu)$ is a complete measure space,

$$\Lambda f = \int f \,\mathrm{d}\mu, \qquad f \in C_c(X),$$

and μ also satisfies the following properties:

(a) $\mu(K) < \infty$ *for all compact sets* $K \subset X$,

(b) $\mu(E) = \inf\{\mu(V)\colon E \subset V,\, V \text{ open}\}$ *for all* $E \in \mathcal{F}$,

(c) $\mu(E) = \sup\{\mu(K)\colon K \subset E,\, K \text{ compact}\}$ *for all open E and also for all $E \in \mathcal{F}$ with $\mu(E) < \infty$.*

Proof. **Uniqueness.** To prove uniqueness, suppose $\mu_1, \mu_2\colon \mathcal{F} \to [0, \infty]$ are two measures both satisfying the regularity properties (a)-(c) and the property that

$$\Lambda f = \int f \,\mathrm{d}\mu_j, \qquad f \in C_c(X),\ j = 1, 2.$$

We want to show that $\mu_1(A) = \mu_2(A)$ for all $A \in \mathcal{F}$. Let first $A = K \subset X$ be compact Let $\epsilon > 0$ and use the regularity property (b) of μ_2 with $E = K$ to find an open set $U \supset K$ with $\mu_2(U) \leq \mu_2(K) + \epsilon$. Using Urysohn's lemma we find $f \in C_c(U)$ so that $1_K \leq f \leq 1_U$. Then we have

$$\mu_1(K) = \int 1_K \,\mathrm{d}\mu_1 \leq \int f \,\mathrm{d}\mu_1 = \Lambda f = \int f \,\mathrm{d}\mu_2 \leq \int 1_U \,\mathrm{d}\mu_2 = \mu_2(U) \leq \mu_2(K) + \epsilon.$$

Letting $\epsilon \to 0$ we get $\mu_1(K) \leq \mu_2(K)$ and by symmetry we then get $\mu_1(K) = \mu_2(K)$. As the property (c) holds for all open sets, it follows that also $\mu_1(U) = \mu_2(U)$ for every open $U \subset X$. Finally, it now follows from the regularity property (b) that $\mu_1(A) = \mu_2(A)$ for every $A \in \mathcal{F}$.

Construction of μ. Whenever $U \subset X$ is open let

$$\mu(U) = \sup\{\Lambda f\colon f \prec U\}.$$

Notice that as $f \prec U$ means in particular that $f \geq 0$, then $\Lambda f \geq 0$, and so $\mu(U) \in [0, \infty]$. For all $E \subset X$ we then set

$$\mu(E) = \inf\{\mu(U)\colon E \subset U,\, U \text{ open}\}.$$

Notice that it is consistent with the open case to make this definition – if E is open the second definition agrees with the first one (think through this). Notice that we have now defined μ on the whole $\mathcal{P}(X)$, and we will first prove that this unrestricted μ is an outer measure.

μ is an outer measure. It is directly clear from the definition that if $A \subset B$ then $\mu(A) \leq \mu(B)$, so μ is monotonic. We then prove subadditivity. We first prove that if $U_1, U_2 \subset X$ are open, then

$$\mu(U_1 \cup U_2) \leq \mu(U_1) + \mu(U_2).$$

Let $f \prec U_1 \cup U_2$ be arbitrary. We need to show that $\Lambda f \leq \mu(U_1) + \mu(U_2)$. Let $K := \operatorname{spt} f$ – K is compact and $K \subset U_1 \cup U_2$. Choose a partition of unity $h_1 \prec U_1$, $h_2 \prec U_2$ with $h_1 + h_2 = 1$ on K. It follows that $f = f(h_1 + h_2) = fh_1 + fh_2$. Hence, as $fh_1 \prec U_1$ and $fh_2 \prec U_2$ we get by linearity of Λ that

$$\Lambda f = \Lambda(fh_1) + \Lambda(fh_1) \leq \mu(U_1) + \mu(U_2)$$

Of course, the version with finitely many (instead of just two) open sets follows inductively or by the same proof.

Let now $A_1, A_2, \ldots \subset X$ be arbitrary. For every i choose (using the definition) open $U_i \supset A_i$ with

$$\mu(U_i) \leq \mu(A_i) + \frac{\epsilon}{2^i}.$$

Set $U = \bigcup_i U_i$ so that $A := \bigcup_i A_i \subset U$ and $\mu(A) \leq \mu(U)$. We claim that

$$\mu(U) \leq \sum_i \mu(A_i) + \epsilon,$$

which then implies subadditivity. By what we have proved, this would be obvious if the union would be finite. Because it is not, we need an extra argument. Consider an arbitrary $f \prec U$. Since $K := \operatorname{spt} f \subset U = \bigcup_i U_i$, where each U_i is open, we have by compactness that $K \subset U_1 \cup \cdots \cup U_n$ for some n. Thus, we have $f \prec U_1 \cup \cdots \cup U_n$ and so

$$\Lambda f \leq \mu(U_1 \cup \cdots \cup U_n) \leq \sum_{i=1}^{n} \mu(U_i) \leq \sum_{i=1}^{n} \left(\mu(A_i) + \frac{\epsilon}{2^i}\right) \leq \sum_{i=1}^{\infty} \left(\mu(A_i) + \frac{\epsilon}{2^i}\right) = \sum_{i=1}^{\infty} \mu(A_i) + \epsilon.$$

Since $f \prec U$ was arbitrary, this shows that $\mu(U) \leq \sum_i \mu(A_i) + \epsilon$, and so we have proved subadditivity.

To see $\mu(\emptyset) = 0$ we can, for example, use the formula (5.0.4) that we prove next:

$$\mu(\emptyset) = \inf\{\Lambda f \colon \emptyset \prec f\}.$$

Notice that $\emptyset \prec 0$ and so

$$0 \leq \mu(\emptyset) \leq \Lambda 0 = 0,$$

where $\Lambda 0 = 0$ follows from linearity. Thus $\mu(\emptyset) = 0$.

Behaviour of μ on compact sets. We prove the useful formula

$$\mu(K) = \inf\{\Lambda f \colon K \prec f\} \tag{5.0.4}$$

whenever $K \subset X$ is compact. First, we prove $\mu(K) \leq \inf\{\Lambda f \colon K \prec f\}$. We fix f with $K \prec f$ and need to prove that $\mu(K) \leq \Lambda f$. Fix $\delta \in (0,1)$ and define the open set $U_\delta = \{f > \delta\}$. For every $x \in K$ we have $f(x) = 1 > \delta$ so $K \subset U_\delta$. Thus, we have $\mu(K) \leq \mu(U_\delta)$, where by definition

$$\mu(U_\delta) = \sup\{\Lambda g \colon g \prec U_\delta\}.$$

But notice that if $g \prec U_\delta$, we have $g \leq 1_{U_\delta} \leq \frac{f}{\delta}$. Notice also that if $h_1, h_2 \in C_c(X)$ with $h_1 \leq h_2$, then $h_2 - h_1 \geq 0$, and so $\Lambda h_2 - \Lambda h_1 = \Lambda(h_2 - h_1) \geq 0$ – that is, $\Lambda h_1 \leq \Lambda h_2$. Applying this we get

$$\mu(K) \leq \mu(U_\delta) \leq \Lambda \frac{f}{\delta} = \frac{1}{\delta} \Lambda f.$$

The desired inequality follows by letting $\delta \to 1$. We now show $\inf\{\Lambda f \colon K \prec f\} \leq \mu(K)$. To do this we let $\epsilon > 0$ and choose open $U \supset K$ with $\mu(U) \leq \mu(K) + \epsilon$. Then we use Urysohn's lemma to choose f with $K \prec f \prec U$. We have

$$\Lambda f \leq \mu(U) \leq \mu(K) + \epsilon$$

showing the claim $\inf\{\Lambda f \colon K \prec f\} \leq \mu(K)$.

One consequence of the identify $\mu(K) = \inf\{\Lambda f \colon K \prec f\}$ is that $\mu(K) < \infty$ whenever K is compact. This is the desired regularity property (a).

We will now prove that

$$\mu(K_1 \cup \cdots K_n) = \sum_{i=1}^{n} \mu(K_i)$$

if $K_1, \ldots, K_n$ are compact and disjoint. It is clearly enough to prove this for $n = 2$, and by subadditivity it is enough to prove $\mu(K_1 \cup K_2) \geq \mu(K_1) + \mu(K_2)$. As $K_1 \cup K_2$ is compact, we know that

$$\mu(K_1 \cup K_2) = \inf\{\Lambda g \colon K_1 \cup K_2 \prec g\}.$$

Fix $\epsilon > 0$ and g with $K_1 \cup K_2 \prec g$ and $\Lambda g \leq \mu(K_1 \cup K_2) + \epsilon$. We will find a suitable f so that $K_1 \prec fg$ and $K_2 \prec (1-f)g$. To this end, choose open $U \supset K_1$ with $K_2 \cap U = \emptyset$ (using property 4 on page 6). By Urysohn's lemma there exists f with $K_1 \prec f \prec U$. In particular $f = 0$ on K_2, so this f works. We thus get

$$\mu(K_1) + \mu(K_2) \leq \Lambda(fg) + \Lambda((1-f)g) = \Lambda g \leq \mu(K_1 \cup K_2) + \epsilon.$$

Letting $\epsilon \to 0$ gives the claim.

An auxiliary collection $\mathcal{F}_0$ and some of its properties. We define

$$\mathcal{F}_0 := \{E \subset X \colon \mu(E) < \infty \text{ and } \mu(E) = \sup\{\mu(K) \colon K \subset E,\, K \text{ compact}\}\}.$$

We just proved that $\mu(K) < \infty$ for a compact K and it is obvious that $\mu(K) = \sup\{\mu(K') \colon K' \subset K,\, K' \text{ compact}\}$, so $\mathcal{F}_0$ contains all compact sets $K \subset X$. We will now show that $\mathcal{F}_0$ contains the open sets with finite measure. It is enough to prove that

$$\mu(U) = \sup\{\mu(K) \colon K \subset U,\, K \text{ compact}\}$$

for every open U (notice that this shows the regularity property (c) for all open sets). Fix open U and consider an arbitrary number α with $\alpha < \mu(U)$. Using the fact that $\mu(U) = \sup\{\Lambda f \colon f \prec U\}$ we choose $f \prec U$ with $\Lambda f \geq \alpha$. We write $K := \operatorname{spt} f$. If we prove $\Lambda f \leq \mu(K)$ we recover the desired regularity, as it then follows $\alpha \leq \mu(K)$ and $\alpha < \mu(U)$ was arbitrary. Suppose $V \subset K$ is an arbitrary open set. Then $f \prec V$ and we have $\Lambda f \leq \mu(V)$. Hence

$$\Lambda f \leq \inf\{\mu(V) \colon V \supset K \text{ open}\} = \mu(K)$$

as desired.

Countable additivity of μ on $\mathcal{F}_0$ We prove that for disjoint $A_1, A_2, \ldots \in \mathcal{F}_0$ we have

$$\mu(A) = \sum_i \mu(A_i), \qquad A := \bigcup_i A_i.$$

Let $\epsilon > 0$. As $A_i \in \mathcal{F}_0$, we can choose compact $K_i \subset A_i$ with $\mu(A_i) \leq \mu(K_i) + \epsilon/2^i$. Then for all n we get

$$\mu(A) \geq \mu(K_1 \cup \cdots \cup K_n) = \sum_{i=1}^{n} \mu(K_i) \geq \sum_{i=1}^{n} \mu(A_i) - \epsilon \tag{5.0.5}$$

using the additivity we proved for finite unions of disjoint compact sets earlier. Letting $n \to \infty$ and $\epsilon \to$ yields

$$\mu(A) \geq \sum_{i=1}^{\infty} \mu(A_i).$$

The desired additivity follows from the converse inequality follows from subadditivity. From what we did above it also follows that if $\mu(A) < \infty$, then $A \in \mathcal{F}_0$. Indeed, by (5.0.5) we get that the measure of the compact set $K_1 \cup \cdots K_n$ can approximate $\mu(A)$ arbitrarily well.

Regularity of $A \in \mathcal{F}_0$. We show here that given $\epsilon > 0$ there exists a compact set $K \subset A$ and an open set $U \supset A$ so that $\mu(U \setminus K) < \epsilon$. By definition, there is an open $U \supset A$ so that $\mu(U) < \mu(A) + \epsilon/2$. As $\mu(A) < \infty$ also $\mu(U) < \infty$. Since $A \in \mathcal{F}_0$ there is also a compact $K \subset A$ so that $\mu(A) \leq \mu(K) + \epsilon/2$. Notice that $U \setminus K$ is open with $\mu(U \setminus K) \leq \mu(U) < \infty$, so $U \setminus K \in \mathcal{F}_0$ as we have showed that $\mathcal{F}_0$ contains the open sets with finite measure. Since also $K \in \mathcal{F}_0$ (recall that $\mathcal{F}_0$ contais all compact sets), the additivity of μ on $\mathcal{F}_0$ yields

$$\mu(K) + \mu(U \setminus K) = \mu(U) < \mu(A) + \epsilon/2 \leq \mu(K) + \epsilon,$$

and so $\mu(U \setminus K) < \epsilon$ as desired.

$\mathcal{F}_0$ is closed under set differences, finite unions and intersections Let $A_1, A_2 \in \mathcal{F}_0$. Let $\epsilon > 0$. By what we proved above there exists compact K_i and open U_i with $K_i \subset A_i \subset U_i$ and $\mu(K_i \setminus A_i) < \epsilon$. Note that now

$$A_1 \setminus A_2 \subset U_1 \setminus K_2 \subset (U_1 \setminus K_1) \cup (K_1 \setminus K_2) \subset (U_1 \setminus K_1) \cup (K_1 \setminus U_2) \cup (U_2 \setminus K_2).$$

This implies that

$$\mu(A_1 \setminus A_2) \leq 2\epsilon + \mu(K_1 \setminus U_2).$$

Notice that $K_1 \setminus U_2 \subset A_1 \setminus A_2$ is compact (closed subsets of a compact set are compact). It follows that $A_1 \setminus A_2 \in \mathcal{F}_0$. Now we also get

$$A_1 \cup A_2 = (A_1 \setminus A_2) \cup A_2,$$

which is a **disjoint** union, where $A_1 \setminus A_2, A_2 \in \mathcal{F}_0$. As $\mu(A_1 \cup A_2) < \infty$, it follows that $A_1 \cup A_2 \in \mathcal{F}_0$ (remember that we showed previously that $\mathcal{F}_0$ is closed under disjoint unions if the union has finite measure). Finally, we have

$$A_1 \cap A_2 = A_1 \setminus (A_1 \setminus A_2) \in \mathcal{F}_0.$$

Definition of the σ-algebra $\mathcal{F}$ containing the Borel sets. As μ is an outer measure, we would readily have the set of μ-measurable sets $\mathcal{M}_\mu(X)$ as a candidate for the σ-algebra $\mathcal{F}$. However, it turns out that the proof benefits from us defining $\mathcal{F}$ with a different definition (after the proof we can actually deduce that the $\mathcal{F}$ we are about to define satisfies $\mathcal{F} \subset \mathcal{M}_\mu(X)$). For the $\mathcal{F}$ we define showing $\mathrm{Bor}(X) \subset \mathcal{F}$ is not hard.

We define

$$\mathcal{F} = \{E \subset X \colon E \cap K \in \mathcal{F}_0 \text{ for all compact } K \subset X\}.$$

This is a σ-algebra. Clearly $X \in \mathcal{F}$, since $\mathcal{F}_0$ contains the compact sets. Suppose then $A \in \mathcal{F}$ and fix a compact set K. Then

$$(X \setminus A) \cap K = K \setminus A = K \setminus (A \cap K).$$

As $A \cap K \in \mathcal{F}_0$ by definition and $K \in \mathcal{F}_0$ as a compact set, and since $\mathcal{F}_0$ is closed under set differences, we also have $(X \setminus A) \cap K \in \mathcal{F}_0$. Thus $X \setminus A \in \mathcal{F}$. Finally, let $A_1, A_2, \ldots \in \mathcal{F}$ and set $A := \bigcup_i A_i$. Let K be compact. Define $B_1 := A_1 \cap K$ and inductively

$$B_n = (A_n \cap K) \setminus (B_1 \cup \cdots \cup B_{n-1}), \qquad n \geq 2.$$

By what we have proved we know that $B_j \in \mathcal{F}_0$ for every j. The sets B_j are also pairwise disjoint. Since $A \cap K = \bigcup_j B_j$ and $\mu(A \cap K) \leq \mu(K) < \infty$, we know that $A \cap K \in \mathcal{F}_0$ (remember that we showed previously that $\mathcal{F}_0$ is closed under countable disjoint unions if the union has finite measure). Thus $A \in \mathcal{F}$ and $\mathcal{F}$ is a σ-algebra.

Since $\mathrm{Bor}(X)$ is the smallest σ-algebra containing the closed sets, the inclusion $\mathrm{Bor}(X)$ follows if we show that every closed $F \subset X$ belongs to $\mathcal{F}$. But if F is closed and K is compact, then $F \cap K$ is compact and hence belongs to $\mathcal{F}_0$. Thus $F \in \mathcal{F}$.

We will next show that in fact we have the exact relationship

$$\mathcal{F}_0 = \{A \in \mathcal{F} \colon \mu(A) < \infty\}.$$

If $A \in \mathcal{F}_0$, by definition $\mu(A) < \infty$, and also $A \in \mathcal{F}$ as $A \cap K \in \mathcal{F}_0$ for every compact K (since both A and K belong to $\mathcal{F}_0$), and so $A \in \{B \in \mathcal{F} \colon \mu(B) < \infty\}$. Suppose then that $A \in \mathcal{F}$ with $\mu(A) < \infty$. We want to show that $A \in \mathcal{F}_0$ – that is, given $\epsilon > 0$ we need to find a compact set $K \subset A$ with $\mu(A) \le \mu(K) + \epsilon$. Choose an arbitrary open $U \supset A$ with $\mu(U) < \infty$. We have showed before that $U \in \mathcal{F}_0$. As $U \in \mathcal{F}_0$ we find a compact $H \subset U$ with $\mu(U \setminus H) < \epsilon/2$. By definition $A \cap H \in \mathcal{F}_0$ and so we can find a compact set $K \subset A \cap H$ with $\mu(A \cap H) \le \mu(K) + \epsilon/2$. Finally, we have

$$\mu(A) \le \mu(A \cap H) + \mu(A \setminus H) \le \mu(K) + \epsilon/2 + \mu(U \setminus H) \le \mu(K) + \epsilon.$$

It follows that $A \in \mathcal{F}_0$.

We now recap where we are on the proof of the properties (a)-(c). We have proved (a) a long time ago. Property (b) holds by definition. We just proved that if $A \in \mathcal{F}$ with $\mu(A) < \infty$, then $A \in \mathcal{F}_0$, so (c) holds for such sets A. We have also proved the regularity property (c) for all open sets. Thus, we have established the claims (a)-(c).

$\mu \colon \mathcal{F} \to [0, \infty]$ **is a measure.** We have already proved that $\mathcal{F}$ is a σ-algebra and $\mu(\emptyset) = 0$. Let then $A_1, A_2, \ldots \in \mathcal{F}$ be pairwise disjoint and $A = \bigcup_i A_i$. We need to prove that

$$\mu(A) = \sum_i \mu(A_i).$$

The claim is trivial by monotonicity if $\mu(A_i) = \infty$ for some i, so we may assume $\mu(A_i) < \infty$ for all i. But then $A_i \in \mathcal{F}_0$ for every i, since $\mathcal{F}_0 = \{B \in \mathcal{F} \colon \mu(B) < \infty\}$. But then it only remains to use the countable additivity of μ on $\mathcal{F}_0$ that we have already proved. Thus, $\mu \colon \mathcal{F} \to [0, \infty]$ is a measure.

Completeness of $(X, \mathcal{F}, \mu)$**.** Let $N \in \mathcal{F}$ with $\mu(N) = 0$ and $A \subset N$. We need to prove that $A \in \mathcal{F}$. For this we need to show that $A \cap K \in \mathcal{F}_0$ for every compact K. But by monotonicity we have $\mu(A \cap K) = 0$ so obviously $\mu(A \cap K) = 0 = \sup\{\mu(K') \colon K' \text{ compact}, K' \subset A \cap K\}$, and so $A \cap K \in \mathcal{F}_0$ and $A \in \mathcal{F}$.

The identity $\Lambda f = \int f \, \mathrm{d}\mu$**.** It only remains to prove this identity. We only need to prove

$$\Lambda f \le \int f \, \mathrm{d}\mu$$

as this then already implies the reverse inequality via

$$-\Lambda f = \Lambda(-f) \le \int (-f) \, \mathrm{d}\mu = -\int f \, \mathrm{d}\mu.$$

We fix $f \in C_c(X)$ and let $[a, b]$ be an interval containing the image of f. Fix $\epsilon > 0$. Divide

$$\tau_0 < a < \tau_1 < \tau_2 < \cdots < \tau_n = b,$$

where $\tau_{j+1} - \tau_j < \epsilon$. Let $K = \mathrm{spt}\ f$ and write $K = \bigcup_j A_j$, where

$$A_j = K \cap f^{-1}(\tau_{j-1}, \tau_j], \qquad 1 \le j \le n.$$

Notice that the sets A_j are disjoint Borel sets. Choose an open set $U_j' \supset A_j$ so that $\mu(U_j') \le \mu(A_j)+\epsilon/n$. Define $U_j = U_j' \cap f^{-1}(\tau_{j-1}, \tau_j + \epsilon)$. Then

$$U_j \supset A_j \cap f^{-1}(\tau_{j-1}, \tau_j + \epsilon) = A_j$$

is open, $\mu(U_j) \le \mu(U_j') \le \mu(A_j) + \epsilon/n$ and $f(x) \le \tau_j + \epsilon$ for $x \in U_j$.

Now $K = \bigcup_{j=1}^n A_j \subset \bigcup_{j=1}^n U_j$ so we can choose a partition of unity $\{h_j\}_{j=1}^n$ sub-ordinate to $\{U_j\}_{j=1}^n$ so that $h_j \prec U_j$ and $K \prec \sum_{j=1}^n h_j$. We now have $f = \sum_{j=1}^n f h_j$ and

$$\Lambda f = \sum_{j=1}^n \Lambda(f h_j).$$

Moreover, using $\mu(K) = \inf\{\Lambda f \colon K \prec f\}$ and $K \prec \sum_{j=1}^n h_j$ we have

$$\mu(K) \le \Lambda\Big(\sum_{j=1}^n h_j\Big) = \sum_{j=1}^n \Lambda h_j.$$

With these preparations done, we are ready to prove $\Lambda f \le \int f \,\mathrm{d}\mu$. Define the simple function $s = \sum_{j=1}^n (\tau_j - \epsilon) 1_{A_j}$. Suppose $x \in K$ is such that $x \in A_j$. Then

$$s(x) = \tau_j - \epsilon < \tau_{j-1} < f(x).$$

So $s \le f$ and it follows that

$$\sum_{j=1}^n (\tau_j - \epsilon)\mu(A_j) = \int s \,\mathrm{d}\mu \le \int f \,\mathrm{d}\mu.$$

As $f h_j \le (\tau_j + \epsilon) h_j$ and so $\Lambda(f h_j) \le (\tau_j + \epsilon)\Lambda h_j$. We now estimate

$$\Lambda f = \sum_{j=1}^n \Lambda(f h_j) \le \sum_{j=1}^n (\tau_j + \epsilon)\Lambda h_j = \sum_{j=1}^n (|a| + \tau_j + \epsilon)\Lambda h_j - |a| \sum_{j=1}^n \Lambda h_j.$$

Here $|a| \sum_{j=1}^n \Lambda h_j \ge |a|\mu(K)$ so $-|a| \sum_{j=1}^n \Lambda h_j \le -|a|\mu(K)$. Notice also that

$$|a| + \tau_j + \epsilon \ge |a| + \tau_0 + \epsilon \ge |a| + \tau_1 \ge |a| + a \ge 0,$$

so we may estimate

$$(|a| + \tau_j + \epsilon)\Lambda h_j \le (|a| + \tau_j + \epsilon)\mu(U_j) \le (|a| + \tau_j + \epsilon)\Big(\mu(A_j) + \frac{\epsilon}{n}\Big).$$

Altogether we get

$$\Lambda f \le \sum_{j=1}^n (|a| + \tau_j + \epsilon)\Big(\mu(A_j) + \frac{\epsilon}{n}\Big) - |a|\mu(K).$$

Notice that

$$\sum_{j=1}^n (|a| + \tau_j + \epsilon)\frac{\epsilon}{n} \le (|a| + |b| + \epsilon)\epsilon.$$

Thus, using this and $\mu(K) = \sum_{j=1}^n \mu(A_j)$ we have

$$\begin{aligned}
\Lambda f &\leq \sum_{j=1}^{n}(|a| + \tau_j + \epsilon)\mu(A_j) + (|a| + |b| + \epsilon)\epsilon - |a|\mu(K) \\
&= \sum_{j=1}^{n}(\tau_j + \epsilon)\mu(A_j) + (|a| + |b| + \epsilon)\epsilon \\
&= \sum_{j=1}^{n}(\tau_j - \epsilon)\mu(A_j) + 2\epsilon\mu(K) + (|a| + |b| + \epsilon)\epsilon \\
&\leq \int f \,\mathrm{d}\mu + 2\epsilon\mu(K) + (|a| + |b| + \epsilon)\epsilon.
\end{aligned}$$

The proof is completed by letting $\epsilon \to 0$. □

An outer measure version of the Riesz representation theorem

We now give some important complements to this theorem – some of the facts will be proved in the exercises. Remember that we actually constructed an **outer** measure μ during the proof – we refer to this particular outer measure μ and the associated σ-algebra $\mathcal{F}$ in what follows (if we talk about some potentially different measure we try to talk about ν or λ instead) – this convention is valid only for the rest of Section 5.

1. Recall that we can associate to the outer measure μ also the collection $\mathcal{M}_\mu(X)$ of μ-measurable sets – see (2.2.3). The constructed σ-algebra $\mathcal{F}$ from the Riesz representation theorem satisfies $\mathcal{F} \subset \mathcal{M}_\mu(X)$ – that is, for every $E \in \mathcal{F}$ we have for every $A \subset X$ that
$$\mu(A) = \mu(A \cap E) + \mu(A \setminus E).$$
It is an exercise to prove this. Thus, we have $\mathrm{Bor}(X) \subset \mathcal{F} \subset \mathcal{M}_\mu(X)$ so we say that $\mu\colon \mathcal{P}(X) \to [0,\infty]$ is a Borel outer measure as every Borel set is μ-measurable (belongs to $\mathcal{M}_\mu(X)$).

2. We say that an outer measure ν is Borel regular if ν is a Borel outer measure and for every $A \subset X$ there exists $B \in \mathrm{Bor}(X)$ with $B \supset A$ and $\nu(A) = \nu(B)$ (a Borel measurable cover of A). This property sometimes allows (combined with some appropriate argument) one to extend results concerning ν that hold for Borel sets to hold for all sets.

   Notice that by the definitions made during the proof, the outer regularity (b) actually holds for all $E \subset X$, not only for $E \in \mathcal{F}$. It is an exercise to prove that this implies that the outer measure μ constructed in the Riesz representation theorem is a Borel regular outer measure.

   We remark the following thing about terminology. If we say that a measure is Borel regular, it means that a Borel cover exists for all measurable sets (not for all sets as in the case of the outer measure).

3. We proved the inner regularity property (c) for all open sets and for all $E \in \mathcal{F}$ with $\mu(E) < \infty$. It is an exercise to prove that (c) actually also holds for all $E \in \mathcal{M}_\mu(X)$ with $\mu(E) < \infty$ – this is a strenghtening as we just proved that $\mathcal{F} \subset \mathcal{M}_\mu(X)$.

4. In the proof we showed that μ is a unique measure on $\mathcal{F}$ satisfying (a)-(c) and
$$\Lambda f = \int f \,\mathrm{d}\mu, \qquad f \in C_c(X).$$
Suppose that ν is another Borel **outer** measure (just like μ) so that it satisfies

(a) $\nu(K) < \infty$ for all compact sets $K \subset X$,
(b) $\nu(E) = \inf\{\nu(V) \colon E \subset V,\, V \text{ open}\}$ for all $E \subset X$,
(c) $\nu(E) = \sup\{\nu(K) \colon K \subset E,\, K \text{ compact}\}$ for all open E and also for all $E \in \mathcal{M}_\nu(X)$ with $\nu(E) < \infty$

and
$$\Lambda f = \int f \, \mathrm{d}\nu, \qquad f \in C_c(X).$$

Does it follow that $\mu(A) = \nu(A)$ for **all** $A \subset X$? Yes, and the argument is easy. We know from the proof that $\mu(U) = \nu(U)$ for all open U. As the regularity property (b) holds for all sets, it follows that $\mu(A) = \nu(A)$ for **all** $A \subset X$.

We can group all of what we did above together, and formulate a nice version of the Riesz representation theorem that explicitly yields outer measures.

5.0.6 Theorem. *Suppose X is a is a locally compact Hausdorff (LCH) space and $\Lambda \colon C_c(X) \to \mathbb{R}$ is an arbitrary positive linear functional. Then there exists a unique Borel regular outer measure $\mu \colon \mathcal{P}(X) \to [0, \infty]$ so that*
$$\Lambda f = \int f \, \mathrm{d}\mu, \qquad f \in C_c(X),$$
and μ also satisfies the following properties:

(a) $\mu(K) < \infty$ *for all compact sets* $K \subset X$,

(b) $\mu(E) = \inf\{\mu(V) \colon E \subset V,\, V \text{ open}\}$ *for all* $E \subset X$,

(c) $\mu(E) = \sup\{\mu(K) \colon K \subset E,\, K \text{ compact}\}$ *for all open E and also for all $E \in \mathcal{M}_\mu(X)$ with $\mu(E) < \infty$.*

About inner regularity

The outer regularity (b) holds more generally than the inner regularity – it is the nature of things that inner regularity (c) requires the measurability of E. However, we will show a setup in which (c) automatically holds somewhat more generally, namely without the restriction $\mu(E) < \infty$.

5.0.7 Remark. There is a technical example, which we omit here (but see exercises), showcasing that under our current minimal assumptions on the space X we cannot in general achieve inner regularity without the assumption $\mu(E) < \infty$. This happens in a space X which is not σ-compact in the sense of the following definition.

5.0.8 Definition. A set $A \subset X$ is called σ-compact if there exists compact sets
$$K_1, K_2, \ldots \subset A$$
so that $A = \bigcup_{i=1}^\infty K_i$. That is, A is a countable union of compact sets.

5.0.9 Remark. We note that if A is σ-compact (notice this implies $A \in \mathcal{M}_\mu(X)$ as A is a union of closed sets), then A also satisfies the inner regularity (c). If $\mu(A) < \infty$ there is nothing to prove as we already know this case. So assume $\mu(A) = \infty$. Write $A = \bigcup_{i=1}^\infty K_i$. By replacing K_i with $K_1 \cup \cdots \cup K_i$ we may assume that $K_1 \subset K_2 \subset \ldots$ (finite unions of compact sets are compact). By the convergence results of measures we know that
$$\infty = \mu(A) = \lim_{i \to \infty} \mu(K_i).$$

Thus, it also follows that

$$\sup\{\mu(K)\colon K \subset A,\ K \text{ compact}\} = \infty = \mu(A),$$

and so the inner regularity holds for A.

If X itself is σ-compact, then the measure μ given by the Riesz representation theorem satisfies somewhat stronger regularity properties – among them the fact that (c) holds for all $E \in \mathcal{M}_\mu(X)$.

5.0.10 Lemma. *Let X be a LCH space which is σ-compact and μ be an outer measure as in Theorem 5.0.6. Then the following holds.*

1. *If $A \in \mathcal{M}_\mu(X)$ and $\epsilon > 0$ there is a closed set F and open set U with $F \subset A \subset U$ and $\mu(U \setminus F) < \epsilon$.*
2. *Inner regularity holds for all $A \in \mathcal{M}_\mu(X)$.*
3. *If $A \in \mathcal{M}_\mu(X)$, there is a $\mathcal{F}_\sigma$-set $H \subset A$ and a $\mathcal{G}_\delta$-set $G \supset A$ so that $\mu(G \setminus H) = 0$.*

Proof. (1): Write $X = \bigcup_i K_i$, where K_i is compact. Let $A \in \mathcal{M}_\mu(X)$ – a difficulty here is that we may have $\mu(A) = \infty$. As $K_i \in \mathcal{M}_\mu(X)$ (as a closed set) we have that $A \cap K_i \in \mathcal{M}_\mu(X)$ and $\mu(A \cap K_i) \leq \mu(K_i) < \infty$. Let $\epsilon > 0$. Using outer regularity we find open $U_i \supset A \cap K_i$ so that

$$\mu(U_i) < \mu(A \cap K_i) + \epsilon/2^{i+1}.$$

Using measurability – $U_i, A \cap K_i \in \mathcal{M}_\mu(X)$ – and the fact that $\mu(A \cap K_i) < \infty$ this implies that

$$\mu(U_i \setminus (A \cap K_i)) < \epsilon/2^{i+1}.$$

Define the open set $U = \bigcup_i U_i \supset \bigcup_i (A \cap K_i) = A$. We have by subadditivity that

$$\mu(U \setminus A) = \mu\Big(\bigcup_i (U_i \setminus A)\Big) \leq \sum_i \mu(U_i \setminus A) \leq \sum_i \mu(U_i \setminus (A \cap K_i)) < \sum_{i=1}^{\infty} \epsilon/2^{i+1} = \frac{\epsilon}{2}.$$

Repeat the above proof with A replaced by $A^c = X \setminus A \in \mathcal{M}_\mu(X)$ – this gives an open set $V \supset A^c$ so that

$$\mu(V \cap A) = \mu(V \setminus A^c) < \frac{\epsilon}{2}.$$

Define the closed set $F = V^c \subset A$. Noticing that

$$U \setminus F = (U \setminus A) \cup (A \setminus F) = (U \setminus A) \cup (A \cap F^c) = (U \setminus A) \cup (A \cap V)$$

we now finally have

$$\mu(U \setminus F) \leq \mu(U \setminus A) + \mu(V \cap A) < \epsilon/2 + \epsilon/2 = \epsilon.$$

(2) Let $A \in \mathcal{M}_\mu(X)$. As the case $\mu(A) < \infty$ is known, we may assume $\mu(A) = \infty$. Using (1) choose a closed $F \subset A$ with $\mu(A \setminus F) < 1$. As

$$\infty = \mu(A) = \mu(F) + \mu(A \setminus F)$$

and $\mu(A \setminus F) < \infty$, this forces $\mu(F) = \infty$. Using the decomposition of $X = \bigcup_i K_i$ write $F = \bigcup_i (K_i \cap F)$, where $K_i \cap F$ is compact. Thus, F is σ-compact and by the remark from above

$$\infty = \mu(F) = \sup\{\mu(K)\colon K \subset F,\ K \text{ compact}\}.$$

Thus

$$\sup\{\mu(K)\colon K \subset A,\ K \text{ compact}\} \geq \sup\{\mu(K)\colon K \subset F,\ K \text{ compact}\} = \infty$$

showing the inner regularity for A.

(3): Simply choose using (1) open sets $U_k \supset A$ and closed sets $F_k \subset A$ so that $\mu(U_k \setminus F_k) < 1/k$, and set $G = \bigcap_k U_k$ and $H = \bigcup_k F_k$. ☐

5.0.11 Remark. Notice that (3) is a characterization of the sets $A \in \mathcal{M}_\mu(X)$. Indeed, if $A \subset X$ is such that (3) holds, then e.g. $A = H \cup (A \setminus H)$, where $H \in \mathrm{Bor}(X) \subset \mathcal{M}_\mu(X)$ and $A \setminus H \in \mathcal{M}_\mu(X)$ as $\mu(A \setminus H) = 0$ and μ is complete. Thus $A \in \mathcal{M}_\mu(X)$.

5.0.12 Remark. Notice that if μ is an outer measure as in Theorem 5.0.6, $A \in \mathcal{M}_\mu(X)$ **and** $\mu(A) < \infty$, then even without σ-compactness we can choose open $U \supset A$ and compact $K \subset A$ so that $\mu(U) < \mu(A) + \epsilon/2$ and $\mu(A) < \mu(K) + \epsilon/2$. Then

$$\mu(U \setminus K) \leq \mu(U \setminus A) + \mu(A \setminus K) < \epsilon.$$

So the fact that we got only a closed set F above (and not a compact one), and that we needed σ-compactness for the argument, was to work with general sets which may have $\mu(A) = \infty$.

Local finiteness implies inner and outer regularity

Next, we want to take a measure ν and use the Riesz representation theorem to show that local finiteness (property (a) i.e. $\nu(K) < \infty$ for compact K) in many spaces automatically implies (b) and (c) directly. This is simply great.

Remember that we started the whole discussion of the Riesz representation theorem by noticing that a locally finite Borel measure ν induces the positive linear functional

$$\Lambda f = \int f \, \mathrm{d}\nu, \qquad f \in C_c(X).$$

So, given the measure ν, the definition of this functional associated with ν does not require any inner or outer regularity to begin with (the regularity properties (b) and (c), respectively). But then the Riesz representation theorem gives us the Borel outer measure μ satisfying (a) to (c) and

$$\int f \, \mathrm{d}\nu = \Lambda f = \int f \, \mathrm{d}\mu, \qquad f \in C_c(X).$$

The measure μ is also unique in a certain sense – thus, thinking quickly, now it may appear that automatically $\mu = \nu$ and so ν has to in fact be inner and outer regular as well. Not quite – remember that we said that μ is the unique measure satisfying (a)-(c) and $\Lambda f = \int f \, \mathrm{d}\mu$. So if ν does not satisfy (b)-(c), we cannot invoke the uniqueness part to conclude $\mu = \nu$. This seems strange – can it really be that $\mu \neq \nu$? Yes, but not in most reasonable settings. This pathology cannot happen if we assume a bit more of the underlying space X – then we can actually prove that $\mu(B) = \nu(B)$ for all Borel sets B, and then it is true that locally finite measures are automatically inner and outer regular for all Borel sets as well. (If μ, ν are both outer measures, be careful with the claim $\mu = \nu$ – without Borel regularity outer measures can easily agree on $\mathrm{Bor}(X)$ but not quite everywhere, see the exercises.) For example, this striking corollary holds in $\mathbb{R}^d$, meaning that many measures in $\mathbb{R}^d$ are automatically regular (inner and outer regular for all Borel sets).

5.0.13 Proposition. *Let X be a LCH space, where every open set $U \subset X$ is σ-compact. Let ν, λ be locally finite Borel measures on X satisfying*

$$\int f \, \mathrm{d}\nu = \int f \, \mathrm{d}\lambda, \qquad f \in C_c(X).$$

Then $\nu(A) = \lambda(A)$ for all Borel sets $A \in \mathrm{Bor}(X)$.

Proof. Let μ be the measure given by the Riesz representation theorem when applied with the functional $\Lambda f := \int f \, \mathrm{d}\nu = \int f \, \mathrm{d}\lambda$, $f \in C_c(X)$. By symmetry it is enough to show that $\nu(A) = \mu(A)$ for every Borel set $A \subset X$. Consider first an open set $U \subset X$. Using the assumption that U is σ-compact write $U = \bigcup_i K_i$, where K_i is compact. By Urysohn's lemma we find f_i with $K_i \prec f_i \prec U$. Define $g_n = \max(f_1, \ldots, f_n)$ so that $g_n \in C_c(X)$, $0 \leq g_1 \leq g_2 \leq \cdots \leq 1_U$ with $1_U(x) = \lim_{n\to\infty} g_n(x)$. By the monotone convergence theorem we have

$$\nu(U) = \int 1_U \, \mathrm{d}\nu = \lim_{n\to\infty} \int g_n \, \mathrm{d}\nu = \lim_{n\to\infty} \Lambda g_n = \lim_{n\to\infty} \int g_n \, \mathrm{d}\mu = \mu(U).$$

Let now A be a Borel set and $\epsilon > 0$. By Lemma 5.0.10 we find closed F and open U with $F \subset A \subset U$ and $\mu(U \setminus F) < \epsilon$. Since $U \setminus F$ is open we have $\nu(U \setminus F) = \mu(U \setminus F) < \epsilon$. Thus, we get

$$\nu(A) \leq \nu(U) = \mu(U) = \mu(F) + \mu(U \setminus F) \leq \mu(F) + \epsilon \leq \mu(A) + \epsilon$$

and similarly

$$\mu(A) \leq \mu(U) = \nu(U) = \nu(F) + \nu(U \setminus F) \leq \nu(F) + \epsilon \leq \nu(A) + \epsilon.$$

Thus $\nu(A) = \nu(A)$. □

5.0.14 Corollary. *Let X be a LCH space, where in addition every open set $U \subset X$ is σ-compact. Let ν be a locally finite Borel measure on X. Then ν is regular in the sense that for every Borel set $E \subset X$ we have*

$$\nu(E) = \inf\{\nu(V) \colon E \subset V, \, V \text{ open}\}$$

and

$$\nu(E) = \sup\{\nu(K) \colon K \subset E, \, K \text{ compact}\}.$$

Proof. By Proposition 5.0.13 the measure ν agrees with the measure μ coming from the Riesz representation theorem at least on all Borel sets E. By Lemma 5.0.10 the measure μ, and thus ν, satisfies the desired regularity properties. □

Chapter 6

Construction of the Lebesgue measure from the Riesz representation theorem

For an elementary construction of the Lebesgue measure, see the optional Appendix A.

One natural viewpoint to the Lebesgue measure on $\mathbb{R}^d$ is that it should be a measure $\mu = m_d$ so that it helps us to generalize Riemann integration in the sense that integration against the Lebesgue measure m_d agrees with the Riemann integration for nice functions. To this end, we can define

$$\Lambda\colon C_c(\mathbb{R}^d) \to \mathbb{R},\ \Lambda f = \int_{\mathbb{R}^d} f(x)\,\mathrm{d}x,$$

where the integral on the right is the Riemann integral – it is well-defined for $f \in C_c(\mathbb{R}^d)$. In fact, we prefer not to explicitly rely on the theory of Riemann integration on $\mathbb{R}^d$ and will instead soon redefine Λf using an elementary construction (but the construction we give is the Riemann integral).

In any case, with Λf defined we can use the Riesz representation theorem, Theorem 5.0.6 – it gives us the Lebesgue outer measure m_d so that

$$\int_{\mathbb{R}^d} f(x)\,\mathrm{d}m_d(x) = \int_{\mathbb{R}^d} f(x)\,\mathrm{d}x$$

at least whenever $f \in C_c(\mathbb{R}^d)$. We also know that m_d is a Borel regular outer measure satisfying

(a) $m_d(K) < \infty$ for all compact sets $K \subset \mathbb{R}^d$,

(b) $m_d(E) = \inf\{m_d(V)\colon E \subset V,\, V \subset \mathbb{R}^d \text{ open}\}$ for all $E \subset \mathbb{R}^d$,

(c) $m_d(E) = \sup\{m_d(K)\colon K \subset E,\, K \subset \mathbb{R}^d \text{ compact}\}$ for all Lebesgue measurable sets

$$E \in \operatorname{Leb}(\mathbb{R}^d) := \mathcal{M}_{m_d}(\mathbb{R}^d).$$

For (c) we used Item 2 of Lemma 5.0.10, which we can do as $\mathbb{R}^d$ is σ-compact – it is also, of course, useful to know that all the rest of the conclusion of Lemma 5.0.10 hold for m_d.

What other properties does m_d have apart from the regularity properties and the fact that it can be used to generalize Riemann integration? Before we go into those, we need some general tools and observations. First, a brief introduction to dyadic cubes. Let

$$\mathcal{D} := \{2^{-k}([0,1)^d + m)\colon k \in \mathbb{Z},\, m \in \mathbb{Z}^d\}$$

be the collection of dyadic cubes in $\mathbb{R}^d$. We can decompose

$$\mathcal{D} = \bigcup_{k \in \mathbb{Z}} \mathcal{D}_k, \qquad \mathcal{D}_k := \{Q \in \mathcal{D} \colon \ell(Q) = 2^{-k}\},$$

where $\ell(Q)$ denotes the side length of Q, and each $\mathcal{D}_k$ partitions $\mathbb{R}^d$:

$$\mathbb{R}^d = \bigcup_{Q \in \mathcal{D}_k} Q$$

is a disjoint union for all $k \in \mathbb{Z}$. Notice also the following critically important nestedness property of the dyadic cubes $Q, R \in \mathcal{D}$: either $Q \cap R = \emptyset$ or $Q \subset R$ or $R \subset Q$.

A final important thing about dyadic cubes concerns the existence and behaviour of maximal cubes in some subcollection $\mathcal{S}$ of $\mathcal{D}$. To this end, let $\mathcal{S} \subset \mathcal{D}$ be a collection of dyadic cubes that does not have an infinite increasing chain – there does not exist an infinite sequence of cubes $Q_i \in \mathcal{S}$, $i \in \mathbb{N}$, with $Q_i \subsetneq Q_{i+1}$ for all i.

1. An element $Q \in \mathcal{S}$ is called maximal if there does not exist $R \in \mathcal{S}$ with $R \supsetneq Q$. The collection of maximal elements in $\mathcal{S}$ is denoted by $\mathcal{S}^*$. It is easy to see that each $Q \in \mathcal{S}$ is contained in some maximal $Q^* \in \mathcal{S}^*$. Indeed, let $Q \in \mathcal{S}$ be arbitrary. If Q is itself maximal, we are done by setting $Q^* = Q$. Otherwise, by definition there exists $Q_1 \in \mathcal{S}$ with $Q_1 \supsetneq Q$. If Q_1 is maximal, we are done by setting $Q^* = Q_1$. Otherwise, we continue this process – there exists $Q_2 \in \mathcal{S}$ with $Q_2 \supsetneq Q_1$. This process has to stop after finitely many steps as, by assumption, $\mathcal{S}$ does not have an infinite increasing chain. Thus, eventually we find a maximal cube Q^* containing Q.

2. Moreover, maximal cubes are disjoint: $Q, R \in \mathcal{S}^*$, $Q \neq R$, implies $Q \cap R = \emptyset$. Aiming for a contradiction, assume that $Q \cap R \neq \emptyset$. By the nestedness property either $Q \subset R$ or $R \subset Q$. By symmetry, we may assume that $Q \subset R$. Since $Q \neq R$ we have $Q \subsetneq R$. But this contradicts the maximality of Q.

6.0.1 Lemma. *Every open set $V \subset \mathbb{R}^d$ is a disjoint union of dyadic cubes.*

Proof. Let

$$\mathcal{S} := \big\{Q \in \mathcal{D} \colon\ Q \subset V \text{ and } \ell(Q) \leq 1\big\}.$$

Obviously $\mathcal{S}$ does not have an infinite increasing chain as we have capped $\ell(Q) \leq 1$ (the choice of the constant 1 is rather arbitrary here). Let $\mathcal{S}^*$ denote the maximal cubes – we know from above that

$$\bigcup_{Q \in \mathcal{S}} Q = \bigcup_{Q^* \in \mathcal{S}^*} Q^*$$

and that $\mathcal{S}^*$ consists of disjoint cubes. It remains to show that given $x \in V$ there exists $Q \in \mathcal{S}$ with $x \in Q$. But this is clear, since V is open and so x belongs to arbitrarily small dyadic cubes Q with $Q \subset V$. ☐

We also need the following later.

6.0.2 Lemma. *Let ν be a locally finite Borel measure on $\mathbb{R}^d$ satisfying for all $Q \in \mathcal{D}$ that*

$$\nu(Q) = m_d(Q).$$

Then $\nu(E) = m_d(E)$ for all Borel sets $E \subset \mathbb{R}^d$.

Proof. By Corollary 5.0.14 we know that

$$\nu(E) = \inf\{\nu(V) \colon E \subset V,\, V \subset \mathbb{R}^d \text{ open}\}$$

for every Borel set $E \subset \mathbb{R}^d$, and we know that m_d satisfies the corresponding outer regularity property as well. If we prove that $\nu(V) = m_d(V)$ for all open sets V, this then implies that $\nu(E) = m_d(E)$ for all Borel sets E. Thus, fix an open set V and write using Lemma 6.0.1 that $V = \bigcup_{Q\in\mathcal{P}} Q$, where $\mathcal{P} \subset \mathcal{D}$ is some collection of disjoint dyadic cubes. By countable additivity

$$\nu(V) = \sum_{Q\in\mathcal{P}} \nu(Q) = \sum_{Q\in\mathcal{P}} m_d(Q) = m_d(V).$$

□

6.0.3 Remark. Clearly we only needed that $\nu(Q) = m_d(Q)$ for all $Q \in \mathcal{D}$ with $\ell(Q) \le 1$.

6.0.4 Definition. A rectangle on $\mathbb{R}^d$ (with sides parallel to the coordinate axes) is a set R of the form

$$R = I_1 \times \cdots \times I_d = \prod_{i=1}^{d} I_i,$$

where I_i is an interval with endpoints $-\infty < a_i < b_i < \infty$. We define

$$\operatorname{vol}(R) := \prod_{i=1}^{d} (b_i - a_i).$$

6.0.5 Remark. In what follows, when we say rectangle, we mean one with sides parallel to the coordinate axes.

We now more carefully define Λf, $f \in C_c(\mathbb{R}^d)$, with an explicit formula (this definition will agree with the Riemann integral of f). If $g \colon \mathbb{R}^d \to \mathbb{R}$ is any function with compact support (not necessarily continuous), define

$$\Lambda_k g = \sum_{Q\in\mathcal{D}_k} g(x^Q)\operatorname{vol}(Q), \qquad k = 0, 1, 2, \ldots,$$

where x^Q is the corner of $Q \in \mathcal{D}_k$ as follows

$$Q = \{x \colon x_i^Q \le x_i < x_i^Q + \ell(Q),\ 1 \le i \le d\} = \{x \colon x_i^Q \le x_i < x_i^Q + 2^{-k},\ 1 \le i \le d\}.$$

Clearly the sum is finite as g has compact support.

We will define

$$\Lambda f = \lim_{k\to\infty} \Lambda_k f$$

whenever $f \in C_c(\mathbb{R}^d)$, but we first need to briefly justify why the limit exists. We make the following preliminary observation. Suppose g has the special form

$$g = \sum_{R\in\mathcal{D}_N} c_R 1_R,$$

that is, g is contant on dyadic cubes of length scale 2^{-N}. Let $k > N$ – we want to argue that $\Lambda_k g = \Lambda_N g$. For a dyadic cube Q we denote by $Q^{(m)}$ the unique dyadic cube with $Q \subset Q^{(m)}$ and $\ell(Q^{(m)}) = 2^m \ell(Q)$ (the mth dyadic parent of Q). We may now write

$$\Lambda_k g = \sum_{Q\in\mathcal{D}_k} g(x^Q)\operatorname{vol}(Q)$$

$$= \sum_{R \in \mathcal{D}_N} \sum_{\substack{Q \in \mathcal{D}_k \\ Q^{(k-N)} = R}} g(x^Q) \operatorname{vol}(Q) = \sum_{R \in \mathcal{D}_N} c_R 2^{-dk} \sum_{\substack{Q \in \mathcal{D}_k \\ Q^{(k-N)} = R}} 1.$$

Given $R \in \mathcal{D}_N$ how many $Q \in \mathcal{D}_k$ are there with $Q^{(k-N)} = R$? Let M be the total number – then we must have

$$M \cdot 2^{-dk} = 2^{-dN}$$

so that $M = 2^{dk} 2^{-dN}$. Thus, we get

$$\Lambda_k g = \sum_{R \in \mathcal{D}_N} c_R 2^{-dk} 2^{dk} 2^{-dN} = \sum_{R \in \mathcal{D}_N} g(x^R) \operatorname{vol}(R) = \Lambda_N g$$

as desired.

Let now $f \in C_c(\mathbb{R}^d)$ and choose some (non-dyadic) open cube W with $\operatorname{spt} f \subset W$. Let $\epsilon > 0$. Notice that f is uniformly continuous (as its support is compact) and so we find N and construct functions g, h which are also supported on W, constant on each $Q \in \mathcal{D}_N$, $g \leq f \leq h$ and $h - g < \epsilon$. For all $k > N$ we get

$$\Lambda_N g = \Lambda_k g \leq \Lambda_k f \leq \Lambda_k h = \Lambda_N h.$$

This shows that

$$\limsup_{k \to \infty} \Lambda_k f - \liminf_{k \to \infty} \Lambda_k f \leq \Lambda_N (h - g) < \epsilon \sum_{R \in \mathcal{D}_N :\ R \subset W} \operatorname{vol}(R) \leq \epsilon \operatorname{vol}(W),$$

where we used that the cubes $R \in \mathcal{D}_N$ are disjoint. Thus, the limit $\Lambda f = \lim_{k \to \infty} \Lambda_k f$ exists. Now we have a good definition of Λf without concretely relying on some theory of Riemann integration. We move on to proving some natural properties for the related Lebesgue measure m_d – the properties suggest we can think of $m_d(A)$ as a rigorous notion of the d-dimensional volume of $A \subset \mathbb{R}^d$.

6.0.6 Lemma. *We have*

$$m_d(R) = \operatorname{vol}(R)$$

whenever R is a rectangle.

Proof. Let R be an open rectangle. Let

$$E_k := \bigcup_{\substack{Q \in \mathcal{D}_k \\ \overline{Q} \subset R}} Q.$$

Choose $E_k \prec g_k \prec R$, $k = 0, 1, \ldots$, and set $f_k = \max(g_1, \ldots, g_k)$. Clearly $0 \leq f_1 \leq f_2 \leq \cdots$ and $1_R(x) = \lim_{k \to \infty} f_k(x)$. By the monotone convergence theorem we have

$$m_d(R) = \int 1_R(x) \, \mathrm{d}m_d(x) = \lim_{k \to \infty} \int f_k(x) \, \mathrm{d}m_d(x) = \lim_{k \to \infty} \Lambda f_k.$$

Moreover, we have

$$\Lambda f_k = \lim_{m \to \infty} \Lambda_m f_k,$$

where

$$\Lambda_m f_k \leq \Lambda_m 1_R = \sum_{Q \in \mathcal{D}_m} 1_R(x^Q) \operatorname{vol}(Q) \leq \sum_{\substack{Q \in \mathcal{D}_m \\ Q \cap R \neq \emptyset}} \operatorname{vol}(Q).$$

Notice that if $Q \in \mathcal{D}_m$ and $Q \cap R \neq \emptyset$, then $Q \subset R_m := \{x\colon\ \operatorname{dist}(x, R) \leq \sqrt{d}2^{-m}\}$, and so

$$\sum_{\substack{Q\in\mathcal{D}_m \\ Q\cap R\neq\emptyset}} \operatorname{vol}(Q) \leq \sum_{\substack{Q\in\mathcal{D}_m \\ Q\subset R_m}} \operatorname{vol}(Q) \leq \operatorname{vol}(R_m) \to \operatorname{vol}(R)$$

as $m \to \infty$. Thus $\Lambda f_k \leq \operatorname{vol}(R)$ and so $m_d(R) \leq \operatorname{vol}(R)$. We also have

$$\Lambda f_k \geq \Lambda g_k = lim_{m\to\infty}\Lambda_m g_k,$$

where

$$\Lambda_m g_k \geq \Lambda_m 1_{E_k} = \sum_{Q\in\mathcal{D}_m} 1_{E_k}(x^Q)\operatorname{vol}(Q).$$

If $m > k$ we may write

$$\begin{aligned}\sum_{Q\in\mathcal{D}_m} 1_{E_k}(x^Q)\operatorname{vol}(Q) &= \sum_{S\in\mathcal{D}_k}\sum_{\substack{Q\in\mathcal{D}_m \\ Q^{(m-k)}=S}} 1_{E_k}(x^Q)\operatorname{vol}(Q)\\ &\geq \sum_{\substack{S\in\mathcal{D}_k \\ S\subset E_k}}\sum_{\substack{Q\in\mathcal{D}_m \\ Q^{(m-k)}=S}} \operatorname{vol}(Q)\\ &= \sum_{\substack{S\in\mathcal{D}_k \\ S\subset E_k}} \operatorname{vol}(S),\end{aligned}$$

and so

$$\Lambda f_k \geq \sum_{\substack{S\in\mathcal{D}_k \\ S\subset E_k}} \operatorname{vol}(S) \to \operatorname{vol}(R)$$

as $k \to \infty$, and so $m_d(R) \geq \operatorname{vol}(R)$. We have proved the case that R is an open rectangle – the general case follows from this by writing a general rectangle as an intersection of a decreasing sequence of open rectangles. □

6.0.7 Lemma. *For an arbitrary $A \subset \mathbb{R}^d$ we have for all $x \in \mathbb{R}^d$ and $t > 0$ that*

$$m_d(A + x) = m_d(A) \qquad \text{and} \qquad m_d(tA) = t^d m_d(A),$$

where

$$A + x = \{a + x\colon a \in A\} \qquad \text{and} \qquad tA := \{ta\colon a \in A\}.$$

Proof. Fix x and define the outer measure $\nu(A) := m_d(A + x)$. Then for a rectangle R we have $\nu(R) = m_d(R + x) = \operatorname{vol}(R + x) = \operatorname{vol}(R) = m_d(R)$. By the argument of Lemma 6.0.2 we have $\nu(A) = m_d(A)$ for every set $A \subset \mathbb{R}^d$. The other claim is proved similarly using $\operatorname{vol}(tR) = t^d \operatorname{vol}(R)$ if R is a rectangle. □

6.0.8 Example. Notice that

$$B(x, r) = x + rB(0, 1)$$

and so

$$m_d(B(x, r)) = m_d(rB(0, 1)) = m_d(B(0, 1))r^d = C(d)r^d,$$

where $C(d) := m_d(B(0,1))$ is some dimensional constant. It can be proved that

$$C(d) = \frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}+1\right)},$$

where the gamma function is defined by

$$\Gamma(a) := \int_0^\infty x^{a-1}e^{-x}\,\mathrm{d}x, \qquad 0 < a < \infty.$$

However, this explicit formula is not necessary for most things.

We call measures in $\mathbb{R}^d$ satisfying $\nu(A+x) = \nu(A)$ translation invariant. There are not many of these.

6.0.9 Lemma. *Let ν be a locally finite Borel measure on $\mathbb{R}^d$ that is translation invariant. Then for some constant C we have $\nu(E) = Cm_d(E)$ for all Borel sets E.*

Proof. Let $C := \nu(Q_0)$, where $Q_0 \in \mathcal{D}_0 = \{Q \in \mathcal{D} : \ell(Q) = 1\}$. Notice that C is independent of the choice of Q_0 as ν is translation invariant. For $k \geq 0$ write

$$Q_0 = \bigcup_{\substack{Q\in\mathcal{D}_k \\ Q\subset Q_0}} Q.$$

As

$$C = \nu(Q_0) = \sum_{\substack{Q\in\mathcal{D}_k \\ Q\subset Q_0}} \nu(Q),$$

$\nu(Q) =: C_k$ is a constant independent of $Q \in \mathcal{D}_k$ and there are exactly 2^{dk} cubes in the union, we must have

$$C = 2^{dk}C_k.$$

Thus, for every $Q \in \mathcal{D}_k$ we have

$$\nu(Q) = C_k = C2^{-dk} = C\,\mathrm{vol}(Q) = Cm_d(Q).$$

The claim follows from Lemma 6.0.2. □

6.0.10 Remark. Suppose $T : \mathbb{R}^d \to \mathbb{R}^d$ is a bijective linear map (if it is not bijective the following is not interesting as then the dimension of $T\mathbb{R}^d$ is smaller than d). Then T is a homeomorphism and we can define the Borel measure $\nu(E) = m_d(TE)$ (TE is a Borel set if E is as T is a homeomorphism). Notice that by linearity and the translation invariance of m_d we have

$$\nu(E+x) = m_d(T(E+x)) = m_d(TE+Tx) = m_d(TE) = \nu(E).$$

The previous lemma then implies

$$m_d(TE) = \nu(E) = C_T m_d(E)$$

for some constant C_T. It can be shown that $C_T = |\det T|$, but this requires a separate argument.

One interesting special case of the above is the case that T is a rotation. Then we have

$$C_T = \frac{m_d(TB(0,1))}{m_d(B(0,1))} = \frac{m_d(B(0,1))}{m_d(B(0,1))} = 1.$$

This shows that m_d is rotation invariant as well.

There are many ways to define the Lebesgue measure. The one we gave is not elementary as it is based on the deep Riesz representation theorem. On the other hand, it is clearly motivated by the desire to generalize Riemann integration, and it is also extremely convenient since we get for free that m_d is a Borel outer measure and satisfies the various inner and outer regularity properties. A more elementary, geometrically motivated, definition is given by the following:

$$m_d(A) = \inf \sum_i \mathrm{vol}(R_i),$$

where the infimum is taken over all countable coverings $A \subset \bigcup_i R_i$ by closed rectangles. With some work and using the previous result, one can show that this agrees with our definition.

It is very usual to write $m_d(A) = |A|$ for $A \subset \mathbb{R}^d$. It is also very usual to write

$$\int_A f(x)\,\mathrm{d}x := \int_A f(x)\,\mathrm{d}m_d(x).$$

When $d = 1$ it is normal to write

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

as well, since it does not matter how we interpret this –

$$\int_{[a,b]} f(x)\,\mathrm{d}m_d(x) = \int_{[a,b)} f(x)\,\mathrm{d}m_d(x) = \cdots$$

and so on as $|\{x\}| = 0$. Everything learned about integration in calculus remains useful: if e.g. f is continuous on $[a,b]$, then the Riemann and Lebesgue integral of f over $[a,b]$ coincide. However, these are also more general Lebesgue integral versions of all relevant results – like the fundamental theorem of calculus. It is not the focus of this course to study all of the aspects of the Lebesgue measure and we content with this introduction, which is sufficient for many purposes.

Chapter 7

Signed measures and the Radon-Nikodym theorem

Up to this point our measures have always been non-negative. We want to generalize this now.

7.0.1 Definition. Let X be a space and $\mathcal{F}$ be a σ-algebra on X (i.e. $(X, \mathcal{F})$ is a measurable space). A function $\mu\colon \mathcal{F} \to \dot{\mathbb{R}}$ is called a signed measure if the following holds.

1. $\mu(\emptyset) = 0$.
2. μ can only receive at most one of the values $\pm\infty$.
3. If $A = \bigcup_{i=1}^{\infty} A_i$, $A_i \in \mathcal{F}$, and $A_i \cap A_j = \emptyset$ for $i \neq j$, then
$$\mu(A) = \sum_i \mu(A_i),$$
where the sum on the right converges absolutely.

The requirement (2) is added to make sure that a scenario $\infty - \infty$ does not appear. Some authors even demand that μ is strictly $\mathbb{R}$-valued (never takes the value ∞ or $-\infty$).

For example, the difference $\mu_1 - \mu_2$ of two measures $\mu_1, \mu_2\colon \mathcal{F} \to [0, \infty]$, one of which is a finite measure, is a signed measure. In fact, all signed measures arise like this as we will show (Jordan decomposition).

7.0.2 Definition. We say that $P \in \mathcal{F}$ is a positive set for μ if $\mu(A) \geq 0$ for all $A \in \mathcal{F}$ with $A \subset P$. We say that $P \in \mathcal{F}$ is negative set for μ if $\mu(A) \leq 0$ for all $A \in \mathcal{F}$ with $A \subset P$.

7.0.3 Lemma. *Let $\mu\colon \mathcal{F} \to \dot{\mathbb{R}}$ be a signed measure.*

1. *If $A \subset B$, $A, B \in \mathcal{F}$ and $|\mu(B)| < \infty$, then $|\mu(A)| < \infty$.*
2. *Let $A_i \in \mathcal{F}$ and $A_1 \subset A_2 \subset \cdots$. Then we have*
$$\mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \lim_{i\to\infty} \mu(A_i).$$
3. *Let $A_i \in \mathcal{F}$, $A_1 \supset A_2 \supset \cdots$ and $|\mu(A_1)| < \infty$. Then we have*
$$\mu\Big(\bigcap_{i=1}^{\infty} A_i\Big) = \lim_{i\to\infty} \mu(A_i).$$

Proof. Optional (easy) exercise – compare to the properties of usual measures. □

7.0.4 Lemma. *Let $\mu\colon \mathcal{F} \to \dot{\mathbb{R}}$ be a signed measure. If the sets $P_1, P_2, \ldots \in \mathcal{F}$ are positive sets for μ, then so is $\bigcup_i P_i$.*

Proof. Let $E \subset P := \bigcup_i P_i$, $E \in \mathcal{F}$. We need to show that $\mu(E) \geq 0$. Define $E_1 = E \cap P_1$ and inductively

$$E_k = E \cap P_k \setminus \bigcup_{i=1}^{k-1} P_i, \qquad k \geq 2.$$

As $E_k \in \mathcal{F}$ and $E_k \subset P_k$ we have $\mu(E_k) \geq 0$ by assumption. Thus, we have by disjointness that

$$\mu(E) = \sum_{k=1}^{\infty} \mu(E_k) \geq 0.$$

□

7.0.5 Lemma. *Let $\mu\colon \mathcal{F} \to \dot{\mathbb{R}}$ be a signed measure, $A \in \mathcal{F}$ and $0 < \mu(A) < \infty$. Then there exists $P \subset A$ so that $P \in \mathcal{F}$ is a positive set for μ and $\mu(P) > 0$.*

Proof. If A is a positive set for μ, we are done. Otherwise, we have

$$L_1 := \inf\{\mu(E)\colon E \in \mathcal{F},\, E \subset A\} < 0.$$

Let n_1 be the smallest natural number so that $L_1 < -1/n_1$. Choose $A_1 \in \mathcal{F}$ with $A_1 \subset A$ and $\mu(A_1) < -1/n_1$. Notice that $|\mu(A_1)| < \infty$ as $A_1 \subset A$ and so

$$\mu(A \setminus A_1) = \mu(A) - \mu(A_1) > \mu(A) + \frac{1}{n_1}.$$

So $0 < \mu(A \setminus A_1) < \infty$ – if this set is a positive set for μ we are done. Otherwise, we repeat the previous argument with A replaced by $A \setminus A_1$, and find $A_2 \subset A \setminus A_1$ with $\mu(A_2) < -1/n_2$, where n_2 is the smallest natural number so that

$$L_2 := \inf\{\mu(E)\colon E \in \mathcal{F},\, E \subset A \setminus A_1\} < -\frac{1}{n_2}.$$

Then

$$\mu(A \setminus (A_1 \cup A_2)) > \mu(A) + \frac{1}{n_1} + \frac{1}{n_2}.$$

If this process ends after finitely many steps, we are done. Otherwise, we produce disjoint sets $A_1, A_2, \ldots \in \mathcal{F}$, numbers $L_1, L_2, \ldots$ and $n_1, n_2, \ldots$ defined in the above way. Define

$$P := A \setminus \bigcup_{j=1}^{\infty} A_j.$$

Notice that

$$\infty > \mu(P) = \mu(A) - \sum_{j=1}^{\infty} \mu(A_j) > \mu(A) + \sum_{j=1}^{\infty} \frac{1}{n_j} > 0.$$

It remains to show that P is a positive set for μ. The above estimate shows that $\sum_{j=1}^{\infty} \frac{1}{n_j} < \infty$ so $n_j \to \infty$ as $j \to \infty$. Aiming for a contradiction, suppose there exists $E \in \mathcal{F}$, $E \subset P$ so that $\mu(E) < 0$. Using $n_j \to \infty$ choose $n_{j_0} > 1$ so that

$$\frac{1}{n_{j_0} - 1} < -\mu(E)$$

i.e. $\mu(E) < -1/(n_{j_0} - 1)$. As we in particular have, since $E \subset P$, that

$$E \subset A \setminus \bigcup_{j=1}^{j_0-1} A_j,$$

we have that

$$L_{j_0} < -1/(n_{j_0} - 1).$$

But we chose n_{j_0} to be the smallest integer so that $L_{j_0} < -1/n_{j_0}$ so the same inequality cannot hold with the smaller integer $n_{j_0} - 1$ in place of n_{j_0} – a contradiction. □

7.0.6 Theorem (Hahn decomposition)**.** *Let $\mu \colon \mathcal{F} \to \dot{\mathbb{R}}$ be a signed measure. Then there exists $P \in \mathcal{F}$ that is a positive set for μ so that $X \setminus P$ is a negative set for μ. The pair (P, P^c) is a Hahn decomposition of the signed measure μ.*

Proof. We may assume $\mu(E) < \infty$ for all $E \in \mathcal{F}$ (if this does not originally hold for μ we can study $-\mu$). Let

$$L := \sup\{\mu(P) \colon P \in \mathcal{F} \text{ is a positive set for } \mu\}.$$

Notice that L is well-defined as at least $\emptyset$ is a positive set for μ. Choose positive sets $P_1, P_2, \ldots$ of μ so that $\mu(P_i) \to L$, and define

$$P := \bigcup_{i=1}^{\infty} P_i.$$

Notice that P is a positive set for μ by Lemma 7.0.4. It remains to show that $X \setminus P = P^c$ is a negative set for μ.

Notice first that $\mu(P) = L$ – indeed, since P is a positive set for μ, we have $\mu(P) \leq L$ and $\mu(P \setminus P_i) \geq 0$ so that

$$L \geq \mu(P) = \mu(P_i) + \mu(P \setminus P_i) \geq \mu(P_i) \to L.$$

As $L = \mu(P)$ we in particular have $L < \infty$ since by our assumption from the beginning $\mu(P) < \infty$.

Aiming for a contradiction, suppose there exists $E \in \mathcal{F}$, $E \subset P^c$ with $\mu(E) > 0$. By Lemma 7.0.5 we find $S \in \mathcal{F}$ so that $S \subset E \subset P^c$, S is a positive set for μ and $\mu(S) > 0$. Now $P \cup S$ is a positive set for μ (by Lemma 7.0.4) and by disjointness

$$\mu(P \cup S) = \mu(P) + \mu(S) = L + \mu(S) > L,$$

which is a contradiction. □

Two (positive) measures $\mu, \nu \colon \mathcal{F} \to [0, \infty]$ are called singular if there are disjoint $N_1, N_2 \in \mathcal{F}$ with $X = N_1 \cup N_2$ and $\mu(N_1) = 0 = \nu(N_2)$. This is denoted $\mu \perp \nu$.

7.0.7 Theorem (Jordan decomposition)**.** *Let $\mu \colon \mathcal{F} \to \dot{\mathbb{R}}$ be a signed measure. Then there are measures $\mu^+, \mu^- \colon \mathcal{F} \to [0, \infty]$ – at least one of which is finite – so that $\mu = \mu^+ - \mu^-$ and $\mu^- \perp \mu^+$. This decomposition is unique.*

Proof. We may again assume that $\mu(E) < \infty$ for all $E \in \mathcal{F}$. Fix a Hahn decomposition (P, P^c) of μ and define

$$\mu^+(E) := \mu(E \cap P)$$
$$\mu^-(E) := -\mu(E \cap P^c)$$

Now $\mu^+ \colon \mathcal{F} \to [0, \mu(P)]$ and $\mu^- \colon \mathcal{F} \to [0, \infty]$ are measures and

$$\mu(E) = \mu^+(E) - \mu^-(E), \qquad E \in \mathcal{F}.$$

Moreover, we have $\mu^+(P^c) = 0 = \mu^-(P)$ so $\mu^+ \perp \mu^-$. The uniqueness part is an exercise. □

7.0.8 Corollary. *Every signed measure $\mu\colon \mathcal{F} \to \dot{\mathbb{R}}$ is either bounded from above or from below.*

Proof. Write the Jordan decomposition $\mu = \mu^+ - \mu^-$ and notice that

$$-\mu^-(X) \le -\mu^-(E) \le \mu^+(E) - \mu^-(E) \le \mu^+(E) \le \mu^+(X), \qquad E \in \mathcal{F},$$

where either $\mu^+(X) < \infty$ or $\mu^-(X) < \infty$. □

7.0.9 Definition. Let $\mu\colon \mathcal{F} \to \dot{\mathbb{R}}$ be a signed measure. The associated total variation measure $|\mu|\colon \mathcal{F} \to [0, \infty]$ is defined by

$$|\mu|(A) := \sup \sum_{i=1}^{k} |\mu(A_i)|,$$

where the supremum is over all finite collections $A_1, \ldots, A_k \in \mathcal{F}$ of disjoint sets with $A = \bigcup_{i=1}^{k} A_i$.

7.0.10 Theorem. *Let $\mu\colon \mathcal{F} \to \dot{\mathbb{R}}$ be a signed measure with the Jordan decomposition $\mu = \mu^+ - \mu^-$. Then the following holds for all $A \in \mathcal{F}$.*

1. $\mu^+(A) = \sup\{\mu(E)\colon E \in \mathcal{F},\, E \subset A\}$.
2. $\mu^-(A) = -\inf\{\mu(E)\colon E \in \mathcal{F},\, E \subset A\}$.
3. $|\mu|(A) = \mu^+(A) + \mu^-(A)$ *– in particular $|\mu|$ really is a measure.*

7.0.11 Remark. Sometimes $\mu^+(A)$ is called the upper variation and $\mu^-(A)$ the lower variation of μ in A.

Proof of Theorem 7.0.10. (1) Take a Hahn decomposition (P, P^c) and remember $\mu^+(A) = \mu(A \cap P)$. Let $L := \sup\{\mu(E)\colon E \in \mathcal{F},\, E \subset A\}$. As $\mu^+(A) := \mu(A \cap P)$, where $A \cap P \subset A$ and $A \cap P \in \mathcal{F}$, it is clear that $\mu^+(A) \le L$. Let now $E \subset A$ where $E \in \mathcal{F}$. Then we have

$$\mu(E) = \mu(E \cap P) + \mu(E \cap P^c) \le \mu(E \cap P),$$

since $\mu(E \cap P^c) \le 0$. As $\mu((A \setminus E) \cap P) \ge 0$ we further have

$$\mu(E) \le \mu(E \cap P) \le \mu(E \cap P) + \mu((A \setminus E) \cap P) = \mu(A \cap P) = \mu^+(A).$$

Taking the supremum over $E \in \mathcal{F}$ with $E \subset A$ shows $L \le \mu^+(A)$, and so $L = \mu^+(A)$ as desired.
(2) Proved similarly as (1).
(3) Notice first that

$$\mu^+(A) + \mu^-(A) = \mu(A \cap P) - \mu(A \cap P^c) = |\mu(A \cap P)| + |\mu(A \cap P^c)| \le |\mu|(A)$$

as $A = (A \cap P) \cup (A \cap P^c)$ is a disjoint union of measurable sets. To prove the converse direction consider disjoint $A_1, \ldots, A_k \in \mathcal{F}$ with $A = \bigcup_{i=1}^{k} A_i$. We get

$$\begin{aligned}\sum_{i=1}^{k} |\mu(A_i)| &= \sum_{i=1}^{k} |\mu^+(A_i) - \mu^-(A_i)| \\ &\le \sum_{i=1}^{k} \mu^+(A_i) + \sum_{i=1}^{k} \mu^-(A_i) \\ &= \mu^+\Big(\bigcup_{i=1}^{k} A_i\Big) + \mu^-\Big(\bigcup_{i=1}^{k} A_i\Big) = \mu^+(A) + \mu^-(A).\end{aligned}$$

Taking the supremum over such decompositions gives $|\mu|(A) \le \mu^+(A) + \mu^-(A)$. □

7.1 Radon-Nikodym theorem

Let $(X, \mathcal{F}, \nu)$ be a measure space (so that ν is a usual non-negative measure) and $\mu\colon \mathcal{F} \to \dot{\mathbb{R}}$ be a signed measure. Similarly as in the case that μ is a usual measure, we say that μ is absolutely continuous with respect to ν – and denote this with $\mu \ll \nu$ – if $\nu(A) = 0$ implies $\mu(A) = 0$.

7.1.1 Example. Suppose $(X, \mathcal{F}, \nu)$ is a measure space and $f\colon X \to \dot{\mathbb{R}}$ is a measurable function so that at least one of the integrals $\int_X f^+ \,\mathrm{d}\nu$ or $\int_X f^- \,\mathrm{d}\nu$ is finite (a slightly more general assumption than the usual integrability assumption where both are finite). Then we can define a signed measure $\mu\colon \mathcal{F} \to \dot{\mathbb{R}}$ by setting

$$\mu(A) := \int_A f^+ \,\mathrm{d}\nu - \int_A f^- \,\mathrm{d}\nu = \int_A f \,\mathrm{d}\nu.$$

This is a signed measure with $\mu \ll \nu$.

There are no other kind of absolutely continuous measure as we prove next – this is the famous Radon–Nikodym theorem. We say that μ is σ-finite if we can write

$$X = \bigcup_i A_i$$

where $A_i \in \mathcal{F}$ and $|\mu(A_i)| < \infty$ for every i.

7.1.2 Theorem (Radon–Nikodym)**.** *Let $(X, \mathcal{F}, \nu)$ be a σ-finite measure space and $\mu\colon \mathcal{F} \to \dot{\mathbb{R}}$ be a σ-finite signed measure with $\mu \ll \nu$. Then there exists a measurable $f\colon X \to \mathbb{R}$ so that*

$$\mu(A) = \int_A f \,\mathrm{d}\nu, \qquad A \in \mathcal{F}.$$

The function f is unique in the following sequences: if g is another function with the above property, then $f = g$ ν-a.e.

7.1.3 Remark. The unique function f is called the Radon–Nikodym derivative of μ with respect to ν. It is often denoted $f = \frac{\mathrm{d}\mu}{\mathrm{d}\nu}$.

Before we prove this we need one more lemma.

7.1.4 Lemma. *Let $\nu, \lambda\colon \mathcal{F} \to [0, \infty)$ be finite measures with $\lambda(X) > 0$ and $\lambda \ll \nu$. Then there exists $\epsilon > 0$ and $P \in \mathcal{F}$ so that $\nu(P) > 0$ and that P is a positive set of $\lambda - \epsilon\nu$.*

Proof. Let (P_n, P_n^c) be a Hahn decomposition of the signed measure $\kappa_n := \lambda - \frac{1}{n}\nu$, and set $P := \bigcup_{n=1}^\infty P_n$. As $P^c = \bigcap_{n=1}^\infty P_n^c \subset P_n^c$ we have $\kappa_n(P^c) \le 0$, that is,

$$\lambda(P^c) \le \frac{1}{n}\nu(P^c).$$

This implies that $\lambda(P^c) = 0$ and so $\lambda(P) = \lambda(X) > 0$. As $\lambda \ll \nu$, we must have $\nu(P) > 0$ as well. But as $P = \bigcup_{n=1}^\infty P_n$ we must have $\nu(P_n) > 0$ for some n. By construction P_n is a positive set of $\lambda - \frac{1}{n}\nu$. □

Proof of Theorem 7.1.2. **Step I.** Assume that $\mu, \nu\colon \mathcal{F} \to [0, \infty)$ are finite measures. Set

$$\mathcal{G} := \left\{ f\colon X \to [0, \infty]\colon\ f \text{ measurable and } \int_A f \,\mathrm{d}\nu \le \mu(A) \text{ for all } A \in \mathcal{F} \right\}$$

and

$$L := \sup_{f \in \mathcal{G}} \int_X f \, \mathrm{d}\nu,$$

where $L \le \mu(X) < \infty$. We want to first find f so that $\int_X f \, \mathrm{d}\nu = L$. To this end, choose $f_n \in \mathcal{G}$ so that $\int_X f_n \, \mathrm{d}\nu \to L$. Define $f := \sup_n f_n$. Notice that

$$\int_X f \, \mathrm{d}\nu \ge \int_X f_n \, \mathrm{d}\nu \to L,$$

so that $\int_X f \, \mathrm{d}\nu \ge L$. We will then show that $f \in \mathcal{G}$ which gives $\int_X f \, \mathrm{d}\nu \le L$. As a tool towards this we set $g_n := \max_{1 \le i \le n} f_n$. We first show that $g_n \in \mathcal{G}$. With A, n fixed set

$$\begin{aligned} A_1 &= \{x \in A \colon g_n(x) = f_1(x)\} \\ A_2 &= \{x \in A \setminus A_1 \colon g_n(x) = f_2(x)\} \\ &\vdots \\ A_n &= \Big\{x \in A \setminus \bigcup_{i=1}^{n-1} A_i \colon g_n(x) = f_n(x)\Big\}. \end{aligned}$$

Notice that

$$\int_A g_n \, \mathrm{d}\nu = \sum_{i=1}^n \int_{A_i} f_i \, \mathrm{d}\nu \le \sum_{i=1}^n \mu(A_i) = \mu(A),$$

where we used that $A = \bigcup_{i=1}^n A_i$ is a disjoint union and $f_i \in \mathcal{G}$ for all i. This shows that $g_n \in \mathcal{G}$. Finally, as $0 \le g_1 \le g_2 \le \cdots$ and $g_n \to f$ pointwise, we have by monotone convergence theorem and the fact that $g_n \in \mathcal{G}$ for all n that

$$\int_A f \, \mathrm{d}\nu = \lim_{n \to \infty} \int_A g_n \, \mathrm{d}\nu \le \mu(A).$$

We have showed that $f \in \mathcal{G}$ and

$$\int_X f \, \mathrm{d}\nu = L < \infty.$$

In particular, this means $f(x) < \infty$ ν-a.e. – we can modify f in a set of measure zero to get that $f(x) < \infty$ everywhere.

To show that f is the desired Radon–Nikodym derivative, we define the finite measure $\lambda \colon \mathcal{F} \to [0, \infty)$ with

$$\lambda(A) := \mu(A) - \int_A f \, \mathrm{d}\nu.$$

We want to show that $\lambda(A) = 0$ for all $A \in \mathcal{F}$. Aiming for a contradiction, suppose that $\lambda(X) > 0$. As $\lambda \ll \nu$ (since $\mu \ll \nu$) we find, using Lemma 7.1.4, $\epsilon > 0$ and $P \in \mathcal{F}$ so that $\nu(P) > 0$ and that P is a positive set of $\lambda - \epsilon\nu$. We define

$$g := f + \epsilon 1_P.$$

This function satisfies

$$\int_X g \, \mathrm{d}\nu = \int_X f \, \mathrm{d}\nu + \epsilon\nu(P) = L + \epsilon\nu(P) > L. \tag{7.1.5}$$

This is our desired contradiction if we show that $g \in \mathcal{G}$. For this, just notice that for all $A \in \mathcal{F}$ there holds $\lambda(A \cap P) - \epsilon\nu(A \cap P) \ge 0$ and so

$$\int_A g \, \mathrm{d}\nu = \int_A f \, \mathrm{d}\nu + \epsilon\nu(A \cap P)$$

$$\begin{aligned}&\leq \int_A f \, \mathrm{d}\nu + \lambda(A \cap P) \\ &= \int_A f \, \mathrm{d}\nu + \mu(A \cap P) - \int_{A\cap P} f \, \mathrm{d}\nu \\ &= \int_{A\setminus P} f \, \mathrm{d}\nu + \mu(A \cap P) \leq \mu(A \setminus P) + \mu(A \cap P) = \mu(A).\end{aligned}$$

Thus, $g \in \mathcal{G}$ – this implies $\int_X g \, \mathrm{d}\nu \leq L$ contradicting (7.1.5). We have proved the existence of f in the Case I.

Step II. We then assume $|\mu(X)| < \infty$ and $\nu(X) < \infty$. We write the Jordan decomposition $\mu = \mu^+ - \mu^-$. By Theorem 7.0.10 we know that

$$\mu^+(A) = \sup\{\mu(E) \colon E \in \mathcal{F},\ E \subset A\}.$$

From this it follows that $\mu^+ \ll \nu$, since $\mu \ll \nu$, and similarly we see $\mu^- \ll \nu$. Applying Step I we find $f^+, f^- \colon X \to [0, \infty)$ so that

$$\mu^+(A) = \int_A f^+ \, \mathrm{d}\nu \qquad \text{and} \qquad \mu^-(A) = \int_A f^- \, \mathrm{d}\nu,$$

and it follows that $f := f^+ - f^-$ satisfies

$$\int_A f \, \mathrm{d}\nu = \mu^+(A) - \mu^-(A) = \mu(A).$$

Step III. Assume now only that μ and ν are σ-finite. Using this it is easy to get **disjoint** sets X_k so that

$$X = \bigcup_k X_k,$$

$|\mu(X_k)| < \infty$ and $\nu(X_k) < \infty$ for all k. Step II gives us functions $f_k \colon X_k \to \mathbb{R}$ so that

$$\mu(A) = \int_A f_k \, \mathrm{d}\nu, \qquad A \in \mathcal{F},\ A \subset X_k.$$

We define $f(x) = f_k(x)$ if $x \in X$ satisfies $x \in X_k$ (remember the sets are disjoint). We need to show that f is the desired Radon–Nikodym derivative. Let $A \in \mathcal{F}$ be arbitrary. Set $X^+ := \{x \in X \colon f(x) \geq 0\}$. It follows that

$$\begin{aligned}\mu(A \cap X^+) &= \sum_k \mu(A \cap X^+ \cap X_k) \\ &= \sum_k \int_{A\cap X^+\cap X_k} f_k \, \mathrm{d}\nu \\ &= \sum_k \int_{A\cap X_k} f^+ \, \mathrm{d}\nu = \int_A f^+ \, \mathrm{d}\nu,\end{aligned}$$

and similarly $\mu(A \cap (X^+)^c) = -\int_A f^- \, \mathrm{d}\nu$. It follows that $\int_A f \, \mathrm{d}\nu$ makes sense and

$$\mu(A) = \mu(A \cap X^+) + \mu(A \cap (X^+)^c) = \int_A f^+ \, \mathrm{d}\nu - \int_A f^- \, \mathrm{d}\nu = \int_A f \, \mathrm{d}\nu.$$

Step IV. It remains to show the uniqueness of f. This follows from the next theorem. □

7.1.6 Theorem. *Suppose ν is a σ-finite measure and $f, g\colon X \to \dot{\mathbb{R}}$ are two measurable functions so that $\int_X f \,\mathrm{d}\nu$ and $\int_X g \,\mathrm{d}\nu$ exist. If*

$$\int_A f \,\mathrm{d}\nu = \int_A g \,\mathrm{d}\nu$$

for all $A \in \mathcal{F}$, then $f = g$ almost everywhere.

Proof. With a simple argument we see that it is enough to prove the claim when ν is finite. As we have done before, we can assume that $\int_E f \,\mathrm{d}\nu > -\infty$ for all $E \in \mathcal{F}$. Define $A := \{f < g\}$ – it is enough to show $\nu(A) = 0$. Decompose

$$A = \bigcup_{n=1}^{\infty} A_n \cup A_\infty = \bigcup_{n=1}^{\infty} E_n,$$

where

$$A_n := A \cap \{g < n\},\ A_\infty = A \cap \{g = \infty\},\ E_n = A \cap \{f < n\}.$$

We now have

$$-\infty < \int_{A_n} f \,\mathrm{d}\nu = \int_{A_n} g \,\mathrm{d}\nu \leq n\nu(X) < \infty,$$

so that $g - f > 0$ is integrable on A_n with $\int_{A_n} (g - f) \,\mathrm{d}\nu = 0$. It follows that we must have $\nu(A_n) = 0$ for all n. To show that $\nu(A) = 0$ it remains to show $\nu(A_\infty) = 0$. For this, it is enough to show $\nu(A_\infty \cap E_n) = 0$ for all n. But we see that

$$\int_{A_\infty \cap E_n} g \,\mathrm{d}\nu = \int_{A_\infty \cap E_n} f \,\mathrm{d}\nu \leq n\nu(X) < \infty.$$

This is only possible if $\nu(A_\infty \cap E_n) = 0$ so we are done. □

We will later use the Radon–Nikodym theorem to prove the fundamental theorem of calculus for the Lebesgue integral.

Chapter 8

Product measures and Fubini's theorem

8.1 Extending measures and uniqueness properties

Before moving on to product measures, we prove a useful theorem which allows us to construct a measure by first defining it on a small algebra of sets, and then use this theorem to guarantee its extension to a σ-algebra.

8.1.1 Definition. A collection $\mathcal{A} \subset \mathcal{P}(X)$ is called an algebra on X if the following holds.

1. $\emptyset \in \mathcal{A}$.
2. $A_1 \cap A_2 \in \mathcal{A}$ if $A_1, A_2 \in \mathcal{A}$.
3. $X \setminus A \in \mathcal{A}$ if $A \in \mathcal{A}$.

Notice that if $A_1, A_2 \in \mathcal{A}$, then $A_1 \cup A_2 = X \setminus ((X \setminus A_1) \cap (X \setminus A_2)) \in \mathcal{A}$. We are not assuming anything about countable unions or intersections as in the case of a σ-algebra. However, we can call μ a measure on an algebra $\mathcal{A}$ if $\mu \colon \mathcal{A} \to [0, \infty]$ is a function on $\mathcal{A}$ with the property that $\mu(\emptyset) = 0$ and $\mu(A) = \sum_{i=1}^{\infty} \mu(A_i)$ whenever $A_1, A_2, \ldots \in \mathcal{A}$ are disjoint and $A := \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$. Notice that here we separately assume $A \in \mathcal{A}$, while this follows from the assumptions in the case of a σ-algebra. We could also call μ a pre-measure – but we use the above terminology "a measure on the algebra $\mathcal{A}$" instead.

Let μ be a measure on an algebra $\mathcal{A}$. We use Lemma 2.2.9 with $\mathcal{S} := \mathcal{A}$ and $h(A) := \mu(A)$ to get an outer measure $\nu \colon \mathcal{P}(X) \to [0, \infty]$ defined by

$$\nu(E) := \inf \Big\{ \sum_{i=1}^{\infty} \mu(A_i) \colon\ E \subset \bigcup_{i=1}^{\infty} A_i,\ A_i \in \mathcal{A} \Big\}, \qquad E \subset X.$$

In this more concrete situation we can say more about ν compared to the situation of Lemma 2.2.9.

8.1.2 Lemma. *Let μ be a measure on an algebra $\mathcal{A}$ and ν be the outer measure constructed from μ as above. Then the following holds.*

1. *$\nu(A) = \mu(A)$ if $A \in \mathcal{A}$.*
2. *$\mathcal{A} \subset \mathcal{M}_\nu(X)$ – i.e. every $A \in \mathcal{A}$ is ν-measurable. Thus, we also have $\sigma(\mathcal{A}) \subset \mathcal{M}_\nu(X)$.*

Proof. Exercise. □

Remember that by Theorem 2.2.7 we have that $\nu|\mathcal{M}_\nu(X)$ is a measure, and so by the above lemma $\lambda := \nu|\sigma(\mathcal{A})$ is a measure on the σ-algebra $\sigma(\mathcal{A})$ generated by the algebra $\mathcal{A}$ and $\lambda(A) = \mu(A)$ whenever $A \in \mathcal{A}$. We now have the following.

8.1.3 Theorem (Carathéodory–Hahn extension theorem)**.** *Let μ be a measure on an algebra $\mathcal{A}$. Then there exists a measure $\lambda\colon \sigma(\mathcal{A}) \to [0,\infty]$ with $\lambda|\mathcal{A} = \mu$. In addition, if μ is σ-finite ($X = \bigcup_i A_i$, $A_i \in \mathcal{A}$, $\mu(A_i) < \infty$), then the extension to $\sigma(\mathcal{A})$ is unique.*

Proof. We showed the existence above. We develop the necessary theory to prove the uniqueness next. □

To prove the uniqueness in the Carathéodory–Hahn extension theorem, we next study the general question that under what conditions on $\mathcal{P}$ does it follow that if $\nu, \lambda\colon \sigma(\mathcal{P}) \to [0,\infty]$ are two measures with $\nu(P) = \lambda(P)$ for all $P \in \mathcal{P}$, then $\nu = \mu$ on the whole σ-algebra $\sigma(\mathcal{P})$.

8.1.4 Definition. We say $\mathcal{P} \subset \mathcal{P}(X)$ is a π-system if $\mathcal{P} \neq \emptyset$ and $P_1 \cap P_2 \in \mathcal{P}$ for $P_1, P_2 \in \mathcal{P}$.

8.1.5 Definition. We say that $D \subset \mathcal{P}(X)$ is a Dynkin system if the following holds.

1. $X \in D$.
2. $B \setminus A \in D$ if $A, B \in D$ and $A \subset B$.
3. $\bigcup_{i=1}^\infty A_i \in D$ if $A_i \in D$ and $A_1 \subset A_2 \subset \cdots$.

Dynkin used the term λ-system for the collections that we now call Dynkin systems. This explains the name of the following theorem.

8.1.6 Theorem (Dynkin's π-λ theorem)**.** *Let $\mathcal{P}$ be a π-system and D be a Dynkin system with $\mathcal{P} \subset D$. Then we have $\sigma(\mathcal{P}) \subset D$.*

Proof. Let

$$D(\mathcal{P}) := \bigcap \Big\{ D\colon\ D \text{ Dynkin system with } D \supset \mathcal{P} \Big\}$$

be the Dynkin system generated by $\mathcal{P}$ (it is easy to check that this is, indeed, a Dynkin system). If we show that $D(\mathcal{P})$ is a σ-algebra, then the claim follows: if D is an arbitrary Dynkin system with $\mathcal{P} \subset D$, then $\sigma(\mathcal{P}) \subset D(\mathcal{P}) \subset D$. So we need to show that a Dynkin system generated by a π-system is in fact a σ-algebra.

As $D(\mathcal{P})$ is a Dynkin system it follows that $X \in D(\mathcal{P})$ and that $A \in D(\mathcal{P})$ implies $X \setminus A \in D(\mathcal{P})$. The interesting part is to show that if $A_1, A_2, \ldots \in D(\mathcal{P})$, then $A := \bigcup_{i=1}^\infty A_i \in D(\mathcal{P})$. We will soon show that $D(\mathcal{P})$ is closed under finite unions – but then it also follows that

$$A = \bigcup_{i=1}^\infty A_i' \in D(\mathcal{P}), \qquad A_i' := A_1 \cup \cdots \cup A_i \in D(\mathcal{P}),\ A_1' \subset A_2' \subset \cdots,$$

since $D(\mathcal{P})$ is a Dynkin system. So it remains to show that $D(\mathcal{P})$ is closed under finite unions. We will show that it is closed under finite intersections (the claim about unions then follows by taking complements).

We first show the weaker claim that $P \in \mathcal{P}$ and $A \in D(\mathcal{P})$ implies $A \cap P \in D(\mathcal{P})$. Define

$$D_1 := \{A \in D(\mathcal{P})\colon\ A \cap P \in D(\mathcal{P}) \text{ for all } P \in \mathcal{P}\}.$$

Notice that $\mathcal{P} \subset D_1$ as $\mathcal{P}$ is a π-system. We show that D_1 is a Dynkin system, which then implies the desired conclusion $D_1 = \mathcal{D}(\mathcal{P})$. Clearly $X \in D_1$. Suppose $A_1, A_2 \in D_1$ and $A_1 \subset A_2$. Then for all $P \in \mathcal{P}$ we have

$$(A_2 \setminus A_1) \cap P = (A_2 \cap P) \setminus (A_1 \cap P) \in D(\mathcal{P}),$$

and so $A_2 \setminus A_1 \in D_1$. Suppose then that $A_i \in D_1$ and $A_1 \subset A_2 \subset \cdots$ – then for all $P \in \mathcal{P}$ we have

$$\bigcup_{i=1}^{\infty} A_i \cap P = \bigcup_{i=1}^{\infty} (A_i \cap P) \in D(\mathcal{P}),$$

and so $\bigcup_{i=1}^{\infty} A_i \in D_1$. Thus D_1 is a Dynkin system.

We now show that $A_1, A_2 \in D(\mathcal{P})$ implies $A_1 \cap A_2 \in D(\mathcal{P})$. Define

$$D_2 := \{A \in D(\mathcal{P})\colon A \cap B \in D(\mathcal{P}) \text{ for all } B \in D(\mathcal{P})\}.$$

By what we just proved $\mathcal{P} \subset D_2$. By the same proof as for D_1 we see that D_2 is a Dynkin system. It follows that $D_2 = D(\mathcal{P})$ and so we are done. □

8.1.7 Theorem. *Let $\mathcal{P}$ be a π-system and $\nu, \lambda\colon \sigma(\mathcal{P}) \to [0, \infty)$ be finite measures with $\nu(X) = \lambda(X)$. If $\nu(P) = \lambda(P)$ for all $P \in \mathcal{P}$, then $\nu = \lambda$.*

Proof. Define

$$D := \{A \in \sigma(\mathcal{P})\colon \nu(A) = \lambda(A)\}.$$

We show that D is a Dynkin system – then it follows from Dynkin's theorem that $D = \sigma(P)$ as desired. By assumption $X \in D$. Let then $A, B \subset D$ with $A \subset B$. We have (as $\nu(A), \lambda(A) < \infty$) that

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

and so $B \setminus A \in D$. Finally, let $A_i \in D$ and $A_1 \subset A_2 \subset \cdots$. If $A := \bigcup_{i=1}^{\infty} A_i$ we have by convergence of measures that

$$\nu(A) = \lim_{i \to \infty} \nu(A_i) = \lim_{i \to \infty} \lambda(A_i) = \lambda(A),$$

and so $A \in D$ – we are done. □

8.1.8 Remark. It is now easy to prove a result like Lemma 6.0.2 using different tools. Let $\mathcal{D}$ be the dyadic cubes on $\mathbb{R}^d$. Suppose $\nu, \lambda\colon \operatorname{Bor}(\mathbb{R}^d) \to [0, \infty]$ are locally finite measures with $\nu(Q) = \lambda(Q)$ for all $Q \in \mathcal{D}$. Fix $Q_0 \in \mathcal{D}$ and define the finite measures $\nu_0(A) := \nu(A \cap Q_0)$ and $\lambda_0(A) := \lambda(A \cap Q_0)$, $A \in \operatorname{Bor}(\mathbb{R}^d)$. Now $\nu_0(\mathbb{R}^d) = \nu(Q_0) = \lambda(Q_0) = \lambda_0(\mathbb{R}^d)$ and for every $Q \in \mathcal{D}$ we also have $\nu_0(Q) = \lambda_0(Q)$. As $\mathcal{D} \cup \{\emptyset\}$ is a π-system and $\sigma(\mathcal{D} \cup \{\emptyset\}) = \sigma(\mathcal{D}) = \operatorname{Bor}(\mathbb{R}^d)$ (remember we showed that open sets are even disjoint unions of dyadic cubes) the above theorem says that $\nu_0 = \lambda_0$. We have shown that $\nu(A \cap Q) = \lambda(A \cap Q)$ for all $Q \in \mathcal{D}$. To prove $\nu(A) = \lambda(A)$ we can, by writing A as a suitable disjoint finite union (intersect with the quadrants like $[0, \infty)^d$), assume that there is a sequence of dyadic cubes $Q_1 \subset Q_2 \subset \cdots$ so that $A = \bigcup_i (A \cap Q_i)$. It then follows from convergence of measures that $\nu(A) = \lambda(A)$.

We now return to prove the uniqueness in the Carathéodory–Hahn extension theorem.

Proof of uniqueness in Theorem 8.1.3. So suppose $\nu, \lambda\colon \sigma(\mathcal{A}) \to [0, \infty]$ are two measures with $\lambda|\mathcal{A} = \mu = \nu|\mathcal{A}$, and that μ is σ-finite – $X = \bigcup_i A_i$, $A_i \in \mathcal{A}$, $\mu(A_i) < \infty$ and $A_1 \subset A_2 \subset \cdots$. Fix i and define the finite measures $\nu_i(E) := \nu(E \cap A_i)$ and $\lambda_i(E) = \lambda(E \cap A_i)$, $E \in \sigma(\mathcal{A})$. We have for all $A \in \mathcal{A}$ (including $A = X$) that $\nu_i(A) = \nu(A \cap A_i) = \mu(A \cap A_i) = \lambda(A \cap A_i) = \lambda_i(A_i)$, since $A \cap A_i \in \mathcal{A}$.

It follows from Theorem 8.1.7 that $\nu_i(E) = \lambda_i(E)$ for all $E \in \sigma(\mathcal{A})$. By convergence of measures we have for all $E \in \sigma(\mathcal{A})$ that

$$\nu(E) = \lim_{i\to\infty} \nu(E \cap A_i) = \lim_{i\to\infty} \lambda(E \cap A_i) = \lambda(E),$$

which is the desired uniqueness. □

There is an additional small detail that we need to address. In the product measure setting we will not quite have an algebra – rather we will have a semi-algebra.

8.1.9 Definition. We say that $\mathcal{A} \subset \mathcal{P}(X)$ is a semi-algebra if the following holds.

1. $\emptyset \in \mathcal{A}$.
2. $A_1 \cap A_2 \in \mathcal{A}$ if $A_1, A_2 \in \mathcal{A}$.
3. If $A \in \mathcal{A}$, then $X \setminus A = \bigcup_{i=1}^n A_i$ for some disjoint $A_1, \ldots, A_n \in \mathcal{A}$.

Here we have weakened the third assumption compared to an algebra. If we have a measure on a semi-algebra (defined similarly as in the algebra case) can we uniquely extend it to a measure on some algebra containing the semi-algebra? If so, Carathéodory–Hahn extension theorem allows us to further extend to a σ-algebra. Luckily the extension can be done and it is straightforward.

8.1.10 Lemma. *Let $\mathcal{A}_0$ be a semi-algebra. Then the collection*

$$\mathcal{A} := \Big\{ \bigcup_{i=1}^n A_i \colon\ A_1, \ldots, A_n \in \mathcal{A}_0 \text{ disjoint} \Big\}$$

is an algebra. Let μ be a measure on the semi-algebra $\mathcal{A}_0$. Then

$$\lambda(A) := \sum_{i=1}^n \mu(A_i), \qquad A = \bigcup_{i=1}^n A_i \in \mathcal{A},$$

where $A_1, \ldots, A_n \in \mathcal{A}_0$ are disjoint, is a well-defined measure on the algebra $\mathcal{A}$, and it is the unique extension ($\mu(A) = \lambda(A)$ for $A \in \mathcal{A}_0$) of μ to the algebra $\mathcal{A}$.

Proof. Exercise. □

8.2 Product measures

Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be σ-finite measure spaces. Define

$$\mathcal{S} := \{A \times B \colon A \in \mathcal{M},\ B \in \mathcal{N}\} \subset \mathcal{P}(X \times Y)$$

be the collection of "measurable rectangles". This is clearly a semi-algebra, and it is then awfully naturally to define $\lambda \colon \mathcal{S} \to [0, \infty]$ by

$$\lambda(A \times B) := \mu(A)\nu(B),$$

and hope that λ is a measure on the semi-algebra $\mathcal{S}$, and then extend using our previous results.

8.2.1 Lemma. *Suppose $A_i \times B_i \in \mathcal{S}$ are disjoint and assume that $\bigcup_{i=1}^\infty (A_i \times B_i) = A \times B \in \mathcal{S}$. Then we have*

$$\lambda(A) = \sum_{i=1}^\infty \lambda(A_i \times B_i).$$

Proof. We need to prove

$$\mu(A)\nu(B) = \sum_{i=1}^{\infty} \mu(A_i)\nu(B_i).$$

Notice that by disjointness

$$1_A(x)1_B(y) = 1_{A\times B}(x,y) = \sum_{i=1}^{\infty} 1_{A_i\times B_i}(x,y) = \sum_{i=1}^{\infty} 1_{A_i}(x)1_{B_i}(y).$$

Taking the $\int_X \, \mathrm{d}\mu(x)$ integral of this and using that a series of non-negative functions can be integrated term by term we get

$$\mu(A)1_B(y) = \sum_{i=1}^{\infty} \mu(A_i)1_{B_i}(y).$$

Taking the $\int_Y \, \mathrm{d}\nu(y)$ integral of this gives the claim. □

8.2.2 Theorem (Existence of the unique product measure)**.** *Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be σ-finite measure spaces. Then there exists a unique measure $\mu \times \nu \colon \sigma(\mathcal{S}) \to [0, \infty]$ with the property that*

$$(\mu \times \nu)(A \times B) = \mu(A)\nu(B), \qquad A \times B \in \mathcal{S}.$$

Proof. First, consider the measure $\lambda(A \times B) = \mu(A)\nu(B)$ in the semi-algebra $\mathcal{S}$, and use Lemma 8.1.10 to extend it uniquely to the algebra $\tilde{S}$ generated by the measurable rectangles $\mathcal{S}$. Notice that this extended measure $\tilde{\mathcal{S}} \to [0, \infty]$ is σ-finite as the measure spaces $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ-finite. Then, use the Carathéodory–Hahn extension theorem, Theorem 8.1.3, to extend this uniquely to $\sigma(\tilde{\mathcal{S}}) = \sigma(\mathcal{S})$. □

8.2.3 Remark. Even if $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are both complete measure spaces, the product measure space $(X \times Y, \sigma(\mathcal{S}), \mu \times \nu)$ is not necessarily complete. Suppose e.g. $A \in \mathcal{M}$, $A \neq \emptyset$, with $\mu(A) = 0$ and $B \notin \mathcal{N}$. Then $A \times B \notin \sigma(S)$, see Lemma 8.2.4 below, but $A \times B \subset A \times Y$, where $(\mu \times \nu)(A \times Y) = \mu(A)\nu(Y) = 0$.

For this reason, e.g. the Lebesgue measure m_2 on $\mathbb{R}^2$ is really the completion of the the product measure $m_1 \times m_1$.

For $E \subset X \times Y$ we denote

$$E_x := \{y \in Y \colon (x, y) \in E\} \subset Y, \qquad x \in X,$$
$$E^y := \{x \in X \colon (x, y) \in E\} \subset X, \qquad y \in Y.$$

8.2.4 Lemma. *If $E \in \sigma(\mathcal{S})$, then $E_x \in \mathcal{N}$ and $E^y \in \mathcal{M}$ for all $x \in X$ and $y \in Y$.*

Proof. Let

$$\mathcal{F} := \{E \in \sigma(S) \colon E_x \in \mathcal{N} \text{ for all } x \in X\}.$$

We will prove that $\mathcal{F}$ is a σ-algebra on $X \times Y$ with $\mathcal{S} \subset \mathcal{F}$ – it then follows that $\sigma(S) = \mathcal{F}$ proving the claim for the sets E_x. The other claim is symmetric.

It is clear that $\mathcal{S} \subset \mathcal{F}$ as $(A \times B)_x = B$ if $x \in A$ and $(A \times B) = \emptyset$ if $x \notin A$. It is obvious that $\mathcal{F}$ is a σ-algebra after one notices that $(E_x)^c = (E^c)_x$ and $\left(\bigcup_i E_i\right)_x = \bigcup_i (E_i)_x$. □

8.3 Fubini's theorem

Fubini's theorem answers when it is possible to compute

$$\int_{X\times Y} f \,\mathrm{d}(\mu\times\nu)$$

as an iterated integral $\int_X \int_Y f(x,y)\,\mathrm{d}\nu(y)\,\mathrm{d}\mu(x)$ or $\int_Y \int_X f(x,y)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y)$.

For $f\colon X\times Y\to\dot{\mathbb{R}}$ define $f_x\colon Y\to\dot{\mathbb{R}}$, $f_x(y)=f(x,y)$, and $f^y\colon X\to\dot{\mathbb{R}}$, $f^y(x)=f(x,y)$. Lemma 8.2.4 directly implies the following.

8.3.1 Lemma. *If $f\colon X\times Y\to\dot{\mathbb{R}}$ is $\sigma(\mathcal{S})$-measurable, then f_x is $\mathcal{N}$-measurable and f^y is $\mathcal{M}$-measurable for all $x\in X$ and $y\in Y$.*

The following key lemma, which is the crux of Fubini's theorem, is a bit complicated to prove.

8.3.2 Lemma. *If $E\in\sigma(\mathcal{S})$, then $x\mapsto\nu(E_x)$ is $\mathcal{M}$-measurable and $y\mapsto\mu(E^y)$ is $\mathcal{N}$-measurable. Moreover, we have*

$$\int_X \nu(E_x)\,\mathrm{d}\mu(x)=\int_Y \mu(E^y)\,\mathrm{d}\nu(y)=(\mu\times\nu)(E).$$

Proof. Let

$$\mathcal{F}:=\Big\{E\in\sigma(S)\colon x\mapsto\nu(E_x)\text{ is }\mathcal{M}\text{-measurable and }\int_X\nu(E_x)\,\mathrm{d}\mu(x)=(\mu\times\nu)(E)\Big\}.$$

Notice that $\nu((A\times B)_x)=1_A(x)\nu(B)$, so it is clear that $\mathcal{S}\subset\mathcal{F}$. We prove that $\mathcal{F}$ is a Dynkin's system – Dynkin's π-λ theorem then implies that $\sigma(S)=\mathcal{F}$, which, by symmetry, is enough. As $\mathcal{S}\subset\mathcal{F}$, we already have $X\times Y\in\mathcal{F}$. Let now $E_1,E_2,\ldots\in\mathcal{F}$ with $E_1\subset E_2\subset\cdots$. We want to show that $E:=\bigcup_i E_i\in\mathcal{F}$. We have that $E_x=\bigcup_i(E_i)_x$, where $(E_1)_x\subset(E_2)_x\subset\cdots$, and so $\nu(E_x)=\lim_{i\to\infty}\nu((E_i)_x)$ is measurable as a limit of measurable functions. By convergence of measures and the monotone convergence theorem we get

$$(\mu\times\nu)(E)=\lim_{i\to\infty}(\mu\times\nu)(E_i)=\lim_{i\to\infty}\int_X\nu((E_i)_x)\,\mathrm{d}\mu(x)=\int_X\nu(E_x)\,\mathrm{d}\mu(x).$$

Thus, we have $E\in\mathcal{F}$ as desired. There is one more property to prove to conclude that $\mathcal{F}$ is a Dynkin's system, but this is trickier.

Using σ-finiteness write $X=\bigcup_i X_i$, where $X_i\in\mathcal{M}$, $X_1\subset X_2\subset\cdots$, and $\mu(X_i)<\infty$. Similarly, write $Y=\bigcup_i Y_i$. Define $S_i:=X_i\times Y_i$. We take a detour and prove that

$$\mathcal{C}:=\{E\in\sigma(S)\colon E\cap S_i\in\mathcal{F}\}$$

satisfies $\mathcal{C}=\sigma(S)$. The proof of this goes in the usual way – we show that $\mathcal{C}$ is a Dynkin's system with $\mathcal{S}\subset\mathcal{C}$, which implies the claim by Dynkin's theorem. Notice that if $A\times B\in\mathcal{S}$, then $(A\times B)\cap S_i\in\mathcal{S}\subset\mathcal{F}$, and so $A\times B\in\mathcal{C}$ showing that $\mathcal{S}\subset\mathcal{C}$. In particular $X\times Y\in\mathcal{S}$. Let now $E_1,E_2,\ldots\in\mathcal{C}$ with $E_1\subset E_2\subset\cdots$. Notice that $E:=\bigcup_j E_j$ satisfies $E\cap S_i=\bigcup_j(E_j\cap S_i)$, where $E_j\cap S_i\in\mathcal{F}$ and $E_1\cap S_i\subset E_2\cap S_i\subset\cdots$. Thus, by the proved union property of $\mathcal{F}$, we have that $E\cap S_i\in\mathcal{F}$ and so $E\in\mathcal{C}$. To complete the proof that $\mathcal{C}$ is a Dynkin's system, let now $E,F\in\mathcal{C}$ with $E\subset F$. We want to show that $(F\setminus E)\cap S_i\in\mathcal{F}$. Notice that $(F\setminus E)\cap S_i=F_i\setminus E_i$, where $E_i:=E\cap S_i\in\mathcal{F}$ and $F_i:=F\cap S_i\in\mathcal{F}$. We write

$$\nu((F_i\setminus E_i)_x)=\nu((F_i)_x\setminus(E_i)_x)=\nu((F_i)_x)-\nu((E_i)_x),$$

which is possible since $\nu((E_i)_x) \le \nu(Y_i) < \infty$. We then get, since $F_i, E_i \in \mathcal{F}$, that

$$x \mapsto \nu((F_i \setminus E_i)_x) = \nu((F_i)_x) - \nu((E_i)_x)$$

is $\mathcal{M}$-measurable and

$$\begin{aligned}\int_X \nu((F_i \setminus E_i)_x)\,\mathrm{d}\mu(x) &= \int_X \nu((F_i)_x)\,\mathrm{d}\mu(x) - \int_X \nu((E_i)_x)\,\mathrm{d}\mu(x)\\ &= (\mu\times\nu)(F_i) - (\mu\times\nu)(E_i) = (\mu\times\nu)(F_i \setminus E_i),\end{aligned}$$

where in the end we used that $(\mu\times\nu)(E_i) \le (\mu\times\nu)(S_i) < \infty$. This proves that $(F\setminus E)\cap S_i = F_i\setminus E_i \in \mathcal{F}$, and so $F \setminus E \in \mathcal{C}$ as desired. We have showed $\mathcal{C} = \sigma(S)$.

Finally, let $E, F \in \mathcal{F}$ with $E \subset F$ – we need to show that $F \setminus E \in \mathcal{F}$. Define $E_i := E \cap S_i$ and $F_i := F \cap S_i$. As $\mathcal{C} = \sigma(S)$ and $\mathcal{F} \subset \sigma(S)$, we know that $F_i, E_i \in \mathcal{F}$. The same proof as in the $\mathcal{C}$ case then shows that $F_i \setminus E_i \in \mathcal{F}$. Finally, the proved union property of $\mathcal{F}$ implies that $F \setminus E = \bigcup_i (F_i \setminus E_i) \in \mathcal{F}$, as $F_i \setminus E_i \in \mathcal{F}$ and $F_i \setminus E_i = (F \setminus E) \cap S_i \subset (F \setminus E) \cap S_{i+1} = F_{i+1} \setminus E_{i+1}$. □

We now prove a version of Fubini for non-negative functions, but we have already done all of the heavy lifting.

8.3.3 Theorem (Fubini for non-negative functions)**.** *Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be σ-finite measure spaces, and let $f\colon X \times Y \to [0,\infty]$ be $\sigma(\mathcal{S})$-measurable. Then the following holds.*

1. *By Lemma 8.3.1 we may form the non-negative function $x \mapsto \int_Y f(x,y)\,\mathrm{d}\nu(y)$, and this function is $\mathcal{M}$-measurable. The symmetric statement also holds.*

2. *By (1) the iterated integral*

$$\int_X \int_Y f(x,y)\,\mathrm{d}\nu(y)\,\mathrm{d}\mu(x)$$

makes sense, and we have

$$\int_X \int_Y f(x,y)\,\mathrm{d}\nu(y)\,\mathrm{d}\mu(x) = \int_{X\times Y} f\,\mathrm{d}(\mu\times\nu).$$

The symmetric statement also holds.

Proof. By Lemma 8.3.2 the claim holds if $f = 1_E$, $E \in \sigma(\mathcal{S})$. This is almost all of the work – the rest is a completely standard limiting argument. By linearity, the claim holds for simple functions $f = \sum_{i=1}^n c_i 1_{E_i}$, $E_i \in \sigma(\mathcal{S})$. In the general case, take a sequence of simple functions s_i so that $f(x) = \lim_{i\to\infty} s_i(x)$ and $s_1(x) \le s_2(x) \le \cdots$. Using monotone convergence theorem and the fact that limits of measurable functions are measurable, the claim follows rather directly. □

If we write $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ for some d_1, d_2 with $d = d_1 + d_2$, this version of Fubini cannot be exactly applied to the measure space $(\mathbb{R}^d, \mathrm{Leb}(\mathbb{R}^d), m_d)$, since, as we have explained, $\mathrm{Leb}(\mathbb{R}^d)$ is the completion of $\sigma(\mathrm{Leb}(\mathbb{R}^{d_1}) \times \mathrm{Leb}(\mathbb{R}^{d_2}))$. This causes a rather minor technical detail that needs to be taken into account – the sections f_x, f^y are measurable only almost everywhere. That is, Fubini holds in the following (almost identical) sense for the Lebesgue measure.

8.3.4 Theorem. *Let $f\colon \mathbb{R}^d \to [0,\infty]$ be Lebesgue measurable. Then the following holds.*

1. *For almost every $x_1 \in \mathbb{R}^{d_1}$ the function $x_2 \mapsto f(x_1, x_2)$ is Lebesgue measurable. The symmetric statement also holds.*

2. *By (1) we may form the non-negative function* $x_1 \mapsto \int_{\mathbb{R}^{d_2}} f(x_1, x_2)\,\mathrm{d}x_2$*, and this function is Lebesgue measurable. The symmetric statement also holds.*

3. *By (2) the iterated integral*
$$\int_{\mathbb{R}^{d_1}} \int_{\mathbb{R}^{d_2}} f(x_1, x_2)\,\mathrm{d}x_2\,\mathrm{d}x_1$$
makes sense, and we have
$$\int_{\mathbb{R}^d} f(x)\,\mathrm{d}x = \int_{\mathbb{R}^{d_1}} \int_{\mathbb{R}^{d_2}} f(x_1, x_2)\,\mathrm{d}x_2\,\mathrm{d}x_1.$$
The symmetric statement also holds.

Proof. Follows from the general version of Fubini and small technical work related to completions. We omit the details. □

The following generalization of Fubini from non-negative functions to integrable functions is straightforward – just carefully apply the existing version to f^+ and f^- and use the definitions and $f = f^+ - f^-$. It is also also key to notice the following. If $f\colon X \times Y \to \dot{\mathbb{R}}$ is $\sigma(\mathcal{S})$-measurable, we already know by the previous version of Fubini's theorem that

$$\int_{X\times Y} |f|\,\mathrm{d}(\mu\times\nu) = \int_X \int_Y |f(x,y)|\,\mathrm{d}\nu(y)\,\mathrm{d}\mu(x) = \int_Y \int_X |f(x,y)|\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y). \tag{8.3.5}$$

Thus, f is $(\mu\times\nu)$-integrable if and only if it is $\sigma(\mathcal{S})$-measurable and one of the above integrals is finite. That is, to check the assumption of integrability in the next theorem, one can calculate any one of the three quantities and show that it is finite. This is often very useful.

8.3.6 Theorem (Fubini for integrable functions)**.** *Let* $(X, \mathcal{M}, \mu)$ *and* $(Y, \mathcal{N}, \nu)$ *be* σ*-finite measure spaces, and let* $f\colon X \times Y \to \dot{\mathbb{R}}$ *be integrable with respect to* $\mu \times \nu$*. Then the following holds.*

1. *For* μ*-a.e.* $x \in X$ *the function* $y \mapsto f(x, y)$ *is* ν*-integrable. The symmetric statement also holds.*

2. *By (1) we may form the function* $x \mapsto \int_Y f(x, y)\,\mathrm{d}\nu(y)$*, and this function is* μ*-integrable. The symmetric statement also holds.*

3. *By (2) the iterated integral*
$$\int_X \int_Y f(x, y)\,\mathrm{d}\nu(y)\,\mathrm{d}\mu(x)$$
makes sense, and we have
$$\int_X \int_Y f(x, y)\,\mathrm{d}\nu(y)\,\mathrm{d}\mu(x) = \int_{X\times Y} f\,\mathrm{d}(\mu\times\nu).$$
The symmetric statement also holds.

We will see examples of Fubini througout our analysis – it is one of the most useful results.

Chapter 9

L^p spaces

We work in a fixed complete measure space $(X, \mathcal{F}, \mu)$. We first define the endpoint $p = \infty$ of the scale of spaces $L^p(\mu)$, $1 \le p \le \infty$. This simply means bounded functions – however, with the distinction that we do not care what happens outside of a set of measure zero. Therefore, the L^∞ norm of a function will be the smallest constant M so that pointwise almost everywhere $|f| \le M$.

9.0.1 Definition. Define the norm

$$\|f\|_{L^\infty(\mu)} = \inf\big\{M \ge 0 \colon \mu(\{|f| > M\}) = 0\big\}$$

and the corresponding space

$$L^\infty(\mu) = \Big\{f \colon X \to \dot{\mathbb{R}} \,\Big|\, f \text{ measurable, } \|f\|_{L^\infty(\mu)} < \infty\Big\}.$$

The following argument proves that, indeed, we have $|f| \le \|f\|_{L^\infty(\mu)}$ almost everywhere. Notice that

$$\{|f| > \|f\|_{L^\infty(\mu)}\} = \bigcup_j \{|f| > \|f\|_{L^\infty(\mu)} + 1/j\}$$

is a countable union of sets of measure zero – and hence of measure zero. Sometimes the notation

$$\|f\|_{L^\infty(\mu)} = \operatorname{ess\,sup} |f|$$

is used for this essential supremum (the notation on the left hand side is better as it specifies the underlying measure μ).

We now define the rest of the L^p spaces.

9.0.2 Definition. For $1 \le p < \infty$ we define

$$\|f\|_{L^p(\mu)} = \Big(\int_X |f|^p \, \mathrm{d}\mu\Big)^{\frac{1}{p}}.$$

The related $L^p(\mu)$ space is then naturally defined as

$$L^p(\mu) = \Big\{f \colon X \to \dot{\mathbb{R}} \,\Big|\, f \text{ measurable, } \|f\|_{L^p(\mu)} < \infty\Big\}.$$

9.0.3 Remark. Actually, there is no problem to make the same definition for all $0 < p < \infty$ (and this is used in the exercises). However, many/most of the interesting results below require $p \ge 1$ for various reasons.

9.0.4 Example. With $L^p(B(0,1))$ we mean here the space defined by

$$\|f\|_{L^p(B(0,1))} = \Big(\int_{B(0,1)} |f(x)|^p \,\mathrm{d}x\Big)^{\frac{1}{p}}, \qquad 1 \le p < \infty.$$

This type of notation $L^p(X)$ instead of $L^p(\mu)$ is also often used – especially whenever the underlying measure is clear from the context. Consider $f(x) = |x|^\alpha$ – we are interested when this belongs to $L^p(B(0,1))$. To estimate the L^p norm we decompose the ball $B(0,1)$ to dyadic annuli as follows

$$B(0,1) = \bigcup_{k=0}^{\infty} \{x\colon\ 2^{-k-1} \le |x| < 2^{-k}\}.$$

Before continuing further notice that $|\{x\colon\ 2^{-k-1} \le |x| < 2^{-k}\}| = |B(0,2^{-k})| - |B(0,2^{-k-1})| = C(2^{-kd} - 2^{-(k+1)d}) \sim 2^{-kd}$. We have

$$\begin{aligned}\int_{B(0,1)} |f(x)|^p \,\mathrm{d}x &= \int_{B(0,1)} |x|^{\alpha p} \,\mathrm{d}x \\ &= \sum_{k=0}^{\infty} \int_{\{x\colon\ 2^{-k-1} \le |x| < 2^{-k}\}} |x|^{\alpha p} \,\mathrm{d}x \\ &\sim \sum_{k=0}^{\infty} 2^{-\alpha p k} 2^{-dk} = \sum_{k=0}^{\infty} 2^{(-\alpha p - d)k}.\end{aligned}$$

We get that $f \in L^p(B(0,1))$ if and only if $-\alpha p - d < 0$ or

$$\alpha > -\frac{d}{p}.$$

9.0.5 Definition. Given an exponent $p \in (1,\infty)$ define the dual exponent $p' \in (1,\infty)$ by

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

This definition also makes sense for $p = 1$ and then $p' = \infty$ – similarly, for $p = \infty$ it makes sense to set $p' = 1$.

9.0.6 Remark. Sometimes the dual exponent is also called the Hölder conjugate. Notice that if $p = 2$ we have $p' = 2 = p$. This makes the case $p = 2$ rather special. The space $L^2(\mu)$ is also a Hilbert space (its norm is given by the inner product $\langle f, g\rangle := \int fg \,\mathrm{d}\mu$) – this is often very important. However, we will not focus on this functional analytic side on this course.

To understand the $L^p(\mu)$ spaces it will be extremely important to prove Hölder's inequality. To this end, we first need the following auxiliary result.

9.0.7 Lemma (Young's inequality). *Let $1 < p < \infty$. Then we have*

$$ab \le \frac{a^p}{p} + \frac{b^{p'}}{p'}, \qquad a, b \ge 0.$$

Proof. Define

$$h(x) = \frac{x^p}{p} + \frac{1}{p'} - x, \qquad x \ge 0,$$

and notice that, by elementary analysis (differentiation),

$$h(x) \geq h(1) = 0, \text{ i.e., } x \leq \frac{x^p}{p} + \frac{1}{p'}.$$

Apply this with $x = ab^{1/(1-p)}$ to get

$$ab^{1/(1-p)} \leq \frac{a^p}{p} b^{-p'} + \frac{1}{p'}.$$

Here we used that

$$p' = \frac{p}{p-1}.$$

Multiply both sides of this inequality with

$$b^{1-\frac{1}{1-p}} = b^{\frac{-p}{1-p}} = b^{p'}$$

to establish the desired result. □

We are ready to prove Hölder's inequality.

9.0.8 Theorem (Hölder's inequality)**.** *Let $1 \leq p \leq \infty$, $f \in L^p(\mu)$ and $g \in L^{p'}(\mu)$. Then $fg \in L^1(\mu)$ and we have*

$$\|fg\|_{L^1(\mu)} \leq \|f\|_{L^p(\mu)} \|g\|_{L^{p'}(\mu)}.$$

Proof. If $p = 1$ we have the pointwise estimate $g \leq \|g\|_{L^\infty(\mu)}$ μ-a.e. and so

$$\|fg\|_{L^1(\mu)} = \int_X |fg| \, \mathrm{d}\mu \leq \|g\|_{L^\infty(\mu)} \int_X |f| \, \mathrm{d}\mu = \|f\|_{L^1(\mu)} \|g\|_{L^\infty(\mu)}.$$

The case $p = \infty$ is symmetric.

The main case $1 < p < \infty$ is an easy corollary of Young's inequality, Lemma 9.0.7. First, we may assume $\|f\|_{L^p(\mu)} > 0$ as otherwise $f = 0$ μ-a.e., and then also $\|fg\|_{L^1(\mu)} = 0$. Similarly, we may assume $\|g\|_{L^{p'}(\mu)} > 0$. Let x be such that $|f(x)| < \infty$ and $|g(x)| < \infty$ and define the scalars

$$a = \frac{|f(x)|}{\|f\|_{L^p(\mu)}} \qquad \text{and} \qquad b = \frac{|g(x)|}{\|g\|_{L^{p'}(\mu)}}.$$

Applying Young's inequality gives

$$\frac{|f(x)|}{\|f\|_{L^p(\mu)}} \frac{|g(x)|}{\|g\|_{L^{p'}(\mu)}} \leq \frac{1}{p} \frac{|f(x)|^p}{\|f\|^p_{L^p(\mu)}} + \frac{1}{p'} \frac{|g(x)|^{p'}}{\|g\|^{p'}_{L^{p'}(\mu)}}$$

This holds μ-a.e. as $f \in L^p(\mu)$ implies $|f(x)| < \infty$ μ-a.e. and similarly for g. Thus, we may take the μ integral over this inequality to get

$$\frac{\|fg\|_{L^1(\mu)}}{\|f\|_{L^p(\mu)} \|g\|_{L^{p'}(\mu)}} \leq \frac{1}{p} \frac{\|f\|^p_{L^p(\mu)}}{\|f\|^p_{L^p(\mu)}} + \frac{1}{p'} \frac{\|g\|^{p'}_{L^{p'}(\mu)}}{\|g\|^{p'}_{L^{p'}(\mu)}} = \frac{1}{p} + \frac{1}{p'} = 1.$$

Multiplying by $\|f\|_{L^p(\mu)} \|g\|_{L^{p'}(\mu)}$ yields the desired inequality. □

9.0.9 Remark. The L^p and L^q spaces need in general not be contained in one another in any particular way. There is one exception, where we have a clear rule. If $\mu(X) < \infty$ and $p < q$ we have by Hölder's inequality that with $s = q/p > 1$ that

$$\begin{aligned}\int_X |f|^p \, \mathrm{d}\mu &\leq \Big(\int_X |f|^{ps} \, \mathrm{d}\mu \Big)^{\frac{1}{s}} \Big(\int_X 1^{s'} \, \mathrm{d}\mu \Big)^{\frac{1}{s'}} \\ &= \Big(\int_X |f|^q \, \mathrm{d}\mu \Big)^{\frac{p}{q}} \mu(X)^{1-\frac{1}{s}} \\ &= \|f\|_{L^q(\mu)}^p \mu(X)^{1-\frac{p}{q}}\end{aligned}$$

and so

$$\|f\|_{L^p(\mu)} \leq \|f\|_{L^q(\mu)} \mu(X)^{\frac{1}{p}-\frac{1}{q}}.$$

So we have the quantitative estimate from above – in particular, we have $L^q(\mu) \subset L^p(\mu)$. It would be possible to establish the inclusion with a more elementary argument as well.

9.0.10 Remark. When applying Hölder's inequality do not be afraid of calculations involving the exponents. Often one has to manipulate the relationship

$$\frac{1}{p} + \frac{1}{p'} = 1$$

in various ways. For example, we can solve

$$p' = \frac{p}{p-1}.$$

We can also get further identities like

$$1 - p' = 1 - \frac{p}{p-1} = -\frac{1}{p-1}$$

and

$$\frac{p}{p'} = p - 1.$$

This dual exponent calculus takes a while to master but is extremely important.

9.0.11 Remark. As summation is integration against the counting measure (as we have seen), we also have

$$\sum_{j=1}^{\infty} |a_j||b_j| \leq \Big(\sum_{j=1}^{\infty} |a_j|^p \Big)^{\frac{1}{p}} \Big(\sum_{j=1}^{\infty} |b_j|^{p'} \Big)^{\frac{1}{p'}}.$$

It is not necessarily easy to appreciate Hölder's inequality at first sight. However, it is one of the most fundamental inequalities and an analyst applies it every day. Next, we present a key application by showing that $\|\cdot\|_{L^p(\mu)}$ is a norm – only the triangle inequality requires some work and this inequality is often also called Minkowski's inequality.

9.0.12 Corollary ($\triangle$-inequality for $\|\cdot\|_{L^p}$ / Minkowski's inequality)**.** *For $p \in [1, \infty]$ we have*

$$\|f + g\|_{L^p(\mu)} \leq \|f\|_{L^p(\mu)} + \|g\|_{L^p(\mu)}, \qquad f, g \in L^p(\mu).$$

Proof. The case $p = 1$ is clear by the usual triangle inequality. The case $p = \infty$ is an exercise. Suppose now $p \in (1, \infty)$.

We have the pointwise estimate

$$|f+g|^p = |f+g||f+g|^{p-1} \le |f||f+g|^{p-1} + |g||f+g|^{p-1}.$$

Integrating this and using Hölder's inequality gives that

$$\begin{aligned}\|f+g\|_{L^p(\mu)}^p &\le \|f|f+g|^{p-1}\|_{L^1(\mu)} + \|g|f+g|^{p-1}\|_{L^1(\mu)} \\ &\le \|f\|_{L^p(\mu)}\||f+g|^{p-1}\|_{L^{p'}(\mu)} + \|g\|_{L^p(\mu)}\||f+g|^{p-1}\|_{L^{p'}(\mu)}.\end{aligned}$$

Recall that

$$p' = \frac{p}{p-1}$$

so that $(p-1)p' = p$. Therefore, we have $\||f+g|^{p-1}\|_{L^{p'}(\mu)} = \|f+g\|_{L^p(\mu)}^{p/p'}$, and we have established

$$\|f+g\|_{L^p(\mu)}^p \le (\|f\|_{L^p(\mu)} + \|g\|_{L^p(\mu)})\|f+g\|_{L^p(\mu)}^{p/p'}. \tag{9.0.13}$$

Notice then the simple algebra

$$p - p/p' = p\Big(1 - \frac{1}{p'}) = \frac{p}{p} = 1.$$

We are essentially done as it only remains to divide by $\|f+g\|_{L^p(\mu)}^{p/p'}$ in (9.0.13) and use the above algebra. However, we have to worry about the possibility that $\|f+g\|_{L^p(\mu)} = \infty$, in which case (9.0.13) says absolutely nothing. We can rule this out by the following very rough estimate. Notice that

$$|f+g|^p \le (|f|+|g|)^p \le (2\max(|f|,|g|))^p \le 2^p(|f|^p + |g|^p)$$

and so by integrating this we get

$$\|f+g\|_{L^p(\mu)}^p \lesssim \|f\|_{L^p(\mu)}^p + \|g\|_{L^p(\mu)}^p < \infty.$$

Now we can perform the division in (9.0.13) and end the proof. □

9.0.14 Corollary. *$(L^p(\mu), \|\cdot\|_{L^p(\mu)})$ is a normed space.*

Proof. Notice that $\|f\|_{L^p(\mu)} = 0$ implies that pointwise μ-a.e. we have $|f|^p = 0$ and so $f = 0$ (recall that we identify functions that agree almost everywhere). Also, it is clear from the definition that $\|af\|_{L^p(\mu)} = |a|\|f\|_{L^p(\mu)}$. The triangle inequality was proved above – it is exactly Minkowski's inequality. □

The following trivial inequality is often useful (recommendation: remember just the proof technique). It is also called Chebyshev's inequality. We have used similar estimates already before.

9.0.15 Lemma. *Let $1 \le p < \infty$ and $f \in L^p(\mu)$. Then for every $\lambda > 0$ we have*

$$\mu(\{|f| > \lambda\}) \le \frac{1}{\lambda^p}\int |f|^p \,\mathrm{d}\mu.$$

Proof. Notice that

$$\mu(\{|f| > \lambda\}) = \frac{1}{\lambda^p}\int_{\{|f|>\lambda\}} \lambda^p \,\mathrm{d}\mu \le \frac{1}{\lambda^p}\int |f|^p \,\mathrm{d}\mu.$$

□

9.0.16 Remark. Written differently the above says that

$$\sup_{\lambda>0} \lambda\mu(\{|f| > \lambda\})^{\frac{1}{p}} \le \|f\|_{L^p(\mu)}.$$

It is then possible to define the weak-L^p space or $L^{p,\infty}$ by asking that

$$\|f\|_{L^{p,\infty}(\mu)} := \sup_{\lambda>0} \lambda\mu(\{|f| > \lambda\})^{\frac{1}{p}} < \infty.$$

The above shows that $L^p(\mu) \subset L^{p,\infty}(\mu)$ and

$$\|f\|_{L^{p,\infty}(\mu)} \le \|f\|_{L^p(\mu)}.$$

9.1 Completeness of L^p

Given a normed space $(Y, \|\cdot\|_Y)$ and a sequence (y_j), $y_j \in Y$, we say that (y_j) is a Cauchy sequence in Y if the following holds. For all $\epsilon > 0$ there is m_ϵ so that

$$\|y_i - y_j\|_Y < \epsilon$$

whenever $i, j \ge m_\epsilon$. Notice that a converging sequence is automatically a Cauchy sequence. We say that Y is complete if every Cauchy sequence in Y converges in Y – that is, whenever (y_j) is a Cauchy sequence in Y there is $y \in Y$ so that

$$\lim_{j\to\infty} \|y_j - y\|_Y = 0.$$

A complete normed space is called a Banach space.

We will show that $L^p(\mu)$, $1 \le p \le \infty$, is a Banach space, which is a fundamental property of these spaces. We already know that $L^p(\mu)$ is a normed space so it remains to prove the completeness.

The following result is quite useful on its own, but will also yield the completeness in the case $p < \infty$ easily.

9.1.1 Theorem. *Suppose (f_j) is a Cauchy sequence in $L^p(\mu)$, $1 \le p < \infty$. Then there is a subsequence (f_{j_k}) that converges pointwise μ-a.e. to some $f \in L^p(\mu)$.*

Proof. Using the fact that (f_j) is a Cauchy sequence choose the indices $j_1 < j_2 < \cdots$ so that

$$\|f_i - f_j\|_{L^p(\mu)} < 2^{-k}$$

whenever $i, j \ge j_k$. Define formally the series

$$g(x) = \sum_{m=1}^{\infty} [f_{j_{m+1}}(x) - f_{j_m}(x)].$$

Does this converge? Notice that by the monotone convergence theorem and the triangle inequality in $L^p(\mu)$ we have

$$\begin{aligned}\Big\| \sum_{m=1}^{\infty} |f_{j_{m+1}} - f_{j_m}| \Big\|_{L^p(\mu)} &= \lim_{k\to\infty} \Big\| \sum_{m=1}^{k} |f_{j_{m+1}} - f_{j_m}| \Big\|_{L^p(\mu)} \\ &\le \lim_{k\to\infty} \sum_{m=1}^{k} \|f_{j_{m+1}} - f_{j_m}\|_{L^p(\mu)} \le \sum_{m=1}^{\infty} 2^{-m} = 1.\end{aligned}$$

It follows that

$$\sum_{m=1}^{\infty} |f_{j_{m+1}}(x) - f_{j_m}(x)| < \infty$$

for μ-a.e. x, and therefore the series defining g converges almost everywhere. Now, notice that we have the telescoping sum identity

$$f_{j_k} = f_{j_1} + \sum_{m=1}^{k-1} [f_{j_{m+1}} - f_{j_m}]$$

and so $f_{j_k}(x) \to f_{j_1}(x) + g(x) =: f(x)$ for μ-a.e. x, $f \in L^p(\mu)$. □

9.1.2 Lemma. *The space $L^p(\mu)$, $1 \le p \le \infty$, is Banach.*

Proof. **Case $p < \infty$:** Suppose (f_j) is a Cauchy sequence in $L^p(\mu)$. By Theorem 9.1.1 there exists $f \in L^p(\mu)$ and a subsequence (f_{j_k}) so that $f_{j_k}(x) \to f(x)$ for μ-a.e. x. It is enough to show that

$$\lim_{j\to\infty} \|f - f_j\|_{L^p(\mu)} = 0.$$

Notice that this claim involves the original sequence, not just the subsequence that converges almost everywhere. Combining the almost everywhere convergence with the fact that the sequence is Cauchy in $L^p(\mu)$ yields the claim quite easily. Indeed, notice that by the a.e. convergence and Fatou's lemma we have for all $\epsilon > 0$ that

$$\begin{aligned}\|f - f_j\|^p_{L^p(\mu)} &= \int |f(x) - f_j(x)|^p \, \mathrm{d}\mu(x) \\ &= \int \lim_{k\to\infty} |f_{j_k}(x) - f_j(x)|^p \, \mathrm{d}\mu(x) \\ &\le \liminf_{k\to\infty} \int |f_{j_k}(x) - f_j(x)|^p \, \mathrm{d}\mu(x) = \liminf_{k\to\infty} \|f_{j_k} - f_j\|^p_{L^p(\mu)} < \epsilon\end{aligned}$$

for all large enough j. This shows the claim.

Case $p = \infty$: This is an exercise – a simpler elementary case based on the completeness of the scalar field. □

9.1.3 Remark. Notice carefully that almost everywhere convergence alone is not enough for L^p convergence (exercise). Moreover, L^p convergence does not imply that the **whole** sequence would converge almost everywhere (exercise).

9.2 Approximation by continuous functions

9.2.1 Lemma. *Suppose X is a locally compact Hausdorff space and $(X, \mathcal{F}, \mu)$ is a measure space like in the Riesz representation theorem 5.0.3. Let $1 \le p < \infty$. Then $C_c(X)$ is dense in $L^p(\mu)$. In other words, given $f \in L^p(\mu)$ for every $\epsilon > 0$ there exists $g \in C_c(X)$ so that $\|f - g\|_{L^p(\mu)} < \epsilon$.*

Proof. We first assume that $f = 1_A$, where $A \in \mathcal{F}$ and $\mu(A) < \infty$. Let $\epsilon > 0$. Then we choose compact $K \subset A$ and open $V \supset A$ so that $\mu(U \setminus K) < \epsilon$. Using Urysohn's lemma we find $g \in C_c(X)$ with $K \prec g \prec U$. This gives that

$$\|f - g\|_{L^p(\mu)} \le \Big(\int_{U\setminus K} 1 \, \mathrm{d}\mu \Big)^{1/p} = \mu(U \setminus K)^{1/p} < \epsilon^{1/p}.$$

Let now $f \in L^p(\mu)$. We may assume that $f \geq 0$ (write $f = f_+ - f_-$ and approximate each piece separately). We then know by Lemma 3.1.2 that there exists simple functions s_i so that $0 \leq s_1 \leq s_2 \leq \ldots \leq f$ and $f(x) = \lim_{i\to\infty} s_i(x)$ for every $x \in X$. By dominated convergence theorem we have $\lim_{i\to\infty} \|f - s_i\|_{L^p(\mu)} = 0$. Given $\epsilon > 0$ fix $s := s_i$ so that $\|f - s\|_{L^p(\mu)} < \epsilon$. Write $s = \sum_{j=1}^N c_j 1_{A_j}$, where the sets $A_j \in \mathcal{F}$, $j = 1, \ldots, N$, are pairwise disjoint and $c_j > 0$. As $s \in L^p(\mu)$ ($s \leq f$, $f \in L^p(\mu)$) and

$$\|s\|_{L^p(\mu)}^p = \sum_{j=1}^N c_j^p \mu(A_j),$$

we can conclude that $\mu(A_j) < \infty$ for every j. By the first part of the proof, we can approximate each 1_{A_j} with a $C_c(X)$ function in the $L^p(\mu)$ norm – it follows right away that we can approximate the linear combination s as well. So choose $g \in C_c(X)$ with $\|s - g\|_{L^p(\mu)} < \epsilon$. Now we have $\|f - g\|_{L^p(\mu)} \leq \|f - s\|_{L^p(\mu)} + \|s - g\|_{L^p(\mu)} < 2\epsilon$, so we are done.

□

The following is the main result of this section – the continuity of translations in L^p. It is extremely useful – however, the proof is almost trivial given the above density of continuous functions.

9.2.2 Theorem. *Suppose $f \in L^p(\mathbb{R}^d)$ for $1 \leq p < \infty$. Then*

$$\lim_{y\to 0} \int_{\mathbb{R}^d} |f(x) - f(x+y)|^p \, \mathrm{d}x = 0.$$

Proof. Fix $f \in L^p$ and denote the translation operator $\tau_y f(x) = f(x+y)$. Let $\epsilon > 0$ and choose $g \in C_c$ so that $\|f - g\|_p < \epsilon$. By translation invariance also $\|\tau_y f - \tau_y g\|_p = \|f - g\|_p < \epsilon$. Hence, by the triangle inequality for the L^p norm (Minkowski's inequality), it is enough to show that

$$\lim_{y\to 0} \|g - \tau_y g\|_p = 0.$$

That is, we have reduced to showing the theorem for continuous functions – a fundamental and often used technique. Take an arbitrary sequence $y_k \to 0$. We may suppose that $|y_k| < 1$. Choose $M > 1$ so that $\operatorname{spt} g \subset B(0, M)$. Notice that if $g(x + y_k) \neq 0$, then $x + y_k \in B(0, M)$ and so $x \in B(0, 2M)$. Thus, we have

$$\|g - \tau_{y_k} g\|_p^p = \int_{B(0,2M)} |g(x) - g(x + y_k)|^p \, \mathrm{d}x.$$

As g is bounded, $|g(x) - g(x + y_k)| \leq C$ and $C \in L^p(B(0, 2M))$, DCT gives

$$\lim_{k\to\infty} \int_{B(0,2M)} |g(x) - g(x + y_k)|^p \, \mathrm{d}x = \int_{B(0,2M)} \lim_{k\to\infty} |g(x) - g(x + y_k)|^p \, \mathrm{d}x = 0,$$

where we used the continuity of g with a fixed x. We are done. □

We will next use this result to study convolutions.

9.3 Convolution and L^p convergence of approximate identities

Convolutions are used in applied and pure mathematics for various approximation arguments. The definition can look rather strange at first, but they are fundamental. We again work with the Lebesgue measure as translation invarience is important here.

For $f, g \in L^1$ we define for $x \in \mathbb{R}^d$ the convolution

$$f * g(x) = \int_{\mathbb{R}^d} f(y)g(x-y)\,\mathrm{d}y.$$

This is well-defined, since $\int_{\mathbb{R}^d} |f(y)g(x-y)|\,\mathrm{d}y < \infty$ for a.e. $x \in \mathbb{R}^d$. The latter follows from Fubini and translation invariance:

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |f(y)g(x-y)|\,\mathrm{d}y\,\mathrm{d}x = \int_{\mathbb{R}^d} |f(y)| \int_{\mathbb{R}^d} |g(x-y)|\,\mathrm{d}x\,\mathrm{d}y = \Big(\int_{\mathbb{R}^d} |f|\Big)\Big(\int_{\mathbb{R}^d} |g|\Big).$$

In fact, we were not very rigorous yet – the use of Fubini requires the measurability of the function

$$(x, y) \mapsto f(x-y)g(y).$$

To show this we use Corollary **??** to choose Borel functions f_0, g_0 so that $f = f_0$ and $g = g_0$ almost everywhere. We show that $(x, y) \mapsto f_0(x-y)g_0(y)$ is a Borel function – this follows easily, since $f_0(x-y)g_0(y) = f_0(u(x, y))g_0(v(x, y))$, where $u(x, y) = x - y$ and $v(x, y) = y$ are continuous (in particular, Borel). Indeed, now $f_0 \circ u$ is Borel – see the discussion near (3.0.7). As also $g_0 \circ v$ is Borel, the product of these functions is Borel, in particular Lebesgue measurable.

We now want to say that $f(x-y)g(y) = f_0(x-y)g_0(y)$ for almost every (x, y) and use Lemma 3.0.15 to conclude that the function $(x, y) \mapsto f(x-y)g(y)$ is measurable. We stop to think, even though it sounds clear, why $f(x-y)g(y) = f_0(x-y)g_0(y)$ for almost every (x, y). First, we choose $N \subset \mathbb{R}^d$ so that $|N| = 0$, $f(x) = f_0(x)$ and $g(x) = g_0(x)$ for $x \in \mathbb{R}^d \setminus N$. Now, by Corollary A.0.15 we may assume that N is a Borel set. It follows that

$$\{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d \colon x - y \in N\} = u^{-1}N$$

is a Borel set. It has zero measure, since by applying Fubini to the corresponding characteristic function (a non-negative measurable function) we have

$$|\{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d \colon x - y \in N\}| = \int_{\mathbb{R}^d} |N + y|\,\mathrm{d}y = \int_{\mathbb{R}^d} |N|\,\mathrm{d}y = 0.$$

Outside this set of measure zero $f(x-y) = f_0(x-y)$. A similar result holds for g, and we get the claim.

What we have proved implies that $f * g \in L^1$ if $f, g \in L^1$ and

$$\|f * g\|_1 \le \|f\|_1 \|g\|_1.$$

The following properties of the convolution are left as an exercise (here $f, g, h \in L^1$):

1. $f * (g + h) = f * g + f * h$;
2. $(\lambda f) * g = \lambda(f * g)$, $\lambda \in \mathbb{R}$;
3. $f * g = g * f$;
4. $f * (g * h) = (f * g) * h$;

9.3.1 Remark. It is important that the convolution of two functions can be defined in more generality. It is an exercise to prove the following result. Let $1 \le p, q, r \le \infty$ satisfy

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

If $f \in L^p$ and $g \in L^q$ we have $f * g \in L^r$ and

$$\|f * g\|_r \leq \|f\|_p \|g\|_q.$$

In particular, $f * g \in L^p$ if $f \in L^p$ and $g \in L^1$. Some of the calculations in the proof of Proposition 9.3.4 give a hint how to do this.

The way that we will use convolutions to approximate general functions follows the following general scheme:

- The convolution $f * g$ is usually as regular as the more regular of the two functions f and g. In particular, if e.g. just g is smooth, then so will be $f * g$ (as we will show).
- We aim to identify general families of functions g_ϵ so that $f * g_\epsilon$ converges to f in L^p. If we can furthermore choose these functions to be e.g. smooth, we will have obtained smooth functions $f * g_\epsilon$ that approximate f on L^p.

We first focus on general abstract conditions on a family of functions φ_ϵ that guarantee when convolutions $f * \varphi_\epsilon$ converge to f.

9.3.2 Definition. A family $\varphi_\epsilon \in L^1$, $\epsilon > 0$, is an approximate identity (as $\epsilon \to 0$) if the following conditions hold.

1. We have $\int_{\mathbb{R}^d} \varphi_\epsilon = 1$ for all $\epsilon > 0$.
2. We have $\sup_\epsilon \|\varphi_\epsilon\|_1 < \infty$.
3. For every $\delta > 0$ we have

$$\lim_{\epsilon \to 0} \int_{|x| \geq \delta} |\varphi_\epsilon(x)| \, \mathrm{d}x = 0.$$

9.3.3 Remark. The following pointers regarding approximate identities are often helpful.

- Notice that if $\varphi_\epsilon \geq 0$, then (2) follows from (1). This is often the case.
- If $\operatorname{spt} \varphi_\epsilon \subset B(0, c(\epsilon))$, where $\lim_{\epsilon \to 0} c(\epsilon) = 0$, then (3) holds.
- If a fixed function $\eta \in L^1$ satisfies $\int \eta = 1$ and $\operatorname{spt} \eta \subset B(0,1)$, then $\eta_\epsilon := \frac{1}{\epsilon^d} \eta(x/\epsilon)$ is an approximate identity. In fact, the condition $\operatorname{spt} \eta \subset B(0,1)$ is not needed (exercise).

Convolutions with approximate identities $f * \varphi_\epsilon$ are a very important way to approximate a given function $f \in L^p$ as $\epsilon \to 0$. Notice that by Remark 9.3.1 $f * \varphi_\epsilon$ is a well-defined L^p function if $f \in L^p$, $1 \leq p < \infty$ (as $\varphi_\epsilon \in L^1$).

9.3.4 Proposition. *Let $1 \leq p < \infty$, $f \in L^p$ and $(\varphi_\epsilon)_{\epsilon > 0}$ be an approximate identity. Then we have*

$$\|f - f * \varphi_\epsilon\|_p \to 0, \qquad \epsilon \to 0.$$

Proof. Fix $f \in L^p$. Using $\int_{\mathbb{R}^d} \varphi_\epsilon = 1$ and $f * \varphi_\epsilon = \varphi_\epsilon * f$ we write the pointwise identity

$$\begin{aligned} f(x) - f * \varphi_\epsilon(x) &= f(x) \int_{\mathbb{R}^d} \varphi_\epsilon(y) \, \mathrm{d}y - \int_{\mathbb{R}^d} f(x-y) \varphi_\epsilon(y) \, \mathrm{d}y \\ &= \int_{\mathbb{R}^d} [f(x) - f(x-y)] \varphi_\epsilon(y) \, \mathrm{d}y. \end{aligned}$$

For the moment let $p > 1$. We get using Hölder's inequality that

$$\begin{aligned}|f(x) - f * \varphi_\epsilon(x)| &\leq \int_{\mathbb{R}^d} |f(x) - f(x-y)||\varphi_\epsilon(y)|^{1/p}|\varphi_\epsilon(y)|^{1/p'} \,\mathrm{d}y \\ &\leq \Big(\int_{\mathbb{R}^d} |f(x) - f(x-y)|^p|\varphi_\epsilon(y)| \,\mathrm{d}y\Big)^{1/p}\Big(\int_{\mathbb{R}^d} |\varphi_\epsilon(y)| \,\mathrm{d}y\Big)^{1/p'} \\ &\lesssim \Big(\int_{\mathbb{R}^d} |f(x) - f(x-y)|^p|\varphi_\epsilon(y)| \,\mathrm{d}y\Big)^{1/p},\end{aligned}$$

where the last step used that $\sup_\epsilon \|\varphi_\epsilon\|_1 \lesssim 1$. Therefore, we have

$$|f(x) - f * \varphi_\epsilon(x)|^p \lesssim \int_{\mathbb{R}^d} |f(x) - f(x-y)|^p|\varphi_\epsilon(y)| \,\mathrm{d}y,$$

which also clearly holds with $p = 1$. We integrate this over $x \in \mathbb{R}^d$, and use Fubini's theorem, to get that

$$\|f - f * \varphi_\epsilon\|_p^p \lesssim \int_{\mathbb{R}^d} |\varphi_\epsilon(y)| \int_{\mathbb{R}^d} |f(x) - f(x-y)|^p \,\mathrm{d}x \,\mathrm{d}y.$$

Let $\gamma > 0$. Using the continuity of translations – Theorem 9.2.2 – we find $\delta > 0$ so that

$$\int_{\mathbb{R}^d} |f(x) - f(x-y)|^p \,\mathrm{d}x < \gamma$$

whenever $|y| < \delta$. Using property (3) of Definition 9.3.2 we find ϵ_0 so that

$$\int_{|y|\geq\delta} |\varphi_\epsilon(y)| \,\mathrm{d}y < \gamma$$

for all $\epsilon \leq \epsilon_0$. For all $\epsilon \leq \epsilon_0$ we therefore have

$$\begin{aligned}&\int_{\mathbb{R}^d} |\varphi_\epsilon(y)| \int_{\mathbb{R}^d} |f(x) - f(x-y)|^p \,\mathrm{d}x \,\mathrm{d}y \\ &\quad= \int_{|y|<\delta} |\varphi_\epsilon(y)| \int_{\mathbb{R}^d} |f(x) - f(x-y)|^p \,\mathrm{d}x \,\mathrm{d}y \\ &\quad+ \int_{|y|\geq\delta} |\varphi_\epsilon(y)| \int_{\mathbb{R}^d} |f(x) - f(x-y)|^p \,\mathrm{d}x \,\mathrm{d}y \\ &\quad\lesssim \gamma \int_{\mathbb{R}^d} |\varphi_\epsilon(y)| \,\mathrm{d}y + \|f\|_p^p \int_{|y|\geq\delta} |\varphi_\epsilon(y)| \,\mathrm{d}y.\end{aligned}$$

Recalling $\sup_\epsilon \|\varphi_\epsilon\|_1 \lesssim 1$ and $\int_{|y|\geq\delta} |\varphi_\epsilon(y)| \,\mathrm{d}y < \gamma$ we get that for all $\epsilon \leq \epsilon_0$ we have

$$\|f - f * \varphi_\epsilon\|_p^p \lesssim \gamma(1 + \|f\|_p^p).$$

This ends the proof as $\gamma > 0$ was arbitrary. □

In the next section we will build suitable smooth approximations of the identity – and this will prove the density of smooth functions.

9.4 Interpolation

Let (X, μ) be a σ-finite measure space ($X = \bigcup_{i=1}^{\infty} X_i$, $\mu(X_i) < \infty$). For $0 < p < \infty$ and a measurable $f\colon X \to \mathbb{R}$ recall/define that

$$\begin{aligned}
\|f\|_{L^p(X)} &= \|f\|_{L^p(\mu)} = \Big(\int_X |f|^p \, \mathrm{d}\mu \Big)^{1/p}, \\
\|f\|_{L^{p,\infty}(X)} &= \|f\|_{L^{p,\infty}(\mu)} = \sup_{\lambda>0} \lambda \mu(\{x \in X\colon |f(x)| > \lambda\})^{1/p}, \\
\|f\|_{L^\infty(X)} &= \|f\|_{L^\infty(\mu)} = \inf\{C \ge 0\colon |f(x)| \le C \text{ for } \mu\text{-a.e. } x \in X\}, \\
\|f\|_{L^{\infty,\infty}(X)} &= \|f\|_{L^\infty(X)}.
\end{aligned}$$

The so-called weak-$L^p(X)$ – denoted $L^{p,\infty}(X)$ – consists of those f for which we have $\|f\|_{L^{p,\infty}(X)} < \infty$. If $f \in L^p(X)$ then for all $\lambda > 0$ we have

$$\mu(\{|f| > \lambda\}) = \frac{1}{\lambda^p} \int_{\{|f|>\lambda\}} \lambda^p \le \frac{1}{\lambda^p} \int |f|^p$$

from which it follows that $\|f\|_{L^{p,\infty}(X)} \le \|f\|_{L^p(X)} < \infty$. That is, we have the natural inclusion $L^p(X) \subset L^{p,\infty}(X)$.

9.4.1 Theorem (Marcinkiewicz interpolation theorem)**.** *Let (X, μ) and (Y, ν) be σ-finite measure spaces and let $0 < p_0 < p_1 \le \infty$. Let T be a **sublinear** operator defined on the space $L^{p_0}(X) + L^{p_1}(X)$ and taking values in the space of measurable functions on Y. Assume that there exists two constants A_0 and A_1 such that*

$$\begin{aligned}
\|Tf\|_{L^{p_0,\infty}(Y)} &\le A_0 \|f\|_{L^{p_0}(X)}, \qquad f \in L^{p_0}(X), \\
\|Tf\|_{L^{p_1,\infty}(Y)} &\le A_1 \|f\|_{L^{p_1}(X)}, \qquad f \in L^{p_1}(X).
\end{aligned}$$

Let $p \in (p_0, p_1)$ and write

$$\frac{1}{p} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}, \qquad \theta \in (0,1).$$

Then we have

$$\|Tf\|_{L^p(Y)} \le 2\Big(\frac{p}{p-p_0} + \frac{p}{p_1 - p} \Big)^{1/p} A_0^{1-\theta} A_1^{\theta} \|f\|_{L^p(X)}.$$

9.4.2 Remark. Sublinearity means that we have the pointwise estimates

$$|T(f+g)| \le |Tf| + |Tg| \qquad \text{and} \qquad |T(\lambda f)| = |\lambda||Tf|,\ \lambda \in \mathbb{R}.$$

Marcinkiewicz interpolation theorem is an easy but very useful interpolation theorem. The good points are:

1. We can assume only $L^q \to L^{q,\infty}$ type estimates at the endpoints $q \in \{p_0, p_1\}$ but conclude strong $L^p \to L^p$ estimates for $p_0 < p < p_1$.
2. T does not need to be linear – this is important in what follows (T will e.g. be a so-called maximal function).

This theorem has a rather simple proof using the important identity

$$\int_X |f|^p \, \mathrm{d}\mu = p \int_0^\infty \lambda^{p-1} \mu(\{x \in X\colon |f(x)| > \lambda\}) \, \mathrm{d}\lambda, \qquad 0 < p < \infty. \tag{9.4.3}$$

The proof of this identity is left as an exercise. The weak point of the Marcinkiewicz interpolation theorem is that we cannot interpolate estimates like $L^{p_0} \to L^{q_0}$ and $L^{p_1} \to L^{q_1}$, but rather need to have $p_0 = q_0$ and $p_1 = q_1$. Such interpolation results do exist (the Riesz-Thorin interpolation theorem), but we will not cover those here.

Proof of Theorem 9.4.1. Assume $p_1 < \infty$ – the case $p_1 = \infty$ is an exercise. Let $f \in L^p$, $p_0 < p < p_1$. Fix some parameter $\lambda > 0$ related to the level sets of the form $\{|g| > \lambda\}$ appearing in (9.4.3) and fix also another technical parameter $\delta > 0$ (which we will later fix in a natural way to recover the claimed quantitative estimate).

Define $f_0 = f 1_{\{|f|>\delta\lambda\}}$ and $f_1 = f - f_0$. It is almost obvious that $f_0 \in L^{p_0}(X)$ (as $p_0 - p < 0$) and $f_1 \in L^{p_1}(X)$ (as $p_1 - p > 0$) – in particular, Tf is defined by assumption and we have by sublinearity that

$$|Tf| \le |Tf_0| + |Tf_1|.$$

Therefore, we have

$$\{|Tf| > \lambda\} \subset \{|Tf_0| > \lambda/2\} \cup \{|Tf_1| > \lambda/2\},$$

and so

$$\begin{aligned}
\nu(\{|Tf| > \lambda\}) &\le \nu(\{|Tf_0| > \lambda/2\}) + \nu(\{|Tf_1| > \lambda/2\}) \\
&\le \Big(\frac{\lambda}{2}\Big)^{-p_0} \|Tf_0\|_{L^{p_0,\infty}(Y)}^{p_0} + \Big(\frac{\lambda}{2}\Big)^{-p_1} \|Tf_1\|_{L^{p_1,\infty}(Y)}^{p_1} \\
&\le \Big(\frac{\lambda}{2}\Big)^{-p_0} A_0^{p_0} \|f_0\|_{L^{p_0}(X)}^{p_0} + \Big(\frac{\lambda}{2}\Big)^{-p_1} A_1^{p_1} \|f_1\|_{L^{p_1}(X)}^{p_1} \\
&= \Big(\frac{\lambda}{2}\Big)^{-p_0} A_0^{p_0} \int_{|f|>\delta\lambda} |f(x)|^{p_0} \,\mathrm{d}\mu(x) + \Big(\frac{\lambda}{2}\Big)^{-p_1} A_1^{p_1} \int_{|f|\le\delta\lambda} |f(x)|^{p_1} \,\mathrm{d}\mu(x).
\end{aligned}$$

In the last estimate we used the main assumption concerning the weak type estimates $L^{p_0}(X) \to L^{p_0,\infty}(Y)$ and $L^{p_1}(X) \to L^{p_1,\infty}(Y)$.

Using (9.4.3) we get that

$$\begin{aligned}
&\|Tf\|_{L^p(Y)}^p = p \int_0^\infty \lambda^{p-1} \nu(\{|Tf| > \lambda\}) \,\mathrm{d}\lambda \\
&\le p(2A_0)^{p_0} \int_0^\infty \lambda^{p-1}\lambda^{-p_0} \int_{|f|>\delta\lambda} |f(x)|^{p_0} \,\mathrm{d}\mu(x) \,\mathrm{d}\lambda \\
&+ p(2A_1)^{p_1} \int_0^\infty \lambda^{p-1}\lambda^{-p_1} \int_{|f|\le\delta\lambda} |f(x)|^{p_1} \,\mathrm{d}\mu(x) \,\mathrm{d}\lambda = I + II.
\end{aligned}$$

By Fubini's theorem we have

$$\begin{aligned}
I &= p(2A_0)^{p_0} \int_X |f(x)|^{p_0} \int_0^{|f(x)|/\delta} \lambda^{p-p_0-1} \,\mathrm{d}\lambda \,\mathrm{d}\mu(x) \\
&= \frac{p(2A_0)^{p_0}}{p-p_0} \frac{1}{\delta^{p-p_0}} \int_X |f(x)|^p \,\mathrm{d}\mu(x)
\end{aligned}$$

and similarly

$$II = \frac{p(2A_1)^{p_1}}{p_1-p} \frac{1}{\delta^{p-p_1}} \int_X |f(x)|^p \,\mathrm{d}\mu(x).$$

Therefore, we have already proved that

$$\|Tf\|_{L^p(Y)}^p \le p\Big(\frac{(2A_0)^{p_0}}{p-p_0}\frac{1}{\delta^{p-p_0}} + \frac{(2A_1)^{p_1}}{p_1-p}\frac{1}{\delta^{p-p_1}}\Big)\|f\|_{L^p(X)}^p.$$

If we want to recover the exact claimed quantitative dependence on the various constants (which will not be important to us in what follows), it is now natural to fix δ so that

$$(2A_0)^{p_0}\frac{1}{\delta^{p-p_0}} = (2A_1)^{p_1}\frac{1}{\delta^{p-p_1}},$$

which gives

$$\delta = \frac{1}{2} A_0^{\frac{p_0}{p_1 - p_0}} A_1^{-\frac{p_1}{p_1 - p_0}}.$$

We then get

$$\|Tf\|_{L^p(Y)} \leq (2A_0)^{p_0/p} \frac{1}{\delta^{1-p_0/p}} \left(\frac{p}{p - p_0} + \frac{p}{p_1 - p} \right)^{1/p} \|f\|_{L^p(X)},$$

where

$$(2A_0)^{p_0/p} \frac{1}{\delta^{1-p_0/p}} = 2A_0^{1 - \frac{p_0 p_1 - p p_1}{p_0 p - p p_1}} A_1^{\frac{p_0 p_1 - p p_1}{p_0 p - p p_1}}.$$

We are done after solving for θ, which gives the desired formula

$$\theta = \frac{1/p - 1/p_0}{1/p_1 - 1/p_0} = \frac{p_0 p_1 - p p_1}{p_0 p - p p_1}.$$

□

Chapter 10

Differentiation

For a locally integrable $f \in L^1_{\mathrm{loc}} = L^1_{\mathrm{loc}}(\mathbb{R}^d; \mathrm{d}x)$ define the (centered) Hardy–Littlewood maximal function

$$Mf(x) := \sup_{r>0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y)| \, \mathrm{d}y.$$

In practical arguments it is often convenient to use the following larger maximal function as well

$$x \mapsto \sup_{B \text{ open ball}} \frac{1_B(x)}{|B|} \int_B |f(y)| \, \mathrm{d}y.$$

Notice that if $x \in B = B(z,r)$, then $B \subset B(x,2r)$, and so (as $|B| \sim r^n \sim |B(x,2r)|$) we have

$$\sup_{B \text{ open ball}} \frac{1_B(x)}{|B|} \int_B |f(y)| \, \mathrm{d}y \lesssim Mf(x).$$

That is, these are pointwise comparable functions, and results that hold for one of them, also hold for the other. We can call this other one the 'non-centred maximal function' and denote it e.g. by $M_{nc}f(x)$.

The maximal function is of fundamental use in analysis as it has good mapping properties and it e.g. dominates many other operators pointwise. We will now prove the mapping properties.

10.0.1 Theorem (Basic covering theorem)**.** *Let $\mathcal{B}$ be a finite family of open (or closed) balls in $\mathbb{R}^d$. Then there exists pairwise disjoint balls $B_1, B_2, \ldots, B_m \in \mathcal{B}$ such that*

$$\bigcup_{B \in \mathcal{B}} B \subset \bigcup_{i=1}^{m} 3B_i.$$

Proof. Let $\mathcal{B} = \{U_j\}_{j=1}^N$, where $U_j = B(x_j, r_j)$. As this is a finite collection, by reordering we may assume that $r_1 \geq r_2 \geq \ldots \geq r_N$. Let $B_1 = U_1$, and then let B_2 be the biggest ball U_j so that $U_j \not\subset 3B_1$ (if it exists). Let then B_3 be the biggest ball U_j so that $U_j \not\subset 3B_1 \cup 3B_2$ (if it exists). We continue this selection process as long as possible – the process finishes after a finite, say m, number of steps. It follows from the construction directly that

$$\bigcup_{B \in \mathcal{B}} B \subset \bigcup_{i=1}^{m} 3B_i.$$

Importantly, the balls B_i, $i = 1, \dots, m$, are disjoint. To see this, suppose that $B_{i_1} \cap B_{i_2} \neq \emptyset$ for some $1 \le i_1 < i_2 \le m$. As the radius of B_{i_1} is also larger than or equal to the radius of B_{i_2}, we must have (by triangle inequality) that $B_{i_2} \subset 3B_{i_1}$. But this is a contradiction with the selection process.

□

10.0.2 Remark. If $f \in L^1$ is non-trivial ($f \neq 0$ on a set of positive measure), then $Mf \notin L^1$. Indeed, in this case in some ball $B_R = B(0, R)$ we must have

$$\int_{B_R} |f| \gtrsim 1.$$

If $|x| > R$, then $B_R \subset B(x, 2|x|)$, and so

$$Mf(x) \ge \frac{1}{|B(x, 2|x|)|} \int_{B(x,2|x|)} |f| \gtrsim \frac{1}{|x|^n}.$$

Notice that

$$\int_{\mathbb{R}^d \setminus B(0,R)} |x|^{-n} \, \mathrm{d}x = \sum_{k=0}^{\infty} \int_{2^k R \le |x| < 2^{k+1} R} |x|^{-n} \, \mathrm{d}x \gtrsim \sum_{k=0}^{\infty} 1 = \infty.$$

Despite the previous remark, we do have the following result. It is typical in analysis that an operator does not map L^1 to L^1 but does map L^1 to $L^{1,\infty}$.

10.0.3 Theorem. *We have that* $M \colon L^1(\mathbb{R}^d) \to L^{1,\infty}(\mathbb{R}^d)$ *boundedly – i.e.,*

$$\|Mf\|_{L^{1,\infty}} \lesssim \|f\|_1.$$

Proof. Fix $f \in L^1$ and $\lambda > 0$. Define

$$\Omega_\lambda := \{x \in \mathbb{R}^d \colon Mf(x) > \lambda\}.$$

Let $K \subset \Omega_\lambda$ be an arbitrary compact set, and for every $x \in K$ choose (using the fact that $Mf(x) > \lambda$) a radius $r_x > 0$ and the related ball $U_x = B(x, r_x)$ so that

$$\frac{1}{|U_x|} \int_{U_x} |f| > \lambda.$$

As $\{U_x \colon x \in K\}$ is an open cover of K, we can use compactness to choose a finite subfamily $U_{x_1}, \dots, U_{x_N}$ so that

$$K \subset \bigcup_{j=1}^{N} U_{x_j}.$$

By the basic covering theorem choose disjoint $B_1, \dots, B_m \in \{U_{x_j} : j = 1, \dots, N\}$ so that

$$K \subset \bigcup_{j=1}^{N} U_{x_j} \subset \bigcup_{i=1}^{m} 3B_i.$$

We now get

$$|K| \le \sum_{i=1}^{m} |3B_i| \lesssim \sum_{i=1}^{m} |B_i| \le \frac{1}{\lambda} \sum_{i=1}^{m} \int_{B_i} |f| \le \frac{1}{\lambda} \int_{\mathbb{R}^d} |f|.$$

As $K \subset \Omega_\lambda$ was an arbitrary compact subset, the same inequality holds with $|K|$ replaced by $|\Omega_\lambda|$, and we are done. □

10.0.4 Corollary. *For all $1 < p < \infty$ and $f \in L^p$ we have*

$$\|Mf\|_p \lesssim \|f\|_p.$$

Proof. As we have $\|Mf\|_{L^{1,\infty}} \lesssim \|f\|_1$ and the trivial estimate $\|Mf\|_\infty \le \|f\|_\infty$, the claim follows from Marcinkiewicz interpolation theorem. □

10.1 General measures

Can we develop a similar theory with the Lebesgue measure replaced with a general locally finite Borel measure on $\mathbb{R}^d$? The main difficulty is the covering theorem we used as for a general measure μ we do not have any estimate linking $\mu(3B)$ to $\mu(B)$. It is possible to do this, but the theory only works for the centered maximal function and requires the much harder Besicovitch covering theorem. Due to time constraints, we do not prove these results in this course.

10.2 Lebesgue's differentiation theorem

10.2.1 Theorem (Lebesgue's differentiation theorem)**.** *For $f \in L^1_{\text{loc}}$ we have*

$$\lim_{r\to 0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y) - f(x)| \, \mathrm{d}y = 0$$

for almost every $x \in \mathbb{R}^d$. In particular, we have

$$\lim_{r\to 0} \frac{1}{|B(x,r)|} \int_{B(x,r)} f(y) \, \mathrm{d}y = f(x)$$

for almost every $x \in \mathbb{R}^d$.

Proof. The latter claim follow from the first as

$$\frac{1}{|B(x,r)|} \int_{B(x,r)} f(y) \, \mathrm{d}y - f(x) = \frac{1}{|B(x,r)|} \int_{B(x,r)} [f(y) - f(x)] \, \mathrm{d}y,$$

and so it is enough to prove the first claim.

This is a local claim, so we can assume without loss of generality that $f \in L^1$ (enough to prove that the claim holds for every k and for a.e. $x \in B(0,k)$ – with a fixed k we can replace f by $f1_{B(0,2k)} \in L^1$).

There is a standard protocol to show almost everywhere convergence for integrable functions. It involves the following two steps: 1) show convergence in some appropriate dense subset; 2) prove the boundedness of the relevant maximal operator (depending on the problem at hand). In this case, the relevant maximal function is Mf, and we already know Theorem 10.0.3 – this gives us 2). But 1) is also clear, as the claim is obvious for continuous functions (which are dense). We now show how the standard protocol pieces these two facts together.

Let

$$\sigma_r f(x) = \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y) - f(x)| \, \mathrm{d}y.$$

It is enough to show that

$$|\{x\colon \limsup_{r\to 0} \sigma_r f(x) > 0\}| = 0.$$

We fix an arbitrary $\lambda > 0$ and show that

$$|\{x\colon \limsup_{r\to 0} \sigma_r f(x) > \lambda\}| = 0,$$

which is enough. Let $\epsilon > 0$. Choose $g \in C_c$ so that

$$\|f - g\|_1 < \epsilon.$$

We know that because g is continuous we have

$$\lim_{r\to 0} \sigma_r g(x) = 0$$

for every $x \in \mathbb{R}^d$. Estimating

$$\sigma_r f(x) \le \sigma_r (f-g)(x) + \sigma_r g(x)$$

we see that

$$\limsup_{r\to 0} \sigma_r f(x) \le \sup_{r>0} \sigma_r (f-g)(x) \le M(f-g)(x) + |f(x) - g(x)|.$$

Therefore, we have by Theorem 10.0.3 that

$$\begin{aligned} &\Big|\Big\{x\colon \limsup_{r\to 0} \sigma_r f(x) > \lambda\Big\}\Big| \\ &\le \Big|\Big\{x\colon M(f-g)(x) > \frac{\lambda}{2}\Big\}\Big| + \Big|\Big\{x\colon |f(x) - g(x)| > \frac{\lambda}{2}\Big\}\Big| \\ &\le \frac{2}{\lambda}\big(\|M(f-g)\|_{L^{1,\infty}} + \|f-g\|_{L^{1,\infty}}\big) \lesssim \frac{1}{\lambda}\|f-g\|_1 < \frac{\epsilon}{\lambda}. \end{aligned}$$

This ends the proof. □

10.2.2 Remark. Notice that Lebesgue's differentiation theorem implies that $|f(x)| \le Mf(x)$ for almost every x.

We present two immediate but important corollaries.

10.2.3 Corollary. *Let $f \in L^1([a,b])$ and define*

$$F(x) = \int_a^x f(y)\,\mathrm{d}y, \qquad x \in [a,b].$$

For almost every $x \in [a,b]$ we have

$$F'(x) = f(x).$$

Proof. Suppose $h > 0$. We have

$$\Big|\frac{F(x+h) - F(x)}{h} - f(x)\Big| \le \frac{1}{h}\int_x^{x+h} |f(y) - f(x)|\,\mathrm{d}y \le \frac{2}{2h}\int_{x-h}^{x+h} |f(y) - f(x)|\,\mathrm{d}y,$$

which goes, for almost every x, to 0 as $h \to 0+$ by Lebesgue's differentiation theorem. We can control the limit $\lim_{h\to 0-}$ similarly, and then the claim follows. □

10.2.4 Corollary. *Let $E \subset \mathbb{R}^d$ be measurable. Then for a.e. $x \in E$ we have*

$$\lim_{r\to 0} \frac{|E \cap B(x,r)|}{|B(x,r)|} = 1$$

and for a.e. $x \in E^c$ we have

$$\lim_{r\to 0} \frac{|E \cap B(x,r)|}{|B(x,r)|} = 0.$$

Proof. Lebesgue's differentiation theorem applied to $1_E \in L^1_{\text{loc}}$ gives that

$$\lim_{r \to 0} \frac{|E \cap B(x,r)|}{|B(x,r)|} = \lim_{r \to 0} \frac{1}{|B(x,r)|} \int_{B(x,r)} 1_E(y)\,\mathrm{d}y = 1_E(x)$$

for almost every $x \in \mathbb{R}^d$. □

10.3 Fundamental theorem of calculus

We will prove a characterization for those $f\colon [a,b] \to \mathbb{R}$ for which $f'(x)$ exists almost everywhere, $f' \in L^1([a,b])$ and we have

$$f(x) - f(a) = \int_a^x f'.$$

This is the fundamental theorem of calculus, in optimal generality, for the Lebesgue integral. To prove the result, we use the Radon–Nikodym theorem.

10.3.1 Definition. A function $f\colon [a,b] \to \mathbb{R}$ is absolutely continuous (AC) on $[a,b]$ if given $\epsilon > 0$ there is $\delta > 0$ so that

$$\sum_{j=1}^{k} |f(b_j) - f(a_j)| < \epsilon$$

whenever $(a_1, b_1), \ldots, (a_k, b_k) \subset [a,b]$ are disjoint and satisfy

$$\sum_{j=1}^{k} (b_j - a_j) < \delta.$$

In the exercises we proved that if $g\colon [a,b] \to \dot{\mathbb{R}}$ is integrable and we define

$$F(x) := \int_a^x g(y)\,\mathrm{d}y,$$

then F is absolutely continuous. We have also proved that $F'(x) = g(x)$ almost everywhere.

So if given f we wish to write

$$f(x) = f(a) + \int_a^x f'(y)\,\mathrm{d}y,$$

then f has to be absolutely continuous. The question is whether or not this necessary condition is also sufficient. It turns out that it is.

10.3.2 Theorem. *Let $f\colon [a,b] \to \mathbb{R}$. Then the following conditions are equivalent.*

1. *The function f is absolutely continuous.*
2. *We have that $f'(x)$ exists almost everywhere, $f' \in L^1([a,b])$ and we have*

$$f(x) - f(a) = \int_a^x f'(y)\,\mathrm{d}y.$$

We first prove this result for non-decreasing functions – i.e. $f(x) \geq f(y)$ if $x \geq y$.

10.3.3 Theorem. *Let $f\colon [a,b] \to \mathbb{R}$ be continuous and non-decreasing. Then the following condition are equivalent.*

1. *f is absolutely continuous.*
2. *f maps sets of measure to sets of measure zero.*
3. *$f'(x)$ exists almost everywhere, $f' \in L^1([a,b])$ and we have*

$$f(x) - f(a) = \int_a^x f'(y)\,\mathrm{d}y.$$

Proof. We first prove that (1) implies (2), so assume f is AC and let $E \subset \mathbb{R}$ with $|E| = 0$. Without loss of generality we may assume $a, b \notin E$. Choose $\epsilon > 0$ and associate $\delta > 0$ to f as in the definition of absolute continuity. There is an open set V with $E \subset V \subset [a,b]$ so that $|V| < \delta$. An open set in $\mathbb{R}$ is a disjoint union of open intervals – $V = \bigcup_i (a_i, b_i)$. We then have

$$\sum_i (b_i - a_i) = |V| < \delta,$$

and so

$$\sum_i (f(b_i) - f(a_i)) < \epsilon.$$

(This is not a finite sum necessarily and the definition of AC involves finite sums – this is not a problem as the definition implies that this holds for all partial sums and thus for the full sum.) It follows that

$$|fE| \le |fV| \le \sum_i (f(b_i) - f(a_i)) < \epsilon,$$

and as this holds for every $\epsilon > 0$ we must have $|fE| = 0$ as desired.

Assume then that (2) holds. For technical reasons that become clear later we want an injective function and for this reason define

$$g(x) := x + f(x), \qquad x \in [a,b].$$

If $y > x$ then $g(y) = y + f(y) \ge y + f(x) > x + f(x) = g(x)$ so g is injective. A short arguments shows that (2) holds also for g. Suppose now $E \in \mathrm{Leb}(\mathbb{R})$, $E \subset [a,b]$ By Lemma 5.0.10 we have that $E = F \cup N$, where F is $\mathcal{F}_\sigma$ and $|N| = 0$. Thus F is a union of compact sets and so is gF by the continuity of g. This proves that $gF \in \mathrm{Leb}(\mathbb{R})$. As $|gN| = 0$ we also have $gN \in \mathrm{Leb}(\mathbb{R})$ – this proves that $gE = gF \cup gN \in \mathrm{Leb}(\mathbb{R})$. This, combined with the injectivity of g, shows that

$$\mu(E) := |gE|, \qquad E \in \mathrm{Leb}(\mathbb{R}),\ E \subset [a,b],$$

is a finite measure. Indeed, given $E_1, E_2, \ldots \in \mathrm{Leb}(\mathbb{R})$ the images $gE_1, gE_2, \ldots \in \mathrm{Leb}(\mathbb{R})$ are also disjoint, and the countable additivity of μ follows from that of the Lebesgue measure.

Notice that (b) implies that $\mu \ll m$ – where m is the Lebesgue measure on $[a,b]$ – and so by the Radon–Nikodym theorem there exists $h \in L^1([a,b]) = L^1(m)$, $h \ge 0$, with

$$|gE| = \mu(E) = \int_E h(x)\,\mathrm{d}x, \qquad E \in \mathrm{Leb}(\mathbb{R}),\ E \subset [a,b].$$

If $E = [a,x]$, then $gE = [g(a), g(x)]$, and we get

$$g(x) - g(a) = |gE| = \int_a^x h(y)\,\mathrm{d}y.$$

It follows that
$$f(x) - f(a) = g(x) - g(a) - (x-a) = \int_a^x [h(y) - 1]\,\mathrm{d}y.$$
Now we get from Corollary 10.2.3 that $f'(x) = h(x) - 1$ a.e. – this gives (3).

We already know that (3) implies (1) so we are done. ☐

The trick to go from Theorem 10.3.3 to Theorem 10.3.2 is to show that if f is AC then $f = f_1 - f_2$, where f_1, f_2 are AC **and** non-decreasing. Notice that this directly gives Theorem 10.3.2. To do this, we associate to f its total variation function
$$F(x) := \sup \sum_{i=1}^{n} |f(t_i) - f(t_{i-1})|, \qquad x \in [a,b],$$
where the supremum is taken over all N and over all choices of t_i with
$$a = t_0 < t_1 < \cdots < t_N = x.$$
Write
$$f = \frac{1}{2}(F + f) - \frac{1}{2}(F - f).$$
The fact that $F + f$ and $F - f$ are non-decreasing AC functions is the content of the next lemma.

10.3.4 Lemma. *If $f\colon [a,b] \to \mathbb{R}$ is AC then F, $F + f$ and $F - f$ are non-decreasing AC functions on $[a,b]$.*

Proof. Let $a = t_0 < t_1 < \cdots < t_N = x$ and $x < y \le b$. Then we have
$$F(y) \ge |f(y) - f(x)| + \sum_{i=1}^{N} |f(t_i) - f(t_{i-1})|,$$
and so also $F(y) \ge |f(y) - f(x)| + F(x)$. In particular, we have
$$F(y) - f(y) \ge F(x) - f(x) \qquad \text{and} \qquad F(y) + f(y) \ge F(x) + f(x).$$
We have showed that $F, F - f$ and $F + f$ are non-decreasing. To show that they are all AC, it suffices to show that F is AC.

Choose $\epsilon > 0$ and associate $\delta > 0$ to f as in the definition of absolute continuity. Let $(a_1, b_1), \ldots, (a_k, b_k) \subset [a,b]$ be disjoint and satisfy
$$\sum_{j=1}^{k} (b_j - a_j) < \delta.$$
Considering one interval (a_j, b_j) we see that
$$F(b_j) - F(a_j) = \sup \sum_{i=1}^{n} |f(t_i) - f(t_{i-1})|,$$
where the supremum is over all n and over all $a_j = t_0 < t_1 < \cdots < t_n = b_j$. Using this choose $a_j = t_{0,j} < t_{1,j} < \cdots < t_{n_j,j} = b_j$ so that
$$F(b_j) - F(a_j) \le \sum_{i=1}^{n_j} |f(t_{i,j}) - f(t_{i-1,j})| + \frac{\epsilon}{2^j}.$$

Thus, we have

$$\sum_{j=1}^{k} |F(b_j) - F(a_j)| \leq \sum_{j=1}^{k} \sum_{i=1}^{n_j} |f(t_{i,j}) - f(t_{i-1,j})| + \sum_{j=1}^{\infty} \frac{\epsilon}{2^j} < 2\epsilon.$$

Here we used that the intervals $(t_{i-1,j}, t_{i,j})$ are disjoint and

$$\sum_{j=1}^{k} \sum_{i=1}^{n_j} (t_{i,j} - t_{i-1,j}) = \sum_{j=1}^{k} (b_j - a_j) < \delta.$$

We are done showing that F is AC and so the lemma is proved. □

We have established Theorem 10.3.2 – the fundamental theorem of calculus for AC functions.

10.3.5 Example. Without any reference to Riemann integration, we will calculate

$$\lim_{j\to\infty} j \int_0^1 x^{-\frac{3}{2}} \sin \frac{x}{j} \, \mathrm{d}x.$$

Define $f_j(x) = jx^{-\frac{3}{2}} \sin \frac{x}{j} \cdot 1_{[0,1]}(x)$. Since $|\sin x| \leq x$ for all $x \geq 0$, we have

$$|f_j(x)| \leq x^{-\frac{1}{2}} \cdot 1_{[0,1]}(x) =: g(x).$$

We need to rigorously show that g is integrable over $[0, 1]$ to be able to use DCT. Of course, we want to calculate the integral of g over $[0, 1]$ using FTC – it seems safer to use FTC on intervals $[1/j, 1]$, since on such intervals $G(x) := 2x^{\frac{1}{2}}$ with $G'(x) = g(x)$ is clearly C^1 and so AC. So we first use MCT (the functions are non-negative) to turn $\int g$ to $\lim_{j\to\infty} \int_{1/j}^{1} G'(x) \, \mathrm{d}x$ and then use FTC and take limits. That is, we have

$$\int g(x) \, \mathrm{d}x = \lim_{j\to\infty} \int_{1/j}^{1} G'(x) \, \mathrm{d}x = \lim_{j\to\infty} (G(1) - G(1/j)) = 2 < \infty$$

by MCT and FTC and so g is integrable.

We need to know the pointwise limits of f_j still. We have

$$f_j(x) = x^{-\frac{1}{2}} \cdot \frac{j}{x} \sin \frac{x}{j} \cdot 1_{[0,1]}(x) \to x^{-\frac{1}{2}} \cdot 1_{[0,1]}(x) =: f(x) = g(x).$$

Here we used that

$$\lim_{y\to 0+} \frac{\sin y}{y} = \lim_{y\to 0+} \frac{\cos y}{1} = \cos 0 = 1$$

by L'Hopital's rule.

Finally, by the dominated convergence theorem we have

$$\lim_{j\to\infty} \int f_j(x) \, \mathrm{d}x = \int g(x) \, \mathrm{d}x = 2.$$

10.4 Integration by parts

Integration by parts is of fundamental importance in modern analysis. We state a general version. Suppose now $f, g\colon [a, b] \to \mathbb{R}$ are absolutely continuous. The product fg is also absolutely continuous and for almost every $y \in [a, b]$ we have

$$(fg)'(y) = f'(y)g(y) + f(y)g'(y).$$

Integrating this over $y \in [a, x]$, where $x \in [a, b]$, we get

$$f(x)g(x) - f(a)g(a) = \int_a^x (fg)' = \int_a^x f'g + \int_a^x fg'.$$

Written in a different order we arrive at the integration by parts formula

$$\int_a^x fg' = [f(x)g(x) - f(a)g(a)] - \int_a^x f'g, \qquad x \in [a, b].$$

Appendix A

Elementary construction of the Lebesgue measure

This is an optional appendix, which gives an alternative and elementary construction of the Lebesgue measure. We only use some basic properties from Section 2, so mainly only that section is required to read this, apart from a few facts regarding dyadic cubes.

A.0.1 Definition. A rectangle on $\mathbb{R}^d$ (with sides parallel to the coordinate axes) is a set R of the form

$$R = I_1 \times \cdots \times I_d = \prod_{i=1}^{d} I_i,$$

where I_i is an interval with endpoints $-\infty < a_i < b_i < \infty$. We define

$$\mathrm{vol}(R) := \prod_{i=1}^{d} (b_i - a_i).$$

A union of rectangles is called almost disjoint if the interiors of the rectangles are disjoint. We give a proof of the following obvious result which is based only on the above algebraic definition of 'volume'.

A.0.2 Lemma. *Suppose R is a closed rectangle and it is an almost disjoint union of closed rectangles $R = \bigcup_{k=1}^{N} R_k$. Then*

$$\mathrm{vol}(R) = \sum_{k=1}^{N} \mathrm{vol}(R_k).$$

Proof. Suppose first that $R = \bigcup_{k=1}^{N} R_k$. The idea is to extend indefinately all of the sides of the rectangles $R_1, \dots, R_N$ – this finer grid yields finitely many closed rectangles $T_1, \dots, T_M$ and a partition $J_1, \dots, J_N$ of the integers $1, \dots, M$ so that all of the unions

$$R = \bigcup_{j=1}^{M} T_j \qquad \text{and} \qquad R_k = \bigcup_{j \in J_k} T_j,\ k = 1, \dots, N,$$

are almost disjoint. The identity

$$\mathrm{vol}(R) = \sum_{j=1}^{M} \mathrm{vol}(T_j)$$

can now be easily proved since the new grid partitions the sides of R (write this down with an example on $\mathbb{R}^2$). The analogous identity holds for $R_1, \ldots, R_N$ and so

$$\operatorname{vol}(R) = \sum_{j=1}^{M} \operatorname{vol}(T_j) = \sum_{k=1}^{N} \sum_{j \in J_k} \operatorname{vol}(T_j) = \sum_{k=1}^{N} \operatorname{vol}(R_k).$$

□

A modification of the argument also gives the following.

A.0.3 Lemma. *If $R, R_1, \ldots, R_N$ are closed rectangles and $R \subset \bigcup_{k=1}^{N} R_k$, then*

$$\operatorname{vol}(R) \leq \sum_{k=1}^{N} \operatorname{vol}(R_k).$$

We will use these results later to prove that the Lebesgue measure of a rectangle agrees with the above natural volume.

A.0.4 Definition. The Lebesgue outer measure $m_d(A) = |A|$ of a set $A \subset \mathbb{R}^d$ is

$$m_d(A) = |A| := \inf \sum_i \operatorname{vol}(R_i) \in [0, \infty],$$

where the infimum is taken over all countable coverings $A \subset \bigcup_i R_i$ by closed rectangles.

Recall Lemma 2.2.9. The above construction corresponds with the choices $\mathcal{S} := \{R \subset \mathbb{R}^d \colon R \text{ closed rectangle}\}$ and $h(R) := \operatorname{vol}(R)$. In particular, m_d is an outer measure.

A.0.5 Lemma. *We have*

$$\operatorname{vol}(R) = |R|$$

whenever R is a closed rectangle.

Proof. If R is a closed rectangle, it is its own cover and by definition $|R| \leq \operatorname{vol}(R)$. Suppose then $R \subset \bigcup_i R_i$, where each R_i is a closed rectangle. If we show that

$$\operatorname{vol}(R) \leq \sum_i \operatorname{vol}(R_i) \tag{A.0.6}$$

then the claim

$$\operatorname{vol}(R) \leq |R|$$

follows by taking the infimum over all covers. This is Lemma A.0.3 except for the fact that the cover is not necessarily finite. We will use compactness to reduce to finite covers – for this, we need to change the cover to be open. Let $\epsilon > 0$ and choose a slightly larger open rectangle $R_j \subset T_j$ so that $\operatorname{vol}(T_j) \leq (1+\epsilon)\operatorname{vol}(R_j)$. Use the compactness of R to choose a finite cover $R \subset \bigcup_{j=1}^{N} T_j$. Now, by Lemma A.0.3 we have

$$\operatorname{vol}(R) \leq \sum_{j=1}^{N} \operatorname{vol}(\overline{T_j}) = \sum_{j=1}^{N} \operatorname{vol}(T_j) \leq (1+\epsilon) \sum_{j=1}^{N} \operatorname{vol}(R_j) \leq (1+\epsilon) \sum_{j=1}^{\infty} \operatorname{vol}(R_j).$$

We are done. □

A.0.7 Remark. The same is true for all rectangles (closed or not).

Because of the above lemma, we may stop using the vol notation. The Lebesgue measurable sets are denoted $\mathrm{Leb}(\mathbb{R}^d) := \mathcal{M}_{m_d}(\mathbb{R}^d)$. By Theorem 2.2.7 we have that $\mathrm{Leb}(\mathbb{R}^d)$ is a σ-algebra and $m_d|\,\mathrm{Leb}(\mathbb{R}^d)$ is a measure.

A.0.8 Lemma. *All rectangles R are Lebesgue measurable.*

Proof. Let R be a closed rectangle and let $A \subset \mathbb{R}^d$ be arbitrary. For an arbitrary $\epsilon > 0$ choose closed rectangles $R_1, R_2, \dots$ so that $A \subset \bigcup_i R_i$ and $\sum_i |R_i| \leq |A| + \epsilon$ (remember that we know $\mathrm{vol}(R_i) = |R_i|$). We know that $R \cap R_i = T_i$, where T_i is a rectangle (or an emptyset). We also know that $R_i \setminus R = \bigcup_k U_i^k$ is a **finite** union of almost disjoint rectangles (draw a picture). By Lemma A.0.2 (and because volume and Lebesgue measure agree for rectangles) we have

$$|R_i| = |T_i| + \sum_k |U_i^k|,$$

and by summing over i and using subadditivity we get

$$\begin{aligned}
|A| + \epsilon \geq \sum_i |R_i| &= \sum_i |T_i| + \sum_{i,k} |U_i^k| \\
&\geq \Big|\bigcup_i T_i\Big| + \Big|\bigcup_{i,k} U_i^k\Big| \\
&= \Big|R \cap \bigcup_i R_i\Big| + \Big|\bigcup_i (R_i \setminus R)\Big| \geq |A \cap R| + |A \setminus R|.
\end{aligned}$$

Letting $\epsilon \to 0$ shows that R is measurable. To show that e.g. an open rectangle is measurable, write it as an increasing union of closed rectangles. □

A.0.9 Corollary. *We have* $\mathrm{Bor}(\mathbb{R}^d) \subset \mathrm{Leb}(\mathbb{R}^d)$.

Proof. It suffices to show that all open sets $V \subset \mathbb{R}^d$ are measurable. We use the elementary fact, Lemma 6.0.1, which implies that V can be written as a countable union of cubes. As $\mathrm{Leb}(\mathbb{R}^d)$ is a σ-algebra and cubes (even rectangles) are measurable, it follows that $V \in \mathrm{Leb}(\mathbb{R}^d)$ as desired. □

This fact that $\mathrm{Leb}(\mathbb{R}^d)$ is large – namely it contais all Borel sets – is given for free by our other construction of the Lebesgue measure. The other construction also gives for free that Lebesgue measure satisfies various regularity properties. These can be proved independently as well, like we do now.

A.0.10 Lemma. *For an arbitrary $A \subset \mathbb{R}^d$ we have*

$$|A| = \inf\{|V| \colon V \text{ open and } A \subset V\}. \tag{A.0.11}$$

If A is measurable we have

$$|A| = \sup\{|K| \colon K \text{ compact and } K \subset A\}. \tag{A.0.12}$$

Proof. We first prove (A.0.11). Using the definition of the Lebesgue outer measure we cover A with closed rectangles R_i so that

$$\sum_i |R_i| \leq |A| + \epsilon.$$

Choose open rectangles $R_i \subset T_i$ so that

$$|T_i| \leq |R_i| + \frac{\epsilon}{2^i}.$$

For the open set $V = \bigcup_i T_i \supset A$ we see that

$$|V| \leq \sum_i |T_i| \leq \sum_i |R_i| + \epsilon \leq |A| + 2\epsilon.$$

This shows that

$$|A| \geq \inf\{|V| \colon V \text{ open and } A \subset V\}$$

and the reverse inequality is obvious by monotonicity. This proves (A.0.11).

We move on to prove (A.0.12). First, suppose A is bounded and measurable. Let $F \supset A$ be an arbitrary compact set containing A. We apply (A.0.11) to find, given $\epsilon > 0$, an open set $V \supset F \setminus A$ so that

$$|V| \leq |F \setminus A| + \epsilon.$$

Define $K := F \setminus V$, which is bounded (as F is) and closed (as the intersection $F \cap (\mathbb{R}^d \setminus V)$ of closed sets), and hence compact. We also have $K \subset A$ by construction. By the measurability of A we have

$$|F| = |F \cap A| + |F \setminus A| = |A| + |F \setminus A| \geq |A| + |V| - \epsilon.$$

Thus, we have

$$|A| - \epsilon \leq |F| - |V|.$$

As $F \subset K \cup V$ we have $|F| \leq |K| + |V|$ and so $|F| - |V| \leq |K|$. We have showed that the compact set $K \subset A$ satisfies

$$|A| - \epsilon \leq |K|,$$

and so

$$|A| \leq \sup\{|K| \colon K \text{ compact and } K \subset A\}.$$

The reverse inequality is trivial so we are done with the case that A is bounded.

In the general case we define the measurable and bounded sets $A_k = A \cap B(0, k)$. By Theorem 2.3.1 we have

$$|A| = \lim_{k \to \infty} |A_k|.$$

Suppose $|A| = \infty$. Choose compact $K_k \subset A_k \subset A$ so that $|K_k| \geq |A_k| - 1$. As $|A_k| \to \infty$ it follows that $|K_k| \to \infty$. This shows the claim in the case $|A| = \infty$.

Finally, suppose $|A| < \infty$ (but that A is unbounded). For $\epsilon > 0$ choose k so that

$$|A_k| \geq |A| - \epsilon.$$

Then choose a compact $K_k \subset A_k$ so that $|K_k| \geq |A_k| - \epsilon$. We then have $K_k \subset A$ and

$$|K_k| \geq |A_k| - \epsilon \geq |A| - 2\epsilon$$

ending the proof. □

Notice that (A.0.11) does not require that A is measurable. Therefore, for any set $A \subset \mathbb{R}^d$ and $\epsilon > 0$ we can find open $V \supset A$ so that $|V| \leq |A| + \epsilon$. This does not imply, without measurability, that $|V \setminus A| < \epsilon$. The following is a characterization of measurability using this condition.

A.0.13 Theorem. *A set $A \subset \mathbb{R}^d$ is Lebesgue measurable if and only if for every $\epsilon > 0$ there is an open $V \supset A$ so that $|V \setminus A| < \epsilon$.*

Proof. Suppose first that A is measurable. The case $|A| < \infty$ follows directly from (A.0.11) and measurability. Suppose that $|A| = \infty$ and write $A = \bigcup_k A_k$, $A_k := A \cap B(0,k)$. Let $\epsilon > 0$ and choose open $V_k \supset A_k$ so that $|V_k \setminus A_k| < \epsilon/2^k$. Now $V = \bigcup_k V_k \supset A$ is open and

$$|V \setminus A| \leq \sum_k |V_k \setminus A| \leq \sum_k |V_k \setminus A_k| \leq \sum_k \epsilon/2^k = \epsilon.$$

Suppose then that $A \subset \mathbb{R}^d$ is arbitrary but satisfies the condition. Let $\epsilon > 0$ and choose open $V \supset A$ so that $|V \setminus A| < \epsilon$. Let $E \subset \mathbb{R}^d$ be an arbitrary test set for measurability. We have by the measurability of V that

$$|E| + \epsilon = |E \cap V| + |E \setminus V| + \epsilon > |E \cap V| + |E \setminus V| + |V \setminus A|.$$

Since

$$E \setminus A = [E \setminus V] \cup [E \cap (V \setminus A)],$$

we have by monotonicity and subadditivity that

$$|E \setminus V| + |V \setminus A| \geq |E \setminus V| + |E \cap (V \setminus A)| \geq |E \setminus A|.$$

On the other hand, trivially by monotonicity $|E \cap V| \geq |E \cap A|$. Thus, we have proved that

$$|E| + \epsilon \geq |E \cap A| + |E \setminus A|.$$

As $\epsilon > 0$ was arbitrary this shows that A is measurable and ends the proof. □

We also formulate a modified version of the above result.

A.0.14 Theorem. *A set $A \subset \mathbb{R}^d$ is Lebesgue measurable if and only if for every $\epsilon > 0$ there is an open set $V \supset A$ and a closed set $F \subset A$ so that $|V \setminus F| < \epsilon$. If $|A| < \infty$ the set F can be chosen to be compact.*

Proof. If $A \subset \mathbb{R}^d$ satisfies the condition of the theorem, then it is clearly, by monotonicity, measurable by Theorem A.0.13.

Suppose now that $A \subset \mathbb{R}^d$ is Lebesgue measurable. We apply, given $\epsilon > 0$, Theorem A.0.13 to the measurable sets A and $\mathbb{R}^d \setminus A$ to find open sets $V \supset A$ and $G \supset \mathbb{R}^d \setminus A$ so that

$$|V \setminus A| < \frac{\epsilon}{2} \qquad \text{and} \qquad |G \setminus (\mathbb{R}^d \setminus A)| < \frac{\epsilon}{2}.$$

Define the closed set $F := \mathbb{R}^d \setminus G \subset A$. We have

$$|V \setminus F| = |V \setminus A| + |A \setminus F| = |V \setminus A| + |G \setminus (\mathbb{R}^d \setminus A)| < \epsilon.$$

It remains to show that F may be chosen compact if $|A| < \infty$. We use (A.0.12) to choose a compact $K \subset A$ so that

$$|A| < |K| + \frac{\epsilon}{2}.$$

Note that the above requires $|A| < \infty$. By measurability $|A \setminus K| = |A| - |K| < \epsilon/2$. Repeating the above argument with F replaced by K yields the claim. □

A.0.15 Corollary. *Suppose $A \subset \mathbb{R}^d$ is measurable. Then there exists a $\mathcal{F}_\sigma$-set $F \subset A$ and a $\mathcal{G}_\delta$-set $G \supset A$ so that*

$$|G \setminus A| = 0 = |A \setminus F|.$$

Proof. Simply choose open sets $V_k \supset A$ and closed sets $F_k \subset A$ so that $|V_k \setminus F_k| < 1/k$, and set $G = \bigcap_k V_k$ and $F = \bigcup_k F_k$. □

A.0.16 Corollary. *The completion of* $(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d), |\cdot|)$ *is* $(\mathbb{R}^d, \mathrm{Leb}(\mathbb{R}^d), |\cdot|)$.

Proof. Let $A \in \mathrm{Leb}(\mathbb{R}^d)$. Then by the above result there is a Borel set $F \subset A$ so that $|A \setminus F| = 0$. As $A \setminus F \in \mathrm{Leb}(\mathbb{R}^d)$ there is, again by the above result, a Borel set $G \supset A \setminus F$ so that $|G| = |G \setminus (A \setminus F)| = 0$. We conclude that

$$A = F \cup (A \setminus F),$$

where F is Borel and $A \setminus F \subset G$, where G is Borel and $|G| = 0$. This shows that $\mathrm{Leb}(\mathbb{R}^d)$ is the completion of $\mathrm{Bor}(\mathbb{R}^d)$, and we are done. □

A.0.17 Remark. Remember also that (A.0.11) does not require measurability. Thus, if $A \subset \mathbb{R}^d$ is arbitrary there are open sets $V_k \supset A$ so that $|V_k| \le |A| + 1/k$. Then the set $V = \bigcap_k V_k$ satisfies for every k that

$$|A| \le |V| \le |V_k| \le |A| + 1/k,$$

and so $V \supset A$ is a Borel set with $|V| = |A|$. This means that the Lebesgue outer measure is Borel regular: all Borel sets are measurable and given an arbitrary A there is a Borel set $B \supset A$ with $|A| = |B|$.

Appendix B

Non-measurable sets and other examples

For a nice measure like the Lebesgue measure, one could hope that $\mathrm{Leb}(\mathbb{R}^d)$ equals $\mathcal{P}(\mathbb{R}^d)$. This is not true. The following is the standard example of a non-Lebesgue measurable set. It is by Vitali. For $x \in \mathbb{R}$ define

$$E(x) = x + \mathbb{Q}$$

so that $E(x) = E(y)$ if and only if $x - y \in \mathbb{Q}$. From each set $E(x)$ we choose exactly one element that belongs to $[0, 1]$, and we call the collection of these elements $A \subset [0, 1]$. The sets $A + r$, $r \in \mathbb{Q}$, are disjoint, since if $(A + r) \cap (A + s) \neq \emptyset$ for some $r, s \in \mathbb{Q}$, then $a_1 + r = a_2 + s$ for some $a_1, a_2 \in A$. It follows that $a_1 - a_2 = s - r \in \mathbb{Q}$ and so $E(a_1) = E(a_2)$. But we chose exactly one element from each $E(x)$, so that we must have $a_1 = a_2$. But this means that $s = r$.

Aiming for a contradiction, assume that A is measurable. Then we have that

$$2 = |[0, 2]| \geq \Big| \bigcup_{n=1}^{\infty} \Big(A + \frac{1}{n}\Big) \Big| = \sum_{n=1}^{\infty} \Big| A + \frac{1}{n} \Big| = \sum_{n=1}^{\infty} |A|,$$

where we used the disjointness of the sets $A + \frac{1}{n} \subset [0, 2]$, the fact that A and so also each $A + \frac{1}{n}$ is measurable, countable additivity for measurable sets and translation invariance. We conclude that we must have $|A| = 0$.

On the other hand, we notice that

$$\mathbb{R} = \bigcup_{q \in \mathbb{Q}} (A + q).$$

This is because for each $x \in \mathbb{R}$ we find (by construction) $a \in A$ so that $a \in E(x)$. It follows that $x - a = r$ for some $r \in \mathbb{Q}$, that is, $x = a + r$. We now get the following contradiction:

$$0 = \sum_{q \in \mathbb{Q}} |A| = \sum_{q \in \mathbb{Q}} |A + q| = \Big| \bigcup_{q \in \mathbb{Q}} (A + q) \Big| = |\mathbb{R}| = \infty.$$

We conclude that A must be non-measurable.

B.0.1 Remark. In fact, a slight modification of the argument gives that whenever $E \subset \mathbb{R}$ is an arbitrary set with $|E| > 0$, then there is $A \subset E$ so that A is not measurable.

We continue with some other examples.

B.1 Cantor set

Let $I = [0,1]$ and $p_1, p_2, \dots$ be constants with $p_i \in (0,1)$. From the middle of I we remove an open interval $I_{1,1}$ of length p_1. This means that

$$I = J_{1,1} \cup I_{1,1} \cup J_{1,2},$$

where $J_{1,1}, J_{1,2}$ are open intervals of length $(1-p_1)/2$. We continue this iteratively as follows. Next, from the middle of $J_{1,k}$ we remove an open interval $I_{2,k}$ of length $\ell(I_{2,k}) = p_2\ell(J_{1,k}) = p_2(1-p_1)/2$. What is left can be written as

$$I \setminus (I_{1,1} \cup I_{2,1} \cup I_{2,2}) = J_{2,1} \cup J_{2,2} \cup J_{2,3} \cup J_{2,4},$$

where each $J_{2,k}$ is an open interval of length

$$\ell(J_{2,k}) = \frac{1-p_1}{2}\frac{1-p_2}{2}.$$

Notice that the total length of all of the intervals $J_{2,k}$, $k = 1, \dots, 2^2$, is

$$(1-p_1)(1-p_2).$$

We continue this and denote the remaining set as

$$C = C_{(p_1,p_2\dots)} = I \setminus \bigcup_{j=1}^{\infty}\bigcup_{k=1}^{2^{j-1}} I_{j,k} = \bigcap_{j=1}^{\infty}\bigcup_{k=1}^{2^j} J_{j,k}.$$

As the sets $C_j := \bigcup_{k=1}^{2^j} J_{j,k}$ satisfy $C_1 \supset C_2 \supset \cdots$ and $|C_1| < \infty$, we have by Theorem 2.3.1 that

$$|C| = \lim_{j\to\infty} |C_j| = \prod_{j=1}^{\infty}(1-p_i).$$

B.1.1 Remark. Given $\alpha \in [0,1)$ it is possible to choose the numbers p_i so that $|C_{(p_1,p_2\dots)}| = \alpha$.

Notice that the set C is compact and does not contain any intervals. Let $p = (p_1, p_2, \dots)$ and $q = (q_1, q_2, \dots)$. Denote the corresponding Cantor sets by C_p and C_q, and denote the corresponding intervals by $J_{j,k}(p)$ and $J_{j,k}(q)$. If $x \in C_p$ notice that there is a unique sequence $J_{1,k_1}(p) \supset J_{2,k_2}(p) \supset \cdots$ so that

$$\{x\} = \bigcap_{j=1}^{\infty} J_{j,k_j}(p).$$

We can define a homeomorphism $f\colon E_p \to E_q$ via the relation that if $\{x\} = \bigcap_{j=1}^{\infty} J_{j,k_j}(p)$ then

$$\{f(x)\} = \bigcap_{j=1}^{\infty} J_{j,k_j}(q).$$

A homeomorphism is a continuous bijection so that f^{-1} is also continuous. It is clear that f is a bijection. The continuity can be shown as follows. If $x, y \in C_p$ and

$$|x-y| < \delta_j := \min\{|I_{i,k}(p)|\colon i \le j\},$$

then x, y belong to the same interval $J_{j,k}(p)$, and so by construction $f(x), f(y) \in J_{j,k}(q)$. But as there 2^j of the disjoint intervals $J_{j,k}(q)$, we have

$$|f(x) - f(y)| \le |J_{j,k}(q)| \le 2^{-j}.$$

This shows that f is continuous and to show that f^{-1} is continuous is similar.

B.1.2 Example. Choose Cantor sets C and C' so that $|C| > 0$ and $|C'| = 0$. Let $f\colon C \to C'$ be a homeomorphism. Choose a set $E \subset C$ that is not Lebesgue measurable (which can be done as $|C| > 0$). Notice that $fE \subset C'$ necessarily satisfies $|fE| = 0$, and so it is Lebesgue measurable. We argue first that fE cannot be a Borel set. Aiming for a contradiction, suppose that fE is Borel. Then it follows that $E = f^{-1}(fE)$ is Borel – but this is a contradiction as E is not even Lebesgue measurable. So, first of all, this gives an example of a set $fE \in \mathrm{Leb}(\mathbb{R}) \setminus \mathrm{Bor}(\mathbb{R})$.

Notice also that this gives an example of a nice function f so that $f^{-1}F$ is not Lebesgue measurable even though $F := fE$ is Lebesgue measurable. This is relevant for the discussion about the definition of measurable functions.

Bibliography

[1] F. Di Plinio, A dyadic approach to Lebesgue theory, Lecture Notes, WUSTL, 2020.

[2] I. Holopainen, Real Analysis I, Lecture Notes, University of Helsinki, 2017.

[3] I. Holopainen, Measure and Integration, Lecture Notes, University of Helsinki, 2004.

[4] I. Holopainen, Modern Real Analysis, Lecture Notes, University of Helsinki, 2005.

[5] J. Kinnunen, Real Analysis, Lecture Notes, Aalto University, 2021.

[6] H. Martikainen, Real Analysis I, Lecture Notes, University of Helsinki, 2019.

[7] H. Martikainen, 4121: Introduction to Lebesgue Integration, Lecture Notes, WashU, 2022.

[8] H. Martikainen, 5051: Measure Theory and Functional Analysis I, Lecture Notes, WashU, 2022.

[9] W. Rudin, Real and complex analysis, McGraw-Hill, Singapore, 1987.

[10] E. Stein, R. Shakarchi, Real Analysis: Measure Theory, Integration, and Hilbert Spaces (Princeton Lectures in Analysis), Princeton University Press, Princeton, New Jersey, 2005.