Analysis IV:
Lebesgue Integration and Fourier Theory

Maria Colombo

These notes contain the material discussed in the "Analysis IV" course at EPFL during the academic year 2023/24.

Thanks to David Wiedemann, for typing down lecture notes from the course in A.Y. 2021/22, Toni Abi Aoun, Mathis Duguin, Giulia Mescolini and Matthieu Protais, for the edition of the current version, and Nicola De Nitti for the revision.

Contents

CHAPTER 1

Measure theory

In this chapter, we follow closely the content of [Tao16, Chapter 7].

1.1. Motivation

Studying integration in several variables, the general question we wish to answer is this:

Given some subset Ω of $\mathbb{R}^d$, and some real-valued function $f : \Omega \to \mathbb{R}$, is it possible to integrate f on Ω and therefore define $\int_\Omega f$?

In the case $d = 1$, we have already developed the notion of Riemann integral, which answers the above question when Ω is an interval $\Omega = [a, b] \subset \mathbb{R}$, and f is Riemann integrable. However, the class of Riemann integrable functions is rather unsatisfactorily small and the extension of this notion to higher dimensions is possible but requires quite a bit of effort. For such reasons, we must look beyond the Riemann integral and introduce the notion of the Lebesgue integral, which will be the central topic of the first two chapters of the course.

Before we turn to the details, we begin with an informal discussion. In order to understand how to compute an integral $\int_\Omega f$, we must first understand a more basic and fundamental question:

How does one compute the length, area, or volume of a subset $E \subset \mathbb{R}^d$?

This question is connected to that of integration, because if one integrates the function 1 on the set E, then one should obtain the length of E (if it is one-dimensional), the area of E (if it is two-dimensional), or the volume of E (if it is three-dimensional). To avoid splitting into cases depending on the dimension, we shall refer to the *measure* of E as either the length, area, volume, or hypervolume of E, depending on what Euclidean space $\mathbb{R}^d$ we are working in.

Ideally, to every subset E of $\mathbb{R}^d$ we would like to associate a nonnegative number $m(E)$, which will be the measure of E. We allow the possibility for $m(E)$ to be zero (that happens, for example, when E is just a single point or is the empty set) or for $m(E)$ to be infinite (e.g., if E is all of $\mathbb{R}^d$). This measure should obey certain reasonable properties, for instance

(i) (Empty set) $m(\varnothing) = 0$.
(ii) (Positivity) $0 \leq m(E) \leq +\infty \quad \forall E \subset \mathbb{R}^d$.
(iii) (Normalization) being $(0,1)^d := \{(x_1, \ldots, x_d) : 0 < x_i < 1\}$ the unit cube, $m((0,1)^d) = 1$.
(iv) (Additivity) $m(E \cup F) = m(E) + m(F)$ if E and F are disjoint.
(v) (Monotonicity) $m(E) \leq m(F)$ whenever $E \subseteq F$,
(vi) (Translation invariance) $m(x + E) = m(E)$ for any $x \in \mathbb{R}^d$ (i.e., if we shift E by the vector x the measure should be the same).

Remarkably, it turns out that such a measure does not exist; one cannot assign a non-negative number to every subset of $\mathbb{R}^d$ which has the above properties. This is quite a surprising fact, as it goes against one's intuitive concept of volume, but we will prove it later in these notes. An even more dramatic example of this failure of intuition is the Banach-Tarski paradox, in which a unit ball in $\mathbb{R}^3$ is decomposed into five pieces, and then the five pieces are reassembled via translations and rotations to form two complete and disjoint unit balls, thus violating any concept of conservation of volume; however we will not discuss this paradox here.

What such paradoxes mean is that it is impossible to find a reasonable way to assign a measure to every single subset of $\mathbb{R}^d$. However, we can rescue the situation by only measuring a certain class of sets in $\mathbb{R}^d$, that we will define *measurable sets*. These are the only sets E for which we will define the measure $m(E)$, and once one restricts one's attention to measurable sets, one recovers all the above properties again.

1.2. The goal: Lebesgue measure

Let $\mathbb{R}^d$ be a Euclidean space. Our goal in this chapter is to define a concept of measurable set, which will be a special category of subset of $\mathbb{R}^d$ and for every such measurable set $E \subset \mathbb{R}^d$, we will then define the Lebesgue measure $m(E)$ to be a certain number in $[0, \infty]$.
The concept of measurable set will obey the following properties:

(i) (Borel property) every open set and every closed set in $\mathbb{R}^d$ are measurable.
(ii) (Complementarity) if E is measurable, then $\mathbb{R}^d \backslash E$ is also measurable.
(iii) (Boolean algebra property) if $(E_j)_{j\in J}$ is any finite collection of measurable sets (with J finite), then the union $\bigcup_{j\in J} E_j$ and intersection $\bigcap_{j\in J} E_j$ are also measurable.
(iv) (σ-algebra property) if $(E_j)_{j\in J}$ is any countable collection of measurable sets (with J countable), then the union $\bigcup_{j\in J} E_j$ and intersection $\bigcap_{j\in J} E_j$ are also measurable.

Remark 1.1. Some of these properties are redundant: for instance, (iv) will imply (iii), and once one knows all open sets are measurable, (ii) will imply that all closed sets are measurable also.

To every measurable set E, we associate the Lebesgue measure $m(E)$ of E, which will obey the following properties:

(i) (Empty set) $m(\varnothing) = 0$.
(ii) (Positivity) $0 \leq m(E) \leq +\infty$ for every measurable set E.
(iii) (Monotonicity) if $E \subseteq F$, and E and F are both measurable, then $m(E) \leq m(F)$
(iv) (Finite sub-additivity) if $(E_j)_{j\in J}$ is a finite collection of measurable sets, then $m\left(\bigcup_{j\in J} E_j\right) \leq \sum_{j\in J} m(E_j)$.
(v) (Finite additivity) if $(E_j)_{j\in J}$ is a finite collection of *disjoint* measurable sets, then $m\left(\bigcup_{j\in J} E_j\right) = \sum_{j\in J} m(E_j)$.
(vi) (Countable sub-additivity) if $(E_j)_{j\in J}$ is a countable collection of measurable sets, then $m\left(\bigcup_{j\in J} E_j\right) \leq \sum_{j\in J} m(E_j)$.
(vii) (Countable additivity) if $(E_j)_{j\in J}$ is a countable collection of *disjoint* measurable sets, then $m\left(\bigcup_{j\in J} E_j\right) = \sum_{j\in J} m(E_j)$.
(viii) (Normalization) The unit cube $[0,1]^d = \left\{(x_1, \ldots, x_d) \in \mathbb{R}^d : 0 \leq x_j \leq 1 \text{ for all } 1 \leq j \leq d\right\}$ has measure $m\left([0,1]^d\right) = 1$.
(ix) (Translation invariance) If E is a measurable set, and $x \in \mathbb{R}^d$, then $x + E \coloneqq \{x + y : y \in E\}$ is also measurable, and $m(x + E) = m(E)$.

Remark 1.2. Many of these properties are redundant; for instance the countable additivity property can be used to deduce the finite additivity property, which in turn can be used to derive monotonicity (when combined with the positivity property). One can also obtain the subadditivity properties from the additivity ones.

Remark 1.3. Note that $m(E)$ can be $+\infty$, and so in particular some of the sums in the above properties may also equal $+\infty$ (and since everything is positive we will never have to deal with indeterminate forms such as $-\infty + +\infty$).

Our goal for this chapter can then be stated in the following:

THEOREM 1.1 (Existence of Lebesgue measure). *There exists a notion of a measurable set, and a way to assign a number $m(E)$ to every measurable subset $E \subseteq \mathbb{R}^d$, which satisfies all of the properties (i)-(ix).*

1.3. First attempt: Outer measure

Before we construct Lebesgue measure, we first discuss a somewhat naive approach to finding the measure of a set - namely, we try to cover the set by boxes, and then add up the volume of each box. This approach will almost work, giving us a concept called outer measure which can be applied to every set and obeys all of the properties (i)-(ix) except for the additivity properties (v), (vii). Later we will have to restrict the outer measure to a class of special sets (called *measurable sets*) to recover the additivity property.

We begin with the notion of an open box.

DEFINITION (Open box). An open box (or box for short) B in $\mathbb{R}^d$ is any set of the form

$$B=\prod_{i=1}^{d}(a_i,b_i) := \left\{(x_1,\ldots,x_d)\in\mathbb{R}^d : x_i\in(a_i,b_i) \text{ for all } 1\le i\le d\right\},$$

where $b_i \geq a_i$ are real numbers. We define the volume $\operatorname{vol}(B)$ of this box to be the number

$$\operatorname{vol}(B) := \prod_{i=1}^{d}(b_i-a_i)=(b_1-a_1)(b_2-a_2)\ldots(b_d-a_d).$$

Remark 1.4. The unit cube $(0,1)^d$ is a box, and has volume 1.
Remark 1.5. In one dimension $d=1$, boxes are the same as open intervals. One can easily check in general dimension that open boxes are indeed open.
Remark 1.6. Note that if we have $b_i=a_i$ for some i, then the box becomes empty, and has volume 0 , but we still consider this to be a box.

We of course expect the measure $m(B)$ of a box to be the same as the volume $\operatorname{vol}(B)$ of that box. This is a natural fact that will be proved below and it is in fact an inevitable consequence of the axioms (i)-(viii).

DEFINITION (Covering by boxes). Let $E \subseteq \mathbb{R}^d$ be a subset of $\mathbb{R}^d$. We say that a collection $(B_j)_{j\in J}$ of boxes cover E iff

$$E\subseteq\bigcup_{j\in J}B_j.$$

Suppose $E \subseteq \mathbb{R}^d$ can be covered by a finite or countable collection of boxes $(B_j)_{j\in J}$. If we wish E to be measurable, and if we wish to have a measure obeying the monotonicity and sub-additivity properties (iii), (iv), (vi) and if we wish $m(B_j)=\operatorname{vol}(B_j)$ for every box j, then we must have

$$m(E)\le m\left(\bigcup_{j\in J}B_j\right)\le\sum_{j\in J}m(B_j)=\sum_{j\in J}\operatorname{vol}(B_j)$$

We thus conclude

$$m(E)\le\inf\left\{\sum_{j\in J}\operatorname{vol}(B_j):(B_j)_{j\in J} \text{ covers } E; J \text{ at most countable}\right\}.$$

Inspired by this, we define

Definition (Outer measure). If E is a set, we define the outer measure $m^*(E)$ of E to be the quantity

$$m^*(E) := \inf \left\{ \sum_{j\in J} \operatorname{vol}(B_j) : (B_j)_{j\in J} \text{ covers } E; J \text{ at most countable} \right\}.$$

Remark 1.7. Since $\sum_{j=1}^{\infty} \operatorname{vol}(B_j)$ is non-negative, we know that $m^*(E) \geq 0$ for all E. However, it is quite possible that $m^*(E)$ equal $+\infty$.
Remark 1.8. We are allowing ourselves to use a countable number of boxes, because every subset of $\mathbb{R}^d$ has at least one countable cover by boxes; in fact $\mathbb{R}^d$ itself can be covered by countably many translates of the unit cube $(0,1)^d$.
Remark 1.9. The outer measure can be defined for every single set (not just the measurable ones), because we can take the infimum of any non-empty set.

The outer measure obeys several of the desired properties of a measure:

Lemma 1.2 (Properties of outer measure)**.** *The outer measure has the following properties:*

(i) (Empty set) The empty set $\varnothing$ has outer measure $m^*(\varnothing) = 0$.
(ii) (Positivity) We have $0 \leq m^*(E) \leq +\infty$ for every measurable set E.
(iii) (Monotonicity) If $A \subseteq B \subseteq \mathbb{R}^d$, then $m^*(A) \leq m^*(B)$.
(iv) (Translation invariance) If E is a subset of $\mathbb{R}^d$, and $x \in \mathbb{R}^d$, then $m^*(x+E) = m^*(E)$.
(v) (Countable sub-additivity) If $(A_j)_{j\in J}$ is a countable collection of subsets of $\mathbb{R}^d$, then $m^*\left(\bigcup_{j\in J} A_j\right) \leq \sum_{j\in J} m^*(A_j)$.
(vi) (Homogeneity) If E is a subset of $\mathbb{R}^d$, and $a \in \mathbb{R}$, then $m^*(aE) = |a|^d m^*(E)$.

Proof. (i) We can cover $\varnothing$ with $(0,\epsilon)^d$ for any $\epsilon > 0$. It means

$$m^*(\varnothing) \leq \operatorname{vol}\left((0,\epsilon)^d\right) = \epsilon^d.$$

We conclude that $m^*(\varnothing) = 0$ by the fact that ϵ is arbitrary.
(ii) Follows from the definition of the volume, as observed in Remark 1.7.
(iii) Follows from the fact that any cover of B is a cover of A.
(iv) Similarly, $(B_j)_{j\in J}$ is a cover for E if and only if $(x+B_j)_{j\in J}$ is a cover for $x+E$. Moreover, $\operatorname{vol}(x+B_j) = \operatorname{vol}(B_j)$. We then deduce the claim, because:

$$\begin{aligned} m^*(E) &= \inf \left\{ \sum_{j\in J} \operatorname{vol}(B_j) : (B_j)_{j\in J} \text{ covers } E \right\} \\ &= \inf \left\{ \sum_{j\in J} \operatorname{vol}(x+B_j) : (B_j)_{j\in J} \text{ covers } E \right\} \\ &= \inf \left\{ \sum_{j\in J} \operatorname{vol}\left(B'_j\right) : \left(B'_j\right)_{j\in J} \text{ covers } x+E \right\} \\ &= m^*(x+E). \end{aligned}$$

(v) For any $j \in \mathbb{N}$, by definition of the outer measure as an infimum, there exists $\left(B_i^j\right)_{i\in I_j}$ a countable cover of A_j by boxes such that

$$\sum_{i\in I_j} \operatorname{vol}(B_i^j) \leq m^*(A_j) + \frac{\epsilon}{2^j}.$$

Taking $\left(B_i^j\right)_{i\in I_j, j\in\mathbb{N}}$ as a cover for $\bigcup_{j\in J} A_j$, it follows

$$m^*\left(\bigcup_{j=1}^{\infty} A_j\right) \le \sum_{j\in\mathbb{N}}\sum_{i\in I_j} \mathrm{vol}(B_i^j) \le \sum_{j\in\mathbb{N}}\left(m^*(A_j)+\frac{\epsilon}{2^j}\right) = \epsilon + \sum_{j\in\mathbb{N}} m^*(A_j).$$

And we conclude by the fact that ϵ is arbitrary and we can take $\varepsilon \to 0$.

(vi) Assume that a is not 0, otherwise the claim follows from Remark 1.10. $(B_j)_{j\in J}$ is a cover for E if and only if $(aB_j)_{j\in J}$ is a cover for aE. Moreover, $\mathrm{vol}(aB_j) = |a|^d \,\mathrm{vol}(B_j)$. We then deduce the claim, because:

$$\begin{aligned}
|a|^d m^*(E) &= |a|^d \inf\left\{\sum_{j\in J} \mathrm{vol}\left(B_j\right) : \left(B_j\right)_{j\in J} \text{ covers } E\right\}\\
&= |a|^d \inf\left\{\frac{1}{|a|^d}\sum_{j\in J} \mathrm{vol}\left(aB_j\right) : \left(B_j\right)_{j\in J} \text{ covers } E\right\}\\
&= \inf\left\{\sum_{j\in J} \mathrm{vol}\left(B'_j\right) : \left(B'_j\right)_{j\in J} \text{ covers } aE\right\}\\
&= m^*(aE).
\end{aligned}$$

□

Remark 1.10. With a proof similar to (i), we can prove that sets containing only one point $x_0 \in \mathbb{R}^d$ have outer measure 0: $m^*(\{x_0\}) = 0 \quad \forall x_0 \in \mathbb{R}^d$. Indeed, we can consider coverings of the form $(x_0-\varepsilon, x_0+\varepsilon)^d \quad \forall \varepsilon > 0$.

The outer measure of a closed box is also what we expect:

Proposition 1.3 (Outer measure of closed box)**.** *For any closed box*

$$\bar{B} = \prod_{i=1}^{d} [a_i, b_i] \coloneqq \left\{(x_1,\dots,x_d) \in \mathbb{R}^d : x_i \in [a_i,b_i] \text{ for all } 1 \le i \le d\right\},$$

we have

$$m^*(\bar{B}) = \prod_{i=1}^{d} (b_i - a_i).$$

Example 1.1. $m^*([0,1]^d) = 1$.

Proof. Clearly, we can cover the closed box $\bar{B} = \prod_{i=1}^{d}[a_i,b_i]$ by the open boxes

$$\prod_{i=1}^{d}(a_i-\varepsilon, b_i+\varepsilon) \quad \forall\varepsilon > 0.$$

Thus we have

$$m^*(\bar{B}) \le \mathrm{vol}\left(\prod_{i=1}^{d}(a_i-\varepsilon, b_i+\varepsilon)\right) = \prod_{i=1}^{d}(b_i-a_i+2\varepsilon)$$

for every $\varepsilon > 0$. Taking limits as $\varepsilon \to 0$, we obtain

$$m^*(\bar{B}) \le \prod_{i=1}^{d}(b_i - a_i).$$

To finish the proof, we need to show that

$$m^*(\bar{B}) \geq \prod_{i=1}^{d} (b_i - a_i) \,.$$

By the definition of $m^*(\bar{B})$, it is equivalent to show that

$$\sum_{j\in J} \operatorname{vol}\left(B^{(j)}\right) \geq \prod_{i=1}^{d} (b_i - a_i)$$

whenever $\left(B^{(j)}\right)_{j\in J}$ is a finite or countable cover of $\bar{B}$. We use the notation with the superscript $B^{(j)}$ because we will need the subscripts to denote components.

Since $\bar{B}$ is closed and bounded, it is compact (by the Heine-Borel theorem), and in particular every open cover has a finite subcover. Thus to prove the above inequality for countable covers, it suffices to do it for finite covers.

To summarize, our goal is now to prove that

$$\sum_{j\in J} \operatorname{vol}\left(B^{(j)}\right) \geq \prod_{i=1}^{d} (b_i - a_i) \tag{1.1}$$

whenever $\left(B^{(j)}\right)_{j\in J}$ is a finite cover of $\prod_{j=1}^{d} [a_i, b_i]$.

To prove the inequality (1.1), we shall use induction on the dimension d. First we consider the base case $d = 1$. Here B is just a closed interval $B = [a, b]$, and each box $B^{(j)}$ is just an open interval $B^{(j)} = (a_j, b_j)$. We have to show that

$$\sum_{j\in J} (b_j - a_j) \geq (b - a)$$

To do this we use the Riemann integral. For each $j \in J$, let $\mathbb{1}_{(a_j,b_j)} : \mathbb{R} \to \mathbb{R}$ be the function such that $\mathbb{1}_{(a_j,b_j)}(x) = 1$ when $x \in (a_j, b_j)$ and $\mathbb{1}_{(a_j,b_j)}(x) = 0$ otherwise. Then we have that $\mathbb{1}_{(a_j,b_j)}$ is Riemann integrable (because it is piecewise constant, and compactly supported) and

$$\int_{-\infty}^{\infty} \mathbb{1}_{(a_j,b_j)} = b_j - a_j$$

Summing this over all $j \in J$, and interchanging the integral with the finite sum, we have

$$\int_{-\infty}^{\infty} \sum_{j\in J} \mathbb{1}_{(a_j,b_j)} = \sum_{j\in J} (b_j - a_j) \,.$$

But since the intervals (a_j, b_j) cover $[a, b]$, we have $\sum_{j\in J} \mathbb{1}_{(a_j,b_j)}(x) \geq 1$ for all $x \in [a, b]$. For all other values if x, we have $\sum_{j\in J} \mathbb{1}_{(a_j,b_j)}(x) \geq 0$. Thus

$$\int_{-\infty}^{\infty} \sum_{j\in J} \mathbb{1}_{(a_j,b_j)} \geq \int_{[a,b]} 1 = b - a$$

and the claim follows by combining this inequality with the previous equality. This proves (1.1) when $d = 1$.

Now assume inductively that $d > 1$, and we have already proven the inequality (1.1) for dimensions $d - 1$. We shall use a similar argument to the preceding one. Each box $B^{(j)}$ is now of the form

$$B^{(j)} = \prod_{i=1}^{d} \left(a_i^{(j)}, b_i^{(j)}\right)$$

We can write this as

$$B^{(j)} = A^{(j)} \times \left(a_d^{(j)}, b_d^{(j)}\right)$$

where $A^{(j)}$ is the $d-1$-dimensional box $A^{(j)} \coloneqq \prod_{i=1}^{d-1} \left(a_i^{(j)}, b_i^{(j)}\right)$. Note that

$$\mathrm{vol}\left(B^{(j)}\right) = \mathrm{vol}_{d-1}\left(A^{(j)}\right)\left(b_d^{(j)} - a_d^{(j)}\right)$$

where we have subscripted vol_{d-1} by $d-1$ to emphasize that this is $d-1$-dimensional volume being referred to here. We similarly write

$$\bar{B} = \bar{A} \times [a_d, b_d]$$

where $\bar{A} \coloneqq \prod_{i=1}^{d-1} [a_i, b_i]$, and again note that

$$\mathrm{vol}(B) = \mathrm{vol}_{d-1}(A)\,(b_d - a_d)\,,$$

where A is the interior of $\bar{A}$.

For each $j \in J$, let $f^{(j)}$ be the function such that $f^{(j)}(x_d) = \mathrm{vol}_{d-1}\left(A^{(j)}\right) \mathbb{1}_{\left(a_d^{(j)}, b_d^{(j)}\right)}(x_d)$. Then $f^{(j)}$ is Riemann integrable and

$$\int_{-\infty}^{\infty} f^{(j)} = \mathrm{vol}_{d-1}\left(A^{(j)}\right)\left(b_d^{(j)} - a_d^{(j)}\right) = \mathrm{vol}\left(B^{(j)}\right)$$

and hence

$$\sum_{j\in J} \mathrm{vol}\left(B^{(j)}\right) = \int_{-\infty}^{\infty} \sum_{j\in J} f^{(j)}. \tag{1.2}$$

Now let $x_d \in [a_d, b_d]$ and $(x_1, \ldots, x_{d-1}) \in A$. Then $(x_1, \ldots, x_d)$ lies in B, and hence lies in one of the $B^{(j)}$. Clearly we have $x_d \in \left(a_d^{(j)}, b_d^{(j)}\right)$, and $(x_1, \ldots, x_{d-1}) \in A^{(j)}$. In particular, we see that for each $x_d \in [a_d, b_d]$, the set

$$\left\{A^{(j)} : j \in J; x_d \in \left(a_d^{(j)}, b_d^{(j)}\right)\right\}$$

of $d-1$-dimensional boxes covers A. Applying the inductive hypothesis (1.1) at dimension $d-1$ we thus see that

$$\sum_{j\in J} f^{(j)}(x_d) = \sum_{j\in J : x_d \in \left(a_d^{(j)}, b_d^{(j)}\right)} \mathrm{vol}_{d-1}\left(A^{(j)}\right) \geq \mathrm{vol}_{d-1}(A).$$

Integrating this over $[a_d, b_d]$, we obtain

$$\int_{[a_d, b_d]} \sum_{j\in J} f^{(j)} \geq \mathrm{vol}_{d-1}(A)\,(b_d - a_d) = \mathrm{vol}(B)$$

and in particular

$$\int_{-\infty}^{\infty} \sum_{j\in J} f^{(j)} \geq \mathrm{vol}_{d-1}(A)\,(b_d - a_d) = \mathrm{vol}(B)$$

since $\sum_{j\in J} f^{(j)}$ is always non-negative. Combining this with (1.2) we obtain (1.1), and the induction is complete.

□

Once we obtain the measure of a closed box, the corresponding result for an open box is easy:

Corollary 1.4. *For any open box*

$$B = \prod_{i=1}^{d} (a_i, b_i) := \left\{ (x_1, \dots, x_d) \in \mathbb{R}^d : x_i \in (a_i, b_i) \text{ for all } 1 \le i \le d \right\},$$

we have

$$m^*(B) = \prod_{i=1}^{d} (b_i - a_i)$$

In particular, outer measure obeys the normalization property.

Proof. Note that we may assume that $b_i > a_i$ for all i, since if $b_i = a_i$ this follows from Lemma 1.2 (i). Now observe that

$$\prod_{i=1}^{d} [a_i + \varepsilon, b_i - \varepsilon] \subset \prod_{i=1}^{d} (a_i, b_i) \subset \prod_{i=1}^{d} [a_i, b_i]$$

for all $\varepsilon > 0$, assuming that ε is small enough that $b_i - \varepsilon > a_i + \varepsilon$ for all i. Applying Proposition 1.3 and Lemma 1.2 (iii) we obtain

$$\prod_{i=1}^{d} (b_i - a_i - 2\varepsilon) \le m^* \left(\prod_{i=1}^{d} (a_i, b_i) \right) \le \prod_{i=1}^{d} (b_i - a_i).$$

Sending $\varepsilon \to 0$ one obtains the result. □

We now compute some examples of outer measure on the real line $\mathbb{R}$.

Example 1.2. Let us compute the one-dimensional measure of $\mathbb{R}$. Since $(-R, R) \subset \mathbb{R}$ for all $R > 0$, we have

$$m^*(\mathbb{R}) \ge m^*((-R, R)) = 2R$$

by Corollary 1.4. Letting $R \to +\infty$ we thus see that $m^*(\mathbb{R}) = +\infty$.

Example 1.3. Now let us compute the one-dimensional measure of $\mathbb{Q}$. From Remark 1.10 we see that for each rational number $\mathbb{Q}$, the point $\{q\}$ has outer measure $m^*(\{q\}) = 0$. Since $\mathbb{Q}$ is clearly the union $\mathbb{Q} = \bigcup_{q \in \mathbb{Q}} \{q\}$ of all these rational points q, and $\mathbb{Q}$ is countable, we have

$$m^*(\mathbb{Q}) \le \sum_{q \in \mathbb{Q}} m^*(\{q\}) = \sum_{q \in \mathbb{Q}} 0 = 0$$

and so $m^*(\mathbb{Q})$ must equal zero. In fact, the same argument shows that every countable set has measure zero. (This, incidentally, gives another proof that the real numbers are uncountable.)

Remark 1.11. One consequence of the fact that $m^*(\mathbb{Q}) = 0$ is that given any $\varepsilon > 0$, it is possible to cover the rationals $\mathbb{Q}$ by a countable number of intervals whose total length is less than ε. For example, writing $\mathbb{Q} = (q_i)_{i \in \mathbb{N}}$, we can take an interval with length $\varepsilon/2^i$ around each $q_i \in \mathbb{Q}$, thus obtaining that the total length of the intervals is $\sum_{i \in \mathbb{N}} \varepsilon/2^i = \varepsilon$.

Example 1.4. Now let us compute the one-dimensional measure of the irrationals $\mathbb{R} \backslash \mathbb{Q}$. From finite sub-additivity we have

$$m^*(\mathbb{R}) \le m^*(\mathbb{R} \backslash \mathbb{Q}) + m^*(\mathbb{Q}).$$

Since $\mathbb{Q}$ has outer measure 0 , and $m^*(\mathbb{R})$ has outer measure $+\infty$, we thus see that the irrationals $\mathbb{R} \backslash \mathbb{Q}$ have outer measure $+\infty$. A similar argument shows that $[0, 1] \backslash \mathbb{Q}$, the irrationals in $[0, 1]$, have outer measure 1.

Example 1.5. By Proposition 1.3, the unit interval $[0, 1]$ in $\mathbb{R}$ has one-dimensional outer measure 1, but the unit interval $\{(x, 0) : 0 \leq x \leq 1\}$ in $\mathbb{R}^2$ has two-dimensional outer measure 0. Thus one-dimensional outer measure and two-dimensional outer measure are quite different. Note that the above remarks and countable additivity imply that the entire x-axis of $\mathbb{R}^2$ has two-dimensional outer measure 0, despite the fact that $\mathbb{R}$ has infinite one-dimensional measure.

1.4. Outer measure is not additive

In light of Lemma 1.2, it would seem now that all we need to do is to verify the additivity properties, and we have everything we need to have a usable measure. Unfortunately, these properties fail for outer measure, even in one dimension, as it can be proved with Vitali's construction.

Proposition 1.5 (Failure of countable additivity)**.** *There exists a countable collection $(A_j)_{j\in J}$ of disjoint subsets of $\mathbb{R}$, such that $m^*\left(\bigcup_{j\in J} A_j\right) \neq \sum_{j\in J} m^*(A_j)$.*

Proof. We shall need some notation. Let $\mathbb{Q}$ be the rationals, and $\mathbb{R}$ be the reals. We say that a set $A \subset \mathbb{R}$ is a coset of $\mathbb{Q}$ if it is of the form $A = x + \mathbb{Q}$ for some real number x. For instance, $\sqrt{2} + \mathbb{Q}$ is a coset of $\mathbb{R}$, as is $\mathbb{Q}$ itself, since $\mathbb{Q} = 0 + \mathbb{Q}$. Note that a coset A can correspond to several values of x; for instance $2 + \mathbb{Q}$ is exactly the same coset as $0 + \mathbb{Q}$. Also observe that it is not possible for two cosets to partially overlap; if $x + \mathbb{Q}$ intersects $y + \mathbb{Q}$ in even just a single point z, then $x - y$ must be rational (use the identity $x - y = (x - z) - (y - z)$), and thus $x + \mathbb{Q}$ and $y + \mathbb{Q}$ must be equal. So any two cosets are either identical or distinct.

We observe that every coset A of the rationals $\mathbb{Q}$ has a non-empty intersection with $[0, 1]$. Indeed, if A is a coset, then $A = x + \mathbb{Q}$ for some real number x. If we then pick a rational number q in $[-x, 1 - x]$ then we see that $x + q \in [0, 1]$, and thus $A \cap [0, 1]$ contains $x + q$.

Let $\mathbb{R}/\mathbb{Q}$ denote the set of all cosets of $\mathbb{Q}$; note that this is a set whose elements are themselves sets (of real numbers). For each coset A in $\mathbb{R}/\mathbb{Q}$, let us pick an element x_A of $A \cap [0, 1]$. (This requires us to make an infinite number of choices, and thus requires the axiom of choice.) Let E be the set of all such x_A, i.e., $E \coloneqq \{x_A : A \in \mathbb{R}/\mathbb{Q}\}$. Note that $E \subseteq [0, 1]$ by constrution.

Now consider the set

$$X = \bigcup_{q\in\mathbb{Q}\cap[-1,1]} (q + E) \tag{1.3}$$

Clearly this set is contained in $[-1, 2]$ (since $q+x \in [-1, 2]$ whenever $q \in [-1, 1]$ and $x \in E \subseteq [0, 1]$). We claim that this set contains the interval $[0, 1]$. Indeed, for any $y \in [0, 1]$, we know that y must belong to some coset A (for instance, it belongs to the coset $y + \mathbb{Q}$). But we also have x_A belonging to the same coset, and thus $y - x_A$ is equal to some rational q. Since y and x_A both live in $[0, 1]$, then q lives in $[-1, 1]$. Since $y = q + x_A$, we have $y \in q + E$, and hence $y \in X$ as desired.

We claim that

$$m^*(X) \neq \sum_{q\in\mathbb{Q}\cap[-1,1]} m^*(q + E)$$

which would prove the claim. To see why this is true, observe that since $[0, 1] \subseteq X \subseteq [-1, 2]$, that we have $1 \leq m^*(X) \leq 3$ by monotonicity and Proposition 1.3. For the right hand side, observe from translation invariance that

$$\sum_{q\in\mathbb{Q}\cap[-1,1]} m^*(q + E) = \sum_{q\in\mathbb{Q}\cap[-1,1]} m^*(E).$$

The set $\mathbb{Q} \cap [-1, 1]$ is countably infinite. Thus the right-hand side is either 0 (if $m^*(E) = 0$) or $+\infty$ (if $m^*(E) > 0$). Either way, it cannot be between 1 and 3, and the claim follows. □

The above proof used the axiom of choice. This turns out to be necessary; one can prove using some advanced techniques in mathematical logic that if one does not assume the axiom of choice, then it is possible to have a mathematical model where outer measure is countably additive.

One can refine the above argument, and show in fact that m^* is not finitely additive either:

Proposition 1.6 (Failure of finite additivity)**.** *There exists a finite collection $(A_j)_{j\in J}$ of disjoint subsets of $\mathbb{R}$, such that*

$$m^*\left(\bigcup_{j\in J} A_j\right) \neq \sum_{j\in J} m^*(A_j).$$

PROOF. This is accomplished by an indirect argument. Suppose for sake of contradiction that m^* was finitely additive. Let E and X be the sets introduced in Proposition 1.5. From countable sub-additivity and translation invariance we have

$$m^*(X) \leq \sum_{q\in\mathbb{Q}\cap[-1,1]} m^*(q+E) = \sum_{q\in\mathbb{Q}\cap[-1,1]} m^*(E).$$

Since we know that $1 \leq m^*(X) \leq 3$, we thus have $m^*(E) \neq 0$, since otherwise we would have $m^*(X) \leq 0$, a contradiction.

Since $m^*(E) \neq 0$, there exists a finite integer $n > 0$ such that $m^*(E) > 1/n$. Now let J be a finite subset of $\mathbb{Q}\cap[-1,1]$ of cardinality 3n. If m^* were finitely additive, then we would have

$$m^*\left(\bigcup_{q\in J} q+E\right) = \sum_{q\in J} m^*(q+E) = \sum_{q\in J} m^*(E) > 3n\frac{1}{n} = 3.$$

But we know that $\bigcup_{q\in J} q+E$ is a subset of X, which has outer measure at most 3. This contradicts monotonicity. Hence m^* cannot be finitely additive. □

Remark 1.12. The examples here are related to the Banach-Tarski paradox, which demonstrates (using the axiom of choice) that one can partition the unit ball in $\mathbb{R}^3$ into a finite number of pieces which, when rotated and translated, can be reassembled to form two complete unit balls! Of course, this partition involves non-measurable sets. We will not present this paradox here as it requires some group theory which is beyond the scope of the course.

1.5. Measurable sets

As we mentioned in the introduction to this chapter and rigorously proved in section 1.4, m^* cannot be countably or finite additive on all subsets of $\mathbb{R}^d$. We need to exclude pathological sets to recover finite and countable additivity. Fortunately, this can be done, thanks to a clever definition by Constantin Carathéodory (1873-1950):

DEFINITION (Lebesgue measurability). Let E be a subset of $\mathbb{R}^d$. We say that E is Lebesgue measurable, or measurable for short, iff we have the identity

$$m^*(A) = m^*(A\cap E) + m^*(A\backslash E)$$

for every subset A of $\mathbb{R}^d$. If E is measurable, we define the Lebesgue measure of E to be $m(E) = m^*(E)$; if E is not measurable, we leave $m(E)$ undefined.

Remark 1.13. In other words, E being measurable means that if we use the set E to divide up an arbitrary set A into two parts, we keep the additivity property.

The above definition is somewhat hard to work with, and in practice one does not verify a set is measurable directly from this definition. Instead, we will use this definition to prove various useful properties of measurable sets (Lemmas 1.7–1.14), and after that we will rely more or less exclusively on the properties in those lemmas, and no longer need to refer to the above definition.

We begin by showing that a large number of sets are indeed measurable. The empty set $E = \varnothing$ and the whole space $E = \mathbb{R}^d$ are clearly measurable:

$$m^*(A) = m^*(A \cap \varnothing) + m^*(A \cap \mathbb{R}^d) = m^*(\varnothing) + m^*(A) = m^*(A).$$

Here is another example of a measurable set:

Lemma 1.7 (Half-spaces are measurable)**.** *The half-space*

$$\big\{(x_1, \dots, x_d) \in \mathbb{R}^d : x_d > 0\big\}$$

is measurable.

Proof. *[This proof is the content of Series* 2*, ex.* $2-3$*]* We first handle the case $d = 1$. Note that since we are working in one dimension the volume of a cube corresponds to the length of an interval. We have already proved that $m^*(A) \le m^*(A \cap (0,\infty)) + m^*(A \setminus (0,\infty))$, by finite subadditivity of m^*. Now, we prove the reverse inequality. Fix $\varepsilon > 0$ and consider $\{B_i\}_{i=1}^\infty$ a countable familly of intervals such that

$$\sum_{i=1}^\infty \operatorname{vol} B_i \le m^*(A) + \varepsilon \quad \text{ and } \quad A \subseteq \bigcup_{i=1}^\infty B_i.$$

Now, for all $i = 1, 2, \dots$, define

$$B_i^1 = B_i \cap (0,\infty),$$
$$B_i^2 = B_i \cap (-\infty, \varepsilon/2^i).$$

Note that B_i^1 and B_i^2 are open intervals and $\operatorname{vol} B_i^1 + \operatorname{vol} B_i^2 \le \operatorname{vol} B_i + \varepsilon/2^i$. Observe that

$$A \cap (0,\infty) \subseteq \Big[\bigcup_{i=1}^\infty B_i\Big] \cap (0,\infty) = \bigcup_{i=1}^\infty [B_i \cap (0,\infty)] = \bigcup_{i=1}^\infty B_i^1.$$

In a similar way, we can prove $A \setminus (0,\infty) \subseteq \bigcup_{i=1}^\infty B_i^2$. Finally, by definition of outer measure,

$$m^*(A \cap (0,\infty)) + m^*(A \setminus (0,\infty)) \le \sum_{i=1}^\infty \operatorname{vol} B_i^1 + \sum_{i=1}^\infty \operatorname{vol} B_i^2 \le \sum_{i=1}^\infty \Big[\operatorname{vol} B_i + \frac{\varepsilon}{2^i}\Big] \le m^*(A) + 2\varepsilon.$$

Since this inequality is true for any $\varepsilon > 0$, we deduce

$$m^*(A \cap (0,\infty)) + m^*(A \setminus (0,\infty)) \le m^*(A)\,.$$

This proves the claim when the dimension is one. We now deal with the case $d > 1$. We have $m^*(A) \le m^*(A \cap E) + m^*(A \setminus E)$ by finite subadditivity of m^*. Then we prove the reverse inequality. Fix $\varepsilon > 0$. There is a countable family of open boxes $\{B_i\}_{i=1}^\infty$ such that

$$\sum_{i=1}^\infty \operatorname{vol} B_i \le m^*(A) + \varepsilon, \quad A \subseteq \bigcup_{i=1}^\infty B_i.$$

Every B_i is an open box of the form $B_i = \prod_{k=1}^d (a_k^{(i)}, b_k^{(i)})$. Define for any $i = 1, 2, \dots$,

$$\varepsilon_i = \frac{\varepsilon}{2^i \prod_{k=1}^{d-1} |a_k^{(i)} - b_k^{(i)}|},$$

and correspondingly,

$$B_i^1 = \prod_{k=1}^{d-1} (a_k^{(i)}, b_k^{(i)}) \times \left[(a_d^{(i)}, b_d^{(i)}) \cap (0, \infty) \right],$$

$$B_i^2 = \prod_{k=1}^{d-1} (a_k^{(i)}, b_k^{(i)}) \times \left[(a_d^{(i)}, b_d^{(i)}) \cap (-\infty, \varepsilon_i) \right].$$

Notice that B_i^1 and B_i^2 are open boxes and

$$\mathrm{vol}\, B_i^1 + \mathrm{vol}\, B_i^2 \leq \prod_{k=1}^{d-1} |a_k^{(i)} - b_k^{(i)}| \left(|a_d^{(i)} - b_d^{(i)}| + \varepsilon_i \right) = \prod_{k=1}^{d} |a_k^{(i)} - b_k^{(i)}| + \varepsilon/2^i = \mathrm{vol}\, B_i + \varepsilon/2^i$$

We can prove that $A \cap E \subseteq \cup_{i=1}^{\infty} B_i^1$ and $A \setminus E \subseteq \cup_{i=1}^{\infty} B_i^2$. Finally,

$$m^*(A \cap E) + m^*(A \setminus E) \leq \sum_{i=1}^{\infty} \mathrm{vol}\, B_i^1 + \sum_{i=1}^{\infty} \mathrm{vol}\, B_i^2 \leq \sum_{i=1}^{\infty} \left[\mathrm{vol}\, B_i + \varepsilon/2^i \right] \leq m^*(A) + 2\varepsilon.$$

Since this inequality is true for any ε, we deduce

$$m^*(A \cap E) + m^*(A \setminus E) \leq m^*(A).$$

□

Remark 1.14. A similar argument will also show that any half-space of the form

$$\left\{ (x_1, \dots, x_d) \in \mathbb{R}^d : x_j > 0 \right\} \text{ or } \left\{ (x_1, \dots, x_d) \in \mathbb{R}^d : x_j < 0 \right\} \tag{1.4}$$

for some $1 \leq j \leq d$ is measurable.

Now we present some more properties of measurable sets.

Lemma 1.8 (Properties of measurable sets)**.** *The following properties hold.*

(i) If E is measurable, then $\mathbb{R}^d \backslash E$ is also measurable.
(ii) (Translation invariance) If E is measurable, and $x \in \mathbb{R}^d$, then $x + E$ is also measurable, and $m(x + E) = m(E)$.
(iii) (Homogeneity) If E is measurable, and $a \in \mathbb{R}$, then aE is also measurable, and $m(aE) = |a|^d m(E)$.
(iv) If E_1 and E_2 are measurable, then $E_1 \cap E_2$ and $E_1 \cup E_2$ are measurable.
(v) (Boolean algebra property) If $E_1, E_2, \dots, E_N$ are measurable, then $\bigcup_{j=1}^{N} E_j$ and $\bigcap_{j=1}^{N} E_j$ are measurable.
(vi) Every open box, and every closed box, is measurable.
(vii) Any set E of outer measure zero (i.e., $m^*(E) = 0$) is measurable.

Proof. (i) We write, for any $A \subset \mathbb{R}^d$, by measurability of E

$$\begin{aligned} m^*(A) &= m^*(A \cap E) + m^*(A \setminus E) = m^*(A \cap (E^c)^c) + m^*(A \setminus E) \\ &= m^*(A \setminus E^c) + m^*(A \cap E^c). \end{aligned}$$

(ii) By translation invariance of m^* (Lemma 1.2, (iv)), we have

$$\begin{aligned} m^*(A) &= m^*(A - x) = m^*((A - x) \cap E) + m^*((A - x) \setminus E) \\ &= m^*(A \cap (x + E)) + m^*(A \setminus (x + E)). \end{aligned}$$

(iii) Assume that a is not 0, otherwise the claim follows from (vii). By homogeneity of m^* (Lemma 1.2, (vi)), we have

$$m^*(A) = |a|^d m^* \left(\frac{1}{a}A\right) = |a|^d \left[m^* \left(\left(\frac{1}{a}A\right) \cap E\right) + m^* \left(\left(\frac{1}{a}A\right) \setminus E\right)\right]$$
$$= m^* \left(A \cap (aE)\right) + m^* \left(A \setminus (aE)\right).$$

(iv) We have to prove that

$$m^*(A) \geq m^*(A \cap (E_1 \cup E_2)) + m^*(A \setminus (E_1 \cup E_2)),$$

the other inequality being always true. We write, by measurability of E_1 and E_2, and by subbaditivity of the outer measure (Lemma 1.2, (v)),

$$\begin{aligned} m^*(A) &= m^*(A \cap E_1) + m^*(A \cap E_1^c) \\ &= m^*(A \cap E_1 \cap E_2) + m^*(A \cap E_1^c \cap E_2) + m^*(A \cap E_1 \cap E_2^c) + m^*(A \cap E_1^c \cap E_2^c) \\ &\geq m^*(A \cap (E_1 \cup E_2)) + m^*(A \cap E_1^c \cap E_2^c), \end{aligned}$$

which yields the claim.

(v) Follows from (iii) by induction.

(vi) Open boxes can be seen as intersections of half-spaces, therefore by Lemma 1.7 and (iii) we prove their measurability. In particular, we define

$$H_{a_j}^{(j)} := \left\{(x_1, \ldots, x_d) \in \mathbb{R}^d : x_j > a_j\right\}, \quad G_{b_j}^{(j)} := \left\{(x_1, \ldots, x_d) \in \mathbb{R}^d : x_j < b_j\right\}.$$

We can then write an open box as:

$$B = \bigcap_{j=1}^{d} H_{a_j}^{(j)} \cap G_{b_j}^{(j)},$$

therefore by the boolean algebra property (iii) B is measurable. The proof for closed boxes is analogue, by taking the half-planes

$$H_{a_j}^{(j)} := \left\{(x_1, \ldots, x_d) \in \mathbb{R}^d : x_j \geq a_j\right\}, \quad G_{a_j}^{(j)} := \left\{(x_1, \ldots, x_d) \in \mathbb{R}^d : x_j \leq b_j\right\}.$$

(vii) Finally, if $m^*(E) = 0$, we have

$$m^*(A \cap E) \leq m^*(E) = 0$$

and

$$m^*(A \cap E^c) \leq m^*(A).$$

Therefore,

$$m^*(A) \geq m^*(A \cap E) + m^*(A \setminus E),$$

which yields the claim as the reverse inequality is always true.

□

With Lemma 1.8, we have proved many properties on our wish list of measurable sets, and we are making progress towards finite additivity. We can actually prove it:

Lemma 1.9 (Finite additivity)**.** *If $(E_j)_{j \in J}$ is a finite collection of disjoint measurable sets and A is any set (not necessarily measurable), we have*

$$m^* \left(A \cap \bigcup_{j \in J} E_j\right) = \sum_{j \in J} m^* \left(A \cap E_j\right). \tag{1.5}$$

Furthermore, we have

$$m\left(\bigcup_{j\in J} E_j\right) = \sum_{j\in J} m\left(E_j\right). \tag{1.6}$$

PROOF. The equality in (1.6) follows from (1.5) by setting $A := \mathbb{R}^d$. To prove (1.5), we show the case $|J| = 2$, the general case following by induction. We know, by the measurability of E_1 that

$$\begin{aligned} m^*\left(A\cap(E_1\cup E_2)\right) &= m^*\left(A\cap(E_1\cup E_2)\cap E_1\right) + m^*\left(A\cap(E_1\cup E_2)\cap E_1^c\right) \\ &= m^*(A\cap E_1) + m^*(A\cap E_2). \end{aligned}$$

For the general case, we reason by induction, proving the case $|J| = N$ and supposing true the case $|J| = N-1$. We write

$$\begin{aligned} m^*\left(A\cap\bigcup_{j=1}^N E_j\right) &= m^*\left(A\cap\bigcup_{j=1}^N E_j\cap E_N\right) + m^*\left(A\cap\bigcup_{j=1}^N E_j\cap E_N^c\right) \\ &= m^*(A\cap E_N) + m^*\left(A\cap\bigcup_{j=1}^{N-1} E_j\right) \\ &= m^*(A\cap E_N) + \sum_{j=1}^{N-1} m^*(A\cap E_j) \\ &= \sum_{j=1}^{N} m^*(A\cap E_j), \end{aligned}$$

which conludes the proof. □

Remark 1.15. Lemma 1.9 and Proposition 1.6 combined can imply that there exist non-measurable sets.

Corollary 1.10. *If $A \subseteq B$ are two measurable sets, then $B\backslash A$ is also measurable and, if in addition $m(A) < +\infty$, we have*

$$m(B\backslash A) = m(B) - m(A)$$

PROOF. Remark that $B = A\cup(B\setminus A)$. By finite additivity (Lemma 1.9) we obtain

$$m(B) = m(A) + m(B\setminus A),$$

which yields the claim. □

Now we show countable additivity.

Lemma 1.11 (Countable additivity)**.** *If $(E_j)_{j\in J}$ is a countable collection of disjoint measurable sets, then $\bigcup_{j\in J} E_j$ is measurable, and*

$$m\left(\bigcup_{j\in J} E_j\right) = \sum_{j\in J} m\left(E_j\right) \tag{1.7}$$

PROOF. Let $E := \bigcup_{j\in J} E_j$. Our first task will be to show that E is measurable. Thus, let A be an arbitrary set (not necessarily measurable); we need to show that

$$m^*(A) = m^*(A\cap E) + m^*(A\backslash E). \tag{1.8}$$

Since J is countable, we can assume without loss of generality $J = \mathbb{N}$. Note that

$$A \cap E = \bigcup_{j=1}^{\infty} (A \cap E_j)$$

and hence by countable sub-additivity

$$m^*(A \cap E) \leq \sum_{j=1}^{\infty} m^*(A \cap E_j).$$

We rewrite this as

$$m^*(A \cap E) \leq \sup_{N \geq 1} \sum_{j=1}^{N} m^*(A \cap E_j).$$

Let F_N be the set $F_N := \bigcup_{j=1}^{N} E_j$. Since the $A \cap E_j$ are all disjoint, and their union is $A \cap F_N$, we see by finite additivity (Lemma 1.9) that

$$\sum_{j=1}^{N} m^*(A \cap E_j) = m^*(A \cap F_N)$$

and hence

$$m^*(A \cap E) \leq \sup_{N \geq 1} m^*(A \cap F_N).$$

Now we look at $A \backslash E$. Since $F_N \subseteq E$, we have $A \backslash E \subseteq A \backslash F_N$. By monotonicity, we thus have

$$m^*(A \backslash E) \leq m^*(A \backslash F_N)$$

for all N. In particular, we see that

$$\begin{aligned} m^*(A \cap E) + m^*(A \backslash E) &\leq \sup_{N \geq 1} m^*(A \cap F_N) + m^*(A \backslash E) \\ &\leq \sup_{N \geq 1} m^*(A \cap F_N) + m^*(A \backslash F_N) \\ &= m^*(A) \end{aligned}$$

where in the last line we used that F_N is measurable thanks to the finite additivity. But from finite sub-additivity we have

$$m^*(A \cap E) + m^*(A \backslash E) \geq m^*(A)$$

and the claim (1.8) follows. This shows that E is measurable.

To finish the lemma, we need to show (1.7). We first observe from countable sub-additivity that

$$m(E) \leq \sum_{j \in J} m(E_j) = \sum_{j=1}^{\infty} m(E_j).$$

On the other hand, by finite additivity and monotonicity we have

$$m(E) \geq \sup_{N \geq 1} m(F_N) = \sup_{N \geq 1} \sum_{j=1}^{N} m(E_j) = \sum_{j=1}^{\infty} m(E_j)$$

and thus we have (1.7) as desired. □

Next, we prove measurability for countable unions and intersections.

Lemma 1.12 (σ-algebra property)**.** *If $(E_j)_{j \in J}$ is any countable collection of measurable sets (so J is countable), then the union $\bigcup_{j \in J} E_j$ and the intersection $\bigcap_{j \in J} E_j$ are also measurable.*

Proof. We remark that

$$\bigcup_{j=1}^{\infty} E_j = E_1 \cup (E_2 \setminus E_1) \cup (E_3 \setminus (E_1 \cup E_2)) \cup \dots ,$$

where the unions are disjoint unions. Combining finite additivity, Corollary 1.10 and countable additivity, we obtain the claim for countable unions of measurable sets. For the case of the intersection, note that

$$\bigcap_{j=1}^{\infty} E_j = \left(\bigcup_{j=1}^{\infty} E_j^c \right)^c .$$

Using the measurability of the complement of a measurable set (Lemma 1.8 (i)) we conclude by the property just obtained for unions. ☐

To finally prove measurability for open and closed sets, first need a preliminary lemma.

Lemma 1.13. *Every open set can be written as a countable or finite union of open boxes.*

Proof. *[This proof is the content of Series* 1*, ex.* 3 *(iii)].* We first need some notation. Call a box $B = \prod_{i=1}^{d} (a_i, b_i)$ rational if all of its components a_i, b_i are rational numbers. Observe that there are only a countable number of rational boxes (this is since a rational box is described by $2d$ rational numbers, and so has the same cardinality as $\mathbb{Q}^{2d}$. But $\mathbb{Q}$ is countable, and the Cartesian product of any finite number of countable sets is countable).

We make the following claim: given any open ball $B(x, r)$, there exists a rational box B which is contained in $B(x, r)$ and which contains x. To prove this claim, write $x = (x_1, \dots, x_d)$. For each $1 \leq i \leq d$, let a_i and b_i be rational numbers such that

$$x_i - \frac{r}{d} < a_i < x_i < b_i < x_i + \frac{r}{d}$$

Then it is clear that the box $\prod_{i=1}^{d} (a_i, b_i)$ is rational and contains x. A simple computation using Pythagoras' theorem (or the triangle inequality) also shows that this box is contained in $B(x, r)$.

Now let E be an open set, and let Σ be the set of all rational boxes B which are subsets of E, and consider the union $\bigcup_{B \in \Sigma} B$ of all those boxes. Clearly, this union is contained in E, since every box in Σ is contained in E by construction. On the other hand, since E is open, we see that for every $x \in E$ there is a ball $B(x, r)$ contained in E, and by the previous claim this ball contains a rational box which contains x. In particular, x is contained in $\bigcup_{B \in \Sigma} B$. Thus we have

$$E = \bigcup_{B \in \Sigma} B$$

as desired; note that Σ is countable or finite because it is a subset of the set of all rational boxes, which is countable. ☐

Lemma 1.14 (Borel property)**.** *Every open set, and every closed set, is Lebesgue measurable.*

Proof. It suffices to do this for open sets, since the claim for closed sets then follows by measurability of complements (Lemma 1.8(i)). Let E be an open set. By Lemma 1.13, E is the countable union of boxes. Since we already know that boxes are measurable, and that the countable union of measurable sets is measurable, the claim follows. ☐

The construction of Lebesgue measure and its basic properties are now complete. Now we make the next step in constructing the Lebesgue integral - describing the class of functions we can integrate.

1.6. Measurable functions

In the theory of the Riemann integral, we are only able to integrate a certain class of functions - the Riemann integrable functions. We will now be able to integrate a much larger range of functions, for instance, non-negative measurable functions.

Definition (Measurable functions). Let Ω be a measurable subset of $\mathbb{R}^d$, and let $f : \Omega \to \mathbb{R}^m$ be a function. A function f is measurable iff $f^{-1}(V)$ is measurable for every open set $V \subseteq \mathbb{R}^m$.

As discussed earlier, most sets that we deal with in real life are measurable, so it is only natural to learn that most functions we deal with in real life are also measurable. For instance, continuous functions are automatically measurable:

Lemma 1.15 (Continuous functions are measurable)**.** *Let Ω be a measurable subset of $\mathbb{R}^d$, and let $f : \Omega \to \mathbb{R}^m$ be continuous. Then f is also measurable.*

Proof. Let V be any open subset of $\mathbb{R}^m$. Then since f is continuous, $f^{-1}(V)$ is open relative to Ω, i.e., $f^{-1}(V) = W \cap \Omega$ for some open set $W \subseteq \mathbb{R}^d$. Since W is open, it is measurable; since Ω is measurable, $W \cap \Omega$ is also measurable. □

Because of Lemma 1.13, we have an easy criterion to test whether a function is measurable or not:

Lemma 1.16. *Let Ω be a measurable subset of $\mathbb{R}^d$, and let $f : \Omega \to \mathbb{R}^m$ be a function. Then f is measurable if and only if $f^{-1}(B)$ is measurable for every open box B.*

Proof. The only if statement is trivial, so we only prove the other direction. By Lemma 1.13, every open set $V \subseteq \mathbb{R}^m$ can be expressed as a countable union of open boxes, namely

$$V = \bigcup_{i=1}^{\infty} B_i.$$

Therefore

$$f^{-1}(V) = f^{-1}\left(\bigcup_{i=1}^{\infty} B_i\right) = \bigcup_{i=1}^{\infty} f^{-1}(B_i)$$

is measurable by Lemma 1.11. □

Corollary 1.17. *Let Ω be a measurable subset of $\mathbb{R}^d$, and let $f : \Omega \to \mathbb{R}^m$ be a function. Suppose that $f = (f_1, \dots, f_m)$, where $f_j : \Omega \to \mathbb{R}$ is the j^{th} co-ordinate of f. Then f is measurable if and only if all of the f_j are individually measurable.*

Proof. If f is measurable, by the fact that

$$f_j^{-1}(a, b) = f^{-1}(\underbrace{\mathbb{R} \times \dots \times \mathbb{R}}_{j-1 \text{ times}} \times (a, b) \times \mathbb{R} \times \dots \times \mathbb{R})$$

we get that $f_j^{-1}((a, b))$ is measurable, so we can conclude by Theorem 1.16 and the fact that

$$\mathbb{R} \times \dots \times \mathbb{R} \times (a, b) \times \mathbb{R} \times \dots \times \mathbb{R}$$

is open.

If f_j is measurable $\forall j$, then

$$f^{-1}\left(\prod_{j=1}^{m} (a_j, b_j)\right) = \bigcap_{j=1}^{m} f_j^{-1}(a_j, b_j)$$

is measurable and again we conclude via Theorem 1.16. □

Unfortunately, it is not true that the composition of two measurable functions is automatically measurable; however we can do the next best thing: a continuous function applied to a measurable function is measurable.

Lemma 1.18. *Let Ω be a measurable subset of $\mathbb{R}^d$, and let W be an open subset of $\mathbb{R}^m$. If $f : \Omega \to W$ is measurable, and $g : W \to \mathbb{R}^p$ is continuous, then $g \circ f : \Omega \to \mathbb{R}^p$ is measurable.*

Proof. Take a open subset $V \subseteq \mathbb{R}^p$. We have

$$(g \circ f)^{-1}(V) = f^{-1}\left(g^{-1}(V)\right).$$

As g is continuous, $g^{-1}(V)$ is open, and as f is measurable, $f^{-1}\left(g^{-1}(V)\right)$ is measurable. □

This has an immediate corollary:

Corollary 1.19. *Let Ω be a measurable subset of $\mathbb{R}^d$. If $f : \Omega \to \mathbb{R}$ is a measurable function, then so is $|f|, \max(f, 0)$, and $\min(f, 0)$.*

Proof. Apply Lemma 1.18 with $g(x) \coloneqq |x|, g(x) \coloneqq \max(x, 0)$, and $g(x) \coloneqq \min(x, 0)$ □

A slightly less immediate corollary:

Corollary 1.20. *Let Ω be a measurable subset of $\mathbb{R}^d$. If $f : \Omega \to \mathbb{R}$ and $g : \Omega \to \mathbb{R}$ are measurable functions, then so is $f + g, f - g, fg$, $\max(f, g)$, and $\min(f, g)$. If $g(x) \neq 0$ for all $x \in \Omega$, then f/g is also measurable.*

Proof. Consider $f + g$. We can write this as $k \circ h$, where $h : \Omega \to \mathbb{R}^2$ is the function $h(x) = (f(x), g(x))$, and $k : \mathbb{R}^2 \to \mathbb{R}$ is the function $k(a, b) \coloneqq a + b$. Since f, g are measurable, then h is also measurable by Corollary 1.17. Since k is continuous, we thus see from Lemma 1.18 that $k \circ h$ is measurable, as desired. A similar argument deals with all the other cases; the only thing concerning the f/g case is that the space $\mathbb{R}^2$ must be replaced with $\{(a, b) \in \mathbb{R}^2 : b \neq 0\}$ in order to keep the map $(a, b) \mapsto a/b$ continuous and well-defined. □

Another characterization of measurable functions is given by

Lemma 1.21. *Let Ω be a measurable subset of $\mathbb{R}^d$, and let $f : \Omega \to \mathbb{R}$ be a function. Then f is measurable if and only if $f^{-1}((a, \infty))$ is measurable for every real number a.*

Proof. The implication $\Rightarrow$ is trivial. To prove the implication $\Leftarrow$, by Theorem 1.16 it suffices to show that $f^{-1}((a, b))$ is measurable $\forall a < b \in \mathbb{R}$. Note that

$$f^{-1}([b, +\infty)) = \bigcap_{n \in \mathbb{N}} \underbrace{f^{-1}\left(\left(b - \frac{1}{n}, +\infty\right)\right)}_{\text{measurable by hypothesis}}.$$

We conclude that $f^{-1}([b, +\infty))$ is measurable by measurability of countable intersections of measurable sets (Theorem 1.8 (iv)). Then, we have that

$$f^{-1}((a, b)) = f^{-1}((a, +\infty)) \cap f^{-1}((-\infty, b) = f^{-1}((a, +\infty)) \cap f^{-1}([b, +\infty))^C,$$

which are both measurable sets. □

Inspired by this lemma, we extend the notion of a measurable function to the extended real number system $\mathbb{R}^* \coloneqq \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$:

Definition (Measurable functions in the extended reals). Let Ω be a measurable subset of $\mathbb{R}^d$. A function $f : \Omega \to \mathbb{R}^*$ is said to be measurable iff $f^{-1}((a, +\infty])$ is measurable for every real number a.

Note that Lemma 1.21 ensures that the notion of measurability for functions taking values in the extended reals $\mathbb{R}^*$ is compatible with that for functions taking values in just the reals $\mathbb{R}$.

Measurability behaves well with respect to limits:

Lemma 1.22 (Limits of measurable functions are measurable). *Let Ω be a measurable subset of $\mathbb{R}^d$. For each positive integer n, let $f_n : \Omega \to \mathbb{R}^*$ be a measurable function. Then the functions $\sup_{n\geq 1} f_n$, $\inf_{n\geq 1} f_n$, $\limsup_{n\to\infty} f_n$, $\liminf_{n\to\infty} f_n$ are also measurable. In particular, if the f_n converge pointwise to another function $f : \Omega \to \mathbb{R}^*$, then f is also measurable.*

Proof. We first prove the claim about $\sup_{n\geq 1} f_n$. Call this function g. We have to prove that $g^{-1}((a, +\infty])$ is measurable for every a. But by the definition of supremum, we have

$$g^{-1}((a,+\infty]) = \bigcup_{n\geq 1} f_n^{-1}((a,+\infty]),$$

and the claim follows since the countable union of measurable sets is again measurable.

A similar argument works for $\inf_{n\geq 1} f_n$. The claim for lim sup and lim inf then follow from the identities

$$\limsup_{n\to\infty} f_n = \inf_{N\geq 1} \sup_{n\geq N} f_n$$

and

$$\liminf_{n\to\infty} f_n = \sup_{N\geq 1} \inf_{n\geq N} f_n.$$

□

1.7. The Cantor set

In this section, we introduce the *Cantor (ternary) set*[1], a famous example of a Borel set which has Lebesgue measure 0 and yet the cardinality of the continuum.

We define it inductively as follows. As a base case, we let

$$\begin{aligned} C_0 &:= [0,1]\,; \\ C_1 &:= [0,1/3] \cup [2/3,1]\,, \\ C_2 &:= [0,1/9] \cup [2/9,1/3] \cup [2/3,7/9] \cup [8/9,1]\,, \end{aligned}$$

that is, for $k \in \mathbb{N}_{\geq 1}$, the set C_k is the union of 2^k disjoint closed intervals of length 3^{-k} and, to obtain C_{k+1} from C_k, we remove the open middle third from each of the intervals in C_k. The interval $I_{1,1} = \left(\frac{1}{3}, \frac{2}{3}\right)$, which is removed from C_0 at the 1st stage to obtain C_1, will be called the *removed (or complementary) interval of* C_1. In general, the 2^k intervals that are removed from C_k to obtain C_{k+1} will be called the *removed (or complementary) intervals* of C_{k+1} and will be labeled, from left to right, as $I_{k+1,j}$, where $j \in \{1, \ldots, 2^k\}$.

We remark that, by induction, it is possible to show that

$$I_{k,j} = \left(\frac{3i+1}{3^k}, \frac{3i+2}{3^k}\right) \qquad \text{for some } i \in \mathbb{N}, \text{ with } 0 \leq i \leq 3^{k-1} - 1.$$

and that the intervals that remain to make up C_k after the $I_{k,j}$ are removed are of the form

$$\left[\frac{3j+0}{3^k}, \frac{3j+1}{3^k}\right] \quad \text{and} \quad \left[\frac{3j+2}{3^k}, \frac{3j+3}{3^k}\right]. \tag{1.9}$$

Now $\{C_k\}_{k\in\mathbb{N}}$ is a nested sequence of non-empty compact sets:

$$\cdots \subset C_{k+1} \subset C_k \subset \cdots \subset C_2 \subset C_1 \subset C_0\,.$$

[1]It was introduced by Henry John Stephen Smith [Smi75], Vito Volterra [Vol81], and Georg Cantor [Can84]. See [Fle94] for a more detailed historical account.

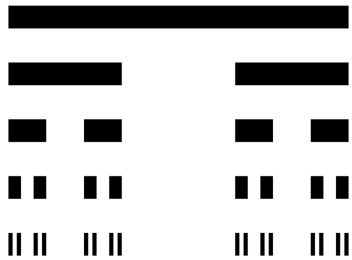


FIGURE 1. First few stages of the construction of the Cantor set.

The *Cantor set* is defined as

$$C := \bigcap_{k=0}^{\infty} C_k\,.$$

If we denote by E_k the union of the open intervals that are removed at the k-th stage, then

$$C := [0,1] \setminus \bigcup_{k=1}^{\infty} E_k := [0,1] \setminus \left(\bigcup_{k=1}^{\infty} \bigcup_{j=1}^{2^{k-1}} I_{k,j} \right).$$

In the following proposition, we present some further properties of the Cantor set.

THEOREM 1.23 (Properties of the Cantor set). *The Cantor set C has the following properties:*

(i) C is compact and non-empty.
(ii) C is measurable and its Lebesgue measure is 0.
(iii) C is nowhere dense[2].
(iv) C is equal to the set of all $x \in [0,1]$ which have a ternary expansion containing only the digits 0 and 2, i.e.,

$$\begin{aligned} C &\equiv \left\{ \sum_{n=1}^{\infty} \frac{a_n}{3^n} : a_n \in \{0,2\} \text{ for } n \in \mathbb{N}^* \right\} \\ &=: \Big\{ a \in [0,1] : a = (0.a_1a_2\ldots)_3 \text{ with } a_i \in \{0,2\} \text{ for } i \in \mathbb{N}^* \Big\}, \end{aligned}$$

where $(0.a_1a_2\ldots)_3$ denotes a possible ternary expansion of $a \in [0,1]$.
(v) C is perfect[3].
(vi) C is totally disconnected[4].
(vii) C is uncountable.

Remark 1.16. The ternary expansion mentioned in (iv) is not unique (see Exercise 5 of Series 2). Notice, however, that while it may be possible for $x \in \mathbb{R}$ to have two different ternary representations (check that $\frac{1}{3} = (0.1000\ldots)_3 = (0.0222\ldots)_3$, as an example), x *cannot* be written in more than one way without using the digit 1. That is, we claim that, if $\sum_{n=1}^{\infty} \frac{\alpha_n}{3n} = \sum_{n=1}^{\infty} \frac{\beta_n}{3n}$, where each of α_n and β_n is either 0 or 2, then $\alpha_n = \beta_n$ for every n. Suppose that there exists an n such that $\alpha_n \neq \beta_n$. Let m be the smallest integer such that $\alpha_m \neq \beta_m$. Then $|\alpha_m - \beta_m| = 2$ and $|\alpha_n - \beta_n| \leq 2$ for every

[2]We say that a set $E \subset \mathbb{R}$ is *nowhere dense* (in $\mathbb{R}$) if the interior of the closure of E (in $\mathbb{R}$) is the empty set.

[3]We say that a set E is *perfect* if it is closed and each point of E is a limit point of E.

[4]We say that a set E is *totally disconnected* if, for each distinct $x \in E$ and $y \in S$, there exist disjoint open sets U and V such that $x \in U$, $y \in V$, and $E = (U \cap E) \cup (V \cap E)$.

n, so that

$$\begin{aligned} 0 &= \left| \sum_{n=m}^{\infty} \frac{\alpha_n}{3^n} - \sum_{n=m}^{\infty} \frac{\beta_n}{3^n} \right| = \left| \sum_{n=m}^{\infty} \frac{\alpha_n - \beta_n}{3^n} \right| \\ &= \left| \frac{\alpha_m - \beta_m}{3^m} + \sum_{n=m+1}^{\infty} \frac{\alpha_n - \beta_n}{3^n} \right| \\ &\geq \frac{|\alpha_m - \beta_m|}{3^m} - \left| \sum_{n=m+1}^{\infty} \frac{\alpha_n - \beta_n}{3^n} \right| \\ &\geq \frac{|\alpha_m - \beta_m|}{3^m} - \sum_{n=m+1}^{\infty} \frac{|\alpha_n - \beta_n|}{3^n} \\ &\geq \frac{1}{3^m} |\alpha_m - \beta_m| - \sum_{n=m+1}^{\infty} \frac{2}{3^n} \\ &= \frac{1}{3^m}, \end{aligned}$$

which yields a contradiction. Hence, $\alpha_n = \beta_n$ for every $n \in \mathbb{N}^*$.

Proof.

(i) For every $k \in \mathbb{N}$, the set C_k is closed. Since any intersection of closed sets is closed, then the Cantor set $C \coloneqq \bigcap_{k=0}^{\infty} C_k$ is closed. Finally, C is compact since it is both closed and bounded (as subset of $[0,1]$).

It remains to show that $C \neq \varnothing$. For each $k \in \mathbb{N}$, let $x_k \in C_k$. Then $\{x_k\}_{k\in\mathbb{N}} \subseteq C_1$. By compactness, there is a convergent subsequence $\left\{x_{k_j}\right\}_{j\in\mathbb{N}}$ with limit $x_0 \in C_1$. However, $\left\{x_{k_j}\right\}_{j=2}^{\infty} \subseteq C_2$. Thus $x_0 \in C_2$. An analogous reasoning shows that $x_0 \in C_k$ for all $k = 1, 2, \ldots$. In conclusion, $x_0 \in \bigcap_{k\in\mathbb{N}} C_k =: C$.

(ii) For every $k \in \mathbb{N}$, the set C_k is measurable. Since any countable intersection of measurable sets is measurable, then the Cantor set $C \coloneqq \bigcap_{k=0}^{\infty} C_k$ is measurable. By construction, C_k is formed by 2^k disjoint intervals of length 3^{-k}, so $m(C_k) = (2/3)^k$. Since $C \subset C_k$ for any $k \in \mathbb{N}$, we deduce

$$0 \leq m(C) \leq m(C_k) = \left(\frac{2}{3}\right)^k.$$

Letting $k \to \infty$, we conclude $m(C) = 0$.

(iii) By (i), C is closed. By contradiction, let us suppose that C contains a (non-empty) open interval I. By the monotonicity of the Lebesgue measure, we have $m(I) \leq m(C)$; however, by (ii), $m(C) = 0$ which yields a contradiction.

(iv) Let us consider the map

$$\begin{array}{rcl} f : \{0,2\}^{\mathbb{N}^*} & \to & [0,1] \\ a & \mapsto & \sum_{n=1}^{+\infty} \frac{a_n}{3^n}. \end{array}$$

To prove the claim of (iv), we need to show that f is a bijection between $\{0,2\}^{\mathbb{N}^*}$ and C. First, we note that f actually takes values in $C \subset [0,1]$. By structure of C_k,

$$\sum_{n=1}^{k} \frac{a_n}{3^n} \in C_k.$$

As a result, for $a \in \{0,2\}^{\mathbb{N}^*}$,

$$\sum_{n=1}^{+\infty} \frac{a_n}{3^n} = \lim_{k\to+\infty} \sum_{n=1}^{k} \frac{a_n}{3^n} \in C.$$

Second, we prove that f is injective. Let a', $a'' \in \{0,2\}^{\mathbb{N}^*}$ such that

$$a_i' = a_i'' \quad \text{for } 1 \le j < N, \quad a_N' < a_N''.$$

Then, we have $a_N' = 0$, $a_N'' = 2$ and

$$\begin{aligned} f(a') = \sum_{i=1}^{+\infty} \frac{a_n'}{3^n} &= \sum_{n=1}^{N-1} \frac{a_n''}{3^n} + \sum_{i=N+1}^{+\infty} \frac{a_n'}{3^n} \\ &\le \sum_{n=1}^{N-1} \frac{a_n''}{3^n} + 3^{-N-1} \frac{2}{1-1/3} \\ &= \sum_{n=1}^{N-1} \frac{a_n''}{3^n} + 3^{-N} \\ &< \sum_{n=1}^{N} \frac{a_n''}{3^n} = f(a''). \end{aligned}$$

Finally, we prove that f is surjective. Let $x \in C$, we need to show that there exists $a \in \{0,2\}^{\mathbb{N}^*}$ such that $x = f(a)$. We start by observing that, since $x \in [0,1]$, we have

$$[0,1] \ni x = \sum_{i=1}^{+\infty} \frac{x_i}{3^i}, \qquad \text{for some } x_i \in \{0,1,2\}. \tag{1.10}$$

We need to show that $x_i \neq 1$ for all $i \in \mathbb{N}^*$. Since $x \in C$, it belongs to an interval of the form (1.9), i.e., for all $n \in \mathbb{N}^*$, there exists $a^{(n)} := (a_1^{(n)}, \dots, a_n^{(n)}) \in \{0,2\}^n$ such that

$$x_{a^{(n)}} \le x \le x_{a^{(n)}} + 3^{-n},$$

with the notation $x_a := \sum_{i=1}^n \frac{a_i}{3^i}$, which implies

$$x = \lim_{n\to+\infty} x_{a^{(n)}} = \lim_{n\to+\infty} \left(\sum_{i=1}^{n} \frac{a_i^{(n)}}{3^i} \right).$$

We now consider three cases. If $x = x_{a^{(n)}}$ for some $n \ge 1$, then

$$x = f\left(a_1^{(n)}, \dots, a_n^{(n)}, 0, 0, \dots\right).$$

If $x = x_{a^{(n)}} + 3^{-n}$ for some $n \ge 1$, then

$$x = \sum_{i=1}^{n} \frac{a_i^{(n)}}{3^j} + \underbrace{\sum_{i=n+1}^{+\infty} \frac{2}{3^i}}_{=3^{-n-1}\frac{2}{1-1/3}=3^{-n}} = f\left(a_1^{(n)}, \dots, a_n^{(n)}, 2, 2, \dots\right),$$

The third case is $x_{a^{(n)}} < x < x_{a^{(n)}} + 3^{-n}$. We first note that (because of (1.10))

$$0 \le x - \sum_{i=1}^{n} \frac{x_i}{3^i} \le \sum_{i=n}^{+\infty} \frac{x_i}{3^i} \le 3^{-n-1} 2 \frac{1}{1-\frac{1}{3}} = 3^{-n},$$

and then (because of $x_{a^{(n)}} < x < x_{a^{(n)}} + 3^{-n}$)

$$\underbrace{3^n \sum_{i=1}^{n} \frac{x_i}{3^i}}_{\in \mathbb{N}} \leq 3^n x \leq 3^n \sum_{i=1}^{n} \frac{x_i}{3^j} + 1, \qquad \underbrace{3^n x_{a^{(n)}}}_{\in \mathbb{N}} < 3^n x < 3^n x_{a^{(n)}} + 1.$$

This implies

$$\sum_{i=1}^{n} x_i 3^{n-i} = 3^n x_{a^{(n)}} = \sum_{i=1}^{n} a_i^{(n)} 3^{n-i}.$$

and, thus,

$$\underbrace{x_1}_{\in \mathbb{N}} + \underbrace{\sum_{i=2}^{n} x_i 3^{1-i}}_{\in \left[0, 6\times 3^{-2} \times \frac{3}{2}\right) = [0,1)} = \underbrace{a_1^{(n)}}_{\in \mathbb{N}} + \underbrace{\sum_{i=1}^{n} a_i^{(n)} 3^{1-i}}_{\in [0,1)},$$

Taking the floor function (or *integer part function*) of each side of the last identity, we get $x_1 = a_1^{(n)}$; similarly, we deduce $x_i = a_i^{(n)}$ for $1 \leq i \leq n$, so that each x_i belongs to $\{0, 2\}$.

(v) To prove that C is perfect, we need to show that it is closed (which follows from (i)) and that every $x \in C$ is a limit point of C. Let $x \in C$ and let $\varepsilon > 0$. We choose an integer n such that $3^{-k} < \varepsilon$. Since $x \in C_k$, there exists a closed interval I of length 3^{-k} such that $x \in I \subseteq C_k$. Let a be an endpoint of I that is distinct from x and note that $a \in C$ and $0 < |x - a| < \varepsilon$. Hence, x is a limit point of C.

(vi) To prove that C is totally disconnected, we argue as follows. Let $x, y \in C$ be distinct and assume, without loss of generality, that $x < y$. Let $\varepsilon = |x - y|$. We choose k so large that $3^{-k} < \varepsilon$. Then $x, y \in C_k$, but x and y cannot both be in the same interval of C_k (since these intervals are of length 3^{-k}). Then there exists t between x and y that does not belong to C_k (and, in particular, is not an element of C). Let us define $U \coloneqq \{s : s < t\}$ and $V \coloneqq \{s : s > t\}$. Then $x \in U \cap C$, hence $U \cap C \neq \varnothing$; analogously, $V \cap C \neq \varnothing$. Moreover, $(U \cap C) \cap (V \cap C) = \varnothing$. As a result, we conclude that $C = (C \cap U) \cup (C \cap V)$ (i.e., that C is totally disconnected).

(vii) The bijectivity of the map f defined in (iv) shows that $\operatorname{card} C = \operatorname{card}\{0, 2\}^{\mathbb{N}^*}$. In turn, we have $\operatorname{card}\{0, 2\}^{\mathbb{N}^*} = \operatorname{card}\{0, 1\}^{\mathbb{N}^*} = \operatorname{card}[0, 1] = \operatorname{card} \mathbb{R}$. Since these sets are uncountable[5], C is also uncountable.

□

Remark 1.17 (Alternative proofs of the uncountability of the Cantor set)**.** We point out that (v) actually implies (vii)[6].

[5] This is the content of Cantor's theorem on the uncountability of the real numbers. For the sake of completeness, let us provide a quick proof of the uncountability of $P \coloneqq \{0, 1\}^{\mathbb{N}^*} = \prod_{n=1}^{\infty} \{0, 1\}$ (which, in turn, is in bijection with the subset of real numbers in $[0, 1]$ whose decimal expansions consist of only digits 0 and 1). We suppose, for the sake of finding a contradiction, that it is countable. Then we can write $P = \{a_n : n \in \mathbb{N}\}$. We can rewrite $a_n \coloneqq \{a_{n,m}\}_{m \in \mathbb{N}}$ for all n. We then construct a sequence $x = \{x_n\}$ such that

$$x_n \coloneqq \begin{cases} 1 & \text{if } a_{n,n} = 0, \\ 0 & \text{if } a_{n,n} = 1. \end{cases}$$

Then x is a sequence with terms either 0 or 1 , but $x \neq a_n$ for any $n \in \mathbb{N}$, i.e., $x \in P \setminus \{a_n : n \in \mathbb{N}\} = \varnothing$, which is a contradiction.

[6] Let us sketch a proof of this fact. Let $E \subset \mathbb{R}$ be a (non-empty) perfect set. Since E has accumulation points, it cannot be finite. Therefore it is either countable or uncountable. We will prove that it is uncountable.

Let us suppose, for the sake of finding a contradiction, that E is countable instead, i.e. $E = \{a_i\}_{i \in \mathbb{N}}$.

Another proof of (vii) is essentially contained in Section 1.9: indeed, the uncountability of C follows from the construction of the Cantor–Lebesgue function. Since it maps the Cantor ternary set C *onto* the interval $[0,1]$, $\operatorname{card} C \geq \operatorname{card}[0,1]$. On the other hand, $\operatorname{card} C \leq \operatorname{card}[0,1]$ because $C \subset [0,1]$. By Cantor–Bernstein–Schröder's theorem, we then deduce that $\operatorname{card} C = \operatorname{card}[0,1]$.

1.8. σ-algebras

Measures are defined for families of sets that satisfy specific properties. For this reason, we introduce the structure of σ-algebra.

Definition (σ-algebra). Let X (usually, $X = \mathbb{R}^d$) be a set, $A \subseteq 2^X$ is a σ-algebra if:

(i) $X \in A$
(ii) $E \in A \Rightarrow E^C \in A$
(iii) $\{E_i\}_{i=1}^{+\infty} \subseteq A \Rightarrow \bigcup_{i=1}^{+\infty} E_i \in A$

Remark 1.18. $\varnothing \in A$ and $\{E_i\}_{i=1}^{+\infty} \subseteq A \Rightarrow \bigcap_{i=1}^{+\infty} E_i \in A$, using the fact that complements and countable unions of elements of A belong to A as well.
Remark 1.19. By the properties in Lemma 1.8, measurable sets form a σ-algebra. Let us denote it with $\mathcal{M}$.

On σ-algebras, we can define measures:

Definition (Measure). Let $(E, \mathcal{A})$ be a measurable space. A map $\mu : \mathcal{A} \to [0, +\infty]$ is a measure on $(E, \mathcal{A})$, if it satisfies the following conditions:

(i) $\mu(\varnothing) = 0$;
(ii) $\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n) \quad \forall \{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$ countable family of pairwise disjoint sets.

Remark 1.20. Positive linear combinations of measures are measures.
Example 1.6. Some examples of measure spaces are:

(i) $(\mathbb{R}^d, \mathcal{M}, m)$ that is the Lebesgue measure space;
(ii) $(\mathbb{R}, \mathcal{P}(\mathbb{R}), \delta_{x_0})$, where δ_0 is the Dirac Delta centered at $x_0 = 0$ (recall that δ_{x_0} is such that $\int_{\mathbb{R}} \phi(x)\delta_{x_0}(x)dx = \phi(x_0) \quad \forall \phi \in C_c^{\infty}(\mathbb{R})$);
(iii) $(\mathbb{R}, \mathcal{P}(\mathbb{R}), \sum_{n=1}^{N} \delta_n)$ is a measure space;
(iv) $(\mathbb{R}, \mathcal{P}(\mathbb{R}), \sum_{n=1}^{\infty} \delta_n)$ is a measure space.

Definition (σ-algebra generated by a collection). Let E be a set and $A \subseteq 2^E$. The σ-algebra generated by A is the smallest σ-algebra containing all elements of A.

This σ-algebra corresponds to the intersection of all the σ-algebras containing all elements of A, and one could check that the intersection of σ-algebras is still a σ-algebra.

Definition (Borel σ-algebra). The Borel σ-algebra $\mathcal{B}$ is the σ-algebra generated by open sets.

Remark 1.21. Equivalently, the Borel σ-algebra $\mathcal{B}$ is the σ-algebra generated by the collection of boxes.

The set $U_1 := (a_1 - 1, a_1 + 1)$ is a neighborhood of a_1. Since a_1 is a limit point of E, there exist infinitely many elements of E belonging to U_1. Next, we take a bounded open interval U_2 such that $\bar{U}_2 \subseteq U_1$, $U_2 \cap E \neq \varnothing$, and $a_1 \notin \bar{U}_2$. Inductively, we find U_{i+1} such that $\bar{U}_{i+1} \subseteq U_i$, $a_i \notin \bar{U}_{i+1}$, and $U_{i+1} \cap E \neq \varnothing$.

For every $i \in \mathbb{N}$, the sets $V_i := \bar{U}_i \cap E$ are compact (closed and bounded) and non-empty (by construction). However, $a_{i-1} \notin V_i$. As a consequence, $a_1 \notin V := \bigcap_{i\in\mathbb{N}} V_i$ (because $a_1 \notin V_2$), $a_2 \notin V$ (because $a_2 \notin V_3$), and, inductively, $a_i \notin V$ for all $i \in \mathbb{N}$. Hence V, being a subset of $E = \{a_i\}_{i\in\mathbb{N}}$, is empty. However, V cannot be empty because it is obtained as the intersection of non-empty nested compact sets (the proof that such an intersection is not empty follows along the same lines as the proof of the second half of (i)); this yields a contradiction.

1.9. $\mathcal{B} \subsetneq \mathcal{M}$

First of all, we know that $\mathcal{B} \subseteq \mathcal{M}$, because $\mathcal{M}$ is a σ-algebra containing open sets. We will now prove that the inclusion is strict, because there exists measurable sets that are not Borel.

Let P be the Cantor set.
We define the Lebesgue function $f : [0,1] \to \mathbb{R}$.
Given $x \in (0,1]$, we can write its binary expansion

$$x = \sum_{i=1}^{\infty} \frac{a_i}{2^i} = 0.a_1a_2a_3\dots \quad a_i \in \{0,1\} \quad \forall i \in \mathbb{N} \tag{1.11}$$

The binary expansion can be made unique if we identify the expansions

$$0.a_1\dots a_{k-1}01\dots1\dots \quad \text{and} \quad 0.a_1\dots a_{k-1}10\dots0\dots,$$

and assume that the expansions are of the first form (therefore, infinitely many a_n are equal to 1, except for $x = 0$). With this convention, we can define:

$$f(x) := \sum_{i=1}^{\infty} \frac{2a_i}{3^n} \tag{1.12}$$

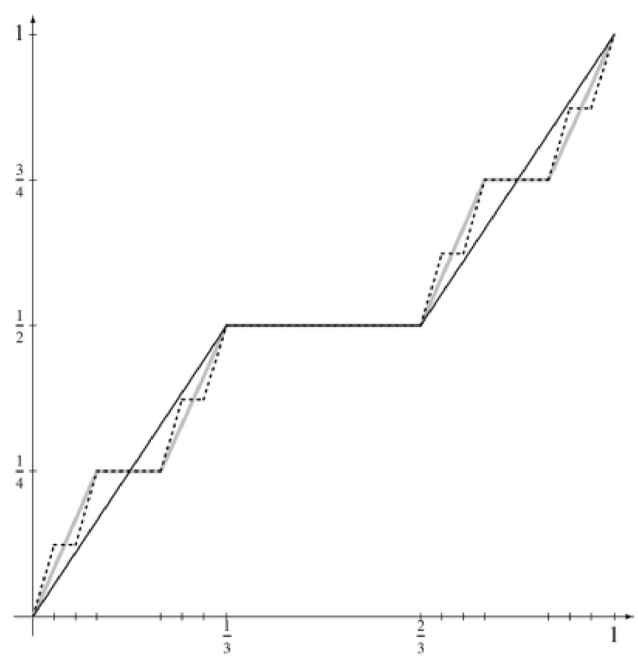


FIGURE 2. First few stages of the construction of the Cantor–Lebesgue function.

Lemma 1.24. *We can prove for f the following properties:*

(i) $f([0,1]) \subseteq P$;
(ii) f *is strictly monotone;*
(iii) f *is measurable.*

PROOF. Let us prove the various points of the lemma:

(i) since $2a_i \in \{0,2\}$, it follows from the definition of the Cantor set that $f([0,1]) \subseteq P$.

(ii) to prove monotonicity, we take $0 < x < y < 1$ and consider their binary expansions (unique with our convention):

$$x = \sum_{n=1}^{\infty} \frac{a_n}{2^n} \qquad y = \sum_{n=1}^{\infty} \frac{b_n}{2^n}.$$

We have that if $x < y$, then there exists $1 \leq k < +\infty$ such that $a_j = b_j \quad \forall j = 1, \ldots, k-1$ and $a_k < b_k$, which implies $a_k = 0$ and $b_k = 1$.
We then have

$$f(y) - f(x) = \frac{2}{3^k} + \sum_{n=k+1}^{\infty} \frac{2(b_n - a_n)}{3^n} \geq \frac{2}{3^k} - \sum_{n=k+1}^{\infty} \frac{2}{3^n} = \frac{1}{3^k} > 0.$$

(iii) from the previous point, we have that f is strictly increasing, and therefore it is measurable.

□

Lemma 1.25. *Let $f : \mathbb{R} \to \mathbb{R}^*$ be measurable. Then, $f^{-1}(B)$ is measurable $\forall B \in \mathcal{B}$.*

Proof. We claim that

$$A_f = \left\{ B \subseteq \mathbb{R} \ : \ f^{-1}(B) \text{ is measurable.} \right\}$$

is a σ-algebra containing intervals.
We can prove the claim using the fact that f is measurable:

(i) $f^{-1}(\mathbb{R}) = \mathbb{R}$, that is measurable;
(ii) $B \in A_f \Rightarrow \mathbb{R} \setminus B \in A_f$ because $f^{-1}(\mathbb{R} \setminus B) = \mathbb{R} \setminus f^{-1}(B)$. Since $B \in A_f$, $f^{-1}(B)$ is measurable, and therefore its complement is measurable as well;
(iii) $\{B_n\}_{n\in\mathbb{N}} \subset A_f \Rightarrow \bigcup_{n\in\mathbb{N}} B_n \in A_f$, because $f^{-1}\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \bigcup_{n\in\mathbb{N}} f^{-1}(B_n)$, and the countable union of measurable sets is measurable.

Then, since $\mathcal{B}$ is the smallest σ-algebra containing open sets, we can conclude the proof.

□

Now we can show that there exists measurable sets which are not Borel:

Theorem 1.26. *There exists a Lebesgue-measurable set $A \in \mathcal{M}$ such that $A \notin \mathcal{B}$.*

Proof. Take the Vitali set $V \subseteq [0, 1]$, that is non-measurable, and the Lebesgue function f. Let

$$B = f(V) \subset f([0, 1]) = P$$

We can prove that B is measurable: indeed, $B \subseteq P$, and P is a measurable set with measure 0. Therefore, its subset B is measurable as well by Lemma 1.8.

Let us now prove that B is not Borel: we assume by contradiction that B is Borel. Then $f^{-1}(B)$ is measurable by Lemma 1.25. However, since f is injective, we have that $f^{-1}(f(V)) = V$, which is not measurable and therefore leads to a contradiction. □

CHAPTER 2

Lebesgue integration

In this chapter, we follow closely the content of [Tao16, Chapter 8].

For the Riemann integral, the typical approach consists in first integrating a particularly simple class of functions (the piecewise constant functions). Once one learns how to integrate them, one can then integrate other Riemann integrable functions by a similar procedure and we shall use a similar philosophy to construct the Lebesgue integral.

We begin by considering a special subclass of measurable functions, called *simple functions*. Then we will show how to integrate simple functions, and then from there we will integrate all measurable functions (or at least the absolutely integrable ones).

2.1. Simple functions

DEFINITION (Simple functions). Let Ω be a measurable subset of $\mathbb{R}^d$, and let $f : \Omega \to \mathbb{R}$ be a measurable function. We say that f is a simple function if the image $f(\Omega)$ is finite. In other words, there exists a finite number of real numbers $c_1, c_2, \ldots, c_N$ such that for every $x \in \Omega$, we have $f(x) = c_j$ for some $1 \leq j \leq N$.

Example 2.1. Let Ω be a measurable subset of $\mathbb{R}^d$, and let E be a measurable subset of Ω. We define the characteristic function $\mathbb{1}_E : \Omega \to \mathbb{R}$ by setting $\mathbb{1}_E(x) := 1$ if $x \in E$, and $\mathbb{1}_E(x) := 0$ if $x \notin E$. Then $\mathbb{1}_E$ is a measurable function because E is measurable, and is a simple function, because the image $\mathbb{1}_E(\Omega)$ is $\{0, 1\}$ (or $\{0\}$ if E is empty, or $\{1\}$ if $E = \Omega$).

We remark on three basic properties of simple functions: they form an algebra, they are linear combinations of characteristic functions, and they approximate non-negative measurable functions. More precisely, we have the following three lemmas:

Lemma 2.1. *Let Ω be a measurable subset of $\mathbb{R}^d$, and let $f : \Omega \to \mathbb{R}$ and $g : \Omega \to \mathbb{R}$ be simple functions. Then $f + g$ and fg are also simple functions. Also, for any scalar $c \in \mathbb{R}$, the function cf is also a simple function.*

PROOF. It follows from the fact that $(f + g)(\Omega) \subset f(\Omega) + g(\Omega)$, $(fg)(\Omega) \subset f(\Omega) \cdot g(\Omega)$, and $(cf)(\Omega) = c \cdot (f(\Omega))$ are finite. ☐

Lemma 2.2. *Let Ω be a measurable subset of $\mathbb{R}^d$, and let $f : \Omega \to \mathbb{R}$ be a simple function. Then there exists a finite number of real numbers $c_1, \ldots, c_N$, and a finite number of disjoint measurable sets $E_1, E_2, \ldots, E_N$ in Ω, such that $f = \sum_{i=1}^N c_i \mathbb{1}_{E_i}$.*

PROOF. Define $\{c_1, \ldots, c_N\} := f(\Omega)$ and $E_j := f^{-1}(\{c_j\})$ to recover the claim. ☐

Lemma 2.3. *Let Ω be a measurable subset of $\mathbb{R}^d$, and let $f : \Omega \to \mathbb{R}$ be a measurable function. Suppose that f is always non-negative, i.e., $f(x) \geq 0$ for all $x \in \Omega$. Then there exists a sequence $f_1, f_2, f_3, \ldots$ of simple functions, $f_n : \Omega \to \mathbb{R}$, such that the f_n are non-negative and increasing,*

$$0 \leq f_1(x) \leq f_2(x) \leq f_3(x) \leq \ldots \text{ for every } x \in \Omega$$

and converge pointwise to f :

$$\lim_{n\to\infty} f_n(x) = f(x) \text{ for every } x \in \Omega$$

PROOF. The functions

$$f_n(x) := \sup_{j\in\mathbb{Z}} \left\{ \frac{j}{2^n} : \frac{j}{2^n} \le \min(f(x), 2^n) \right\} = \min\left(\frac{\lfloor 2^n f(x) \rfloor}{2^n}, 2^n \right)$$

are non-negative, increasing, and $\lim_{n\to\infty} f_n(x) = f(x)$.
First we prove that for all $a \in \mathbb{R}$ and all $n \in \mathbb{N}$ we have $\sup_{j\in\mathbb{Z}} \left\{ \frac{j}{2^n} : \frac{j}{2^n} \le a \right\} = \frac{\lfloor 2^n a \rfloor}{2^n}$. Let $b = \sup_{j\in\mathbb{Z}} \left\{ \frac{j}{2^n} : \frac{j}{2^n} \le a \right\}$, then $2^n b \in \mathbb{Z}$ and $2^n b \le 2^n a$, so $2^n b \le \lfloor 2^n a \rfloor$. For the converse inequality, we have that $\lfloor 2^n a \rfloor \le 2^n a$, so $\lfloor 2^n a \rfloor \le 2^n b$. This proves the formula.
To prove monotonicity, we prove that for all $a \in \mathbb{R}$ and all $n \in \mathbb{N}$ we have $\frac{\lfloor 2^n a \rfloor}{2^n} \le \frac{\lfloor 2^{n+1} a \rfloor}{2^{n+1}}$. To do this, notice that $\lfloor 2^n a \rfloor \le 2^n a$ and so $2\lfloor 2^n a \rfloor \le 2^{n+1} a$. Since $2\lfloor 2^n a \rfloor \in \mathbb{Z}$, we get $2\lfloor 2^n a \rfloor \le \lfloor 2^{n+1} a \rfloor$ and dividing by 2^{n+1} gives the claim.
To prove convergence, it suffices to see that for all $a \in \mathbb{R}$ and $n \in \mathbb{N}$ we have $0 \le a - \frac{\lfloor 2^n a \rfloor}{2^n} < 2^{-n}$. This can be seen through the inequality $\lfloor 2^n a \rfloor \le 2^n a < \lfloor 2^n a \rfloor + 1$. Since $2^n \to +\infty$ as $n \to +\infty$, we get pointwise convergence. ☐

Remark 2.1. One can actually approximate a non-negative measurable function via simple functions having compact support. Indeed, we can consider the sequence

$$\{\mathbb{1}_{B_k}\phi_k\}_{k\in\mathbb{N}} \text{ s.t. } \mathbb{1}_{B_k}\phi_k \uparrow f,$$

where ϕ_k are simple functions that approach f given by Lemma 2.3 and B_k are closed balls of radius k .

We now show how to compute the integral of simple functions.

DEFINITION (Lebesgue integral of simple functions). Let Ω be a measurable subset of $\mathbb{R}^d$, and let $f : \Omega \to \mathbb{R}$ be a simple function which is non-negative; thus f is measurable and the image $f(\Omega)$ is finite and contained in $[0, \infty)$. We then define the Lebesgue integral $\int_\Omega f$ of f on Ω by

$$\int_\Omega f := \sum_{\lambda\in f(\Omega);\lambda>0} \lambda m(\{x \in \Omega : f(x) = \lambda\}).$$

We will also sometimes write $\int_\Omega f$ as $\int_\Omega f dm$ (to emphasize the role of Lebesgue measure m) or use a dummy variable such as x, e.g., $\int_\Omega f(x)dx$.
Example 2.2. Let $f : \mathbb{R} \to \mathbb{R}$ be the function which equals 3 on the interval $[1, 2]$, equals 4 on the integral $(2, 4)$, and is zero everywhere else. Then

$$\int_\Omega f := 3 \times m([1, 2]) + 4 \times m((2, 4)) = 3 \times 1 + 4 \times 2 = 11.$$

Or if $g : \mathbb{R} \to \mathbb{R}$ is the function which equals 1 on $[0, \infty)$ and is zero everywhere else, then

$$\int_\Omega g = 1 \times m([0, \infty)) = 1 \times +\infty = +\infty$$

Thus the simple integral of a simple function can equal $+\infty$. (The reason why we restrict this integral to non-negative functions is to avoid ever encountering the indefinite form $+\infty + (-\infty)$).
Remark 2.2. Note that this definition of integral corresponds to one's intuitive notion of integration (at least of non-negative functions) as the area under the graph of the function (or volume, if one is in higher dimensions).

Another formulation of the integral for non-negative simple functions is as follows.

Lemma 2.4. *Let Ω be a measurable subset of $\mathbb{R}^d$, and let $E_1, \ldots, E_N$ are a finite number of disjoint measurable subsets in Ω. Let $c_1, \ldots, c_N$ be non-negative numbers (not necessarily distinct). Then* *we have*

$$\int_\Omega \sum_{j=1}^N c_j \mathbb{1}_{E_j} = \sum_{j=1}^N c_j m(E_j).$$

Proof. We can assume that none of the c_j are zero, since we can just remove them from the sum on both sides of the equation. Let $f := \sum_{j=1}^N c_j \mathbb{1}_{E_j}$. Then $f(x)$ is either equal to one of the c_j (if $x \in E_j$) or equal to 0 (if $x \notin \bigcup_{j=1}^N E_j$). Thus f is a simple function, and $f(\Omega) \subseteq \{0\} \cup \{c_j : 1 \le j \le N\}$. Thus, by the definition,

$$\int_\Omega f = \sum_{\lambda \in \{c_j : 1 \le j \le N\}} \lambda m(\{x \in \Omega : f(x) = \lambda\})$$

$$= \sum_{\lambda \in \{c_j : 1 \le j \le N\}} \lambda m \left(\bigcup_{1 \le j \le N : c_j = \lambda} E_j \right).$$

But by the finite additivity property of Lebesgue measure, this is equal to

$$\sum_{\lambda \in \{c_j : 1 \le j \le N\}} \lambda \sum_{1 \le j \le N : c_j = \lambda} m(E_j)$$

$$\sum_{\lambda \in \{c_j : 1 \le j \le N\}} \sum_{1 \le j \le N : c_j = \lambda} c_j m(E_j).$$

Each j appears exactly once in this sum, since c_j is only equal to exactly one value of λ. So the above expression is equal to $\sum_{1 \le j \le N} c_j m(E_j)$ as desired. □

Some basic properties of Lebesgue integration of non-negative simple functions:

Proposition 2.5. *Let Ω be a measurable set, and let $f : \Omega \to \mathbb{R}$ and $g : \Omega \to \mathbb{R}$ be non-negative simple functions.*

(i) We have $0 \le \int_\Omega f \le +\infty$. Furthermore, we have $\int_\Omega f = 0$ if and only if $m(\{x \in \Omega : f(x) \neq 0\}) = 0$.
(ii) We have $\int_\Omega (f + g) = \int_\Omega f + \int_\Omega g$.
(iii) For any positive number c, we have $\int_\Omega cf = c \int_\Omega f$.
(iv) If $f(x) \le g(x)$ for almost every $x \in \Omega$, then we have $\int_\Omega f \le \int_\Omega g$.

We make a very convenient notational convention: if a property $P(x)$ holds for all points in Ω, except for a set of measure zero, then we say that P holds for almost every point in Ω. Thus (i) asserts that $\int_\Omega f = 0$ if and only if f is zero for almost every point in Ω.

Proof. From Lemma 2.2 or from the formula

$$f = \sum_{\lambda \in f(\Omega) \setminus \{0\}} \lambda \mathbb{1}_{\{x \in \Omega : f(x) = \lambda\}}$$

we can write f as a combination of characteristic functions, say

$$f = \sum_{j=1}^N c_j \mathbb{1}_{E_j}$$

where $E_1, \dots, E_N$ are disjoint subsets of Ω and the c_j are positive. Similarly we can write

$$g = \sum_{k=1}^M d_k \mathbb{1}_{F_k}$$

where $F_1, \ldots, F_M$ are disjoint subsets of Ω and the d_k are positive.

(i) Since $\int_\Omega f = \sum_{j=1}^N c_j m(E_j)$ it is clear that the integral is between 0 and infinity. If f is zero almost everywhere, then all of the E_j must have measure zero and so $\int_\Omega f = 0$. Conversely, if $\int_\Omega f = 0$, then $\sum_{j=1}^N c_j m(E_j) = 0$, which can only happen when all of the $m(E_j)$ are zero (since all the c_j are positive). But then $\bigcup_{j=1}^N E_j$ has measure zero, and hence f is zero almost everywhere in Ω.

(ii) Write $E_0 := \Omega \backslash \bigcup_{j=1}^N E_j$ and $c_0 := 0$, then we have $\Omega = E_0 \cup E_1 \cup \ldots \cup E_N$ and

$$f = \sum_{j=0}^N c_j \mathbb{1}_{E_j}.$$

Similarly if we write $F_0 := \Omega \backslash \bigcup_{k=1}^M F_k$ and $d_0 := 0$ then

$$g = \sum_{k=0}^M d_k \mathbb{1}_{F_k}$$

Since $\Omega = E_0 \cup \ldots \cup E_N = F_0 \cup \ldots \cup F_M$, we have

$$f = \sum_{j=0}^N \sum_{k=0}^M c_j \mathbb{1}_{E_j \cap F_k}$$

and

$$g = \sum_{k=0}^M \sum_{j=0}^N d_k \mathbb{1}_{E_j \cap F_k}$$

and hence

$$f + g = \sum_{0 \le j \le N; 0 \le k \le M} (c_j + d_k) \mathbb{1}_{E_j \cap F_k}$$

By Lemma 2.4, we thus have

$$\int_\Omega (f+g) = \sum_{0 \le j \le N; 0 \le k \le M} (c_j + d_k) m(E_j \cap F_k).$$

On the other hand, we have

$$\int_\Omega f = \sum_{0 \le j \le N} c_j m(E_j) = \sum_{0 \le j \le N; 0 \le k \le M} c_j m(E_j \cap F_k)$$

and similarly

$$\int_\Omega g = \sum_{0 \le k \le M} d_k m(F_k) = \sum_{0 \le j \le N; 0 \le k \le M} d_k m(E_j \cap F_k)$$

and the claim (ii) follows.

(iii) Since $cf = \sum_{j=1}^N c c_j \mathbb{1}_{E_j}$, we have

$$\int_\Omega cf = \sum_{j=1}^N c c_j m(E_j).$$

Since $\int_\Omega f = \sum_{j=1}^N c_j m(E_j)$, the claim follows.

(iv) First assume that $f(x) \leq g(x)$ for every $x \in \Omega$. Set $h = g - f$, then h is a non-negative simple function. By (ii), we thus have $\int_\Omega g = \int_\Omega f + \int_\Omega h$. From (i), we know that $\int_\Omega h \geq 0$ and so $\int_\Omega f \leq \int_\Omega g$.
To treat the general case, let $A = \{x \in \Omega : f(x) > g(x)\}$, then by assumption we have $m(A) = 0$ and $f(x)\mathbb{1}_{A^c}(x) \leqslant g(x)\mathbb{1}_{A^c}(x)$ for every $x \in \Omega$. Since the product of two non-negative simple functions is a non-negative simple function (Lemma 2.1), by the preceding discussion, we get $\int_\Omega f\mathbb{1}_{A^c} \leq \int_\Omega g\mathbb{1}_{A^c}$. To conclude, it suffices to prove that $\int_\Omega f = \int_\Omega f\mathbb{1}_{A^c}$ and similarly for g. To see this, note that $f\mathbb{1}_{A^c} = \sum_{j=1}^N c_j \mathbb{1}_{E_j \cap A^c}$. Since $E_1, \dots, E_N$ are disjoint, we have that $E_1 \cap A^c, \dots, E_N \cap A^c$ are also disjoint, thus

$$\int_\Omega f = \sum_{j=1}^N c_j m(E_j) = \sum_{j=1}^N c_j m(E_j \cap A^c) = \int_\Omega f\mathbb{1}_{A^c}.$$

□

2.2. Integration of non-negative measurable functions

We now pass from the integration of non-negative simple functions to the integration of non-negative measurable functions. We will allow our measurable functions to take the value of $+\infty$ sometimes.

Definition (Majorization). Let $f : \Omega \to \mathbb{R}$ and $g : \Omega \to \mathbb{R}$ be functions. We say that f majorizes g, or g minorizes f, iff we have $f(x) \geq g(x)$ for every $x \in \Omega$.

We sometimes use the phrase "f dominates g" instead of "f majorizes g".

Definition (Lebesgue integral for non-negative functions). Let Ω be a measurable subset of $\mathbb{R}^d$, and let $f : \Omega \to [0, \infty]$ be measurable and non-negative. Then we define the Lebesgue integral $\int_\Omega f$ of f on Ω to be

$$\int_\Omega f := \sup \left\{ \int_\Omega s : s \text{ is simple and non-negative, and minorizes } f \right\}.$$

Remark 2.3. We can compare this notion to that of a lower Riemann integral, but interestingly we will not need to match this lower integral with an upper integral here.

Remark 2.4. Note that if Ω' is any measurable subset of Ω, then we can define $\int_{\Omega'} f$ as well by restricting f to Ω', thus $\int_{\Omega'} f := \int_{\Omega'} f\big|_{\Omega'}$.

We have to check that this definition is consistent with our previous notion of Lebesgue integral for non-negative simple functions; in other words, if $f : \Omega \to \mathbb{R}$ is a non-negative simple function, then the value of $\int_\Omega f$ given by this definition should be the same as the one given in the previous definition. But this is clear because f certainly minorizes itself, and any other non-negative simple function s which minorizes f will have an integral $\int_\Omega s$ less than or equal to $\int_\Omega f$, thanks to Proposition 2.5 (iv).

Remark 2.5. Note that $\int_\Omega f$ is always at least 0 , since 0 is simple, non-negative, and minorizes f. Of course, $\int_\Omega f$ could equal $+\infty$. Some basic properties of the Lebesgue integral on non-negative measurable functions (which supersede Proposition 2.5):

Proposition 2.6. *Let Ω be a measurable set, and let $f : \Omega \to [0, +\infty]$ and $g : \Omega \to [0, +\infty]$ be non-negative measurable functions.*

(i) We have $0 \leq \int_\Omega f \leq +\infty$. Furthermore, we have $\int_\Omega f = 0$ if and only if $f(x) = 0$ for almost every $x \in \Omega$.

(ii) For any positive number c, we have $\int_\Omega cf = c\int_\Omega f$.

(iii) If $f(x) \leq g(x)$ for almost every $x \in \Omega$, then we have $\int_\Omega f \leq \int_\Omega g$.

(iv) If $f(x) = g(x)$ for almost every $x \in \Omega$, then $\int_\Omega f = \int_\Omega g$.
(v) If $\Omega' \subseteq \Omega$ is measurable, then $\int_{\Omega'} f = \int_\Omega f \mathbb{1}_{\Omega'} \leq \int_\Omega f$.

PROOF. (i) Observe that $\int_\Omega f \geq 0$ because $s \equiv 0$ is a simple function that minorizes f. Furthermore, if $\int_\Omega f = 0$ then $f = 0$ almost everywhere. Suppose for contradiction this is not the case. Let say $f \geq \delta > 0$ on a subset $E \subset \Omega$ of positive measure. Then $h := \delta \mathbb{1}_E$ is a simple function that minorizes f. Therefore

$$\int_\Omega f \geq \int_\Omega h > 0,$$

which is a contradiction.

(ii) Notice that:

$$\begin{aligned} c\int_\Omega f &= c \sup \left\{ \int_\Omega s : s \text{ is simple and non-negative, and minorizes } f \right\} \\ &= \sup \left\{ \int_\Omega cs : s \text{ is simple and non-negative, and minorizes } f \right\} \\ &= \sup \left\{ \int_\Omega s : s \text{ is simple and non-negative, and minorizes } cf \right\} = \int_\Omega cf, \end{aligned}$$

where we have used the fact that if s is a simple function that minorizes f, then cs is a simple function that minorizes cf.

(iii) Let $A = \{x \in \Omega : f(x) > g(x)\}$ then by assumption we have $m(A) = 0$. For any s non-negative simple function minorizing f, we have that $s\mathbb{1}_{\Omega \setminus A}$ is a non-negative simple function minorizing g. From the proof of Proposition 2.5 (iv), we see that $\int_\Omega s = \int_\Omega s \mathbb{1}_{A^c} \leqslant \int_\Omega g$. Since s was arbitrary we deduce that $\int_\Omega f \leq \int_\Omega g$.

(iv) Comes by applying (iii) in both directions.

(v) Comes from (iii) applied to $f\mathbb{1}_{\Omega'} \leq f\mathbb{1}_\Omega$:

$$\int_{\Omega'} f = \int_\Omega f \mathbb{1}_{\Omega'} \leq \int_\Omega f \mathbb{1}_\Omega = \int_\Omega f.$$

□

Remark 2.6. Proposition 2.6 (iv) is quite interesting; it says that one can modify the values of a function on any measure zero set (e.g., you can modify a function on every rational number), and not affect its integral at all. It is as if no individual point, or even a measure zero collection of points, has any “vote" in what the integral of a function should be; only the collective set of points has an influence on an integral.

Remark 2.7. Note that we do not yet try to interchange sums and integrals. From the definition it is fairly easy to prove that $\int_\Omega (f+g) \geq \int_\Omega f + \int_\Omega g$, but to prove equality requires more work and will be done later.

With the Lebesgue integral it is possible to interchange an integral with a limit if the functions are increasing:

THEOREM 2.7 (Monotone convergence theorem). *Let Ω be a measurable subset of $\mathbb{R}^d$, and let $(f_n)_{n=1}^\infty$ be a sequence of non-negative measurable functions from Ω to $\mathbb{R}$ which are increasing in the sense that*

$$0 \leq f_1(x) \leq f_2(x) \leq f_3(x) \leq \dots \text{ for almost every } x \in \Omega.$$

(Note we are assuming that $f_n(x)$ is increasing with respect to n; this is a different notion from $f_n(x)$ increasing with respect to x.) Then we have

$$0 \leq \int_\Omega f_1 \leq \int_\Omega f_2 \leq \int_\Omega f_3 \leq \dots$$

and

$$\int_\Omega \sup_n f_n = \sup_n \int_\Omega f_n.$$

PROOF. Let $A = \{x \in \Omega : (f_n(x))_{n\in\mathbb{N}} \text{ is not increasing}\}$, then by assumption $m(A) = 0$. Using Proposition 2.6 (iv), we have $\int_\Omega f_n = \int_\Omega f_n \mathbb{1}_{\Omega\setminus A}$ for every $n \in \mathbb{N}$. Since $\sup_{n\in\mathbb{N}} f_n = \sup_{n\in\mathbb{N}} f_n \mathbb{1}_{\Omega\setminus A}$ almost everywhere on Ω, we have by Proposition 2.6 (iv) that $\int_\Omega \sup_{n\in\mathbb{N}} f_n = \int_\Omega \sup_{n\in\mathbb{N}} f_n \mathbb{1}_{\Omega\setminus A}$. So we may assume that $(f_n(x))_{n\in\mathbb{N}}$ is increasing for every $x \in \Omega$.
The first conclusion is clear from Proposition 2.6 (iii). Now we prove the second conclusion. From Proposition 2.6 (iii) again we have

$$\int_\Omega \sup_m f_m \geq \int_\Omega f_n$$

for every n; taking suprema in n we obtain

$$\int_\Omega \sup_m f_m \geq \sup_n \int_\Omega f_n$$

which is one half of the desired conclusion. To finish the proof we have to show

$$\int_\Omega \sup_m f_m \leq \sup_n \int_\Omega f_n$$

From the definition of $\int_\Omega \sup_m f_m$, it will suffice to show that

$$\int_\Omega s \leq \sup_n \int_\Omega f_n$$

for all simple non-negative functions which minorize $\sup_m f_m$.

Fix s. We will show that

$$(1-\varepsilon) \int_\Omega s \leq \sup_n \int_\Omega f_n$$

for every $0 < \varepsilon < 1$; the claim then follows by taking limits as $\varepsilon \to 0$. Fix ε. By construction of s, we have

$$s(x) \leq \sup_n f_n(x)$$

for every $x \in \Omega$. Hence, for every $x \in \Omega$ there exists an N (depending on x) such that

$$f_N(x) \geq (1-\varepsilon)s(x)$$

Since the f_n are increasing, this will imply that $f_n(x) \geq (1-\varepsilon)s(x)$ for all $n \geq N$. Thus, if we define the sets E_n by

$$E_n := \{x \in \Omega : f_n(x) \geq (1-\varepsilon)s(x)\}$$

then we have $E_1 \subset E_2 \subset E_3 \subset \dots$ and $\bigcup_{n=1}^\infty E_n = \Omega$.

From Proposition 2.6 (v) we have

$$(1-\varepsilon) \int_{E_n} s = \int_{E_n} (1-\varepsilon)s \leq \int_{E_n} f_n \leq \int_\Omega f_n$$

so to finish the argument it will suffice to show that

$$\sup_n \int_{E_n} s = \int_\Omega s$$

Since s is a simple function, we may write $s = \sum_{j=1}^{N} c_j \mathbb{1}_{F_j}$ for some measurable F_j and positive c_j. Since

$$\int_\Omega s = \sum_{j=1}^{N} c_j m(F_j)$$

and

$$\int_{E_n} s = \int_{E_n} \sum_{j=1}^{N} c_j \mathbb{1}_{F_j \cap E_n} = \sum_{j=1}^{N} c_j m(F_j \cap E_n)$$

it thus suffices to show that

$$\sup_n m(F_j \cap E_n) = m(F_j).$$

This follows because $F_j \cap E_n \leq F_j$ and $F_j \cap E_n \uparrow F_j$. □

We can now interchange addition and integration thanks to this theorem:

Lemma 2.8 (Interchange of addition and integration)**.** *Let Ω be a measurable subset of $\mathbb{R}^d$, and let $f : \Omega \to [0, +\infty]$ and $g : \Omega \to [0, +\infty]$ be measurable functions. Then $\int_\Omega (f + g) = \int_\Omega f + \int_\Omega g$.*

Proof. By Lemma 2.3, there exists a sequence $0 \leq s_1 \leq s_2 \leq \ldots \leq f$ of simple functions such that $\sup_n s_n = f$, and similarly a sequence $0 \leq t_1 \leq t_2 \leq \ldots \leq g$ of simple functions such that $\sup_n t_n = g$. Since the s_n are increasing and the t_n are increasing, it is then easy to check that $s_n + t_n$ is also increasing and $\sup_n (s_n + t_n) = f + g$. By the monotone convergence theorem (Theorem 2.7) we thus have

$$\int_\Omega f = \sup_n \int_\Omega s_n$$

$$\int_\Omega g = \sup_n \int_\Omega t_n$$

$$\int_\Omega (f + g) = \sup_n \int_\Omega (s_n + t_n).$$

But by Proposition 2.5 (ii) we have $\int_\Omega (s_n + t_n) = \int_\Omega s_n + \int_\Omega t_n$. By Proposition 2.5 (iv), $\int_\Omega s_n$ and $\int_\Omega t_n$ are both increasing in n, so

$$\sup_n \left(\int_\Omega s_n + \int_\Omega t_n \right) = \left(\sup_n \int_\Omega s_n \right) + \left(\sup_n \int_\Omega t_n \right)$$

and the claim follows. □

Of course, once one can interchange an integral with a sum of two functions, one can handle an integral and any finite number of functions by induction. More surprisingly, one can handle infinite sums as well of non-negative functions:

Corollary 2.9. *If Ω is a measurable subset of $\mathbb{R}^d$, and $g_1, g_2, \ldots$ are a sequence of non-negative measurable functions from Ω to $[0, +\infty]$, then*

$$\int_\Omega \sum_{n=1}^{\infty} g_n = \sum_{n=1}^{\infty} \int_\Omega g_n.$$

Proof. We have

$$\sum_{n=1}^{\infty} \int_\Omega g_n = \lim_{N \to \infty} \sum_{n=1}^{N} \int_\Omega g_n = \lim_{N \to \infty} \int_\Omega \sum_{n=1}^{N} g_n = \int_\Omega \sum_{n=1}^{\infty} g_n,$$

where the last equality is given by Theorem 2.7 (by non-negativity of g_n). □

Remark 2.8. Note that we do not need to assume anything about the convergence of the above sums; it may well happen that both sides are equal to $+\infty$. However, we do need to assume non-negativity.

One could similarly ask whether we could interchange limits and integrals; in other words, is it true that

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

Unfortunately, this is not true, as the following “moving bump" example shows.

Example 2.3. For each $n = 1, 2, 3\ldots$, let $f_n : \mathbb{R} \to \mathbb{R}$ be the function $f_n = \mathbb{1}_{[n,n+1)}$. Then $\lim_{n\to\infty} f_n(x) = 0$ for every x, but $\int_\mathbb{R} f_n = 1$ for every n, and hence $\lim_{n\to\infty} \int_\mathbb{R} f_n = 1 \neq 0$. In other words, the limiting function $\lim_{n\to\infty} f_n$ can end up having significantly smaller integral than any of the original integrals.

However, the following very useful lemma of Fatou shows that the reverse cannot happen - there is no way the limiting function has larger integral than the limit of the original integrals:

Lemma 2.10 (Fatou's lemma)**.** *Let Ω be a measurable subset of $\mathbb{R}^d$, and let $f_1, f_2, \ldots$ be a sequence of measurable function from Ω to $\mathbb{R}$ such that $f_n \geq 0$ almost everywhere on Ω for all $n \in \mathbb{N}$. Then*

$$\int_\Omega \liminf_{n\to\infty} f_n \leq \liminf_{n\to\infty} \int_\Omega f_n.$$

Proof. For every $n \in \mathbb{N}$, let $A_n = \{x \in \Omega : f_n(x) < 0\}$, then by assumption $m(A_n) = 0$ and by countable subadditivity, we deduce that $m(A) = 0$ where $A = \bigcup_{n\in\mathbb{N}} A_n$. By Proposition 2.6 (iv), we have that $\int_\Omega f_n = \int_\Omega f_n \mathbb{1}_{\Omega\setminus A}$ for every $n \in \mathbb{N}$ and $\int_\Omega \liminf_{n\to\infty} f_n = \int_\Omega \liminf_{n\to\infty} f_n \mathbb{1}_{\Omega\setminus A}$. So we may assume that $f_n(x) \geq 0$ for every $x \in \Omega$ and every $n \in \mathbb{N}$.
Recall that

$$\liminf_{n\to\infty} f_n = \sup_n \left(\inf_{m\geq n} f_m \right)$$

and hence by the monotone convergence theorem

$$\int_\Omega \liminf_{n\to\infty} f_n = \sup_n \int_\Omega \left(\inf_{m\geq n} f_m \right).$$

By Proposition 2.6 (iii) we have

$$\int_\Omega \left(\inf_{m\geq n} f_m \right) \leq \int_\Omega f_j$$

for every $j \geq n$; taking infima in j we obtain

$$\int_\Omega \left(\inf_{m\geq n} f_m \right) \leq \inf_{j\geq n} \int_\Omega f_j.$$

Thus

$$\int_\Omega \liminf_{n\to\infty} f_n \leq \sup_n \inf_{j\geq n} \int_\Omega f_j = \liminf_{n\to\infty} \int_\Omega f_n$$

as desired. ☐

Note that we are allowing our functions to take the value $+\infty$ at some points. It is even possible for a function to take the value $+\infty$ but still have a finite integral; for instance, if E is a measure zero set, and $f : \Omega \to \mathbb{R}$ is equal to $+\infty$ on E but equals 0 everywhere else, then $\int_\Omega f = 0$ by Proposition 2.6 (i). However, if the integral is finite, the function must be finite almost everywhere:

Lemma 2.11. *Let Ω be a measurable subset of $\mathbb{R}^d$, and let $f : \Omega \to [0, +\infty]$ be a non-negative measurable function such that $\int_\Omega f$ is finite. Then f is finite almost everywhere (i.e., the set $\{x \in \Omega : f(x) = +\infty\}$ has measure zero).*

PROOF. Suppose for contradiction this is not the case. Denote

$$E := \{x \in \Omega : f(x) = +\infty\}$$

and suppose that $m(E) = \delta > 0$. Then, the function

$$g(x) = \begin{cases} +\infty & \text{if } x \in E \\ 0 & \text{otherwise} \end{cases}$$

is dominated by f. Therefore

$$\int_\Omega f \geq \int_\Omega g = +\infty,$$

which is a contradiction. □

Form Corollary 2.9 and Lemma 2.11 one has a useful lemma.

Lemma 2.12 (Borel-Cantelli lemma). *Let $\Omega_1, \Omega_2, \ldots$ be measurable subsets of $\mathbb{R}^d$ such that $\sum_{n=1}^\infty m(\Omega_n)$ is finite. Then the set*

$$\Omega := \left\{x \in \mathbb{R}^d : x \in \Omega_n \text{ for infinitely many } n\right\}$$

is a set of measure zero. In other words, almost every point belongs to only finitely many Ω_n.

PROOF. We observe that

$$\Omega = \bigcap_{n=1}^\infty E_n, \qquad \text{with } E_n := \bigcup_{k=n}^\infty \Omega_k.$$

Let $\varepsilon > 0$ be given. We claim that $m(\Omega) < \varepsilon$. By the arbitrariness of $\varepsilon > 0$, this will suffice to conclude the proof.

Since $\sum_{k=1}^\infty m(\Omega_k) < \infty$, we can find a sufficiently large $n \in \mathbb{N}$ such that $\sum_{k=n}^\infty m(\Omega_k) < \varepsilon$. Therefore, by subadditivity,

$$m(E_n) \leq \sum_{k=n}^\infty m(\Omega_k) < \varepsilon.$$

On the other hand, we have $\Omega \subset E_n$ for all $n \in \mathbb{N}$, so we conclude that $m(\Omega) < \varepsilon$. □

2.3. Integration of absolutely integrable functions

We have now completed the theory of the Lebesgue integral for nonnegative functions. Now we consider how to integrate functions which can be both positive and negative. However, we do wish to avoid the indefinite expression $+\infty + (-\infty)$, so we will restrict our attention to a subclass of measurable functions - the absolutely integrable functions.

DEFINITION (Absolutely integrable functions). Let Ω be a measurable subset of $\mathbb{R}^d$. A measurable function $f : \Omega \to \mathbb{R}^*$ is said to be absolutely integrable if the integral $\int_\Omega |f|$ is finite.

Of course, $|f|$ is always non-negative, so this definition makes sense even if f changes sign. Absolutely integrable functions are also known as $L^1(\Omega)$ functions.

If $f : \Omega \to \mathbb{R}^*$ is a function, we define the positive part $f^+ : \Omega \to [0, +\infty]$ and negative part $f^- : \Omega \to [0, +\infty]$ by the formulae

$$f^+ := \max(f, 0); \quad f^- := -\min(f, 0).$$

From Corollary 1.19 we know that f^+ and $f^-$are measurable. Observe also that $f^+$and f^- are non-negative, that $f = f^+ - f^-$, and $|f| = f^+ + f^-$.

Definition (Lebesgue integral). Let $f : \Omega \to \mathbb{R}^*$ be an absolutely integrable function. We define the Lebesgue integral $\int_\Omega f$ of f to be the quantity

$$\int_\Omega f := \int_\Omega f^+ - \int_\Omega f^-$$

Note that since f is absolutely integrable, $\int_\Omega f^+$and $\int_\Omega f^-$are less than or equal to $\int_\Omega |f|$ and hence are finite. Thus $\int_\Omega f$ is always finite; we are never encountering the indeterminate form $+\infty - (+\infty)$.

Note that this definition is consistent with our previous definition of the Lebesgue integral for non-negative functions, since if f is nonnegative then $f^+ = f$ and $f^- = 0$. We also have the useful triangle inequality

$$\left| \int_\Omega f \right| \leq \int_\Omega f^+ + \int_\Omega f^- = \int_\Omega |f|.$$

Proposition 2.13. *Let Ω be a measurable set, and let $f : \Omega \to \mathbb{R}$ and $g : \Omega \to \mathbb{R}$ be absolutely integrable functions.*

(i) For any real number c (positive, zero, or negative), we have that cf is absolutely integrable and $\int_\Omega cf = c \int_\Omega f$.
(ii) The function $f + g$ is absolutely integrable, and $\int_\Omega (f + g) = \int_\Omega f + \int_\Omega g$.
(iii) If $f(x) \leq g(x)$ for almost every $x \in \Omega$, then we have $\int_\Omega f \leq \int_\Omega g$.
(iv) If $f(x) = g(x)$ for almost every $x \in \Omega$, then $\int_\Omega f = \int_\Omega g$.

Proof. (i) First of all, notice that since f is absolutely integrable, cf is also absolutely integrable. Now, if $c = 0$, the result is obvious. If c is positive, we have using the linearity of the integral for nonnegative functions that

$$\int_\Omega (cf)\, dx = \int_\Omega (cf)^+\, dx - \int_\Omega (cf)^-\, dx = \int_\Omega cf^+\, dx - \int_\Omega cf^-\, dx = c \int_\Omega f\, dx. \tag{2.1}$$

If c is negative instead, we have

$$\int_\Omega (cf)\, dx = \int_\Omega (cf)^+\, dx - \int_\Omega (cf)^-\, dx = \int_\Omega |c| f^-\, dx - \int_\Omega |c| f^+\, dx = -\left(\int_\Omega |c| f^+\, dx - \int_\Omega |c| f^-\, dx \right)$$
$$= -\int_\Omega |c| f\, dx \overset{(2.1)}{=} -|c| \int_\Omega f\, dx = c \int_\Omega f\, dx.$$

(ii) We begin by showing that $f+g$ is absolutely integrable. Indeed, by the triangular inequality and the monotonicity of the integral for nonnegative functions, we have

$$\int_\Omega |f + g|\, dx \leq \int_\Omega |f|\, dx + \int_\Omega |g|\, dx < +\infty.$$

Note that

$$(f + g)^+ - (f + g)^- = f + g = f^+ - f^- + g^+ - g^-,$$

and so

$$(f + g)^+ + f^- + g^- = (f + g)^- + f^+ + g^+.$$

Therefore, using linearity of the integral for nonnegative functions,

$$\int_\Omega (f + g)^+ dx + \int_\Omega f^- dx + \int_\Omega g^- dx = \int_\Omega (f + g)^- dx + \int_\Omega f^+ dx + \int_\Omega g^+ dx. \tag{2.2}$$

Thus,

$$\int_\Omega (f+g)\,dx \overset{\text{def}}{=} \int_\Omega (f+g)^+\,dx - \int_\Omega (f+g)^-\,dx$$
$$\overset{(2.2)}{=} \int_\Omega f^+\,dx - \int_\Omega f^-\,dx + \int_\Omega g^+\,dx - \int_\Omega g^-\,dx$$
$$\overset{\text{def}}{=} \int_\Omega f\,dx + \int_\Omega g\,dx.$$

(iii) The assumption $f(x) \le g(x)$ guarantees $f^+(x) \le g^+(x)$ and $g^-(x) \le f^-(x)$ for almost every $x \in \Omega$ and therefore, by the monotonicity of the integral for nonnegative functions,

$$\int_\Omega f\,dx = \int_\Omega f^+\,dx - \int_\Omega f^-\,dx \le \int_\Omega g^+\,dx - \int_\Omega g^-\,dx = \int_\Omega g\,dx.$$

(iv) Comes by applying (iii) in both directions.

□

As mentioned in the previous section, one cannot necessarily interchange limits and integrals, $\lim \int f_n = \int \lim f_n$, as the “moving bump example" showed. However, it is possible to exclude the moving bump example, and successfully interchange limits and integrals, if we know that the functions f_n are all majorized by a single absolutely integrable function. This important theorem is known as the Lebesgue dominated convergence theorem, and is extremely useful:

THEOREM 2.14 (Lebesgue dominated convergence thm). *Let Ω be a measurable subset of $\mathbb{R}^d$, and let $f_1, f_2, \dots$ be a sequence of measurable functions from Ω to $\mathbb{R}^*$ which converge pointwise almost everywhere. Suppose also that there is an absolutely integrable function $F : \Omega \to [0, +\infty]$ such that $|f_n(x)| \le F(x)$ for almost every $x \in \Omega$ and all $n = 1, 2, 3, \dots$. Then*

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

PROOF. Let $f : \Omega \to \mathbb{R}^*$ be the function $f(x) := \limsup_{n\to\infty} f_n(x)$. By Lemma 1.22, f is measurable. Also, since $|f_n(x)| \le F(x)$ for all n and almost all $x \in \Omega$, we see that each f_n is absolutely integrable, and by taking limits we obtain $|f(x)| \le F(x)$ for almost all $x \in \Omega$, so f is also absolutely integrable. Let us now define $\tilde{F}$ that dominates f_n *everywhere*: to do so, consider the set

$$A := \bigcup_{n\in\mathbb{N}} \{x \in \Omega \text{ s.t. } F(x) < f_n(x)\},$$

and set $\tilde{F} = +\infty$ on A, $\tilde{F} = F$ on $\Omega \setminus A$. Note that $m(A) = 0$ because it is a countable union of sets with zero measure, and therefore $\tilde{F}$ is still absolutely integrable.
Our task is now to show that $\lim_{n\to\infty} \int_\Omega f_n = \int_\Omega f$. Note that the functions $\tilde{F} + f_n$ are non-negative *everywhere* and converge pointwise almost everywhere to $\tilde{F} + f$. We can apply Fatou's lemma (Lemma 2.10) combined with Proposition 2.6 (iv):

$$\int_\Omega \tilde{F} + f \le \liminf_{n\to\infty} \int_\Omega \tilde{F} + f_n$$

and thus

$$\int_\Omega f \le \liminf_{n\to\infty} \int_\Omega f_n$$

But the functions $\tilde{F}-f_n$ are also non-negative *everywhere* and converge pointwise almost everywhere to $\tilde{F}-f$. We can apply Fatou's lemma combined with Proposition 2.6 (iv) again:

$$\int_\Omega \tilde{F}-f \le \liminf_{n\to\infty} \int_\Omega \tilde{F}-f_n$$

Since the right-hand side is $\int_\Omega \tilde{F} - \limsup_{n\to\infty} \int_\Omega f_n$ (why did the lim inf become a lim sup?), we thus have

$$\int_\Omega f \ge \limsup_{n\to\infty} \int_\Omega f_n$$

Thus the lim inf and lim sup of $\int_\Omega f_n$ are both equal to $\int_\Omega f$, as desired. □

Finally, we record a lemma which is not particularly interesting in itself, but will have some useful consequences later in these notes.

DEFINITION (Upper and lower Lebesgue integral). Let Ω be a measurable subset of $\mathbb{R}^d$, and let $f:\Omega\to\mathbb{R}$ be a function (not necessarily measurable). We define the upper Lebesgue integral $\overline{\int_\Omega} f$ to be

$$\overline{\int_\Omega} f := \inf\Big\{ \int_\Omega g : g \text{ is an absolutely integrable function from } \Omega \text{ to } \mathbb{R} \text{ that majorizes } f \Big\}$$

and the lower Lebesgue integral $\underline{\int_\Omega} f$ to be

$$\underline{\int_\Omega} f := \sup\Big\{ \int_\Omega g : g \text{ is an absolutely integrable function from } \Omega \text{ to } \mathbb{R} \text{ that minorizes } f \Big\}.$$

Here we adopt the convention $\inf\varnothing = +\infty$ and $\sup\varnothing = -\infty$.
It is easy to see that $\underline{\int_\Omega} f \le \overline{\int_\Omega} f$ (use Proposition 2.13 (iii)). When f is absolutely integrable then equality occurs. The converse is also true:

Lemma 2.15. *Let Ω be a measurable subset of $\mathbb{R}^d$, and let $f:\Omega\to\mathbb{R}$ be a function (not necessarily measurable). Let A be a real number, and suppose $\overline{\int_\Omega} f = \underline{\int_\Omega} f = A$. Then f is absolutely integrable, and*

$$\int_\Omega f = \overline{\int_\Omega} f = \underline{\int_\Omega} f = A$$

PROOF. By definition of upper Lebesgue integral, for every integer $n\ge 1$ we may find an absolutely integrable function $f_n^+:\Omega\to\mathbb{R}$ which majorizes f such that

$$\int_\Omega f_n^+ \le A + \frac{1}{n}$$

Similarly we may find an absolutely integrable function $f_n^-:\Omega\to\mathbb{R}$ which minorizes f such that

$$\int_\Omega f_n^- \ge A - \frac{1}{n}$$

Let $F^+ := \inf_n f_n^+$ and $F^- := \sup_n f_n^-$. Then F^+ and F^- are measurable (by Lemma 1.22) and absolutely integrable (because they are squeezed between the absolutely integrable functions f_1^+

and f_1^-, for instance). Also, F^+ majorizes f and F^- minorizes f. Finally, we have

$$\int_\Omega F^+ \leq \int_\Omega f_n^+ \leq A + \frac{1}{n}$$

for every n, and hence

$$\int_\Omega F^+ \leq A$$

Similarly we have

$$\int_\Omega F^- \geq A$$

but F^+ majorizes F^-, and hence $\int_\Omega F^+ \geq \int_\Omega F^-$. Hence we must have

$$\int_\Omega F^+ = \int_\Omega F^- = A$$

In particular

$$\int_\Omega F^+ - F^- = 0$$

By Proposition 2.6 (i), we thus have $F^+(x) = F^-(x)$ for almost every x. But since f is squeezed between $F^-$and F^+, we thus have $f(x) = F^+(x) = F^-(x)$ for almost every x. In particular, f differs from the absolutely integrable function F^+ only on a set of measure zero and is thus measurable and absolutely integrable, with

$$\int_\Omega f = \int_\Omega F^+ = \int_\Omega F^- = A$$

as desired. □

2.4. Consequences of the dominated convergence theorem

In this chapter, we will see how the dominated convergence theorem enables us to extend the continuity of the integrand function to its Lebesgue integral, and how to interchange derivatives and integrals. Note that, throughout the section, we will consider continuity or derivatives in a variable *that is not the variable of integration.*

THEOREM 2.16. *Let $\Omega \subseteq \mathbb{R}^d$ be a measurable set and X be a metric space. Let $x_0 \in X$ and $f : \Omega \times X \to \mathbb{R}$ be a function such that*

(1) $\forall x \in X$, the function $\omega \mapsto f(\omega, x)$ is measurable,
(2) for almost all $\omega \in \Omega$, the function $x \mapsto f(\omega, x)$ is continuous at x_0,
(3) there exists $g \in L^1(\Omega)$ such that

$$|f(\omega, x)| \leq g(\omega)$$

for almost every $\omega \in \Omega$ and all $x \in X$.

Then the function $F : X \to \mathbb{R}$ defined by

$$F(x) = \int_\Omega f(\omega, x) d\omega, \ \forall x \in X$$

is well defined and continous at x_0.

PROOF. From the bound $|f(\omega,x)| \leq g(\omega)$ for almost every $\omega \in \Omega$ and all $x \in X$, we see that for all $x \in X$ the function $\omega \mapsto f(\omega,x)$ is in $L^1(\Omega)$ and so F is well defined.
To prove continuity at x_0, consider a sequence $(x_n)_{n\in\mathbb{N}} \subset X$ such that $x_n \to x_0$ as $n \to \infty$.
For all $n \in \mathbb{N}$ define the function $f_n : \Omega \to \mathbb{R}$ by $f_n(\omega) = f(\omega, x_n)$. We see that $f_n \in L^1(\Omega)$ for all $n \in \mathbb{N}$ and $f_n(\omega) \to f(\omega, x_0)$ for almost every $\omega \in \Omega$.
Since each f_n satisfies $|f_n| \leq g$ almost everywhere, we can apply the dominated convergence theorem to get

$$\lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} \int_\Omega f(\omega, x_n) d\omega = \int_\Omega f(\omega, x_0) d\omega = F(x_0).$$

This is true for any sequence $(x_n)_{n\in\mathbb{N}} \subset X$ converging to x_0 which proves that F is continous at x_0. □

THEOREM 2.17. *Let $\Omega \subseteq \mathbb{R}^d$ be a measurable set and $I \subseteq \mathbb{R}$ be an open interval. Let $t_0 \in I$ and $f : \Omega \times I \to \mathbb{R}$ be a function such that*

(1) $\forall t \in I$, the function $\omega \mapsto f(\omega,t)$ is in $L^1(\Omega)$,
(2) for almost all $\omega \in \Omega$, the function $t \mapsto f(\omega,t)$ is differentiable at t_0,
(3) there exists $g \in L^1(\Omega)$ such that

$$|f(\omega,t) - f(\omega,t_0)| \leq g(\omega)|t - t_0|$$

for almost every $\omega \in \Omega$ and all $t \in I$.

Then the function $F : I \to \mathbb{R}$ defined by

$$F(t) = \int_\Omega f(\omega,t) d\omega, \ \forall t \in I$$

is well defined, differentiable at t_0 and satisfies

$$F'(t_0) = \int_\Omega \frac{\partial f}{\partial t}(\omega, t_0) d\omega.$$

PROOF. The first condition insures that F is well defined.
Set $t_n = t_0 + n^{-1}$. Since the function

$$\omega \mapsto \frac{\partial f}{\partial t}(\omega, t_0)$$

is the pointwise almost everywhere limit of the sequence f_n defined on Ω by

$$f_n(\omega) = \frac{f(\omega,t_n) - f(\omega,t_0)}{t_n - t_0}, \ \forall \omega \in \Omega,$$

then it is measurable.
Now consider an arbitrary sequence $(t_n)_{n\in\mathbb{N}} \subset I\backslash\{t_0\}$ converging to t_0: for almost every $\omega \in \Omega$, we have

$$\lim_{n\to\infty} \frac{f(\omega,t_n) - f(\omega,t_0)}{t_n - t_0} = \frac{\partial f}{\partial t}(\omega, t_0)$$

and the convergence is dominated by g by hypothesis. Therefore, we can use the dominated convergence theorem and we get

$$\lim_{n\to\infty} \frac{F(t_n) - F(t_0)}{t_n - t_0} = \lim_{n\to\infty} \int_\Omega \frac{f(\omega,t_n) - f(\omega,t_0)}{t_n - t_0} d\omega = \int_\Omega \frac{\partial f}{\partial t}(\omega, t_0) d\omega.$$

Notice that as a result of the dominated convergence theorem, we get that the function

$$\omega \mapsto \frac{\partial f}{\partial t}(\omega, t_0)$$

is in $L^1(\Omega)$.
Since the sequence $(t_n)_{n\in\mathbb{N}}$ was arbitrary, we get the result. □

Corollary 2.18. *Let $\Omega \subset \mathbb{R}$ be a measurable set and $I \subset \mathbb{R}$ be an open interval. Let $f : \Omega \times I \to \mathbb{R}$ be a function such that*

(1) $\forall t \in I$, the function $\omega \mapsto f(\omega, t)$ is in $L^1(\Omega)$,
(2) for almost all $\omega \in \Omega$, the function $t \mapsto f(\omega, t)$ is C^1,
(3) there exists $g \in L^1(\Omega)$ such that

$$\left| \frac{\partial f}{\partial t}(\omega, t) \right| \leq g(\omega)$$

for almost every $\omega \in \Omega$ and all $t \in I$.

Then the function $F : I \to \mathbb{R}$ defined by

$$F(t) = \int_\Omega f(\omega, t) d\omega, \ \forall t \in I$$

is C^1 and satisfies

$$F'(t) = \int_\Omega \frac{\partial f}{\partial t}(\omega, t) d\omega, \ \forall t \in I.$$

2.5. Comparison with the Riemann integral

We have spent a lot of effort constructing the Lebesgue integral, but have not yet addressed the question of how to actually compute any Lebesgue integrals, and whether Lebesgue integration is any different from the Riemann integral (say for integrals in one dimension). Now we show that the Lebesgue integral is a generalization of the Riemann integral. To clarify the following discussion, we shall temporarily distinguish the Riemann integral from the Lebesgue integral by writing the Riemann integral $\int_I f$ as $R. \int_I f$.

Our objective here is to prove

Proposition 2.19. *Let $I \subseteq \mathbb{R}$ be an interval, and let $f : I \to \mathbb{R}$ be a Riemann integrable function. Then f is also absolutely integrable, and $\int_I f = R. \int_I f$*

Proof. Write $A := R. \int_I f$. Since f is Riemann integrable , we know that the upper and lower Riemann integrals are equal to A. Thus, for every $\varepsilon > 0$, there exists a partition $\mathbb{P}$ of I into smaller intervals J such that

$$A - \varepsilon \leq \sum_{J \in \mathbb{P}} |J| \inf_{x \in J} f(x) \leq A \leq \sum_{J \in \mathbb{P}} |J| \sup_{x \in J} f(x) \leq A + \varepsilon$$

where $|J|$ denotes the length of J. Note that $|J|$ is the same as $m(J)$, since J is a box.

Let $f_\varepsilon^- : I \to \mathbb{R}$ and $f_\varepsilon^+ : I \to \mathbb{R}$ be the functions

$$f_\varepsilon^-(x) = \sum_{J \in \mathbb{P}} \inf_{x \in J} f(x) \mathbb{1}_J(x)$$

and

$$f_\varepsilon^+(x) = \sum_{J \in \mathbb{P}} \sup_{x \in J} f(x) \mathbb{1}_J(x);$$

these are simple functions and hence measurable and absolutely integrable. By Lemma 2.4 we have

$$\int_I f_\varepsilon^- = \sum_{J \in \mathbb{P}} |J| \inf_{x \in J} f(x)$$

and

$$\int_I f_\varepsilon^+ = \sum_{J \in \mathbb{P}} |J| \sup_{x \in J} f(x)$$

and hence

$$A - \varepsilon \leq \int_I f_\varepsilon^- \leq A \leq \int_I f_\varepsilon^+ \leq A + \varepsilon$$

Since f_ε^+majorizes f, and f_ε^-minorizes f, we thus have

$$A - \varepsilon \leq \underline{\int_\Omega} f \leq \overline{\int_\Omega} f \leq A + \varepsilon$$

for every ε, and thus

$$\underline{\int_\Omega} f = \overline{\int_\Omega} f = A$$

and hence by Lemma 2.15, f is absolutely integrable with $\int_I f = A$, as desired. □

Thus every Riemann integrable function is also Lebesgue integrable, at least on bounded intervals.

Remark 2.9. The converse is not true: take for instance the function $f : [0, 1] \to \mathbb{R}$ defined by $f(x) := 1$ when x is rational, and $f(x) := 0$ when x is irrational. Then we know that f is not Riemann integrable. On the other hand, f is the characteristic function of the set $\mathbb{Q} \cap [0, 1]$, which is countable and hence measure zero. Thus f is Lebesgue integrable and $\int_{[0,1]} f = 0$.

2.6. Fubini's theorem

In one dimension we have shown that the Lebesgue integral is connected to the Riemann integral. Now we will try to understand the connection in higher dimensions. To simplify the discussion we shall just study two-dimensional integrals, although the arguments we present here can easily be extended to higher dimensions.

We shall study integrals of the form $\int_{\mathbb{R}^2} f$. Note that once we know how to integrate on $\mathbb{R}^2$, we can integrate on measurable subsets Ω of $\mathbb{R}^2$, since $\int_\Omega f$ can be rewritten as $\int_{\mathbb{R}^2} f \mathbb{1}_\Omega$.

Let $f(x, y)$ be a function of two variables. In principle, we have three different ways to integrate f on $\mathbb{R}^2$. First of all, we can use the two-dimensional Lebesgue integral, to obtain $\int_{\mathbb{R}^2} f$. Secondly, we can fix x and compute a one-dimensional integral in y, and then take that quantity and integrate in x, thus obtaining $\int_\mathbb{R} \left(\int_\mathbb{R} f(x, y) dy\right) dx$. Secondly, we could fix y and integrate in x, and then integrate in y, thus obtaining $\int_\mathbb{R} \left(\int_\mathbb{R} f(x, y) dx\right) dy$.

Fortunately, if the function f is absolutely integrable on f, then all three integrals are equal:

Theorem 2.20 (Fubini's theorem). *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be an absolutely integrable function. Then there exists absolutely integrable functions $F : \mathbb{R} \to \mathbb{R}$ and $G : \mathbb{R} \to \mathbb{R}$ such that for almost every $x, f(x, y)$ is absolutely integrable in y with*

$$F(x) = \int_\mathbb{R} f(x, y)\, dy$$

and for almost every $y, f(x, y)$ is absolutely integrable in x with

$$G(y) = \int_\mathbb{R} f(x, y)\, dx.$$

Finally, we have

$$\int_\mathbb{R} F(x) dx = \int_{\mathbb{R}^2} f = \int_\mathbb{R} G(y)\, dy.$$

Remark 2.10. Very roughly speaking, Fubini's theorem says that

$$\int_\mathbb{R} \left(\int_\mathbb{R} f(x, y) dy\right) dx = \int_{\mathbb{R}^2} f = \int_\mathbb{R} \left(\int_\mathbb{R} f(x, y) dx\right) dy.$$

This allows us to compute two-dimensional integrals by splitting them into two one-dimensional integrals. The reason why we do not write Fubini's theorem this way, though, is that it is possible that the integral $\int_{\mathbb{R}} f(x,y)dy$ does not actually exist for every x, and similarly $\int_{\mathbb{R}} f(x,y)dx$ does not exist for every y; Fubini's theorem only asserts that these integrals only exist for almost every x and y. For instance, if $f(x,y)$ is the function which equals 1 when $y>0$ and $x=0$, equals -1 when $y<0$ and $x=0$, and is zero otherwise, then f is absolutely integrable on $\mathbb{R}^2$ and $\int_{\mathbb{R}^2} f=0$ (since f equals zero almost everywhere in $\mathbb{R}^2$), but $\int_{\mathbb{R}} f(x,y)dy$ is not absolutely integrable when $x=0$ (though it is absolutely integrable for every other x).

Proof. The proof of Fubini's theorem is quite complicated and we will only give a sketch here. We begin with a series of reductions.

Roughly speaking (ignoring issues relating to sets of measure zero), we have to show that

$$\int_{\mathbb{R}}\left(\int_{\mathbb{R}} f(x,y)dy\right)dx=\int_{\mathbb{R}^2} f$$

together with a similar equality with x and y reversed. We shall just prove the above equality, as the other one is very similar. We perform some reductions:

(i) First of all, it suffices to prove the theorem **for non-negative functions**, since the general case then follows by writing a general function f as a difference $f^+-f^-$of two non-negative functions, applying Fubini's theorem to $f^+$and f^-separately and then using linearity of the integral (Proposition 2.13 (i) and (ii)). Thus we will henceforth assume that f is non-negative.

(ii) Next, it suffices to prove the theorem for non-negative functions f **supported on a bounded set** such as $[-N,N]\times[-N,N]$ for some positive integer N. Indeed, once one obtains Fubini's theorem for such functions, one can then write a general function f as the supremum of such compactly supported functions as

$$f=\sup_{N>0} f\mathbb{1}_{[-N,N]\times[-N,N]}$$

apply Fubini's theorem to each function $f\mathbb{1}_{[-N,N]\times[-N,N]}$ separately, and then take suprema using the monotone convergence theorem. Thus we will henceforth assume that f is supported on $[-N,N]\times[-N,N]$.

(iii) By another similar argument, it suffices to prove the theorem for nonnegative **simple** functions supported on $[-N,N]\times[-N,N]$, since one can use Lemma 2.2 to write f as the supremum of simple functions (which must also be supported on $[-N,N]$), apply Fubini's theorem to each simple function, and then take suprema using the monotone convergence theorem. Thus we may assume that f is a non-negative simple function supported on $[-N,N]\times[-N,N]$.

(iv) Next, we see that it suffices to prove the theorem for **characteristic functions** supported in $[-N,N]\times[-N,N]$. This is because every simple function is a linear combination of characteristic functions, and so we can deduce Fubini's theorem for simple functions from Fubini's theorem for characteristic functions. Thus we may take $f=\mathbb{1}_E$ for some measurable $E\subseteq[-N,N]\times[-N,N]$.

Our task is then to show (ignoring sets of measure zero) that

$$\int_{[-N,N]}\left(\int_{[-N,N]}\mathbb{1}_E(x,y)dy\right)dx=m(E)$$

It will suffice to show the upper Lebesgue integral estimate

$$\overline{\int_{[-N,N]}} \left(\overline{\int_{[-N,N]}} \mathbb{1}_E(x,y)\, dy \right) \, dx \leq m(E) \tag{2.3}$$

We will prove this estimate later. Once we show this for every set E, we may substitute E with $[-N,N] \times [-N,N] \backslash E$ and obtain

$$\overline{\int_{[-N,N]}} \left(\overline{\int_{[-N,N]}} (1 - \mathbb{1}_E(x,y))\, dy \right) dx \leq 4N^2 - m(E).$$

But the left-hand side is equal to

$$\overline{\int_{[-N,N]}} \left(2N - \underline{\int_{[-N,N]}} \mathbb{1}_E(x,y) dy \right) dx$$

which is in turn equal to

$$4N^2 - \underline{\int_{[-N,N]}} \left(\underline{\int_{[-N,N]}} \mathbb{1}_E(x,y) dy \right) dx$$

and thus we have

$$\underline{\int_{[-N,N]}} \left(\underline{\int_{[-N,N]}} \mathbb{1}_E(x,y) dy \right) dx \geq m(E)$$

In particular we have

$$\underline{\int_{[-N,N]}} \left(\overline{\int_{[-N,N]}} \mathbb{1}_E(x,y) dy \right) dx \geq m(E)$$

and hence by Lemma 2.15 we see that $\overline{\int_{[-N,N]}} \mathbb{1}_E(x,y) dy$ is absolutely integrable and

$$\int_{[-N,N]} \left(\overline{\int_{[-N,N]}} \mathbb{1}_E(x,y) dy \right) dx = m(E).$$

A similar argument shows that

$$\int_{[-N,N]} \left(\underline{\int_{[-N,N]}} \mathbb{1}_E(x,y) dy \right) dx = m(E)$$

and hence

$$\int_{[-N,N]} \left(\overline{\int_{[-N,N]}} \mathbb{1}_E(x,y) dy - \underline{\int_{[-N,N]}} \mathbb{1}_E(x,y) dy \right) dx = 0.$$

Thus by Proposition 2.6 (i) we have

$$\overline{\int_{[-N,N]}} \mathbb{1}_E(x,y) dy = \underline{\int_{[-N,N]}} \mathbb{1}_E(x,y) dy$$

for almost every $x \in [-N,N]$. Thus $\mathbb{1}_E(x,y)$ is absolutely integrable in y for almost every x, and $\int_{[-N,N]} \mathbb{1}_E(x,y)$ is thus equal (almost everywhere) to a function $F(x)$ such that

$$\int_{[-N,N]} F(x) dx = m(E)$$

as desired.

It remains to prove the bound (2.3). Let $\varepsilon > 0$ be arbitrary. Since $m(E)$ is the same as the outer measure $m^*(E)$, we know that there exists an at most countable collection $(B_j)_{j\in J}$ of boxes such that $E \subseteq \bigcup_{j\in J} B_j$ and

$$\sum_{j\in J} m(B_j) \leq m(E) + \varepsilon$$

Each box B_j can be written as $B_j = I_j \times I'_j$ for some intervals I_j and I'_j. Observe that

$$\begin{aligned} m(B_j) = |I_j|\,|I'_j| &= \int_{I_j} |I'_j|\, dx = \int_{I_j}\left(\int_{I'_j} dy\right) dx \\ &= \int_{[-N,N]}\left(\int_{[-N,N]} \mathbb{1}_{I_j\times I'_j}(x,y)dx\right) dy \\ &= \int_{[-N,N]}\left(\int_{[-N,N]} \mathbb{1}_{B_j}(x,y)dx\right) dy. \end{aligned}$$

Adding this over all $j \in J$ (using Corollary 2.9) we obtain

$$\sum_{j\in J} m(B_j) = \int_{[-N,N]}\left(\int_{[-N,N]} \sum_{j\in J} \mathbb{1}_{B_j}(x,y)dx\right) dy.$$

In particular we have

$$\overline{\int_{[-N,N]}}\left(\overline{\int_{[-N,N]}} \sum_{j\in J} \mathbb{1}_{B_j}(x,y)dx\right) dy \leq m(E) + \varepsilon.$$

But $\sum_{j\in J} \mathbb{1}_{B_j}$ majorizes $\mathbb{1}_E$ and thus

$$\overline{\int_{[-N,N]}}\left(\overline{\int_{[-N,N]}} \mathbb{1}_E(x,y)dx\right) dy \leq m(E) + \varepsilon.$$

But ε is arbitrary, and so we have (2.3) as desired. This completes the proof of Fubini's theorem. □

We have a similar result in the case in which the integrand is non-negative:

THEOREM 2.21 (Tonelli's theorem). *Let $f : \mathbb{R}^2 \to \mathbb{R}$ such that $f \geq 0$. Then, defining $F : \mathbb{R} \to \mathbb{R}$ and $G : \mathbb{R} \to \mathbb{R}$ as:*

$$F(x) = \int_{\mathbb{R}} f(x,y)\, dy, \quad G(y) = \int_{\mathbb{R}} f(x,y)\, dx,$$

we have

$$\int_{\mathbb{R}} F(x)dx = \int_{\mathbb{R}^2} f = \int_{\mathbb{R}} G(y)\, dy.$$

Remark 2.11. Note that in this case we do not have neither hypothesis nor conclusions on absolute integrability: whenever we have a non-negative integrand, it is possible to swap the order of integration even if the integral is infinite.

2.7. Change of variables

In Chapter 1 we have seen that the Lebesgue measure is translation invariant and homogeneous. We start this section with a change of variable formula carrying these properties into integration. The method used in the proof of the following proposition is very important and should be learned. Before stating the result, we need the following lemma:

Lemma 2.22. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a measurable function, and $x \in \mathbb{R}^d$, and $\alpha \in \mathbb{R}\backslash\{0\}$, then the functions $g, h : \mathbb{R}^d \to \mathbb{R}$ defined by $g(y) = f(x + y)$ and $h(y) = f(\alpha y)$ are measurable.*

Proof. For any $a \in \mathbb{R}$, we have

$$g^{-1}(a, +\infty) = f^{-1}(a, +\infty) - x \quad \text{and} \quad h^{-1}(a, +\infty) = \frac{1}{\alpha} f^{-1}(a, +\infty).$$

From Lemma 1.8 we know these sets are measurable for every $a \in \mathbb{R}$ so g and h are measurable by Lemma 1.21. □

Proposition 2.23 (Translation and Dilation)**.** *Let $f : \mathbb{R}^d \to [0, +\infty]$ be a measurable function, and $x \in \mathbb{R}^d$, and $\alpha \in \mathbb{R}\backslash\{0\}$, then*

$$\int_{\mathbb{R}^d} f(y)dy = \int_{\mathbb{R}^d} f(x + y)dy \quad \text{and} \quad \int_{\mathbb{R}^d} f(y)dy = |\alpha|^d \int_{\mathbb{R}^d} f(\alpha y)dy.$$

The result holds also for $f \in L^1(\mathbb{R}^d)$.

Proof. It suffices to prove the result for non-negative functions as the general case follows by writing a general function f as a difference $f^+ - f^-$ of two non-negative functions, applying the change of variable formula to f^+ and f^- separately and then concluding using linearity of the integral. We assume henceforth that f is non-negative.
We claim that it suffices to prove the theorem for non-negative simple functions. Indeed, any non-negative function f may be written as the increasing limit of a sequence of non-negative simple functions $f = \lim_{n\to\infty} \uparrow f_n$. Using the monotonce convergence theorem, we get

$$\int_{\mathbb{R}^d} f(y)dy = \lim_{n\to\infty} \int_{\mathbb{R}^d} f_n(y)dy = \lim_{n\to\infty} \int_{\mathbb{R}^d} f_n(x + y)dy = \int_{\mathbb{R}^d} f(x + y)dy,$$

since a function h is simple if and only if $h(x + \cdot)$ is simple, and $\lim_{n\to\infty} \uparrow f_n(x + \cdot) = f(x + \cdot)$. The reduction for the dilation formula is proved similarly.
By linearity, it thus suffices to prove the formula for indicator of measurable sets, i.e. $f = \mathbb{1}_E$ for $E \subset \mathbb{R}^d$ some measurable set. Notice that in this case, the formulas read

$$m(E) = m(E - x) \quad \text{and} \quad m(E) = |\alpha|^d m\left(\frac{1}{\alpha}E\right)$$

which were proven in Lemma 1.8 (ii) and (iii). □

The remainder of this section is not examinable.
Notice that the above two formulas address the following problem:

Given an adequate function g, find an expression for $\int_{\mathbb{R}^d} f \circ g$ of the form $\int_{\mathbb{R}^d} fk$ for some function k (which depends only on g).

In measure theoretic terms, this problem can be succinctly stated as characterizing the density of the pushforward measure through (an adequate) g of the Lebesgue measure.
Proposition 2.23 gives a solution to this problem in the cases $g_1(y) = x + y$ and $g_2(y) = \alpha y$. Our goal by the end of this section is to prove a similar result for g being a C^1 diffeomorphism between two open sets.
An intermediate step in our quest is for linear automorphisms of $\mathbb{R}^d$, namely for $g(x) = Mx$ where M is an invertible $d \times d$ matrix with real coefficients.

Lemma 2.24. *Let D be a diagonal $d \times d$ invertible matrix, then for any $f : \mathbb{R}^d \to [0, +\infty]$ Borel function, we have*

$$\int_{\mathbb{R}^d} f(x)dx = \int_{\mathbb{R}^d} f(Dx)|\det(D)|dx.$$

PROOF. Since D is assumed to be invertible, it induces a homeomorphism of $\mathbb{R}^d$. In particular, D^{-1} is a continous linear map and since f is taken to be Borel the composition $x \mapsto f(Dx)$ is Borel measurable.
Using the standard reduction method, we see that it suffices to prove this formula for indicator functions of Borel sets, namely for $f = \mathbb{1}_A$ for some $A \subset \mathbb{R}^d$ a Borel set. In this case the formula reads

$$m(A) = \int_{\mathbb{R}^d} \mathbb{1}_A(Dx)|\det(D)|dx = \int_{\mathbb{R}^d} \mathbb{1}_{D^{-1}A}(x)|\det(D)|dx = |\det(D)|m(D^{-1}A).$$

By setting $E = D^{-1}A$, which is also a Borel set, it suffices to prove that for all Borel sets E, we have

$$m(DE) = |\det(D)|m(E).$$

As D is a diagonal matrix, for any box $B \subset \mathbb{R}^d$, DB is also a box and $\mathrm{vol}(DB) = |\det(D)| \mathrm{vol}(B)$. Moreover, a collection of boxes $(B_j)_{j\in J}$ covers E if and only if $(DB_j)_{j\in J}$ covers DE. We then deduce the claim, because:

$$\begin{aligned}
|\det(D)|m(E) &= \inf\left\{\sum_{j\in J} \mathrm{vol}\,(B_j) : (B_j)_{j\in J} \text{ covers } E\right\} \\
&= |\det(D)| \inf\left\{\frac{1}{|\det(D)|}\sum_{j\in J} \mathrm{vol}\,(DB_j) : (B_j)_{j\in J} \text{ covers } E\right\} \\
&= \inf\left\{\sum_{j\in J} \mathrm{vol}(B'_j) : \left(B'_j\right) \text{ covers } DE\right\} \\
&= m(DE).
\end{aligned}$$

□

Lemma 2.25. *Let P be an orthogonal $d \times d$ matrix, then for any $f : \mathbb{R}^d \to [0, +\infty]$ Borel function, we have*

$$\int_{\mathbb{R}^d} f(x)dx = \int_{\mathbb{R}^d} f(Px)dx.$$

PROOF. As in Lemma 2.24, we have that $x \mapsto f(Px)$ is measurable, and using the standard reduction method, it suffices to prove the formula for $f = \mathbb{1}_A$ for $A \subset \mathbb{R}^d$ a Borel set, namely

$$m(A) = \int_{\mathbb{R}^d} \mathbb{1}_A(Px)dx = \int_{\mathbb{R}^d} \mathbb{1}_{P^{-1}A}(x)dx = m(P^{-1}A).$$

By setting $E = P^{-1}A$, which is also a Borel set, it suffices to that for all Borel sets E, we have

$$m(E) = m(PE).$$

To do this, consider the function μ defined on $\mathcal{B}$, the Borel sets of $\mathbb{R}^d$, by $\mu(E) = m(PE)$. It is easy to verify that μ is a measure. We claim that μ is also translation invariant and finite on compact sets. To see this, for any $x \in \mathbb{R}^d$ we have

$$\mu(E + x) = m(P(E + x)) = m(PE + Px) = m(PE) = \mu(E).$$

Notice that we used the linearity of P and the translation invariance of the Lebesgue measure.
Since P is a continuous function, it maps compact sets to compact sets. Combining this with the fact that the Lebesgue measure is finite on compact sets, we get that μ is finite on compact sets.

By Exercise 1 of the appendix, we know that $\exists\lambda < +\infty$ such that $\mu = \lambda m$.
Let B denote the unit ball of $\mathbb{R}^d$. As P is an orthogonal matrix, we have $PB = B$ and so

$$\lambda m(B) = \mu(B) = m(PB) = m(B),$$

from which we deduce that $\lambda = 1$ (because $m(B) > 0$ is finite). □

We have proved a formula for particular cases of linear change of variables. We now give a result from linear algebra which will allow us to treat the general linear case.

Lemma 2.26 (Polar decomposition)**.** *Let M be an invertible $d \times d$ matrix then there exists P a $d \times d$ orthogonal matrix and S a $d \times d$ symmetric positive matrix such that $M = PS$.*

Proof. M^*M is a symmetric positive definite $d \times d$, so by the spectral theorem, we can find $\lambda_1, \dots, \lambda_d > 0$ positive numbers and $\{u_1, \dots, u_d\}$ an orthonormal basis of $\mathbb{R}^d$ such that

$$M^*M = \sum_{k=1}^{d} \lambda_k u_k u_k^*.$$

Since $\langle Mu_j, Mu_k\rangle = u_j^* M^* M u_k = \lambda_k u_j^* u_k = \lambda_k \delta_{jk}$, the family $\left\{\frac{1}{\sqrt{\lambda_1}} Mu_1, \dots, \frac{1}{\sqrt{\lambda_d}} Mu_d\right\}$ is an orthonormal basis of $\mathbb{R}^d$.
Set

$$S = \sum_{k=1}^{d} \sqrt{\lambda_k} u_k u_k^*$$

and define P such that

$$Pu_k = \frac{1}{\sqrt{\lambda_k}} Mu_k, \quad \forall k = 1, \dots, d.$$

As P maps an orthonormal basis into an orthonormal basis, it is an orthogonal matrix. It remains to check that $M = PS$. To do this it suffices to check the equality on some basis. For every $j = 1, \dots, d$ we have

$$PSu_j = P \sum_{k=1}^{d} \sqrt{\lambda_k} u_k u_k^* u_j = \sum_{k=1}^{d} \sqrt{\lambda_k} \delta_{kj} Pu_k = Mu_j.$$

□

Proposition 2.27. *Let $f : \mathbb{R}^d \to [0, +\infty]$ be a Borel function, and M be an invertible $d \times d$ matrix, then*

$$\int_{\mathbb{R}^d} f(x)dx = \int_{\mathbb{R}^d} f(Mx)|\det(M)|dx.$$

We have a similar formula if $f \in L^1(\mathbb{R}^d)$ is a Borel function.

Proof. The measurability of $x \mapsto f(Mx)$ follows from the fact that M is a homeomorphism and that f is a Borel function. Using the standard reduction argument, it suffices to prove the formula for $f = \mathbb{1}_A$ for some $A \subset \mathbb{R}^d$ a Borel set. In this case the formula reads

$$m(A) = \int_{\mathbb{R}^d} \mathbb{1}_A(Mx)|\det(M)|dx = \int_{\mathbb{R}^d} \mathbb{1}_{M^{-1}A}(x)|\det(M)|dx = |\det(M)|m(M^{-1}A).$$

By setting $E = M^{-1}A$, which is a Borel set, it suffices to prove that for all Borel sets E, we have

$$m(ME) = |\det(M)|m(E).$$

Using Lemma 2.26, we can write $M = PS$ for some orthogonal matrix P and some symmetric positive definite matrix S. By the spectral theorem, we can write $S = QDQ^*$ for some orthogonal matrix Q and some diagonal matrix D. Using Lemma 2.25 and Lemma 2.24, we get

$$m(ME) = m(PQDQ^*E) = m(QDQ^*E) = m(DQ^*E) = |\det(D)|m(Q^*E) = |\det(D)|m(E).$$

Since the determinant is multiplicative, we have $\det(M) = \det(P)\det(Q)\det(D)\det(Q^*)$, in particular $|\det(M)| = |\det(D)|$ since $P, Q \in O(d)$. ☐

Having proved the change of variables in the linear case, we can now attack the C^1 diffeomorphism case. The idea is to approximate locally a C^1 map by an affine map which we know how treat by the preceding discussion. We still need to make this approximation quantitative which is the content of Lemma 2.28 and to have a 'localization' principle which will allow us to 'reassemble' the various pieces of our approximation, see Lemma 2.31.

Lemma 2.28. *Let $\Omega \subset \mathbb{R}^d$ be an open set and $\varphi : \Omega \to \mathbb{R}^d$ be a C^1 map. Suppose that $D\varphi(x)$ is invertible for some $x \in \Omega$, then $\forall \varepsilon \in (0,1)$, $\exists \delta > 0$ such that $\forall \eta \in (0,\delta)$, we have*

$$\varphi(x) + D\varphi(x)\bar{B}\left(0,(1-\varepsilon)\eta\right) \subset \varphi\left(\bar{B}(x,\eta)\right) \subset \varphi(x) + D\varphi(x)\bar{B}\left(0,(1+\varepsilon)\eta\right).$$

Proof. First assume that $D\varphi(x) = I_d$ the identity matrix. By definition of differentiability, we can find $\delta > 0$ with $B(x,\delta) \subset \Omega$ such that

$$|\varphi(y) - \varphi(x) - (y-x)| \le \varepsilon|y-x|, \quad \forall y \in B(x,\delta).$$

In particular, we get

$$(1-\varepsilon)|y-x| \le |\varphi(y) - \varphi(x)| \le (1+\varepsilon)|y-x|, \quad \forall y \in B(x,\delta).$$

Fix $\eta \in (0,\delta)$, then for any $y \in \bar{B}(x,\eta)$, we have

$$|\varphi(y) - \varphi(x)| \le (1+\varepsilon)|y-x| \le (1+\varepsilon)\eta$$

which yields the inclusion

$$\varphi(\bar{B}(x,\eta)) \subset \varphi(x) + \bar{B}(0,(1+\varepsilon)\eta).$$

To get the other inclusion, note that by the inverse function theorem, φ is an open map on $B(x,\delta)$. For any $\eta \in (0,\delta)$, we have that

$$\varphi^{-1}\left[\partial\varphi(\bar{B}(x,\eta))\right] \cap \bar{B}(x,\eta) \subset \partial\bar{B}(x,\eta).$$

Informally, this relation is saying that the boundary of the image is included only in the image of the boundary.

From this, $\forall z \in \partial\varphi(\bar{B}(x,\eta))$, we have $z = \varphi(y)$ for some $y \in \partial\bar{B}(x,\eta)$ and so

$$|z - \varphi(x)| = |\varphi(y) - \varphi(x)| \ge (1-\epsilon)|y-x| = (1-\epsilon)\eta.$$

We claim that this implies the desired inclusion, namely

$$\varphi(x) + \bar{B}(0,(1-\epsilon)\eta) \subset \varphi(\bar{B}(x,\eta)).$$

Suppose this is not the case, then $\exists w \in B(\varphi(x),(1-\epsilon)\eta)$ such that $w \notin \varphi(\bar{B}(x,\eta))$. Take $t := \sup\{s \in (0,1) : (1-s)\varphi(x) + sw \notin \varphi(\bar{B}(x,\eta))\}$. Since $\varphi(\bar{B}(x,\eta))$ is compact and φ is an open map in $B(x,\delta)$, we have $(1-t)\varphi(x) + tw \in \partial\varphi(\bar{B}(x,\eta))$. But it holds that

$$|(1-t)\varphi(x) + tw - \varphi(x)| = t|w - \varphi(x)| < (1-\varepsilon)\eta,$$

which is a contradiction. To treat the general case, consider the auxiliary map $\psi = D\varphi(x)^{-1}\varphi$. Then ψ is C^1 and satisfies $D\psi(x) = I_d$. By the previous discussion, we can thus find $\delta > 0$ such that for any $\eta \in (0,\delta)$, we have

$$\psi(x) + \bar{B}(0,(1-\varepsilon)\eta) \subset \psi(\bar{B}(x,\eta)) \subset \psi(x) + \bar{B}(0,(1+\varepsilon)\eta).$$

Applying the map $D\varphi(x)$ to this sequence of inclusions and using the linearity we get

$$\varphi(x) + D\varphi(x)\bar{B}(0,(1-\varepsilon)\eta) \subset \varphi(\bar{B}(x,\eta)) \subset \varphi(x) + D\varphi(x)\bar{B}(x,(1+\varepsilon)\eta).$$

$\square$

DEFINITION (Dyadic cubes). For every $n \in \mathbb{N}$, the dyadic cubes of order n in $\mathbb{R}^d$ are

$$\mathcal{Q}_n = \left\{ \prod_{i=1}^{d} \left[\frac{k_i}{2^n}, \frac{k_i+1}{2^n} \right) : k_1, \ldots, k_d \in \mathbb{Z} \right\}.$$

The dyadic cubes in $\mathbb{R}^d$ are

$$\mathcal{Q} = \bigcup_{n\in\mathbb{N}} \mathcal{Q}_n.$$

Proposition 2.29 (Properties of the dyadic cubes)**.** *The following properties hold.*

(i) For every $n \in \mathbb{N}$, $\mathcal{Q}_n$ is a partition of $\mathbb{R}^d$,
(ii) The diameters of dyadic cubes of order n goes to 0 as $n \to \infty$,
(iii) For $n \le m$, $Q \in \mathcal{Q}_n$ and $\tilde{Q} \in \mathcal{Q}_m$, either $\tilde{Q} \subset Q$ or $\tilde{Q} \cap Q = \varnothing$.

PROOF.

(i) $\forall z \in \mathbb{R}$ and $\forall n \in \mathbb{N}$, $\exists! k \in \mathbb{Z}$ given by $k = \lfloor 2^n z \rfloor$ such that $z \in [2^{-n}k, 2^{-n}(k+1))$. So $\forall x \in \mathbb{R}^d$, $\exists!(k_1, \ldots, k_d) \in \mathbb{Z}^d$ such that $x \in \prod_{k=1}^{d} [2^{-n}k_i, 2^{-n}(k_i+1))$.
(ii) $\forall n \in \mathbb{N}$ and $\forall Q \in \mathcal{Q}_n$, we have $\operatorname{diam}(Q) = 2^{-n}\sqrt{d}$ which goes to 0 as $n \to \infty$.
(iii) For $n \le m$, take $Q \in \mathcal{Q}_n$ and $\tilde{Q} \in \mathcal{Q}_m$. Assume $\tilde{Q} \cap Q \neq \varnothing$, we will show that $\tilde{Q} \subset Q$. The result follows easily from the case $d = 1$, so assume that $d = 1$. Then we have $Q = [2^{-n}k, 2^{-n}(k+1))$ and $\tilde{Q} = [2^{-m}j, 2^{-m}(j+1))$ for some $k, j \in \mathbb{Z}$. Since $\tilde{Q} \cap Q \neq \varnothing$, we have $2^{-n}k < 2^{-m}(j+1)$ which implies $2^{m-n}k < j+1$. Since $2^{m-n}k \in \mathbb{Z}$, we deduce $2^{m-n}k \le j$ and so $2^{-n}k \le 2^{-m}j$. Similarly, $2^{-m}j < 2^{-n}(k+1)$ gives $2^{-m}(j+1) \le 2^{-n}(k+1)$ and so $\tilde{Q} \subset Q$.

$\square$

Lemma 2.30. *Let $\Omega \subset \mathbb{R}^d$ be a nonempty open set, then for any $N \in \mathbb{N}$, Ω can be written as the union of a countable disjoint collection of dyadic cubes of order at least N.*

PROOF. Define

$$I_N = \{Q \in \mathcal{Q}_N : Q \subset \Omega\}$$

and inductively for $n > N$

$$I_n = \left\{ Q \in \mathcal{Q}_n : Q \subset \Omega \setminus \bigcup_{k=N}^{n-1} \bigcup_{\tilde{Q}\in I_k} \tilde{Q} \right\}.$$

The subcollection of dyadic cubes given by $I \coloneqq \bigcup_{n\ge N} I_n$ is disjoint by Proposition 2.29 (iii) and is countable since $\mathcal{Q}$ is countable. Moreover, by construction, we have that $\bigcup_{Q\in I} Q \subset \Omega$, so it only remains to prove the converse inclusion.
Take $x \in \Omega$, then $\exists \delta > 0$ such that $B(x,\delta) \subset \Omega$. Using Proposition 2.29 (i) and (ii), we know that there exists $n \ge N$ sufficiently large such that $\exists Q \in \mathcal{Q}_n$ with $x \in Q \subset B(x,\delta)$. In the inductive construction, if

$$x \in \bigcup_{k=N}^{n-1} \bigcup_{\tilde{Q}\in I_k} \tilde{Q}$$

there is nothing to prove. If this is not the case, then $Q \in I_n$. We thus have $\Omega = \bigcup_{Q\in I} Q$. $\square$

Lemma 2.31. *Let $\Omega \subset \mathbb{R}^d$ be a non empty open set and $\delta > 0$ be a positive number, then there exists $(B_n)_{n\in\mathbb{N}}$ a countable collection of disjoint closed balls such that*

$$\operatorname{diam}(B_n) < \delta \quad \forall n \in \mathbb{N} \quad \textit{and} \quad m\left(\Omega \backslash \bigcup_{n\in\mathbb{N}} B_n\right) = 0.$$

Proof. Set

$$\Theta \coloneqq \frac{\omega_d}{4^d}$$

where ω_d is the Lebesgue measure of the unit ball in $\mathbb{R}^d$.
For a dyadic cube $Q = \prod_{i=1}^d [2^{-n}k_i, 2^{-n}(k_i+1))$ in $\mathbb{R}^d$, let $c(Q) \coloneqq= (2^{-n}(k_1+1/2), \ldots, 2^{-n}(k_d+1/2))$ denote its center and $r(Q) \coloneqq 2^{-n}$ denote its sidelenght. Finally, set $B(Q) = \bar{B}(c(Q), r(Q)/4)$ be the closed ball with the same center as Q and radius $1/4$ of the side length of Q. Notice that

$$\Theta = \frac{m(B(Q))}{m(Q)} \in (0,1).$$

Assume now that Ω is bounded. Take $N \in \mathbb{N}$ large enough so that $2^{-n}\sqrt{d} < \delta$. By Lemma 2.30, we can find a countable collection of dyadic cubes $(Q_n^{(0)})_{n\in\mathbb{N}}$ of order at least N such that $\Omega = \bigcup_{n\in\mathbb{N}} Q_n^{(0)}$. By the assumption on the order of the dyadic cubes, we get $\operatorname{diam}(Q_n^{(0)}) < \delta$ for all $n \in \mathbb{N}$.
Let $F_1 = \bigcup_{n\in\mathbb{N}} B(Q_n^{(0)})$, then

$$m(F_1) = \sum_{n\in\mathbb{N}} m(B(Q_n^{(0)})) = \Theta \sum_{n\in\mathbb{N}} m(Q_n^{(0)}) = \Theta m(\Omega).$$

Since $m(\Omega) < \infty$, we can find $N_1 \in \mathbb{N}$ such that

$$m\left(\bigcup_{n=1}^{N_1} B(Q_n^{(0)})\right) \geq \frac{\Theta}{2} m(\Omega).$$

Set $\Omega_1 = \Omega \backslash \bigcup_{n=1}^{N_1} B(Q_n^{(0)})$, which is open and satisfies

$$m(\Omega_1) = m\left(\Omega \backslash \bigcup_{n=1}^{N_1} B(Q_n^{(0)})\right) = m(\Omega) - \sum_{n=1}^{N_1} m(B(Q_n^{(0)})) \leq \left(1 - \frac{\Theta}{2}\right) m(\Omega) < \infty.$$

By the same procedure, we can find finitely many disjoint closed balls $(B(Q_n^{(1)}))_{n=1,\ldots,N_2}$ of diameter less than δ included in Ω_1 such that

$$m\left(\bigcup_{n=1}^{N_2} B(Q_n^{(1)})\right) \geq \frac{\Theta}{2} m(\Omega_1).$$

Set $\Omega_2 = \Omega_1 \backslash \bigcup_{n=1}^{N_2} B(Q_n^{(1)})$, which is open and satisfies

$$m(\Omega_2) = m\left(\Omega_1 \backslash \bigcup_{n=1}^{N_2} B(Q_n^{(1)})\right) \leq \left(1 - \frac{\Theta}{2}\right) m(\Omega_1) \leq \left(1 - \frac{\Theta}{2}\right)^2 m(\Omega).$$

Continue this procedure inductively so that at step k, we have finitely many disjoint closed balls $(B(Q_n^{(k)})_{n=1,\ldots,N_k}$ of radius at most δ included in $\Omega_k = \Omega_{k-1} \backslash \bigcup_{n=1}^{N_{k-1}} B(Q_n^{k-1})$ which satisfy

$$m(\Omega_k) = m\left(\Omega_{k-1} \backslash \bigcup_{n=1}^{N_k} B(Q_n^k)\right) \leq \left(1 - \frac{\Theta}{2}\right) m(\Omega_{k-1}) \leq \left(1 - \frac{\Theta}{2}\right)^k m(\Omega).$$

By construction, the family of closed balls $\{B(Q_n^j) : j \in \mathbb{N}_0,\ n = 1, \dots, N_j\}$ is disjoint and included in Ω and satisfies (using the previous bounds)

$$m\left(\Omega \backslash \bigcup_{j=0}^{\infty} \bigcup_{n=1}^{N_j} B(Q_n^{(j)})\right) \leq \left(1 - \frac{\Theta}{2}\right)^k m(\Omega) \quad \forall k \in \mathbb{N}.$$

To treat the general case, i.e. Ω unbounded, write $\Omega = A \cup \bigcup_{n \geq 1} \Omega_n$ where $\Omega_n = \Omega \cap \{x \in \mathbb{R}^d : n-1 < |x| < n\}$ for every $n \in \mathbb{N}$ and A is a null set. Each Ω_n is open and bounded and so the previous result applies, putting together all the closed balls and using countable subadditivity of the Lebesgue measure gives the result. □

Lemma 2.32. *Let $f : \mathbb{R}^d \to \mathbb{R}^d$ be a k-Lipschitz function for some $k > 0$ and $A \subset \mathbb{R}^d$ be some set then $m^*(f(A)) \leq \left(2k\sqrt{d}\right)^d m^*(A)$.*

Proof. In the definition of outer measure, we used cover with boxes. Had we defined the outer measure using only covers with cubes, we would have gotten the same outer measure, we leave this as an exercise.
Now take a collection of cubes $(B_n)_{n \in \mathbb{N}}$ that cover A, then for every n, we can find a cube $\tilde{B_n}$ such that $f(B_n) \subset \tilde{B_n}$ and $\text{vol}(\tilde{B_n}) \leq \left(2k\sqrt{d}\right)^d \text{vol}(B_n)$. This implies

$$m^*(f(A)) \leqslant \sum_{n \in \mathbb{N}} \text{vol}(\tilde{B_n}) \leqslant \left(2k\sqrt{d}\right)^d \sum_{n \in \mathbb{N}} \text{vol}(B_n).$$

Since the cover of A $(B_n)_{n \in \mathbb{N}}$ was arbitrary, we deduce the result. □

Remark 2.12. In particular if A is a null set, then $f(A)$ is a null set and so measurable.

Theorem 2.33 (Change of variables formula). *Let U, D be two open subsets of $\mathbb{R}^d$ and $g : U \to D$ be a C^1 diffeomorphism. Let $f : D \to [0, +\infty]$ be a Borel measurable function, then*

$$\int_D f(x)dx = \int_U f(g(u))|\det Dg(u)|du.$$

We have a similar formula if $f \in L^1(U)$ is a Borel function.

Proof. $|\det DG|$ is a polynomial of the partial derivatives of g and so is continuous hence measurable. Since f is Borel and g is continuous, we have that $f \circ g$ is Borel measurable. Using the standard reduction argument, it suffices to prove the result for indicator functions of Borel sets, namely $f = \mathbb{1}_A$ for $A \subset D$ a Borel set. In this case, the formula reads

$$m(A) = \int_U \mathbb{1}_A(g(u))|\det Dg(u)|du = \int_U \mathbb{1}_{g^{-1}(A)}(u)|\det Dg(u)|du = \int_{g^{-1}(A)} |\det Dg(u)|du.$$

By setting $E = g^{-1}(A)$ which is a Borel set, it suffices to prove that for all Borel sets, we have

$$m(g(E)) = \int_E |\det Dg(u)|du.$$

We will first prove this for E an open set of U such that $\bar{E} \subset U$ is compact. Take $\varepsilon \in (0,1)$, by Lemma 2.28 we can find $\delta > 0$ such that for any $\eta \in (0, \delta)$ and any $x \in E$, we have

$$g(x) + Dg(x)\bar{B}(0, (1-\varepsilon)\eta) \subset g(\bar{B}(x, \eta)) \subset g(x) + Dg(x)\bar{B}(0, (1+\varepsilon)\eta).$$

The uniformity in $x \in E$ comes from the fact that Dg is continuous on E which is assumed to be compact.
Take $\eta \in (0, \delta)$ such that $(1+\varepsilon)^{-1} \leq |\frac{\det Dg(x)}{\det Dg(y)}| \leq (1+\varepsilon)$ for all $x, y \in E$ with $|x - y| \leq \eta$.

By Lemma 2.31, we can find countably many disjoint closed balls $(B_n)_{n\in\mathbb{N}}$ in U of radius less than η such that $m(U\setminus\bigcup_{n\in\mathbb{N}} B_n)=0$. In particular, we get $m(E\setminus\bigcup_{n\in\mathbb{N}} B_n)=0$ and by Remark 2.12, $m(g(E\setminus\bigcup_{n\in\mathbb{N}} B_n))=0$. Denote x_n the center of the ball B_n and r_n its radius, from Proposition 2.27, we get

$$\begin{aligned} m(g(E)) &= m\left(g\left(\bigcup_{n\in\mathbb{N}} B_n\right)\right) \\ &= \sum_{n\in\mathbb{N}} m(g(B_n)) \\ &\leq \sum_{n\in\mathbb{N}} m(Dg(x_n)\bar{B}(0,(1+\varepsilon)r_n)) \\ &= (1+\epsilon)^d \sum_{n\in\mathbb{N}} m(\bar{B}(x_n,r_n))|\det Dg(x_n)| \\ &\leq (1+\varepsilon)^{d+1}\sum_{n\in\mathbb{N}}\int_{B_n}|\det Dg(x)|dx \\ &= (1+\varepsilon)^{d+1}\int_{\bigcup_{n\in\mathbb{N}} B_n}|\det Dg(x)|dx \\ &= (1+\varepsilon)^{d+1}\int_E|\det Dg(x)|dx. \end{aligned}$$

Similarly

$$\begin{aligned} \int_E|\det Dg(x)|dx &= \sum_{n\in\mathbb{N}}\int_{B_n}|\det Dg(x)|dx \\ &\leq (1+\varepsilon)\sum_{n\in\mathbb{N}} m(B_n)|\det Dg(x_n)| \\ &\leq \frac{1+\varepsilon}{(1-\varepsilon)^d}\sum_{n\in\mathbb{N}} m(g(x_n)+Dg(x_n)\bar{B}(0,(1-\varepsilon)r_n)) \\ &\leq \frac{1+\varepsilon}{(1-\varepsilon)^d}\sum_{n\in\mathbb{N}} m(g(B_n)) \\ &= \frac{1+\varepsilon}{(1-\varepsilon)^d} m\left(g\left(\bigcup_{n\in\mathbb{N}} B_n\right)\right) \\ &= \frac{1+\varepsilon}{(1-\varepsilon)^d} m(g(E)). \end{aligned}$$

Letting $\varepsilon\to 0$ gives the result in the case where E is open and $\bar{E}\subset U$ is compact. For a general open set $E\subset U$, we can write E as an increasing limit of open sets compactly contained in U and use the monotone convergence theorem to conclude.

To treat a general Borel set $E\subset U$, assume first that Dg is bounded on U. Take $\varepsilon>0$, use the outer regularity of the Lebesgue measure and Lemma 2.32 to find $A\subset U$ an open set such that $m(A\setminus E)\leq\varepsilon$ and $m(g(A\setminus E))\leq\varepsilon$, then we get

$$\left|m(g(E))-\int_E|\det Dg(x)|dx\right|\leq m(g(A\setminus E))+m(A\setminus E)\|Dg\|_\infty\leq\varepsilon(1+\|Dg\|_\infty).$$

To remove the assumption that Dg is bounded, we can write E as an increasing limit of sets compactly contained in U and use the monotone convergence theorem. $\square$

Remark 2.13. We have proved the change of variable formulas for Borel function. The result remains valid in the case of Lebesgue measurable functions but we first need to prove the measurability of $f \circ g$. We do not this here and simply state that this can be done by proving that C^1 diffeomorphisms preserve Lebesgue measurability (use Remark 2.12 and the fact that every Lebesgue measurable set is almost Borel measurable).

CHAPTER 3

L^p Spaces

This chapter is inspired by [Dac, Chapter 16] and [Buf22, Chapter 2].

In this chapter, we generalize the notion of absolutely integrable functions, trying to understand what are the implications of the integrability of a general power $p \geq 1$ of the modulus of the function. We begin with the following defintion:

Definition (L^p norm). Let $\Omega \subseteq \mathbb{R}^d$ be a measurable set, let $f : \Omega \to \mathbb{R}^*$ and $p \in [1, \infty)$. We define

$$\|f\|_{L^p(\Omega)} := \left(\int_\Omega |f|^p \right)^{\frac{1}{p}}. \tag{3.1}$$

From the above, we then naturally define at the first attempt the space of functions with finite L^p norm:

$$L^p(\Omega) := \left\{ f : \Omega \to \mathbb{R}^* \text{ measurable, s.t. } \|f\|_{L^p(\Omega)} < +\infty \right\}. \tag{3.2}$$

Remark 3.1. In the case $p = 1$, $L^1(\Omega)$ is the space of absolutely integrable functions on Ω. Moreover, it holds that

$$f \in L^p(\Omega) \iff |f|^p \in L^1(\Omega).$$

We would like to introduce a normed vector space from the definitions (3.1)-(3.2); in order to do so, we need to enforce the vanishing property of the norm (that means, $\|f\|_{L^p(\Omega)} = 0 \iff f = 0$). In our case, since $\|f\|_{L^p(\Omega)} = 0 \Rightarrow f = 0$ a.e., we wish to identify in L^p functions that coincide almost everywhere. To do so, we introduce the equivalence relation:

$$f \sim g \iff f = g \text{ a.e.}$$

and define L^p spaces as follows:

Definition (L^p space).

$$(L^p(\Omega); \|\cdot\|_{L^p}) = (\left\{ f : \Omega \to \mathbb{R}^* \text{ measurable s.t. } \|f\|_{L^p(\Omega)} < +\infty \right\} / \sim \ ; \|\cdot\|_{L^p}).$$

We now introduce a space that morally contains functions that are bounded (up to a set of measure 0).

Definition (L^∞ norm). Let us define

$$\|f\|_{L^\infty(\Omega)} := \operatorname*{ess\,sup}_{x \in \Omega} |f(x)| = \inf \{\alpha \geq 0 : |f| < \alpha \text{ a.e.}\}$$

Remark 3.2. If f is continuous, sup and ess sup coincide.

Remark 3.3. Notice that $\|\cdot\|_{L^\infty(\Omega)}$ is well defined on an equivalence class, because

$$f = g \text{ a.e.} \Longrightarrow \operatorname{ess\,sup} |f| = \operatorname{ess\,sup} |g|.$$

Example 3.1. $\operatorname{ess\,sup} \mathbb{1}_{\mathbb{Q}} = 0$, because $\mathbb{1}_{\mathbb{Q}} = 0$ a.e.

Definition (L^∞ space). $L^\infty(\Omega) := \{f : \Omega \to \mathbb{R}^* \text{ measurable, s.t. } \|f\|_{L^\infty} < +\infty\}$.

We will now introduce the Hölder inequality, a fundamental tool to work with L^p spaces. To do so, we need the following definition.

DEFINITION (Conjugate Hölder exponents). Let $p \in [1, +\infty]$. The conjugate Hölder exponent is p' such that

$$\frac{1}{p} + \frac{1}{p'} = 1 \qquad \Longleftrightarrow \qquad p' = \frac{p}{p-1},$$

with the convention $1/+\infty = 0$.

Example 3.2. 1 is the conjugate of $+\infty$, 2 is the conjugate of itself.

THEOREM 3.1 (Hölder inequality). *Let Ω be measurable, $p \in [1, +\infty]$, then*

$$\|fg\|_{L^1(\Omega)} \leq \|f\|_{L^p(\Omega)} \|g\|_{L^{p'}(\Omega)} \quad \forall f, g \text{ measurable},$$

where p' is the conjugate Hölder exponent of p.

PROOF. If $p = 1$ (symmetrically, $p = +\infty$), we get that $p' = +\infty$ ($p' = 1$). Therefore,

$$\|fg\|_{L^1(\Omega)} \leq \int_\Omega |f||g| \leq \int_\Omega |f| \|g\|_{L^\infty(\Omega)} \leq \left(\int_\Omega |f| \right) \|g\|_{L^\infty(\Omega)} = \|f\|_{L^1(\Omega)} \|g\|_{L^\infty(\Omega)}.$$

If $p, p' \neq 1, +\infty$, we observe that the inequality is invariant under multiplications by constants: the Hölder inequality therefore holds if and only if

$$\|\lambda_1 f \lambda_2 g\|_{L^1(\Omega)} = \lambda_1 \lambda_2 \|fg\|_{L^1(\Omega)} \leq \lambda_1 \|f\|_{L^p(\Omega)} \lambda_2 \|g\|_{L^{p'}(\Omega)} \leq \|\lambda_1 f\|_{L^p(\Omega)} \|\lambda_2 g\|_{L^{p'}(\Omega)} \text{ for all } \lambda_1, \lambda_2 > 0.$$

Therefore, we may reduce ourselves to the case

$$\|f\|_{L^p(\Omega)} = \|g\|_{L^{p'}(\Omega)} = 1.$$

Indeed, assume

$$\|fg\|_{L^1(\Omega)} \leq 1 \quad \forall f, g \text{ s.t. } \|f\|_{L^p(\Omega)} = \|g\|_{L^{p'}(\Omega)} = 1. \tag{3.3}$$

By applying the inequality (3.3) to

$$f = \frac{F}{\|F\|_{L^p(\Omega)}} \quad \text{and} \quad g = \frac{G}{\|G\|_{L^{p'}(\Omega)}},$$

we retrieve that

$$\frac{\|FG\|_{L^1(\Omega)}}{\|F\|_{L^p(\Omega)} \|G\|_{L^{p'}(\Omega)}} \leq 1 \quad \Longrightarrow \quad \|FG\|_{L^1(\Omega)} \leq \|F\|_{L^p(\Omega)} \|G\|_{L^{p'}(\Omega)}.$$

To prove (3.3), we can use the Young's inequality for real numbers $X, Y > 0$, $p \in (1, +\infty)$:

$$XY \leq \frac{X^p}{p} + \frac{Y^{p'}}{p'},$$

in order to obtain that

$$\int_\Omega |fg| \leq \int_\Omega \frac{|f|^p}{p} + \frac{|g|^{p'}}{p'} = \frac{1}{p} + \frac{1}{p'} = 1,$$

which concludes the proof. □

Remark 3.4. The Hölder inequality with $p = p' = 2$ is known as Cauchy-Schwartz inequality:

$$\int_\Omega fg \leq \|f\|_{L^2(\Omega)} \|g\|_{L^2(\Omega)}.$$

We can generalize the Cauchy-Schwartz inequality to the case in which f, g are complex-valued: indeed, we have that

$$\left| \int_\Omega f\overline{g} \right| \leq \int_\Omega |f||g| \leq \|f\|_{L^2(\Omega)} \|g\|_{L^2(\Omega)}.$$

Remark 3.5. Via Hölder inequality, we can also prove that

$$\|fg\|_{L^r} \leq \|f\|_{L^p}\|g\|_{L^q} \quad \forall r,p,q \quad \text{s.t} \quad \frac{1}{r} = \frac{1}{p} + \frac{1}{q}.$$

Indeed, setting $F = f^r, G = g^r$ and applying Hölder inequality with conjugate exponents p/r and q/r, we can prove that:

$$\left(\int_\Omega |fg|^r\right)^{1/r} = \left(\int_\Omega |FG|\right)^{1/r} \leq \left(\int_\Omega |F|^{p/r}\right)^{1/p} \left(\int_\Omega |G|^{q/r}\right)^{1/q} = \left(\int_\Omega |f|^p\right)^{1/p} \left(\int_\Omega |g|^q\right)^{1/q}.$$

We now state and prove some useful properties of L^p spaces.

Proposition 3.2. *Let $\Omega \subset \mathbb{R}^d$ be measurable and $1 \leq p \leq q \leq +\infty$, then*

(i) $L^p(\Omega)$ is a vector space.

(ii) If $m(\Omega) < +\infty$, then $\|f\|_{L^p(\Omega)} \leq K\|f\|_{L^q(\Omega)}$ for all $f : \Omega \to \mathbb{R}^*$ measurable, where K depends on $m(\Omega)$, p and q. In particular,

$$L^q(\Omega) \subseteq L^p(\Omega).$$

(iii) If $m(\Omega) < +\infty$, then $\lim_{p\to+\infty}\|f\|_{L^p(\Omega)} = \|f\|_{L^\infty(\Omega)}$.

(iv) The Minkowski inequality holds:

$$\|f+g\|_{L^p(\Omega)} \leq \|f\|_{L^p(\Omega)} + \|g\|_{L^p(\Omega)} \quad \forall f,g \in L^p(\Omega)$$

This property enables us to conclude that $\|\cdot\|_{L^p(\Omega)}$ is a norm on L^p, because it satisfies the triangular inequality.

Proof. [The proof is contained in Series 6, ex. 7]

(i) Let $f, g \in L^p(\Omega)$ and $\lambda, \mu \in \mathbb{R}$. Notice that $\lambda f + \mu g$ is measurable, and we need to prove that $\lambda f + \mu g \in L^p(\Omega)$.
If $p = +\infty$, we have that $\lambda f + \mu g \in L^\infty(\Omega)$, because

$$\begin{aligned}\{x \in \Omega : |f(x)+g(x)| > \|f\|_{L^\infty(\Omega)} + \|g\|_{L^\infty(\Omega)}\} &\subseteq \{x \in \Omega : |f(x)| + |g(x)| > \|f\|_{L^\infty(\Omega)} + \|g\|_{L^\infty(\Omega)}\} \\ &\subseteq \{x \in \Omega : |f(x)| > \|f\|_{L^\infty(\Omega)}\} \cup \{x \in \Omega : |g(x)| > \|g\|_{L^\infty(\Omega)}\}.\end{aligned}$$

Since

$$m(\{x \in \Omega : |f(x)| > \|f\|_{L^\infty(\Omega)}\}) = m(\{x \in \Omega : |g(x)| > \|g\|_{L^\infty(\Omega)}\}) = 0,$$

by definition of the essential supremum, we conclude that

$$m(\{x \in \Omega : |f(x)+g(x)| > \|f\|_{L^\infty(\Omega)} + \|g\|_{L^\infty(\Omega)}\}) = 0.$$

Therefore,

$$\|f+g\|_{L^\infty(\Omega)} \leq \|f\|_{L^\infty(\Omega)} + \|g\|_{L^\infty(\Omega)}.$$

Assume now $1 \leq p < +\infty$. The function $x \to |x|^p$ is convex for this choice of p, hence

$$\forall x,y \in \mathbb{R} \quad |x+y|^p \leq 2^{p-1}(|x|^p + |y|^p).$$

Therefore,

$$\|\lambda f + \mu g\|^p_{L^p(\Omega)} = \int_\Omega |\lambda f(x) + \mu g(x)|^p dx \leq 2^{p-1}\left(|\lambda|^p\|f\|^p_{L^p(\Omega)} + |\mu|^p\|g\|^p_{L^p(\Omega)}\right) < +\infty,$$

which proves that $\lambda f + \mu g \in L^p(\Omega)$.

(ii) Let Ω be bounded and $1 \le p < q \le +\infty$. Let $f \in L^q(\Omega)$, and first assume that $1 < q < +\infty$. By applying Hölder inequality with exponents q/p and $q/(q-p)$ (admissible choice provided that $q > p$), we get that

$$\|f\|^p_{L^p(\Omega)} = \int_\Omega |f|^p dx \le \left(\int_\Omega (|f|^p)^{q/p} dx\right)^{p/q} \left(\int_\Omega 1 dx\right)^{1-p/q} = (\|f\|_{L^q(\Omega)})^p m(\Omega)^{1-p/q}.$$

By taking the $p-$th root, we get:

$$\|f\|_{L^p(\Omega)} \le \|f\|_{L^q(\Omega)} m(\Omega)^{1/p-1/q}.$$

When $q = +\infty$, we have that:

$$\|f\|_{L^p(\Omega)} = \left(\int_\Omega |f|^p dx\right)^{1/p} \le \left(\int_\Omega \|f\|^p_{L^\infty(\Omega)} dx\right)^{1/p} = m(\Omega)^{1/p} \|f\|_{L^\infty(\Omega)}. \tag{3.4}$$

(iii) Let $f \in L^\infty(\Omega)$. We can prove using (3.4) that

$$\limsup_{p\to+\infty} \|f\|_{L^p(\Omega)} \le \|f\|_{L^\infty(\Omega)},$$

because $m(\Omega)^{1/p} \to 1$ as $p \to +\infty$.
Now, we need to prove that $\liminf_{p\to+\infty} \|f\|_{L^p(\Omega)} \ge \|f\|_{L^\infty(\Omega)}$.
To do so, fix $0 < \varepsilon < \|f\|_{L^\infty(\Omega)}$ and consider the set

$$A_\varepsilon := \{x \in \Omega : |f(x)| \ge \|f\|_{L^\infty(\Omega)} - \varepsilon\}.$$

By definition of the essential supremum, we have that $m(A_\varepsilon) > 0$. Thus,

$$\int_\Omega |f|^p dx \ge \int_{A_\varepsilon} |f|^p dx \ge m(A_\varepsilon)(\|f\|_{L^\infty(\Omega)} - \varepsilon)^p > 0,$$

and therefore, taking the p-th root, we have that

$$\|f\|_{L^p(\Omega)} \ge m(A_\varepsilon)^{1/p}(\|f\|_{L^\infty(\Omega)} - \varepsilon).$$

Since $m(A_\varepsilon) > 0$, we have that $m(A_\varepsilon)^{1/p} \to 1$ as $p \to +\infty$, and we get that:

$$\liminf_{p\to+\infty} \|f\|_{L^p(\Omega)} \ge \|f\|_{L^\infty(\Omega)} - \varepsilon.$$

By arbitrarity of ε, we conclude that

$$\liminf_{p\to+\infty} \|f\|_{L^p(\Omega)} \ge \|f\|_{L^\infty(\Omega)}.$$

Now, we can prove that the limit exists and is equal to the desired quantity, because

$$\|f\|_{L^\infty(\Omega)} \le \liminf_{p\to+\infty} \|f\|_{L^p(\Omega)} \le \limsup_{p\to+\infty} \|f\|_{L^p(\Omega)} \le \|f\|_{L^\infty(\Omega)}.$$

(iv) Notice that $|f+g|^p = |f+g||f+g|^{p-1} \le (|f|+|g|)|f+g|^{p-1}$. Therefore, we have that

$$\begin{aligned}\|f+g\|^p_{L^p(\Omega)} &= \int_\Omega |f+g|^p dx \le \int_\Omega |f||f+g|^{p-1} dx + \int_\Omega |g||f+g|^{p-1} dx \\ &\overset{\star}{\le} \|f\|_{L^p(\Omega)} \||f+g|^{p-1}\|_{L^{p'}(\Omega)} + \|g\|_{L^p(\Omega)} \||f+g|^{p-1}\|_{L^{p'}(\Omega)}. \end{aligned} \tag{3.5}$$

The $\star$ inequality follows from an application of the Hölder inequality with exponents p and $p' = p/(p-1)$. Then, we have that

$$\||f+g|^{p-1}\|_{L^{p'}(\Omega)} = \left(\int_\Omega |f+g|^{(p-1)\frac{p}{p-1}}\right)^{(p-1)/p} = \|f+g\|^{p-1}_{L^p(\Omega)}.$$

Dividing both members in (3.5) by $\|f+g\|^{p-1}_{L^p(\Omega)}$, we get the desired inequality.

□

3.1. Completeness of L^p

We can prove that L^p spaces are complete with the following theorem.

THEOREM 3.3. *Let $\Omega \subseteq \mathbb{R}^d$ be measurable, $p \in [1, +\infty]$. Then, $L^p(\Omega)$ is complete. Namely, if $\{f_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence, then $\exists f \in L^p(\Omega)$ such that*

$$\lim_{n\to+\infty} ||f_n - f||_{L^p(\Omega)} = 0.$$

Moreover, there exists a subsequence $\{f_{m_k}\}_{k\in\mathbb{N}}$ s.t.

$$f_{m_k}(x) \to f(x) \text{ a.e. in } \Omega,$$
$$|f_{m_k}(x)| \leq g(x) \text{ for some } g \in L^p(\Omega).$$

PROOF. [The case $p = +\infty$ is contained in ex. 3 of Series 7.]
First of all, recall that $\{f_n\}_{n\in\mathbb{N}}$ is Cauchy means that

$$\lim_{m,n\to\infty} \|f_n - f_m\|_{L^p(\Omega)} = 0,$$

i.e.

$$\forall \varepsilon > 0 \quad \exists n_0(\varepsilon) \text{ large enough s.t. } \|f_n - f_m\|_{L^p(\Omega)} \leq \varepsilon \quad \forall n, m \geq n_0(\varepsilon).$$

We begin the proof from the easier case $p = +\infty$: for $m, n \in \mathbb{N}$ with $n \neq m$, define

$$A_{m,n} := \{x \in \Omega : |f_n(x) - f_m(x)| > \|f_n - f_m\|_{L^\infty(\Omega)}\},$$
$$B_n = \{x \in \Omega : |f_n(x)| \geq \|f_n\|_{L^\infty(\Omega)}\}.$$

By definition of L^∞ norm, these sets have measure 0. Therefore, their countable union

$$E = \left(\bigcup_{n\neq m} A_{m,n}\right) \cup \left(\bigcup_{n=1}^{\infty} B_n\right)$$

has measure 0 as well.
We claim that, for $x \in E^C$, $\{f_n(x)\}_{n\in\mathbb{N}}$ is a Cauchy sequence. Indeed, $x \in E^C$ means that

$$x \in \bigcap_{m\neq n} A_{m,n}^C,$$

and therefore $|f_n(x) - f_m(x)| \leq \|f_n - f_m\|_{L^\infty(\Omega)} \quad \forall n, m \in \mathbb{N}$, hence it is a Cauchy sequence. We can then define, for $x \in E^C$, $f(x)$ as the limit of the Cauchy sequence $\{f_n(x)\}_{n\in\mathbb{N}}$, and for $x \in E$ we set $f(x)$ to an arbitrary value.
We now prove that $f \in L^\infty(\Omega)$ and $f_n \to f$ in $L^\infty(\Omega)$; since f_n is a Cauchy sequence in $L^\infty(\Omega)$, for any $\varepsilon > 0 \quad \exists N \in \mathbb{N}$ such that

$$\|f_n - f_m\|_{L^\infty(\Omega)} \leq \varepsilon \quad \forall n, m \geq N.$$

Thus, $\forall x \in E^C$ and if $n, m \geq N$, $|f_n(x) - f_m(x)| \leq \|f_n - f_m\|_{L^\infty(\Omega)} \leq \varepsilon$. Letting $n \to +\infty$,

$$|f(x) - f_m(x)| = \lim_{n\to+\infty} |f_n(x) - f_m(x)| \leq \varepsilon \quad \forall m \geq N.$$

Thus, $|f - f_m| \leq \varepsilon$ a.e., which implies

$$|f| \leq |f - f_m| + |f_m| \leq |f_m| + \varepsilon \text{ a.e.}$$

As a consequence, we get $f \in L^\infty(\Omega)$. Moreover, since $|f - f_m| \leq \varepsilon$ a.e. $\forall m \geq N$, we conclude that

$$\|f - f_m\|_{L^\infty(\Omega)} \leq \varepsilon \quad \forall m \geq N,$$

which finishes the proof for this case.

We now prove the result for $p < +\infty$: we want to prove that $\{f_m\}_{m\in\mathbb{N}}$ is Cauchy in L^p implies that, up to a subsequence, $\{f_m\}_{m\in\mathbb{N}}$ converges to a certain f both in L^p and almost everywhere. Indeed, since $\{f_n\}_{n\in\mathbb{N}}$ is Cauchy

$$\lim_{k\to\infty} \|f_{n_k} - f\|_{L^p(\Omega)} = 0 \iff \lim_{n\to\infty} \|f_n - f\|_{L^p(\Omega)} = 0.$$

To find the candidate limit f, we look for a speedy converging subsequence.

We know from the hypothesis that there exists a subsequence $\{n_k\}_{k\in\mathbb{N}}$ such that

$$\|f_{n_k} - f_{n_{k+1}}\|_{L^p(\Omega)} \le 2^{-k} \quad \forall k \in \mathbb{N}.$$

The existence of this sequence $\{n_k\}_{k\in\mathbb{N}}$ is guaranteed: for example, take $\varepsilon = 2^{-k}$ and $n_k = \max(n_0(2^{-k}), n_{k-1} + 1)$ (with $n_0 = 0$). Now, define

$$f(x) := f_{n_1}(x) + \sum_{k=1}^{\infty} f_{n_{k+1}}(x) - f_{n_k}(x). \tag{3.6}$$

This series is absolutely convergent for almost every x: consider the partial sum of the absolute values up to $h \in \mathbb{N}$:

$$g_h(x) := |f_{n_1}(x)| + \sum_{k=1}^{h} |f_{n_{k+1}}(x) - f_{n_k}(x)|.$$

Thanks to Minkowski's inequality, we have that

$$\|g_h\|_{L^p(\Omega)} \le \|f_{n_1}\|_{L^p(\Omega)} + \sum_{k=1}^{h} \underbrace{\|f_{n_{k+1}} - f_{n_k}\|_{L^p(\Omega)}}_{\le 2^{-k}} \le \|f_{n_1}\|_{L^p(\Omega)} + 1.$$

Let $g(x) := \lim_{h\to\infty} g_h(x)$; this limit exists, because the sequence of $\{g_h\}_{h\in\mathbb{N}}$ is increasing. By monotone convergence theorem, we have that

$$0 \le \int_\Omega g^p = \lim_{h\to+\infty} \int_\Omega g_h^p \le (\|f_{n_1}\|_{L^p(\Omega)} + 1)^p < +\infty.$$

Then, g is finite a.e., hence also f in (3.6) is well defined for a.e. x. For a.e. x,

$$f(x) = \lim_{h\to\infty} f_{n_1}(x) + \sum_{k=1}^{h} f_{n_{k+1}}(x) - f_{n_k}(x) = \lim_{h\to\infty} f_{n_{h+1}}(x),$$

because of the telescopic sum. This means that the speedy subsequence converges pointwise and it is dominated by $g \in L^p(\Omega)$.

To prove L^p convergence, we can use the dominated convergence theorem, because $f_n - f \to 0$ a.e. and $|f_n - f|^p \le [|f_n| + |f|]^p \le 2^p g^p \in L^1(\Omega)$. Hence, we obtain

$$\|f_n - f\|_{L^p(\Omega)}^p = \int_\Omega |f_n - f|^p \to 0 \quad \text{as } n \to +\infty$$

□

Remark 3.6. Let $h \in L^{p_0} \cap L^{p_1}$, with $p_0 < p_1$. Prove that, if $m(\Omega) < +\infty$, we have that $h \in L^p \quad \forall p \in [p_0, p_1]$. Indeed, if $p_1 = \infty$, then $\int_\Omega |h|^p \le \int_\Omega |h|^{p_0} \|h\|_{L^\infty(\Omega)}^{p-p_0}$.

In general, $\forall p \in [p_0, p_1]$, we can write $p = \theta p_0 + (1-\theta)p_1 \quad$ for $\theta \in [0,1]$. Therefore,

$$\int_\Omega |h|^p = \int_\Omega |h|^{\theta p_0} |h|^{(1-\theta)p_1} \overset{\star}{\le} \left(\int_\Omega |h|^{\theta p_0 \frac{1}{\theta}}\right)^{\theta} \left(\int_\Omega |h|^{(1-\theta)p_1 \frac{1}{1-\theta}}\right)^{(1-\theta)}$$

where $\star$ follows from Hölder inequality with $p = 1/\theta$, $p' = 1/(1-\theta)$. Hence,

$$||h||^p_{L^p(\Omega)} \leq ||h||^{\theta p_0}_{L^{p_0}(\Omega)} ||h||^{(1-\theta)p_1}_{L^{p_1}(\Omega)} < +\infty.$$

Remark 3.7. $f_n \to f$ in $L^1(\Omega)$, $m(\Omega) < +\infty$. Then, $\sqrt{1+f_n^2} \to \sqrt{1+f^2}$ in $L^1(\Omega)$. Indeed, suppose by contradiction that

$$\left\| \sqrt{1+f_{n_k}^2} - \sqrt{1+f^2} \right\|_{L^1(\Omega)} > \varepsilon \tag{3.7}$$

for some $\varepsilon > 0$ and a subsequence $\{n_k\}_{k\in\mathbb{N}}$. Since $f_n \to f$ in $L^1(\Omega)$, we can take another subsequence n_{k_h} such that $f_{n_{k_h}} \to f$ a.e. and the convergence is dominated by $g \in L^1(\Omega)$. Then,

$$\sqrt{1+f_{n_{k_h}}^2} - \sqrt{1+f^2} \to 0 \text{ a.e.}$$

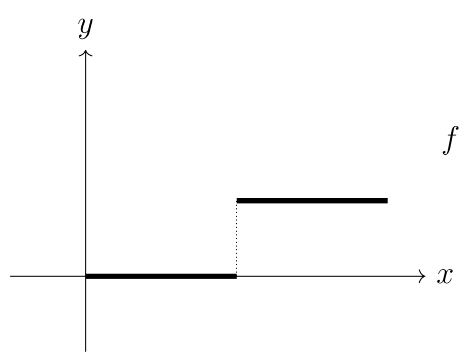
and the convergence is dominated, because

$$\left| \sqrt{1+f_{n_{k_h}}^2} - \sqrt{1+f^2} \right| \leq \sqrt{1+f_{n_{k_h}}^2} + \sqrt{1+f^2} \leq 2\sqrt{1+g^2} \leq 2(1+g) \in L^1(\Omega).$$

Hence, by the dominated convergence theorem, $\int_\Omega \left| \sqrt{1+f_{n_{k_h}}^2} - \sqrt{1+f^2} \right| \to 0$ as $h \to \infty$, but this contradicts (3.7).

3.2. Approximation of L^p functions with $C_c^\infty(\Omega)$ functions

Let us now consider the problem of approximating functions in L^p spaces with C^∞ functions. We will consider this task for the case $p < +\infty$, because we can see with a trivial counterexample that problems arise in L^∞. Indeed, let us consider the function $f(x) = \mathbb{1}_{[1,2]}(x) \in L^\infty([0,2])$: since the uniform limit of continuous functions is continuous, there is no chance to approximate uniformly (even up to throwing away a set of measure 0) this discontinuous function with functions in $C^\infty([0,2])$.



$$f(x) = \begin{cases} 0, & x \in [0,1) \\ 1, & x \in [1,2] \end{cases}$$

FIGURE 1. $f(x) = \mathbb{1}_{[1,2]}(x)$.

For the case $p < +\infty$, instead, we have the possibility to look for approximations in the space of C^∞ functions with compact support, which are defined in the following way:

DEFINITION (Compactly supported functions). If $f : \Omega \to \mathbb{R}^*$, then:

(i) $supp(f) \coloneqq \overline{\{x : f(x) \neq 0\}}$;
(ii) $C_c^0(\Omega) \coloneqq \{f \in C^0(\Omega) : supp(f) \Subset \Omega\}$, where $\Subset$ stands for "compactly contained"[1]};
(iii) $C_c^k(\Omega) \coloneqq C_c^0(\Omega) \cap C^k(\Omega)$.

[1] $A \Subset B$, B open $\Rightarrow \overline{A} \subset B$ and $\overline{A}$ is compact.

THEOREM 3.4. *Let $\Omega \subseteq \mathbb{R}^d$ be an open set, $1 \le p < \infty$, $f \in L^p(\Omega)$. Then, $\exists \{g_k\}_{k\in\mathbb{N}} \subset C_c^\infty(\Omega)$ such that*

$$\lim_{k\to+\infty} \|g_k - f\|_{L^p(\Omega)} = 0.$$

The proof of this result is articulated in 3 parts:

(I) we prove the statement for $\Omega = \mathbb{R}^d$ and approximating functions in $C_c^0(\mathbb{R}^d)$;
(II) we prove the statement for $\Omega \subset \mathbb{R}^d$ and approximating functions in $C_c^0(\Omega)$;
(III) we prove the statement of Theorem 3.4.

PROOF - PART I. We prove the result for $\Omega = \mathbb{R}^d$, and we first work with $g_k \in C_c^0(\mathbb{R}^d)$. We will prove the statement in 5 steps:

(i) We prove the statement for $f(x) = I_B(x)$, being B a box.

For the case $n = 1$, we have that $B = [a, b]$. Define

$$g_\varepsilon(x) = \begin{cases} 1 \quad x \in [a, b) \\ \frac{x}{\varepsilon} + 1 - \frac{a}{\varepsilon} \quad x \in [a-\varepsilon, a) \\ -\frac{x}{\varepsilon} + 1 + \frac{b}{\varepsilon} \quad x \in [b, b+\varepsilon) \\ 0 \quad \text{else} \end{cases} \tag{3.8}$$

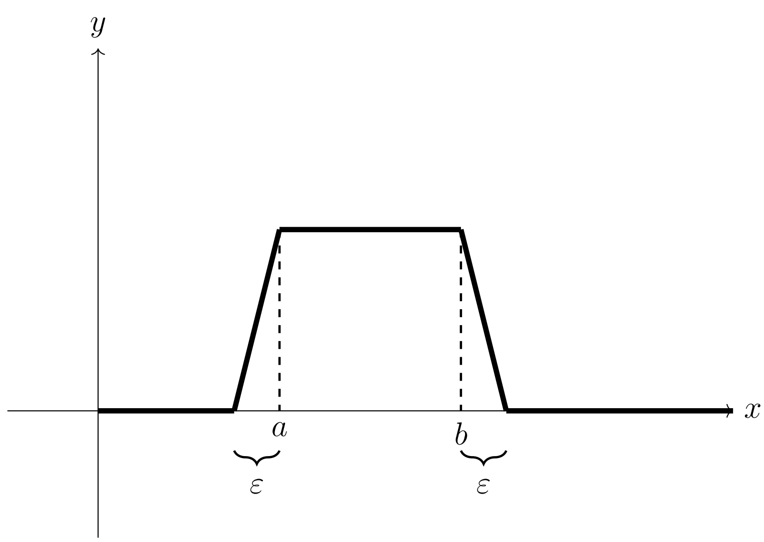


FIGURE 2. $\tilde{g}_\varepsilon(x)$.

We have that $g_\varepsilon \to f$ a.e., and, by the dominated convergence theorem, $\|g_\varepsilon - f\|_{L^p(\Omega)} \to 0$.

For $n > 1$, define $\tilde{g}_\varepsilon(x_1, ..., x_d) = \prod_{i=1}^d g_\varepsilon(x_i)$, where g_ε are defined like (3.8). Again, $\tilde{g}_\varepsilon \to f$ a.e., and, by the dominated convergence theorem, $\|\tilde{g}_\varepsilon - f\|_{L^p(\Omega)} \to 0$.

(ii) We prove the statement for $f(x) = \mathbb{1}_E(x)$, being E measurable and $\overline{E}$ compact[2].

Let $\varepsilon > 0$ and let $\{B_i\}_{i\in\mathbb{N}}$ be a cover made by boxes of E, such that:

$$m(E) \le \sum_{i=1}^{\infty} m(B_i) \le m(E) + \varepsilon, \tag{3.9}$$

[2]Note that, since we are set in $\mathbb{R}^d$, this hypothesis could be replaced by $m(E) < +\infty$.

which means that

$$\int_\Omega \left|\mathbb{1}_E - \sum_{i=1}^\infty \mathbb{1}_{B_i}\right| = \int_\Omega \sum_{i=1}^\infty \mathbb{1}_{B_i} - \mathbb{1}_E = \sum_{i=1}^\infty \int_\Omega \mathbb{1}_{B_i} - \mathbb{1}_E = \sum_{i=1}^\infty m(B_i) - m(E) \le \varepsilon, \tag{3.10}$$

where integral and series have been swapped with the monotone convergence theorem (see Corollary 2.9). Let $N \in \mathbb{N}$ be such that

$$\sum_{i=N+1}^\infty m(B_i) < \varepsilon. \tag{3.11}$$

By (i), we have that $\exists h^i \in C_c^0(\mathbb{R}^d)$ such that:

$$\|h^i - \mathbb{1}_{B_i}\|_{L^p(\mathbb{R}^d)} < \frac{\varepsilon}{N}. \tag{3.12}$$

Now take $h = \sum_{i=1}^N h^i \in C_c^0(\mathbb{R}^d)$. For $p = 1$, we have

$$\begin{aligned}
\|\mathbb{1}_E - h\|_{L^1(\mathbb{R}^d)} &= \|\mathbb{1}_E - \sum_{i=1}^N \mathbb{1}_{B_i} + \sum_{i=1}^N \mathbb{1}_{B_i} - \sum_{i=1}^N h^i\|_{L^1(\mathbb{R}^d)} \\
&\le \|\mathbb{1}_E - \sum_{i=1}^N \mathbb{1}_{B_i}\|_{L^1(\mathbb{R}^d)} + \sum_{i=1}^N \|\mathbb{1}_{B_i} - h^i\|_{L^1(\mathbb{R}^d)} \\
&\overset{(3.12)}{<} \|\mathbb{1}_E - \sum_{i=1}^\infty \mathbb{1}_{B_i}\|_{L^1(\mathbb{R}^d)} + \|\sum_{i=N+1}^\infty \mathbb{1}_{B_i}\|_{L^1(\mathbb{R}^d)} + N\frac{\varepsilon}{N} \\
&\overset{(3.10),(3.11)}{<} 2\varepsilon + \varepsilon = 3\varepsilon.
\end{aligned}$$

with an application of Minkowski's inequality in the first inequality.

If $p > 1$, define $\tilde{h} = (h \wedge 1) \vee 0$ (where $\wedge$ denotes the minimum and $\vee$ denotes the maximum between the two quantities):

$$\|\mathbb{1}_E - \tilde{h}\|^p_{L^p(\Omega)} \le \int_\Omega \underbrace{|\mathbb{1}_E - \tilde{h}|^p}_{\in[0,1]} \le \int_\Omega |\mathbb{1}_E - \tilde{h}| \le \int_\Omega |\mathbb{1}_E - h| \le 3\varepsilon,$$

using the result for $p = 1$ in the last inequality.

(iii) We prove the statement for f simple function.

If f is a simple function, then $f(x) = \sum_{i=1}^N a_i \mathbb{1}_{E_i}(x)$, where E_i are disjoint measurable sets and $\overline{E}_i$ is compact. By (ii), we know that

$$\forall i \quad \exists g_k^i \in C_c^0(\mathbb{R}^d) \text{ s.t. } g_k^i \to \mathbb{1}_{E_i} \text{ in } L^p.$$

Define

$$g_k := \sum_{i=1}^N a_i g_k^i \in C_c^0(\mathbb{R}^d);$$

we have that $g_k \to f$ in L^p because

$$\|g_k - f\|_{L^p(\mathbb{R}^d)} = \|\sum_{i=1}^N a_i(\mathbb{1}_{E_i} - g_k^i)\|_{L^p(\mathbb{R}^d)} \le \sum_{i=1}^N |a_i| \|\mathbb{1}_{E_i} - g_k^i\|_{L^p(\mathbb{R}^d)} \to 0 \quad \text{as } k \to \infty.$$

(iv) We prove the statement for $f \in L^p(\mathbb{R}^d)$ non-negative. Let $\{\phi_k\}_{k\in\mathbb{N}}$ be the sequence of simple functions approximating f given by Lemma 2.3. As mentioned in Remark 2.1, we can choose compactly supported simple functions:

$$\{\mathbb{1}_{B_k}\phi_k\}_{k\in\mathbb{N}} \text{ s.t. } \mathbb{1}_{B_k}\phi_k \uparrow f$$

By step (iii),

$$\forall k \in \mathbb{N} \quad \exists g_k \in C^0_k(\mathbb{R}^d) \text{ s.t. } \|g_k - \mathbb{1}_{B_k}\phi_k\|_{L^p(\mathbb{R}^d)} \le \frac{1}{k}.$$

Now we claim that $g_k \to f$ in L^p:

$$\begin{aligned}\|g_k - f\|_{L^p(\mathbb{R}^d)} &\le \|g_k - \mathbb{1}_{B_k}\phi_k\|_{L^p(\mathbb{R}^d)} + \|\mathbb{1}_{B_k}\phi_k - f\|_{L^p(\mathbb{R}^d)} \\ &\le \frac{1}{k} + \left(\int |\mathbb{1}_{B_k}\phi_k - f|^p\right)^{1/p}\end{aligned}$$

and the last term in the RHS converges to 0 thanks to the dominated convergence theorem.

(v) We prove the statement for a general $f \in L^p(\mathbb{R}^d)$.

If $f = f^+ - f^-$, by (iv) $\exists g_k^+, g_k^- \in C^0_c(\mathbb{R}^d)$ such that

$$\|g_k^+ - f^+\|_{L^p(\mathbb{R}^d)} \to 0, \qquad \|g_k^- - f^-\|_{L^p(\mathbb{R}^d)} \to 0.$$

We claim that $g_k^+ - g_k^- \to f$ in L^p (notice that $g_k^+ - g_k^- \in C^0_c(\mathbb{R}^d)$:

$$\|g_k^+ - g_k^- - f^+ + f^-\|_{L^p(\mathbb{R}^d)} \le \|g_k^+ - f^+\|_{L^p(\mathbb{R}^d)} + \|g_k^- - f^-\|_{L^p(\mathbb{R}^d)} \to 0 \text{ as } k \to \infty,$$

thanks to Minkowski's inequality.

□

Proof - part II. We now need to consider a general open domain $\Omega \subset \mathbb{R}^d$, and we prove that $C^0_c(\Omega)$ is dense in $L^p(\Omega)$.

To do so, consider $f \in L^p(\Omega)$, and extend it to $f \in L^p(\mathbb{R}^d)$ by setting $f = f\mathbb{1}_\Omega$. By the previous part of the proof, we can find

$$\{f_n\}_{n\in\mathbb{N}} \subset C^0_c(\mathbb{R}^d) \text{ s.t. } f_n \to f.$$

Set $g_n(x) := \phi(n\,\mathrm{d}(x, \Omega^C))$, where

$$\phi(y) := \begin{cases} 1, & y \ge 1 \\ 0, & y \le 1/2 \\ 2y - 1, & y \in (1/2, 1). \end{cases}$$

We have that $\{f_n g_n\}_{n\in\mathbb{N}} \subset C^0_c(\Omega)$ and

$$\begin{aligned}\|f_n g_n - f\|_{L^p(\Omega)} &\le \|f(g_n - 1)\|_{L^p(\Omega)} + \|(f_n - f)g_n\|_{L^p(\mathbb{R}^d)} \\ &\le \|f(g_n - 1)\|_{L^p(\Omega)} + \|f_n - f\|_{L^p(\mathbb{R}^d)} \underbrace{\|g_n\|_{L^\infty(\mathbb{R}^d)}}_{\le 1},\end{aligned}$$

where both terms in the RHS go to 0 (the first by dominated convergence theorem with dominant $2f \in L^1(\Omega)$, the second by the first part of the proof). □

3.2.1. Approximation in L^p of C_c^0 functions with C_c^∞ functions. To approximate functions in $C_c^0(\mathbb{R}^d)$ with functions in $C_c^\infty(\mathbb{R}^d)$, we use convolutions.

Let us introduce $\phi \in C_c^\infty(\mathbb{R}^d)$ such that $\phi \geq 0, \phi = 0$ outside B_1, $\int_{\mathbb{R}^d} \phi = 1$; for instance, we could take $\phi(x) = ce^{\frac{1}{|x|^2-1}}$ for $|x| < 1$ and 0 elsewhere (Figure 3).

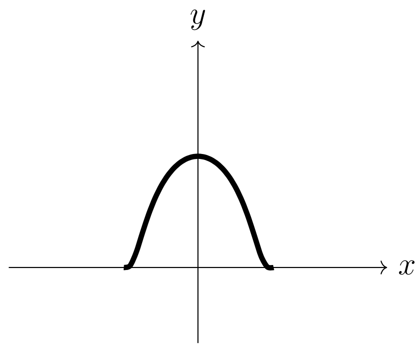


FIGURE 3. $\phi(x) = ce^{\frac{1}{|x|^2-1}}$ for $|x| < 1$ and 0 elsewhere.

DEFINITION (Standard convolution kernel). The *standard mollifier* is

$$\phi_\varepsilon(x) = \varepsilon^{-d}\phi\left(\frac{x}{\varepsilon}\right) \quad \forall \varepsilon > 0.$$

Remark 3.8. Note that by Proposition 2.23

$$\int_{\mathbb{R}^d} \phi_\varepsilon(x)dx = \int_{\mathbb{R}^d} \varepsilon^{-d}\phi(\varepsilon^{-1}x_1, ..., \varepsilon^{-1}x_d)dx \underset{y_i=\varepsilon^{-1}x_i}{=} \int_{\mathbb{R}^d} \phi(y)dy = 1$$

and that ϕ_ε is supported in B_ε.

DEFINITION (Convolution). Now, let $f \in C_c^0(\mathbb{R}^d), g \in C_c^0(\mathbb{R}^d)$ and define the convolution of f and g as

$$(f * g)(x) := \int_{\mathbb{R}^d} f(x-y)g(y)dy.$$

We will call $f_\varepsilon := f * \phi_\varepsilon$.

Lemma 3.5. *Let $f \in C_c^0(\mathbb{R}^d)$; then, $\forall \varepsilon < 1$ we have that*

(i) $supp(f_\varepsilon) \subset supp(f) + B_\varepsilon(0)$;
(ii) $f_\varepsilon \in C_c^\infty(\mathbb{R}^d)$;
(iii) $\|f_\varepsilon\|_{L^1(\mathbb{R}^d)} \leq \|f\|_{L^1(\mathbb{R}^d)}$;
(iv) $f_\varepsilon \to f$ uniformly.

PROOF. (i) $f_\varepsilon(x) = \int_{\mathbb{R}^d} \underbrace{f(x-y)}_{\substack{=0\ \forall y \in B_\varepsilon \\ \text{if } d(x, supp(f)) > \varepsilon}} \underbrace{\phi_\varepsilon(y)}_{=0\ \forall |y| > \varepsilon} dy = 0$

(ii) We can conclude that

$$\lim_{h \to 0} \frac{1}{h}(f_\varepsilon(x+hv) - f_\varepsilon(x)) = \int_{\mathbb{R}^d} f(z)\partial\phi_\varepsilon(x-z)dz$$

by applying Theorem 2.17 on each partial derivative.

As an alternative that does not rely on that statement, observe that, with a change of variables and an application of Fubini-Tonelli, we have

$$f_\varepsilon(x) \overset{x_i - y_i = z_i}{=} \int_{\mathbb{R}^d} f(z)\phi_\varepsilon(x-z)dy$$

Now, we can compute:

$$\frac{1}{h}(f_\varepsilon(x+hv)-f_\varepsilon(x)) = \int_{\mathbb{R}^d} f(z)\frac{\phi_\varepsilon(x+hv-z)-\phi_\varepsilon(x-z)}{h}dz.$$

With the dominated convergence theorem, we can swap the limit for $h \to 0$ with the integral, and recalling that

$$\frac{\phi_\varepsilon(x+hv-z)-\phi_\varepsilon(x-z)}{h} \to \nabla\phi_\varepsilon(x-z)\cdot v \text{ pointwise,}$$

we obtain

$$\lim_{h\to 0}\frac{1}{h}(f_\varepsilon(x+hv)-f_\varepsilon(x)) = \int_{\mathbb{R}^d} f(z)\partial\phi_\varepsilon(x-z)dz.$$

Since we can compute each partial derivative in this way, we get:

$$\nabla(f_\varepsilon) = \nabla(f*\phi_\varepsilon) = f*\nabla\phi_\varepsilon. \tag{3.13}$$

(iii) By definition, we have that:

$$\begin{aligned}\|f_\varepsilon\|_{L^1(\mathbb{R}^d)} = \int_{\mathbb{R}^d}|f_\varepsilon(x)|dx &\le \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}|f(y)|\phi_\varepsilon(x-y)dydx \\ &= \int_{\mathbb{R}^d}|f(y)|\underbrace{\int_{\mathbb{R}^d}\phi_\varepsilon(x-y)dx}_{=1}\,dy = \|f\|_{L^1(\mathbb{R}^d)}\end{aligned}$$

using Tonelli's theorem.

(iv) Note that f is uniformly continuous, therefore

$$\forall\varepsilon'>0\quad \exists\delta>0 \text{ such that } |x-y|<\delta \implies |f(x)-f(y)|<\varepsilon'.$$

Hence, $\forall\varepsilon'>0$,

$$\begin{aligned}|f(x)-f_\varepsilon(x)| &= \left|f(x)-\int_{\mathbb{R}^d}f(x-y)\phi_\varepsilon(y)dy\right| \\ &\overset{\spadesuit}{=} \left|\int_{\mathbb{R}^d}(f(x)-f(x-y))\phi_\varepsilon(y)dy\right| \\ &\le \int_{B_\varepsilon(0)}|f(x)-f(x-y)|\phi_\varepsilon(y)dy \overset{\clubsuit}{\le} \varepsilon'\int_{\mathbb{R}^d}\phi_\varepsilon(y)dy = \varepsilon'\end{aligned}$$

where $\spadesuit$ follows from the fact that $\int_{\mathbb{R}^d}\phi_\varepsilon(y)dy = 1$, and $\clubsuit$ holds if $\varepsilon \le \delta$.

□

We now have all the tools to complete the proof of Theorem 3.4.

Proof - part III. By part II of the proof, we have that $\exists\{g_k\}_{k\in\mathbb{N}} \in C_c^0(\Omega)$ such that

$$\lim_{k\to\infty}\|g_k-f\|_{L^p(\Omega)} = 0.$$

In particular, $\exists K\in\mathbb{N}$ such that

$$\|g_k-f\|_{L^p(\Omega)} < \varepsilon/2 \quad \forall k>K.$$

Now, for any g_k, by (iv) of Lemma 3.5, we have that $\exists h_k \in C_c^\infty(\Omega)$ such that

$$\lim_{n\to\infty}\|g_k-h_k\|_{C^\infty(\Omega)} < \varepsilon/2.$$

Given the compact support, It is then possible to conclude that $\exists K \in \mathbb{N}$ such that

$$\|f - h_k\|_{L^p(\Omega)} \leq \|f - g_k\|_{L^p(\Omega)} + \|g_k - h_k\|_{L^p(\Omega)} < \varepsilon \quad \forall k > K.$$

□

Remark 3.9. Note that (iv) implies that $\|f_\varepsilon - f\|_{L^p(\mathbb{R}^d)} \to 0$ as $\varepsilon \to 0$ for every p, because

$$\|f_\varepsilon - f\|_{L^p(\mathbb{R}^d)} \leq m(supp(f) + B_1)^{1/p} \|f_\varepsilon - f\|_{L^\infty(\mathbb{R}^d)}.$$

Remark 3.10. The space L^2 has a special Hilbert structure.

Define for $f, g \in L^2(\Omega; \mathbb{R})$ or $L^2(\Omega; \mathbb{C})$ the scalar product

$$\langle f, g \rangle = \int_\Omega f(x)\overline{g(x)} dx \in \mathbb{C} \tag{3.14}$$

This quantity is well defined, because by Hölder's inequality we have that:

$$\int_\Omega |f\overline{g}| \leq \left(\int_\Omega |f|^2\right)^{1/2} \left(\int_\Omega |g|^2\right)^{1/2}.$$

The scalar product with the definition in (3.14) has the following properties:

- $\langle f, f \rangle = \int_\Omega |f|^2 = \|f\|^2_{L^2(\Omega)}$
- Hermitian property : $\langle f, g \rangle = \overline{\langle g, f \rangle}$
- It is linear in its first component and anti-linear in the second one: given $c \in \mathbb{C}$, we have:

$$\langle f + g, h \rangle = \langle f, h \rangle + \langle g, h \rangle$$

$$\langle f, g + h \rangle = \langle f, g \rangle + \langle f, h \rangle$$

$$\langle cf, h \rangle = c\langle f, h \rangle$$

$$\langle f, ch \rangle = \overline{c}\langle f, h \rangle$$

- Pythagoras theorem: if $\langle f, g \rangle = 0$, then

$$\|f + g\|^2_{L^2(\Omega)} = \|f\|^2_{L^2(\Omega)} + \|g\|^2_{L^2(\Omega)},$$

  in analogy with vectors in $\mathbb{R}^2$. Indeed,

$$\|f + g\|^2_{L^2(\Omega)} = \langle f, f + g \rangle + \langle g, f + g \rangle = \langle f, f \rangle + \langle g, g \rangle + \langle f, g \rangle + \langle g, f \rangle.$$

3.3. Complementary results in measure theory

Littlewood principles. Speaking of the theory of functions of a real variable, J. E. Littlewood stated in 1944 three principles on the Lebesgue measure, roughly expressible as:

- Every measurable set with finite measure is *nearly* a finite union of boxes.
- Every pointwise converging sequence of functions is *nearly* uniformly convergent.
- Every measurable function is *nearly* continuous.

The adverb *nearly* should be intended in the sense of measure theory, and therefore stands for *up to a set of small measure.*

THEOREM 3.6 (Egorov theorem). *Let $\Omega \subset \mathbb{R}^d$ measurable, $m(\Omega) < +\infty$, $\{f_k\}_{k\in\mathbb{N}}$ measurable, $f_k, f : \Omega \to \mathbb{R}$, such that $f_k \to f$ a.e. Then we have that, given $\varepsilon > 0, \exists C_\varepsilon \subseteq \Omega$ closed such that*

$$m(\Omega \setminus C_\varepsilon) \leq \varepsilon \quad \textit{and } f_k \to f \textit{ uniformly in } C_\varepsilon.$$

PROOF. Without loss of generality, we can consider that $f_k(x) \to f(x) \quad \forall x \in \Omega$ (at least, up to discarding a set of measure 0). $\forall n, k \in \mathbb{N}$ define

$$\Omega_k^n = \left\{ x \in \Omega : |f_j(x) - f(x)| \le \frac{1}{n} \quad \forall j > k \right\}.$$

For fixed n, we have that

$$\Omega_k^n \subset \Omega_{k+1}^n \text{ and } \Omega_k^n \uparrow \Omega \text{ as } k \to \infty.$$

Then,

$$m(\Omega \setminus \Omega_k^n) = \int_\Omega \mathbb{1}_{\Omega \setminus \Omega_k^n} \to 0 \text{ as } k \to \infty,$$

thanks to the dominated convergence theorem using as dominant $\mathbb{1}_\Omega$. This means that $\forall n$ we can fix k_n such that $m(\Omega \setminus \Omega_{k_n}^n) \le 2^{-n}$.
Fix $\varepsilon > 0$ as in the statement, then there exists $N \in \mathbb{N}$ such that

$$\sum_{n=N}^{\infty} 2^{-n} \le \frac{\varepsilon}{2}.$$

Define

$$C_\varepsilon = \bigcap_{n=N}^{\infty} \Omega_{k_n}^n.$$

We have that

$$m(\Omega \setminus C_\varepsilon) = m\left(\bigcup_{n=N}^{\infty} \Omega \setminus \Omega_{k_n}^n \right) \le \sum_{n=N}^{\infty} m(\Omega \setminus \Omega_{k_n}^n) \le \sum_{n=N}^{\infty} 2^{-n} < \varepsilon.$$

Moreover, in C_ε, $f_j \to f$ uniformly; indeed, $\forall \delta > 0$ choose n to be such that $\frac{1}{n} < \delta$. Therefore,

$$|f_j(x) - f(x)| \le \frac{1}{n} < \delta \quad \forall j > k_n, x \in C_\varepsilon \subseteq \Omega_{k_n}^n,$$

which concludes the proof provided that the set C_ε is closed. Even if this is not the case, we can find a closed subset of C_ε up to losing an ε of measure via Proposition 3.7. □

Proposition 3.7. *Let $E \subset \mathbb{R}^d$ be a measurable set. For any $\varepsilon > 0$, there exists an open set U and a closed set C such that $C \subset E \subset U$, $m(U \setminus E) \le \varepsilon$ and $m(E \setminus C) \le \varepsilon$.*

PROOF. We start by proving outer regularity, i.e. $\exists U$ open such that $E \subset U$ and $m(U \setminus E) \le \varepsilon$. First, suppose $m(E) < +\infty$. By definition of Lebesgue measure, we can find countably many open boxes $(B_n)_{n \in \mathbb{N}}$ such that

$$E \subset \bigcup_{n \in \mathbb{N}} B_n \text{ and } \sum_{n \in \mathbb{N}} m(B_n) \le m(E) + \varepsilon.$$

Using subadditivity we deduce

$$m\left(\bigcup_{n \in \mathbb{N}} B_n \right) \le m(E) + \varepsilon.$$

Since E is measurable and of finite measure we get

$$m\left(\bigcup_{n \in \mathbb{N}} B_n \setminus E \right) \le \varepsilon.$$

Notice that we used the inclusion $E \subset \bigcup_{n \in \mathbb{N}} B_n$ to get this last inequality.
To treat the general case with $m(E) = +\infty$, write E as a countable union $(E_n)_{n \in \mathbb{N}}$ of measurable

sets of finite measure (for example E_n is the intersection of E with the ball centered at 0 of radius n) and consider for every $n \in \mathbb{N}$ and open set U_n such that

$$E_n \subset U_n \text{ and } m(U_n \backslash E_n) \leq 2^{-n}\varepsilon.$$

Write $U = \bigcup_{n\in\mathbb{N}} U_n$. Using the inclusion $U\backslash E \subset \bigcup_{n\in\mathbb{N}}(U_n\backslash E_n)$ and subadditivity, we get

$$m(U\backslash E) \leq \sum_{n\in\mathbb{N}} m(U_n\backslash E_n) \leq \varepsilon.$$

Applying this result to E^C, we can find an open set U containing E^C such that $m(U\backslash E^C) \leq \varepsilon$ and setting $C = U^C$ which is closed, we get

$$m(E\backslash C) = m(U\backslash E^C) \leq \varepsilon.$$

□

Remark 3.11. In general, we cannot approximate measurable set from the inside by open sets. For example, take $E = \mathbb{R} \setminus \mathbb{Q}$. Then, $int(E) = \varnothing$ but $m(E) = +\infty$.

Corollary 3.8. *Lebesgue measurable sets are Borel sets up to a set of measure zero. Indeed, let $E \subset \mathbb{R}^d$ be a measurable set then there exists A, B two Borel sets such that $A \subset E \subset B$ and $m(B\backslash A) = 0$.*

Proof. Using the previous proposition, for every $n \in \mathbb{N}$ we can find an open set U_n and a closed set C_n such that $C_n \subset E \subset U_n$ and

$$m(U_n\backslash E) \leq 2^{-n} \text{ and } m(E\backslash C_n) \leq 2^{-n}.$$

Set $A = \bigcup_{n\in\mathbb{N}} C_n$ and $B = \bigcap_{n\in\mathbb{N}} U_n$. These two sets are Borel sets satisfying

$$A \subset E \subset B \text{ and } m(B\backslash A) \leq m(U_n\backslash C_n) \leq 2^{-n+1},\ \forall n \in \mathbb{N}.$$

Letting $n \to \infty$, we get the desired result. □

Remark 3.12. Even if $f_n \to f$ in $L^1((0,1))$, not necessarily we have that $f_n \to f$ uniformly. Indeed, we can consider $f_n(x) = \mathbb{1}_{[0,1/n]}(x)$ and $f(x) = 0$; they are such that $f_n \to f$ in L^1, because $\int_0^1 |f_n| = 1/n$, but the convergence is not uniform, because

$$|f_n(x) - f(x)| = 1 \text{ on } [0, 1/n].$$

Hence, we can conclude that the L^1 convergence is weaker than the L^∞ convergence.

Theorem 3.9 (Lusin's theorem). *Let Ω be a measurable set with $m(\Omega) < +\infty$ and $f : \Omega \to \mathbb{R}$ measurable.*
Then $\forall\varepsilon > 0\ \exists F_\varepsilon \subseteq \Omega$ closed s.t. $m(\Omega \setminus F_\varepsilon) \leq \varepsilon$ and the restriction of f to F_ε, denoted as $f|_{F_\varepsilon}$, is continuous.

Remark 3.13. Note that the thesis of the theorem is that f is continuous when viewed as function defined on the set F_ε, not that f is continuous at the points of F_ε.
Example 3.3. Consider the following examples:

(i) $\mathbb{1}_{[0,+\infty)}$ is continuous when restricted to $[0, +\infty)$, but $x = 0$ is not a point where $\mathbb{1}_{[0,+\infty)}$ is continuous.
(ii) $\mathbb{1}_{\mathbb{Q}\cap[0,1]}$ is continuous when restricted to $\mathbb{Q} \cap [0, 1]$ (because it is identically equal to 1), but seen as a function in $\mathbb{R}$ it is discontinuous everywhere.

Remark 3.14. F_ε cannot be chosen to be open: take as example the function $\mathbb{1}_{\mathbb{Q}\cap[0,1]}$.

PROOF OF LUSIN'S THEOREM. We choose M s.t. $m(\{|f| > M\}) < \varepsilon/2$. Using the approximation of L^1 functions with smooth functions of Theorem 3.4,

$$\exists f_n \to f \mathbb{1}_{\{|f| \leq M\}} \in L^\infty(\Omega) \subset L^1(\Omega).$$

Up to a subsequence, by Theorem 3.3 this convergence holds also almost everywhere.
Now we can apply Egorov's Theorem to make the convergence uniform: let $C_\varepsilon \subseteq \Omega$ from Egorov's Theorem, with $m(C_\varepsilon) \geq m(\Omega) - \varepsilon/2$.
Consider $\tilde{C}_\varepsilon := C_\varepsilon \setminus \{|f| > M\}$: we have that

$$m(\tilde{C}_\varepsilon) \geq m(C_\varepsilon) - \varepsilon/2 \geq m(\Omega) - \varepsilon.$$

We have that $f_n|_{\tilde{C}_\varepsilon}$ is continuous and $f_n \to f$ uniformly in $\tilde{C}_\varepsilon$. Therefore, f is continuous on $\tilde{C}_\varepsilon$.
To make the set closed, take $\tilde{\tilde{C}}_\varepsilon \subset \tilde{C}_\varepsilon$ closed, with $m(\tilde{C}_\varepsilon \setminus \tilde{\tilde{C}}_\varepsilon) < \varepsilon$. □

Remark 3.15. The uniform limit of continuous functions is continuous, but the L^1 limit of continuous functions is not necessarily continuous.

3.4. Comparison between notions of convergence

In Table 1 we summarize the implications among the various types of convergences in a domain with finite measure. Let us consider $\Omega \subset \mathbb{R}^d$ measurable with $m(\Omega) = 1$ (to discard constants in the inequalities, but the results hold in general for the case of Ω with finite measure). Note that we can reduce ourselves to study the convergence to 0, by setting $g_n := f_n - f$ and considering

$$f_n \to f \iff g_n \to 0 \quad \text{as } n \to +\infty.$$

$\Rightarrow$	L^∞ convergence	L^2 convergence	L^1 convergence	a.e. convergence
L^∞ convergence		YES $\|g_n\|_{L^2} \leq \|g_n\|_{L^\infty}$	YES $\|g_n\|_{L^1} \leq \|g_n\|_{L^\infty}$	YES $\forall \varepsilon > 0\ \exists n_0 : \forall n > n_0 \|g_n\| < \varepsilon$ up to a set of measure 0
L^2 convergence	NO* take $\mathbb{1}_{[0,1/n]}$		YES $\|g_n\|_{L^1} \leq \|g_n\|_{L^2}$	NO, in general YES, up to a subsequence
L^1 convergence	NO* take $\mathbb{1}_{[0,1/n]}$	NO take $\sqrt{n}\mathbb{1}_{[0,1/n]}$		NO, in general YES, up to a subsequence
a.e. convergence	NO	NO take $n\mathbb{1}_{[0,1/n]}$	NO take $n\mathbb{1}_{[0,1/n]}$	

TABLE 1. Convergence implications for the case $m(\Omega) < +\infty$.

* : recall that Egorov's theorem states that L^1 and L^2 convergence imply L^∞ convergence up to a set of small measure and up to a subsequence.

In Table 2, we summarize the results for the case $m(\Omega) = +\infty$.

$\nearrow$	L^∞ convergence	L^2 convergence	L^1 convergence	a.e. convergence
L^∞ convergence		NO take $\mathbb{1}_{[0,n]}1/\sqrt{n}$	NO take $\mathbb{1}_{[0,n]}1/n$	YES $\forall\varepsilon > 0\ \exists n_0 : \forall n > n_0 \lvert g_n\rvert < \varepsilon$ up to a set of measure 0
L^2 convergence	NO take $\mathbb{1}_{[0,1/n]}$		NO take $\mathbb{1}_{[0,n]}1/n$	NO, in general YES, up to a subsequence
L^1 convergence	NO take $\mathbb{1}_{[0,1/n]}$	NO take $\sqrt{n}\mathbb{1}_{[0,1/n]}$		NO, in general YES, up to a subsequence
a.e. convergence	NO	NO take $n\mathbb{1}_{[0,1/n]}$	NO take $n\mathbb{1}_{[0,1/n]}$	

TABLE 2. Convergence implications for the case $m(\Omega) = +\infty$.

Note that in this case Egorov's Theorem does not hold, because it requires $m(\Omega) < +\infty$, nor can we apply Hölder's inequality to conclude bounds for norms, because each time we integrate $\int_\Omega 1$ we have $+\infty$.

Remark 3.16. Uniform convergence implies L^∞ convergence, because

$$\operatorname{ess\,sup} |g_n| \leq \sup |g_n|.$$

However, the converse is not true, take for example $g_n(x) = (-1)^n \mathbb{1}_{\{x=0\}}(x)$. Note that this implication is false only for a set of measure 0.

Remark 3.17. Pointwise convergence implies convergence almost everywhere, while the converse is not true in general.

3.5. Application: l^p spaces

In this section, we introduce the l^p spaces of sequence and show how we can use the theory developed so far to study some of their properties.

DEFINITION (l^p norm). Let $(\xi_n)_{n\in\mathbb{N}}$ be a real sequence and $p \in [1, \infty)$. We define

$$\|\xi\|_{l^p} = \left(\sum_{n=1}^{\infty} |\xi_n|^p\right)^{\frac{1}{p}}.$$

DEFINITION (l^p space). Let $(\xi_n)_{n\in\mathbb{N}}$ be a real sequence and $p \in [1, \infty)$. We define

$$(l^p, \|\cdot\|_{l^p}) = \left(\{(\xi_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} : \|\xi\|_{l^p} < +\infty\}, \|\cdot\|_{l^p}\right).$$

DEFINITION (l^∞ norm). Let $(\xi_n)_{n\in\mathbb{N}}$ be a real sequence. We define

$$\|\xi\|_{l^\infty} = \sup_{n\in\mathbb{N}} |\xi_n|.$$

DEFINITION (l^∞ space). Let $(\xi_n)_{n\in\mathbb{N}}$ be a real sequence. We define

$$(l^\infty, \|\cdot\|) = \left(\{(\xi_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} : \|\xi\|_{l^\infty} < +\infty\}, \|\cdot\|_{l^\infty}\right).$$

It is a simple matter to show that l^p spaces are vector spaces for all $p \in [1, +\infty]$. We treat the case $p \in [1, +\infty)$, take $\xi, \eta \in l^p$, then

$$\sum_{n=1}^{\infty} |\xi_n + \eta_n|^p \leqslant 2^{p-1} \left(\sum_{n=1}^{\infty} |\xi_n|^p + \sum_{n=1}^{\infty} |\eta_n|^p\right) < +\infty.$$

We will show that l^p are actually complete normed vector spaces (Banach spaces). One can do this directly, but we will embed these spaces into $L^p(1, \infty)$ and study this embedding to deduce some

properties.
To do this, for every sequence $(\xi_n)_{n\in\mathbb{N}}$, consider the function $f_\xi : (1,+\infty) \to \mathbb{R}$ defined by $f_\xi(x) = \xi_n$ if $n < x \le n+1$ for every $n \in \mathbb{N}$.

Proposition 3.10. *The application $\xi \mapsto f_\xi$ is linear and for every $p \in [1,+\infty]$, and we have the following properties*

(i) $\xi \in l^p$ if and only if $f_\xi \in L^p(1,+\infty)$, in which case $\|\xi\|_{l^p} = \|f_\xi\|_{L^p(1,+\infty)}$,
(ii) The space $\{f_\xi : \xi \in l^p\}$ is a closed subspace of $L^p(1,+\infty)$, hence l^p is a Banach space,
(iii) (Hölder) if $\xi \in l^p$ and $\eta \in l^{p'}$, then

$$\sum_{n=1}^{\infty} |\xi_n \eta_n| \leqslant \|\xi\|_{l^p}\|\eta\|_{l^{p'}},$$

(iv) (Minkowski) if $\xi, \eta \in l^p$ then

$$\|\xi+\eta\|_{l^p} \le \|\xi\|_{l^p} + \|\eta\|_{l^p}.$$

Proof. We first prove that $\xi \mapsto f_\xi$ is a linear map.
Take ξ, η two real sequences and $\alpha \in \mathbb{R}$ then for every $n \in \mathbb{N}$ and $n < x \le n+1$, we have

$$f_{\alpha\xi+\eta}(x) = (\alpha\xi+\eta)_n = \alpha\xi_n + \eta_n = \alpha f_\xi(x) + f_\eta(x).$$

(i) For $p \in [1,+\infty)$ and ξ a real sequence, we have

$$\int_1^\infty |f_\xi(x)|^p dx = \sum_{n=1}^{\infty} \int_n^{n+1} |f_\xi(x)|^p dx = \sum_{n=1}^{\infty} |\xi_n|^p,$$

from which we deduce $\|\xi\|_{l^p} = \|f\|_{L^p(1,+\infty)}$ and $\xi \in l^p$ if and only if $f \in L^p(1,+\infty)$.
For $p = +\infty$, for every $\varepsilon > 0$ and $n \in \mathbb{N}$, we have $m(\{x \in \mathbb{R} : |f_\xi(x)| > |\xi_n| - \epsilon\}) \ge 1 > 0$ from which we get $\|f_\xi\|_{L^\infty(1,+\infty)} \ge \|\xi\|_{l^\infty}$. Conversely, for every $\varepsilon > 0$ and every $x \in \mathbb{R}$ we have $|f_\xi(x)| \le \|\xi\|_{l^\infty} + \varepsilon$ from which we get $\|f_\xi\|_{L^\infty(1,+\infty)} \le \|\xi\|_{l^\infty}$.
This proves that the map $\xi \mapsto f_\xi$ is a linear isometry from l^p to $L^p(1,+\infty)$.

(ii) Take $(f_{\xi_n})_{n\in\mathbb{N}}$ a sequence in $L^p(1,+\infty)$ converging in $L^p(1,+\infty)$ to f. We need to prove that there exists $\xi \in l^p$ such that $f = f_\xi$. Since $\lim_{n\to\infty} f_{\xi_n} = f$ in $L^p(1,+\infty)$, we can find a subsequence that converges pointwise almost everywhere to f. For every $k \in \mathbb{N}$ and every $n \in \mathbb{N}$, f_{ξ_n} is constant on $(k, k+1]$. This implies that for every $k \in \mathbb{N}$, we can find a constant $\xi_k \in \mathbb{R}$ such that $f(x) = \xi_k$ for almost every $x \in (k, k+1]$. Indeed, if this were not the case, we could find $k \in \mathbb{N}$ and $\alpha, \beta \in \mathbb{R}$ such that $\alpha < \beta$ and $m(\{x \in (k,k+1) : f(x) < \alpha\})$, $m(\{x \in (k,k+1) : f(x) > \beta\}) > 0$ which would contradict the convergence pointwise almost everywhere. So we have that $f = f_\xi$ almost everywhere on $(1,+\infty)$, in particular $f = f_\xi$ in $L^p(1,+\infty)$ and $\{f_\xi : \xi \in l^p\}$ is a closed subspace of $L^p(1,+\infty)$. From (i), we deduce that $\xi \in l^p$ and that l^p is a Banach space.

(iii) (Hölder)

$$\sum_{n=1}^{\infty} |\xi_n\eta_n| = \int_1^\infty |f_\xi f(\eta)| \le \|f_\xi\|_{L^p(1,+\infty)}\|f_\eta\|_{L^{p'}(1,+\infty)} = \|\xi\|_{l^p}\|\eta\|_{l^{p'}}.$$

(iv) (Minkowski)

$$\|\xi+\eta\|_{l^p} = \|f_\xi + f_\eta\|_{L^p(1,+\infty)} \le \|f_\xi\|_{L^p(1,+\infty)} + \|f_\eta\|_{L^p(1,+\infty)} = \|\xi\|_{l^p} + \|\eta\|_{l^p}.$$

□

We conclude this section by saying that while this identification of l^p as a subspace of $L^p(1,+\infty)$ is useful, it does not allow us to deduce all properties of l^p spaces. To state an example, we know that when $\Omega \subset \mathbb{R}^d$ is of infinite measure, we do not have inclusions $L^p(\Omega) \subset {}^q(\Omega)$ for $p, q \in [1,+\infty]$ in general. However the following holds

Proposition 3.11. *For $1 \leq p \leq q \leq +\infty$ and ξ a real sequence, we have*

$$\|\xi\|_{l^q} \leq \|\xi\|_{l^p},$$

and so we have the inclusions $l^p \subset l^q$.

Proof. The case $q = \infty$ is easy and we do not prove it, so assume $q < \infty$.
We may assume $\xi \in l^p$ and $\xi \neq 0$. Notice that the inequality is homogeneous and we may suppose without loss of generality that $\|\xi\|_{l^p} = 1$. In particular, for every $n \in \mathbb{N}$, we have $|\xi_n| \leq 1$. Since $p \leq q$, we have $|\xi_n|^q \leq |\xi_n|^p$ for every $n \in \mathbb{N}$, which gives

$$\sum_{n=1}^{\infty} |\xi_n|^q \leqslant \sum_{n=1}^{\infty} |\xi_n|^p = 1.$$

□

CHAPTER 4

Fourier Analysis

This chapter is inspired by [Dac, Chapter 17], [SS03, Chapter 1, 2, 3], [Tao16, Chapter 5].

The development of Fourier Analysis dates back to the XVIII century, and it was encouraged by the investigation on physical phenomena regulated by partial differential equations, such as vibrating strings and heat flows.

The laws describing the above experimental settings are, respectively, the wave and the heat equation; their solutions were sought using linear combinations of sinusoidal terms - that is, the underlying concept of Fourier series.

In 1807, J.Fourier was the first to study systematically the properties of infinite sums of harmonics, but preliminary investigations were carried out for example even by D.Bernoulli, who in 1753 wrote to Euler to propose the solution to certain partial differential equations given by the Fourier series. However, Euler was not entirely convinced of its full generality, because the result could hold only if the function could be expanded in Fourier series. Such doubts, shared also with other mathematicians, were then solved by Fourier in 1807 in his study of the heat equation, which eventually led others to a complete proof that a general function could be represented as a Fourier series.

4.1. Derivation of the heat equation and solution of the Laplace problem in a disk

To see the emergence of Fourier series from physical problems, we will now derive the heat equation, a partial differential equation that formalizes Newton's law of cooling, according to which the heat flows from higher to lower temperatures, at a rate proportional to the difference of the temperatures in the regions.

Let us consider a metal plate $\Omega \subset \mathbb{R}^2$, characterized by a certain specific heat $\sigma > 0$, quantity describing the heat capacity of the material, and conductivity $\kappa > 0$; our aim is to study the evolution of the temperature $u(x, y, t)$, starting from a given initial distribution at time $t = 0$.

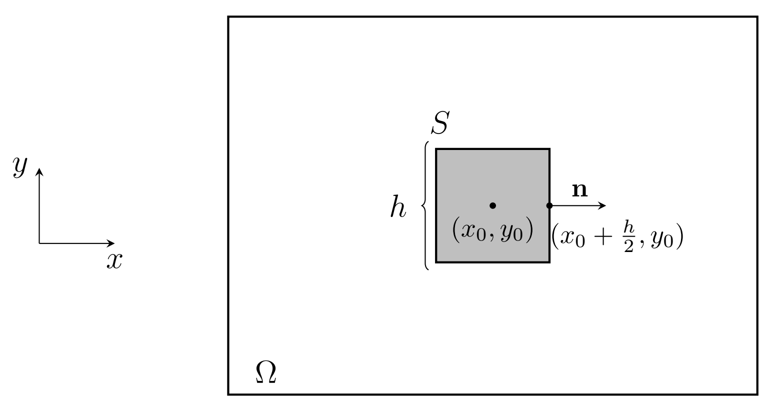


FIGURE 1. Sketch of the square S.

Consider a small square centered in (x_0, y_0) inside the plate $S \subset \Omega$, with edges of length $h \ll 1$ parallel to the axes (see Figure 1), and define the following quantities:

- the amount of heat in S at time t: $H(t) := \sigma \int \int_S u(x,y,t)$;
- the heat flow into S: $\frac{\partial H}{\partial t}(t) := \frac{\partial}{\partial t}\sigma \int \int_S u(x,y,t) = \sigma \int \int_S \partial_t u(x,y,t) \simeq \sigma h^2 \partial_t u(x_0,y_0,t)$; this approximation can be made because we can assume the integrand function to be constant on the small square S, and the area of the square is h^2.
- the incoming heat flow through the boundary ∂S, considered to be positive in the direction given by the vector $\mathbf{n}$:

$$\kappa[h\partial_x u(x_0+\tfrac{h}{2},y_0,t) - h\partial_x u(x_0-\tfrac{h}{2},y_0,t) + h\partial_y u(x_0,y_0+\tfrac{h}{2},t) - h\partial_y u(x_0,y_0-\tfrac{h}{2},t)]$$
$$\simeq \kappa[h^2\partial_{xx}u(x_0,y_0,t) + h^2\partial_{yy}u(x_0,y_0,t)].$$

Newton's law of cooling relates the rate of the heat flow to the difference of the temperatures, to be interpreted as a gradient:

$$\sigma h^2 \partial_t u(x_0,y_0,t) = \kappa h^2(\partial_{xx}u(x_0,y_0,t) + \partial_{yy}u(x_0,y_0,t)). \tag{4.1}$$

From Equation 4.1, by simplifying h^2, we finally obtain the heat equation:

$$\partial_t u(x,y,t) = \frac{\kappa}{\sigma}(\partial_{xx}u(x,y,t) + \partial_{yy}u(x,y,t)). \tag{4.2}$$

After sufficiently long time, the heat exchange will be over and thermal equilibrium will be reached. Therefore, $\partial_t u \sim 0$, and the phenomenon will be described by the steady state version of Equation 4.2, known as Laplace equation:

$$\Delta u(x,y) := \partial_{xx}u(x,y) + \partial_{yy}u(x,y) = 0. \tag{4.3}$$

Functions satisfying Equation 4.3, and therefore having null Laplacian Δ, are called *harmonic* functions.

Let us now consider the Dirichlet problem for the Laplace equation in the unit disk

$$D = \left\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\right\}.$$

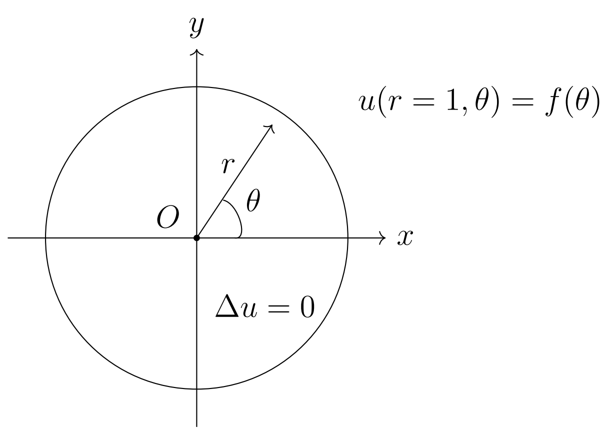


FIGURE 2. Laplace problem in the unit disk in $\mathbb{R}^2$.

Passing to polar coordinates with the usual change of variables $x = r\cos(\theta), y = r\sin(\theta)$, we can reformulate the domain expression as

$$D = \left\{(r,\theta) \in \mathbb{R}^2 : 0 \leq r \leq 1, \theta \in [0,2\pi)\right\}.$$

We then fix Dirichlet boundary conditions in polar coordinates:

$$u(r=1,\theta) = f(\theta),$$

being f a given function that imposes the value of u on the disk boundary, that, in the physical setting described before, corresponds to the temperature distribution at the edge.

We will now rewrite Equation 4.3 in polar coordinates, recalling that the expression of the Laplacian for a function in polar coordinates reads as:

$$\Delta u(x,y) = \Delta u(r\cos(\theta), r\sin(\theta)) = \partial_{rr}\omega(r,\theta) + \frac{1}{r}\partial_r\omega(r,\theta) + \frac{1}{r^2}\partial_{\theta\theta}\omega(r,\theta). \tag{4.4}$$

where $\omega(r,\theta) = u(r\cos(\theta), r\sin(\theta))$.

Proof of (4.4). Let us define $\omega(r,\theta) \coloneqq u(r\cos(\theta), r\sin(\theta))$, and let us relate its partial derivatives with the ones of u, by using the chain rule:

$$\partial_r\omega(r,\theta) = \partial_x u(r\cos(\theta), r\sin(\theta))\cos(\theta) + \partial_y u(r\cos(\theta), r\sin(\theta))\sin(\theta)$$

$$\partial_\theta\omega(r,\theta) = -\partial_x u(r\cos(\theta), r\sin(\theta))r\sin(\theta) + \partial_y u(r\cos(\theta), r\sin(\theta))r\cos(\theta)$$

$$\begin{aligned}\partial_{rr}\omega(r,\theta) = {} & [\partial_{xx}u(r\cos(\theta), r\sin(\theta))\cos(\theta) + \partial_{xy}u(r\cos(\theta), r\sin(\theta))\sin(\theta)]\cos(\theta) + \\ & + [\partial_{xy}u(r\cos(\theta), r\sin(\theta))\cos(\theta) + \partial_{yy}u(r\cos(\theta), r\sin(\theta))\sin(\theta)]\sin(\theta)\end{aligned}$$

$$\begin{aligned}\partial_{\theta\theta}\omega(r,\theta) = {} & [\partial_{xx}u(r\cos(\theta), r\sin(\theta))r\sin(\theta) - \partial_{xy}u(r\cos(\theta), r\sin(\theta))r\cos(\theta)]r\sin(\theta) + \\ & - \partial_x u(r\cos(\theta), r\sin(\theta))r\cos(\theta) + \\ & + [-\partial_{xy}u(r\cos(\theta), r\sin(\theta))r\sin(\theta) + \partial_{yy}u(r\cos(\theta), r\sin(\theta))r\cos(\theta)]r\cos(\theta) + \\ & - \partial_y u(r\cos(\theta), r\sin(\theta))r\sin(\theta)\end{aligned}$$

After these computations, one can verify that

$$\begin{aligned}\frac{1}{r^2}[r^2\partial_{rr}\omega(r,\theta) + r\partial_r\omega(r,\theta) + & \partial_{\theta\theta}\omega(r,\theta)] \\ & = [\partial_{xx}u(r\cos(\theta), r\sin(\theta)) + \partial_{yy}u(r\cos(\theta), r\sin(\theta))](\cos^2(\theta) + \sin^2(\theta)) \\ & = \partial_{xx}u(r\cos(\theta), r\sin(\theta)) + \partial_{yy}u(r\cos(\theta), r\sin(\theta))\end{aligned}$$

□

By inserting Equation 4.4 into Equation 4.3, we finally obtain the Laplace problem with Dirichlet boundary conditions in polar coordinates:

$$\begin{cases} r^2\partial_{rr}\omega(r,\theta) + r\partial_r\omega(r,\theta) = -\partial_{\theta\theta}\omega(r,\theta) \\ \omega(1,\theta) = f(\theta). \end{cases} \tag{4.5}$$

We will look for solutions using the method of separation of variables; namely, we ask for solutions of the form:

$$\omega(r,\theta) = F(r)G(\theta),$$

where G must be periodic with period 2π, as its variable θ represents the angle on the circle. Plugging such solutions inside the first equation of the system in (4.5), we get

$$r^2F''(r)G(\theta) + rF'(r)G(\theta) = F(r)G''(\theta),$$

then, by dividing both members for $F(r)G(\theta)$, we obtain

$$\frac{r^2F''(r) + rF'(r)}{F(r)} = -\frac{G''(\theta)}{G(\theta)},$$

where the LHS depends only on r and the RHS only on θ. Since the two sides depend on independent variables, both the terms have to be equal to some constant $\lambda \in \mathbb{R}$, and we get the system

$$\begin{cases} G''(\theta) + \lambda G(\theta) = 0 \\ r^2 F''(r) + rF'(r) - \lambda F(r) = 0. \end{cases} \tag{4.6}$$

Remark 4.1. For an equation of the form

$$u''(t) + au'(t) + bu(t) = 0, \quad t \in I \subseteq \mathbb{R}, \tag{4.7}$$

where $a, b \in \mathbb{R}$ are constant coefficients (do not depend on t), the solutions can be found as follows: we look for exponential solutions $z(t) = e^{\lambda t}$. Substituting into the equation gives

$$\lambda^2 e^{\lambda t} + a\lambda e^{\lambda t} + be^{\lambda t} = 0,$$

which leads to solving the characteristic equation:

$$\lambda^2 + a\lambda + b = 0.$$

We distinguish three cases based on the sign of the discriminant $\Delta := a^2 - 4b$.

- **Case** $\Delta > 0$: two distinct real roots:

  $$\lambda_1 = -\frac{a}{2} + \frac{\sqrt{a^2 - 4b}}{2}, \quad \lambda_2 = -\frac{a}{2} - \frac{\sqrt{a^2 - 4b}}{2}.$$

  Then $z_1(t) = e^{\lambda_1 t}$ and $z_2(t) = e^{\lambda_2 t}$ are linearly independent solutions. The general solution of the homogeneous equation is:

  $$u(t) = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t}, \tag{4.8}$$

  with arbitrary constants $C_1, C_2 \in \mathbb{R}$.
- **Case** $\Delta < 0$: two complex conjugate roots:

  $$\lambda_1 = -\frac{a}{2} + i\mu, \quad \lambda_2 = -\frac{a}{2} - i\mu, \quad \text{with } \mu = \frac{\sqrt{4b - a^2}}{2}.$$

  This leads to complex-valued functions $e^{\lambda_1 t}$, $e^{\lambda_2 t}$. Define real-valued solutions:

  $$z_1(t) = \frac{e^{\lambda_1 t} + e^{\lambda_2 t}}{2} = e^{-\frac{a}{2}t}\cos(\mu t),$$
  $$z_2(t) = \frac{e^{\lambda_1 t} - e^{\lambda_2 t}}{2i} = e^{-\frac{a}{2}t}\sin(\mu t).$$

  These are linearly independent. Therefore, the general solution is:

  $$u(t) = e^{-\frac{a}{2}t}(C_1 \cos(\mu t) + C_2 \sin(\mu t)). \tag{4.9}$$
- **Case** $\Delta = 0$: a single real root $\lambda = -\frac{a}{2}$.

  Then $z_1(t) = e^{-\frac{a}{2}t}$ is a solution. We seek a second linearly independent solution of the form $z_2(t) = C(t)z_1(t)$, and we find $z_2(t) = te^{-\frac{a}{2}t}$. Therefore, the general solution is:

  $$u(t) = (C_1 + C_2 t)e^{-\frac{a}{2}t}. \tag{4.10}$$

Note that if (4.7) is set in $I = (0, L)$ and it comes with a boundary condition such as $u(0) = u(L) = 0$, we can find solutions only in the case $\Delta < 0$. Indeed, with solutions from the family (4.8), we would need to impose $u(0) = C_1 + C_2 = 0$ and $u(L) = C_1 e^{\lambda_1 L} + C_2 e^{\lambda_2 L} = 0$, which cannot be satisfied, and with the family (4.10) we would need to impose $u(0) = C_1 = 0$, $u(L) = (C_1 + C_2 L)e^{-aL/2} = 0$ which again does not lead to any non-zero solution.

By Remark 4.1, the solutions to the ODE in θ are:

- $e^{\pm i\sqrt{\lambda}\theta}$ (that is, $\cos(\sqrt{\lambda}\theta), \sin(\sqrt{\lambda}\theta)$) if $\lambda \geq 0$;
- $e^{\sqrt{-\lambda}\theta}$ if $\lambda < 0$.

However, the second category of solutions is not periodic, so we want to discard them. Therefore, the periodicity constraint implies that $\lambda = m^2$, for some $m \in \mathbb{N}$, and

$$G(\theta) = \tilde{A}\cos(m\theta) + \tilde{B}\sin(m\theta) \overset{1}{=} Ae^{im\theta} + Be^{-im\theta}$$

for some constants $A, B \in \mathbb{R}$.

On the other hand, the solutions to $F(r)$ are of the form

$$F(r) = \begin{cases} r^m \\ r^{-m} \end{cases} \quad \text{if } m > 0, \quad F(r) = \begin{cases} 1 \\ \log r \end{cases} \quad \text{if } m = 0.$$

But we can reject r^{-m} and $\log(r)$ as they are unbounded and they blow up in the origin. Hence, the solutions obtained via separation of variables are $\omega_m = r^{|m|}e^{im\theta}$ with $m \in \mathbb{Z}$.
Note that ω_m can assume complex values due to the presence of $e^{im\theta}$, and the fact that it is a solution implies that its real and imaginary part $\mathrm{Re}(\omega_m) = r^{|m|}\cos(m\theta)$ and $\mathrm{Im}(\omega_m) = r^{|m|}\sin(m\theta)$ are solutions as well.

Remark 4.2. Since Equation 4.3 is a linear partial differential equation, the superposition principle holds: if u_1, u_2 are solutions to the equation, then a linear combination $a_1u_1 + a_2u_2$ is solution, too, $\forall a_1, a_2 \in \mathbb{C}$.

PROOF.

$$\begin{aligned} \partial_{xx}(a_1u_1 + a_2u_2) + \partial_{yy}(a_1u_1 + a_2u_2) &= a_1\partial_{xx}u_1 + a_2\partial_{xx}u_2 + a_1\partial_{yy}u_1 + a_2\partial_{yy}u_2 = \\ &= a_1(\partial_{xx}u_1 + \partial_{yy}u_1) + a_2(\partial_{xx}u_2 + \partial_{yy}u_2) = \\ &= 0 \end{aligned}$$

□

Thanks to the remark, we can conclude that a finite sum $\omega(r,\theta) = \sum_{m=-N}^{N} a_m r^{|m|}e^{im\theta}$ (where $N \in \mathbb{N}$) is a solution.

Let us now consider the Dirichlet boundary condition: on the disk edge, corresponding to $r = 1$, we have that the boundary value of the solution must match with the boundary datum f:

$$\omega(1,\theta) = \sum_{m=-N}^{N} a_m e^{im\theta} = f(\theta).$$

The previous discussion proves the following theorem:

THEOREM 4.1. *If $f(\theta)$ can be written as*

$$f(\theta) = \sum_{m=-N}^{N} a_m e^{im\theta} = \sum_{m=-N}^{N} a_m(\cos(m\theta) + i\sin(m\theta))$$

for some coefficients $\{a_m\}_{m\in\mathbb{N}} \subseteq \mathbb{C}$, then the solution of the Laplace equation in the unit disk D is given by

$$\omega(r,\theta) = \sum_{m=-N}^{N} a_m r^{|m|}e^{im\theta}. \tag{4.11}$$

[1] Recall Euler's identity $e^{ix} = \cos(x) + i\sin(x)$.

Remark 4.3. The solution of Equation 4.5 in the disk with boundary condition

$$f(\theta) = \sum_{m=-N}^{N} a_m \cos(m\theta) + b_m \sin(m\theta)$$

is a polynomial in (x, y). This can be proved by expanding $\cos(m\theta)$ and $\sin(m\theta)$ into combinations of $\sin(\theta)$ and $\cos(\theta)$, and switching back to cartesian coordinates (recall that $e^{i\theta} = \cos(\theta) + i\sin(\theta)$). Taking $N = 0$ in (4.11), we get

$$\omega(r, \theta) = a_0.$$

For $N = 1$, we get:

$$\begin{aligned}\omega(r,\theta) &= a_0 + a_1 re^{i\theta} + a_{-1}re^{-i\theta} = a_0 + a_1(r\cos(\theta) + ir\sin(\theta)) + a_{-1}(r\cos(\theta) - i\sin(\theta)) \\ &= a_0 + (a_1 + a_{-1})x + i(a_1 - a_{-1})y\end{aligned}$$

For $N = 2$, we get:

$$\begin{aligned}\omega(r,\theta) &= a_0 + a_1 re^{i\theta} + a_{-1}re^{-i\theta} + a_2 re^{i2\theta} + a_{-2}re^{-i2\theta} \\ &= a_0 + a_1(r\cos(\theta) + ir\sin(\theta)) + a_{-1}(r\cos(\theta) - i\sin(\theta)) \\ &\quad + a_2(r^2\cos(2\theta) + ir^2\sin(2\theta)) + a_{-2}(r^2\cos(2\theta) - ir^2\sin(2\theta)) \\ &= a_0 + (a_1 + a_{-1})x + i(a_1 - a_{-1})y \\ &\quad + a_2(r(\cos(\theta) + \sin(\theta))r(\cos(\theta) - \sin(\theta)) + 2ir^2\sin(\theta)\cos(\theta)) \\ &\quad + a_{-2}(r(\cos(\theta) + \sin(\theta))r(\cos(\theta) - \sin(\theta)) - 2ir^2\sin(\theta)\cos(\theta)) \\ &= a_0 + (a_1 + a_{-1})x + i(a_1 - a_{-1})y + a_2((x+y)(x-y) + 2ixy) + a_{-2}((x+y)(x-y) - 2ixy).\end{aligned}$$

We can give evidence for the statement with the following example.

Example 4.1. Consider $f(\theta) = 1 + \sin(2\theta) = 1 + \frac{1}{2i}e^{2i\theta} - \frac{1}{2i}e^{-2i\theta}$; we know that its solution is

$$\omega(r,\theta) = 1r^0e^{i0\theta} + \frac{1}{2i}r^2e^{i2\theta} - \frac{1}{2i}r^2e^{-i2\theta} = 1 + r^2\sin(2\theta).$$

Expressing $1 + r^2\sin(2\theta)$ in cartesian coordinates, we get that

$$1 + r^2\sin(2\theta) = 1 + 2r^2\sin(\theta)\cos(\theta) = 1 + 2(r\sin(\theta))(r\cos(\theta)) = 1 + 2xy,$$

which solves the Laplace equation in cartesian coordinates: $\Delta(1+2xy) = \partial_{xx}(1+2xy)+\partial_{yy}(1+2xy) = 0$.

The superposition principle can be then extended to the case of an infinite sum: therefore, if $f(\theta) = \sum_{m\in\mathbb{Z}} a_m e^{im\theta}$ for some complex coefficients $\{a_m\}_m \subset \mathbb{C}$, then a solution of the heat equation is

$$\omega(r,\theta) = \sum_{m\in\mathbb{Z}} a_m r^{|m|} e^{im\theta}. \tag{4.12}$$

This heuristic motivates the leading question of Fourier analysis, that is:

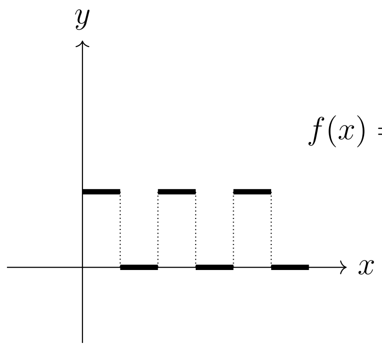
Given $f : [0, 2\pi] \to \mathbb{R}$ such that $f(0) = f(2\pi)$, when can we find coefficients a_m to write it as $f(\theta) = \sum_{m\in\mathbb{Z}} a_m e^{im\theta}$?

The path to answer this question starts from the investigation of periodic functions.

4.2. Periodic Functions

Definition (Periodic function). Let $L > 0$, $f : \mathbb{R} \to \mathbb{C}$ is L-periodic if $f(x+L) = f(x) \quad \forall x \in \mathbb{R}$.

Example 4.2. $f(x) = \sin(x)$ is $2\pi-$periodic, $g(x) = \sin(2\pi x)$ and the square wave (see Figure 3) are $1-$periodic (also called $\mathbb{Z}-$periodic). The constant function $h(x) = 1$ is $L-$periodic $\forall L > 0$.



$$f(x) = \begin{cases} 1, & x \in [n, n+1/2) \\ 0, & x \in [n+1/2, n+1) \end{cases} \qquad n \in \mathbb{N}$$

Figure 3. A $1-$periodic function: the square wave.

Remark 4.4. For f, g 1-periodic, we have that $\|f - g\|_{L^p} = \left(\int_0^1 |f-g|^p\right)^{\frac{1}{p}}$. For $p = 2$, we can define the scalar product to be the quantity $\langle f, g\rangle = \int_0^1 f\overline{g}dx$.

Definition. The space of continuous 1-periodic functions is denoted by $C^0(\mathbb{R}/\mathbb{Z};\mathbb{C})$.

Lemma 4.2. *The following basic properties for periodic functions hold:*

(i) If $f \in C^0(\mathbb{R}/\mathbb{Z};\mathbb{C})$, then f is bounded:

$$\exists M > 0 \ s.t. \ |f(x)| < M \quad \forall x \in \mathbb{R}.$$

(ii) $\mathbb{C}^0(\mathbb{R}/\mathbb{Z};\mathbb{C})$ is a vector space and an algebra:

$$f, g \in \mathbb{C}^0(\mathbb{R}/\mathbb{Z};\mathbb{C}), c \in \mathbb{C} \Rightarrow f+g, cf, fg \in \mathbb{C}^0(\mathbb{R}/\mathbb{Z};\mathbb{C}).$$

(iii) The space is is closed under uniform limits:

$$\{f_n\}_{n\in\mathbb{N}} \subset \mathbb{C}^0(\mathbb{R}/\mathbb{Z};\mathbb{C}), f_n \to f \ \textit{uniformly} \ \Rightarrow f \in \mathbb{C}^0(\mathbb{R}/\mathbb{Z};\mathbb{C}).$$

(iv) The space $\mathbb{C}^0(\mathbb{R}/\mathbb{Z};\mathbb{C})$ is dense in $L^2((0,1);\mathbb{C})$.

Remark 4.5. $\mathbb{C}^0(\mathbb{R}/\mathbb{Z};\mathbb{C})$ is a complete normed vector space (Banach space) with the norm

$$\|f\|_{L^\infty} = \sup_{x\in[0,1]} |f(x)|.$$

4.3. Trigonometric polynomials

Polynomials are combinations of monomials x^n; analogously, we can define trigonometric polynomials as combinations of the functions $e^{2\pi inx}$, called *characters*.

Definition (Character). $\forall n \in \mathbb{Z}$, the character with frequency n is defined as

$$e_n(x) := e^{2\pi inx} = \cos(2\pi nx) + i\sin(2\pi nx).$$

Remark 4.6. $\forall n \in \mathbb{Z}, e_n \in \mathbb{C}^0(\mathbb{R}/\mathbb{Z};\mathbb{C})$.

Definition (Trigonometric polynomial). A function $f \in C^0(\mathbb{R}/\mathbb{Z}, \mathbb{C})$ is a trigonometric polynomial if we can write

$$f(x) = \sum_{n=-N}^{N} c_n e^{2\pi i n x}$$

for some $N \geq 0$.

Example 4.3. $\forall n \in \mathbb{N}$, the function $f(x) = \cos(2\pi n x)$ and $g(x) = \sin(2\pi n x)$ are trigonometric polynomials, because:

$$\cos(2\pi n x) = \frac{e^{2\pi i n x} + e^{-2\pi i n x}}{2} = \frac{1}{2} e_{-n} + \frac{1}{2} e_n,$$

$$\sin(2\pi n x) = \frac{e^{2\pi i n x} - e^{-2\pi i n x}}{2i} = -\frac{1}{2i} e_{-n} + \frac{1}{2i} e_n.$$

Remark 4.7. f is a trigonometric polynomial $\iff$ f is a finite linear combination of terms $\cos(2\pi n x), \sin(2\pi n x)$, for some $n \in \mathbb{N}$.

Lemma 4.3. *The family of* $\{e_n\}_{n \in \mathbb{Z}}$ *is an orthonormal system, i.e.*

$$\langle e_n, e_m \rangle \overset{2}{=} \delta_{nm} \quad \text{and } \|e_n\| = 1 \quad \forall n, m \in \mathbb{Z}.$$

Proof. This proof is proposed in Exercise 1 of Series 8. Using that $\overline{e^{2\pi i m x}} = e^{-2\pi i m x}$, we have

$$\langle e_n, e_m \rangle = \int_0^1 e^{2\pi i n x} \overline{e^{2\pi i m x}}\, dx = \int_0^1 e^{2\pi i (n-m) x}\, dx = \begin{cases} \int_0^1 dx = 1 & \text{if } n = m, \\ \int_0^1 e^{2\pi i (n-m) x}\, dx = 0 & \text{if } n \neq m. \end{cases}$$

□

Corollary 4.4. *Let* $f = \sum_{n=-N}^{N} c_n e_n$ *be a trigonometric polynomial. Then,*

$$\langle f, e_n \rangle = \begin{cases} c_n & \forall -N \leq n \leq N \\ 0 & \forall n < -N \text{ or } n > N. \end{cases}$$

Moreover, a version of Pythagoras Theorem holds:

$$\|f\|_{L^2}^2 = \sum_{n=-N}^{N} |c_n|^2.$$

Proof.

$$\langle f, e_m \rangle = \langle \sum_{n=-N}^{N} c_n e_n, e_m \rangle = \sum_{n=-N}^{N} c_n \langle e_n, e_m \rangle = \sum_{n=-N}^{N} c_n \delta_{nm} = c_m$$

Proceeding in the same way, we can also prove that

$$\|f\|_{L^2}^2 = \langle f, f \rangle = \langle \sum_{n=-N}^{N} c_n e_n, \sum_{j=-N}^{N} c_j e_j \rangle = \sum_{n=-N}^{N} \sum_{j=-N}^{N} c_n \bar{c}_j \langle e_n, e_j \rangle = \sum_{n=-N}^{N} c_n \bar{c}_n = \sum_{n=-N}^{N} |c_n|^2.$$

□

[2] We recall the definition of the Dirac Delta: $\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$.

DEFINITION (Fourier Coefficients). Let $f \in L^2(\mathbb{R}/\mathbb{Z};\mathbb{C})$, $n \in \mathbb{N}$. The n-th Fourier coefficient is

$$\hat{f}(n) := \langle f, e_n \rangle = \int_0^1 f(x) e^{-2\pi i n x} dx.$$

The function $\hat{f} : \mathbb{Z} \to \mathbb{C}$ is called Fourier transform of f.

Thanks to Corollary 4.4, we can state the following properties for trigonometric polynomials:

Corollary 4.5. *Let f be a trigonometric polynomial, then we have the Fourier inversion formula:*

$$f(x) = \sum_{n=-N}^{N} \langle f, e_n \rangle e_n = \sum_{n=-\infty}^{\infty} \langle f, e_n \rangle e_n = \sum_{n=-\infty}^{\infty} \hat{f}(n) e_n = \sum_{n=-\infty}^{\infty} \hat{f}(n) e^{2\pi i n x},$$

and the Plancherel formula (also known as Parseval formula):

$$\|f\|_{L^2}^2 = \sum_{n=-N}^{N} |\langle f, e_n \rangle|^2 = \sum_{n=-\infty}^{\infty} |\langle f, e_n \rangle|^2 = \sum_{n=-\infty}^{\infty} |\hat{f}(n)|^2.$$

For the moment, the result holds only for trigonometric polynomials; we will then extend this result to an arbitrary function $f \in C^0(\mathbb{R}/\mathbb{Z};\mathbb{C})$ and even to the class of 1-periodic, $L^2((0,1))$ functions.

4.4. T-periodic functions and their (complex and real) Fourier coefficients

Let $f \in L^2(0,T)$ be a T-periodic function. The complex Fourier coefficients of f are given by

$$c_n := \frac{1}{T} \int_0^T f(x) e^{\frac{-i2\pi n x}{T}} \, dx.$$

The trigonometric Fourier coefficients of f are given by

$$a_n := \frac{2}{T} \int_0^T f(x) \cos\left(\frac{2\pi n x}{T}\right) dx, \qquad b_n := \frac{2}{T} \int_0^T f(x) \sin\left(\frac{2\pi n x}{T}\right) dx.$$

We remark that

$$a_n = \int_0^1 \left(e^{\frac{i2\pi n x}{T}} + e^{\frac{-i2\pi n x}{T}}\right) f(x) \, dx = c_n + c_{-n}$$

$$b_n = 2 \int_0^1 \frac{1}{2i} \left(e^{\frac{i2\pi n x}{T}} + e^{\frac{-i2\pi n x}{T}}\right) f(x) \, dx = -i c_{-n} + i c_n.$$

That is, $c_0 = a_0/2$ and for $n \geq 1$

$$c_{-n} = \frac{a_n + i b_n}{2},$$

$$c_n = \frac{a_n - i b_n}{2}.$$

The partial Fourier sums of f are

$$F_N f(x) := \sum_{n=-N}^{N} c_n e^{\frac{i2\pi n x}{T}} = \frac{a_0}{2} + \sum_{n=1}^{N} a_n \cos\left(\frac{2\pi n x}{T}\right) + b_n \sin\left(\frac{2\pi n x}{T}\right).$$

Indeed, we note that

$$\sum_{n=-N}^{N} c_n e^{i2\pi n x} = \frac{a_0}{2} + \sum_{n=1}^{N} \left(\frac{a_n - i b_n}{2}\right) e^{i2\pi n x} + \sum_{n=1}^{N} \left(\frac{a_n + i b_n}{2}\right) e^{-i2\pi n x}.$$

Remark 4.8. For $T = 1$, $c_n = \hat{f}(n)$. For a general period T, c_n is the Fourier coefficient of $f(Tx) =: g(x)$:

$$\int_0^1 f(Tx)e^{-i2\pi nx}\,\mathrm{d}x = \frac{1}{T}\int_0^T f(y)e^{-\frac{i2\pi ny}{T}}\,\mathrm{d}y.$$

We also note that

$$F_N f(x) = F_N g(T^{-1}x). \tag{4.13}$$

In the next sections, we will prove that $F_N g \to g$ in various senses (with g being 1-periodic). Owing to (4.13), these results imply the same type of convergence $F_N f(x) \to f(x)$ for a T-periodic function f.

4.5. Uniform approximation of continuous, periodic functions with trigonometric polynomials

We will now present the following result for the approximation of arbitrary functions in $C^0(\mathbb{R}/\mathbb{Z};\mathbb{C})$.

Theorem 4.6 (Weierstrass approximation theorem). *Let $f \in C^0(\mathbb{R}/\mathbb{Z};\mathbb{C})$, and let $\varepsilon > 0$. Then there exists a trigonometric polynomial P such that $\|f - P\|_\infty \leq \varepsilon$.*

To prove it, we first need to introduce several tools:

Definition. Let $f, g \in C^0(\mathbb{R}/\mathbb{Z},\mathbb{C})$. The periodic convolution of f and g is the function $f * g : \mathbb{R} \to \mathbb{C}$ defined by

$$f * g\ (x) = \int_0^1 f(y)g(x-y)dy.$$

Remark 4.9. This notion of convolution is different from the one introduced in L^1 because we integrate on $[0,1]$; it is in fact the same notion adapted to periodic functions. There is a conflict of notation but in fact it is applied to a completely different class of objects, because remember that $L^1(\mathbb{R}) \cap C^0(\mathbb{R}/\mathbb{Z},\mathbb{C}) = \{0\}$.

Lemma 4.7 (Basic properties of periodic convolution.)**.** *Let $f, g, h \in C^0(\mathbb{R}/\mathbb{Z},\mathbb{C})$ and $c \in \mathbb{C}$, then*

(i) (closure) $f * g \in C^0(\mathbb{R}/\mathbb{Z},\mathbb{C})$,
(ii) (commutativity) $f * g = g * f$,
(iii) (bilinearity) $(f+g)*h = f*h+g*h$, $f*(g+h) = f*g+f*h$, $(cf)*g = c(f*g) = f*(cg)$.

Proof. (i) For every $x \in \mathbb{R}$, we have

$$f * g\ (x+1) = \int_0^1 f(y)g(x+1-y)dy = \int_0^1 f(y)g(x-y)dy = f * g\ (x),$$

we used the fact that g is 1-periodic.
To prove continuity, notice that $[0,1]$ is of finite measure and $|f(y)g(x-y)| \leqslant \|f\|_\infty \|g\|_\infty < +\infty$ for every $y \in [0,1]$ and every $x \in \mathbb{R}$. Now, using the continuity of g and the dominated convergence, we get

$$\lim_{z\to x} f * g\ (z) = \lim_{z\to x}\int_0^1 f(y)g(z-y)dy = \int_0^1 f(y)g(x-y)dy = f * g\ (x).$$

(ii) To prove commutativity, remark that $y \mapsto f(y)g(x-y)$ is 1-periodic for every $x \in \mathbb{R}$, so

$$f * g\ (x) = \int_0^1 f(y)g(x-y)dy = \int_{x-1}^x f(x-z)g(z)dz = \int_0^1 f(x-z)g(z)dy = g * f\ (x).$$

The second equality is justified by the change of variable $z = x - y$. For the third equality, we used the fact that the integral of a 1-periodic function over any interval of length 1 is equal to its integral over $[0, 1]$.

(iii) Note that

$$\begin{aligned}(cf + g) * h\ (x) &= \int_0^1 (cf(y) + g(y))h(x - y)dy \\ &= c\int_0^1 f(y)h(x-y)dy + \int_0^1 g(y)h(x-y)dy \\ &= c(f * h)(x) + g * h\ (x).\end{aligned}$$

Thus we have $(cf + g) * h = c(f * h) + g * h$.
Setting $c = 1$, we get $(f + g) * h = f * h + g * h$.
Setting $g = 0$, we get $(cf) * h = c(f * h)$.
Using commutativity, we get $f * (g + h) = (g + h) * f = g * f + h * f = f * g + f * h$.
Similarly, we get $f * (cg) = (cg) * f = c(g * f) = c(f * g) = (cf) * g$.

□

Moreover, we will need the following result on the approximation of the identity via trigonometric polynomials:

Definition (Periodic approximation to the identity). Let $\varepsilon > 0$ and $0 < \delta < 1/2$. A function $f \in C^0(\mathbb{R}/\mathbb{Z}; \mathbb{C})$ is said to be a periodic (ε, δ) approximation to the identity if the following properties are true:

(1) $f(x) \geq 0$ for all $x \in \mathbb{R}$, and $\int_{[0,1]} f = 1$.
(2) We have $f(x) < \varepsilon$ for all $\delta \leq |x| \leq 1 - \delta$.

Lemma 4.8. *For every $\varepsilon > 0$ and $0 < \delta < 1/2$, there exists a trigonometric polynomial P which is an (ε, δ) approximation to the identity.*

Proof. Let $N \geq 1$ be an integer. We define the *Fejér kernel* F_N to be the function

$$F_N = \sum_{n=-N}^{N} \left(1 - \frac{|n|}{N}\right) e_n.$$

Clearly F_N is a trigonometric polynomial. We observe the identity

$$F_N = \frac{1}{N} \left| \sum_{n=0}^{N-1} e_n \right|^2.$$

But from the geometric series formula, we have

$$\sum_{n=0}^{N-1} e_n(x) = \frac{e_N - e_0}{e_1 - e_0} = \frac{e^{\pi i(N-1)x}\sin(\pi N x)}{\sin(\pi x)}$$

when x is not an integer, and hence we have the formula

$$F_N(x) = \frac{\sin(\pi N x)^2}{N \sin(\pi x)^2}.$$

When x is an integer, the geometric series formula does not apply, but one has $F_N(x) = N$ in that case, as one can see by direct computation. In either case we see that $F_N(x) \geq 0$ for any x. Also, we have

$$\int_{[0,1]} F_N(x)dx = \sum_{n=-N}^{N} \left(1 - \frac{|n|}{N}\right) \int_{[0,1]} e_n = \left(1 - \frac{|0|}{N}\right) 1 = 1.$$

Finally, since $|\sin(\pi N x)| \leq 1$, we have

$$F_N(x) \leq \frac{1}{N\sin(\pi x)^2} \leq \frac{1}{N\sin(\pi\delta)^2}$$

whenever $\delta < |x| < 1-\delta$ (this is because sin is increasing on $[0,\pi/2]$ and decreasing on $[\pi/2,\pi]$). Thus by choosing N large enough, we can make $F_N(x) \leq \varepsilon$ for all $\delta < |x| < 1-\delta$. □

Proof of Weierstrass Theorem. Let f be any element of $C^0(\mathbb{R}/\mathbb{Z};\mathbb{C})$; we know that f is bounded, so that we have some $M > 0$ such that $|f(x)| \leq M$ for all $x \in \mathbb{R}$.

Let $\varepsilon > 0$ be arbitrary. Since f is uniformly continuous, there exists a $\delta > 0$ such that $|f(x) - f(y)| \leq \varepsilon$ whenever $|x - y| \leq \delta$. Now use Lemma 4.8 to find a trigonometric polynomial P which is a (ε, δ) approximation to the identity. Then $f * P$ is also a trigonometric polynomial, because P is a trigonometric polynomial, so we can write

$$P = \sum_{n=-N}^{N} a_n e_n$$

for some $a_{-N}, \ldots, a_N \in \mathbb{C}$ and some $N \in \mathbb{N}$.
Using the linearity property of periodic convolution, we have

$$f * P = \sum_{n=-N}^{N} a_n (f * e_n).$$

But for every $n \in \mathbb{Z}$, we have

$$f * e_n(x) = \int_0^1 f(y) e^{2\pi i n(x-y)} dy = \int_0^1 f(y) e^{-2\pi i n y} dy e^{2\pi i n x} = \hat{f}(n) e_n(x).$$

Combining this with our previous observation we get

$$f * P = \sum_{n=-N}^{N} a_n \hat{f}(n) e_n.$$

We now estimate $\|f - f * P\|_\infty$. Let x be any real number. We have

$$\begin{aligned} |f(x) - f * P(x)| &= |f(x) - P * f(x)| \\ &= \left| f(x) - \int_{[0,1]} f(x-y)P(y)dy \right| \\ &= \left| \int_{[0,1]} f(x)P(y)dy - \int_{[0,1]} f(x-y)P(y)dy \right| \\ &= \left| \int_{[0,1]} (f(x) - f(x-y))P(y)dy \right| \\ &\leq \int_{[0,1]} |f(x) - f(x-y)| P(y)dy. \end{aligned}$$

The right-hand side can be split as

$$\int_{[0,\delta]} |f(x) - f(x-y)|P(y)dy + \int_{[\delta,1-\delta]} |f(x) - f(x-y)|P(y)dy + \int_{[1-\delta,1]} |f(x) - f(x-y)|P(y)dy$$

which we can bound from above by

$$\begin{aligned} &\le \int_{[0,\delta]} \varepsilon P(y)dy + \int_{[\delta,1-\delta]} 2M\varepsilon dy \\ &\quad + \int_{[1-\delta,1]} |f(x-1) - f(x-y)|P(y)dy \\ &\le \int_{[0,\delta]} \varepsilon P(y)dy + \int_{[\delta,1-\delta]} 2M\varepsilon dy + \int_{[1-\delta,1]} \varepsilon P(y)dy \\ &\le \varepsilon + 2M\varepsilon + \varepsilon \\ &= (2M+2)\varepsilon. \end{aligned}$$

Thus we have $\|f - f * P\|_\infty \le (2M+2)\varepsilon$. Since M is fixed and ε is arbitrary, we can thus make $f * P$ arbitrarily close to f in sup norm, which proves the Weierstrass approximation theorem.

□

4.6. L^2–convergence of Fourier Series

THEOREM 4.9 (Fourier Theorem). *For any $f \in L^2(\mathbb{R}/\mathbb{Z};\mathbb{C})$, the series $\sum_{n=-\infty}^{\infty} \hat{f}(n)e_n$ converges in L^2 to f. In other words, we have*

$$\lim_{N\to\infty} \left\| f - \sum_{n=-N}^{N} \hat{f}(n)e_n \right\|_{L^2} = 0.$$

PROOF. **Step 1:** proof for $f \in C^0(\mathbb{R}/\mathbb{Z};\mathbb{C})$.

Let $\varepsilon > 0$. We have to show that there exists an N_0 such that $\left\| f - \sum_{n=-N}^{N} \hat{f}(n)e_n \right\|_{L^2} \le \varepsilon$ for all $N \ge N_0$.

By the Weierstrass approximation theorem (Theorem 4.6), we can find a trigonometric polynomial $P = \sum_{n=-N_0}^{N_0} c_n e_n$ such that $\|f - P\|_\infty \le \varepsilon$, for some $N_0 > 0$. In particular we have $\|f - P\|_2 \le \varepsilon$.

Now let $N > N_0$, and let $F_N := \sum_{n=-N}^{N} \hat{f}(n)e_n$. We claim that $\|f - F_N\|_{L^2} \le \varepsilon$. First observe that for any $|m| \le N$, we have

$$\langle f - F_N, e_m\rangle = \langle f, e_m\rangle - \sum_{n=-N}^{N} \hat{f}(n)\langle e_n, e_m\rangle = \hat{f}(m) - \hat{f}(m) = 0.$$

In particular we have

$$\langle f - F_N, F_N - P\rangle = 0$$

since we can write $F_N - P$ as a linear combination of the e_m for which $|m| \le N$. By Pythagoras' theorem we therefore have

$$\|f - P\|_2^2 = \|f - F_N\|_{L^2}^2 + \|F_N - P\|_{L^2}^2$$

and in particular

$$\|f - F_N\|_{L^2} \leq \|f - P\|_2 \leq \varepsilon \tag{4.14}$$

as desired.

Step 2: proof for $f \in L^2(\mathbb{R}/\mathbb{Z};\mathbb{C})$.

Let $f \in L^2(\mathbb{R}/\mathbb{Z};\mathbb{C})$, set $\varepsilon > 0$. By density, $\exists g \in C^0(\mathbb{R}/\mathbb{Z};\mathbb{C})$ such that

$$\|f - g\|_{L^2} \leq \varepsilon.$$

For N large, $\|g - F_N g\|_{L^2} \leq \varepsilon$ by the previous step. Finally, by the best approximation of f with $F_N f$ (as observed for (4.14)), we have

$$\|f - F_N f\|_{L^2} \leq \|f - F_N g\|_{L^2} \leq \|f - g\|_{L^2} + \|g - F_N g\|_{L^2} \leq 2\varepsilon. \tag{4.15}$$

□

As a corollary of the Fourier theorem, we obtain

THEOREM 4.10 (Plancherel (Parseval's identity)). *For any $f \in L^2(\mathbb{R}/\mathbb{Z};\mathbb{C})$, the series $\sum_{n=-\infty}^{\infty} |\hat{f}(n)|^2$ is absolutely convergent, and*

$$\|f\|_2^2 = \sum_{n=-\infty}^{\infty} |\hat{f}(n)|^2.$$

PROOF. Let $\varepsilon > 0$. By the Fourier theorem we know that

$$\left\| f - \sum_{n=-N}^{N} \hat{f}(n) e_n \right\|_{L^2} \leq \varepsilon$$

if N is large enough (depending on ε). In particular, by the triangle inequality this implies that

$$\|f\|_2 - \varepsilon \leq \left\| \sum_{n=-N}^{N} \hat{f}(n) e_n \right\|_{L^2} \leq \|f\|_2 + \varepsilon.$$

On the other hand, by Corollary 4.5 we have

$$\left\| \sum_{n=-N}^{N} \hat{f}(n) e_n \right\|_{L^2} = \left(\sum_{n=-N}^{N} |\hat{f}(n)|^2 \right)^{1/2}$$

and hence

$$\left(\|f\|_2 - \varepsilon\right)^2 \leq \sum_{n=-N}^{N} |\hat{f}(n)|^2 \leq \left(\|f\|_2 + \varepsilon\right)^2.$$

Taking lim sup, we obtain

$$\left(\|f\|_2 - \varepsilon\right)^2 \leq \limsup_{N\to\infty} \sum_{n=-N}^{N} |\hat{f}(n)|^2 \leq \left(\|f\|_2 + \varepsilon\right)^2$$

Since ε is arbitrary, we thus obtain by the squeeze test that

$$\limsup_{N\to\infty} \sum_{n=-N}^{N} |\hat{f}(n)|^2 = \|f\|_2^2$$

and the claim follows. □

Remark 4.10. If f is, instead, T-periodic and $f \in L^2(0,T)$, then

$$F_N f = \frac{a_0}{2} + \sum_{n=1}^{N} a_n \cos\left(\frac{2\pi n x}{T}\right) + b_n \sin\left(\frac{2\pi n x}{T}\right) \to f \text{ in } L^2$$

and

$$\frac{2}{T}\int_0^T |f|^2 \, \mathrm{d}x = \frac{a_0^2}{2} + \sum_{n=1}^{+\infty} a_n^2 + b_n^2.$$

Indeed, we can apply Plancherel's theorem to the 1-periodic function $f(Tx) =: g(x)$, recalling Remark 4.8 and compute

$$\begin{aligned}\|g\|_{L^2(0,1)}^2 = \frac{1}{T}\int_0^T |f|^2 \, \mathrm{d}x = \sum_{n\in\mathbb{Z}} |\hat{g}(n)|^2 &= \sum_{n\in\mathbb{Z}} c_n^2 \\ &= \left(\frac{a_0}{2}\right)^2 + \sum_{n\in\mathbb{N}} \left|\frac{a_n + ib_n}{2}\right|^2 + \left|\frac{a_n - ib_n}{2}\right|^2 \\ &= \frac{a_0^2}{4} + \frac{1}{2}\sum_{n\in\mathbb{N}} |a_n|^2 + |b_n|^2.\end{aligned}$$

Remark 4.11. Non rigorously, we can compute the derivative of $f(x) = \sum_{n=-\infty}^{\infty} \hat{f}(n) e^{i2\pi n x}$ as

$$f'(x) = \sum_{n=-\infty}^{\infty} \hat{f}(n) i2\pi n e^{i2\pi n x}.$$

In fact, it is true that $\hat{f}'(n) = i2\pi n \hat{f}(n)$, because rigourously we can prove it with an integration by parts:

$$\hat{f}'(n) = \int_0^1 f'(x) e^{-i2\pi n x} dx = \underbrace{|e^{i2\pi n x} f(x)|_0^1}_{=0} - \int_0^1 f(x)(-i2\pi n) e^{-i2\pi n x} dx = i2\pi n \hat{f}(n) \tag{4.16}$$

where the term evaluated at the extrema of the integral vanishes due to the fact that $f(0) = f(1)$ by periodicity. Moreover, from (4.16), it is clear that if we have a function $f \in C^1$ (for which $\hat{f}'(n)$ are finite), then its Fourier coefficients decay like n^{-1}.

Iterating the procedure, we can apply the same result and consideration on the decay of the coefficients to functions in C^k, $k > 1$.

4.7. Pointwise convergence of Fourier Series

We now prove a result on pointwise convergence of Fourier Series. To give pointwise meaning, we need functions that are more regular than L^2, and at least continuous. Therefore, let us define the following space:

Definition. Let $\Omega \subseteq \mathbb{R}$ be a bounded set and $\alpha \in (0,1])$. We define:

$$C^{0,\alpha} := \left\{ f \in C^0(\Omega) : \sup_{x,y\in\Omega} \frac{|f(x)-f(y)|}{|x-y|^\alpha} < +\infty \right\}$$

The norm that we define on this space is:

$$\|f\|_{C^{0,\alpha}} := \sup_{x\in\Omega} |f(x)| + \sup_{x,y\in\Omega} \frac{|f(x)-f(y)|}{|x-y|^\alpha}. \tag{4.17}$$

Remark 4.12. Note that $C^{0,\alpha}(\Omega) \supset C^{0,1}(\Omega) = \mathrm{Lip}(\Omega)$. However also non-Lipschitz functions can be C^α, for instance $|x|^{1/2} \in C^{1/2}([0,1])$.

Theorem 4.11 (Dirichlet). *(i) If f is 1−periodic, piecewise C^1, then $\forall x$ we have that $F_N f(x) \to (f(x^-)+f(x^+))/2$.*

(ii) Let $\alpha \in (0,1]$. If $f \in C^{0,\alpha}([0,1])$, 1−periodic (then, $f \in C^{0,\alpha}(\mathbb{R})$ as well), then $\forall x$ we have that $F_N f(x) \to f(x)$.

(iii) If $f \in L^1(\mathbb{R}/\mathbb{Z})$ and for $a \in \mathbb{R}$ we have that $\exists \alpha \in (0,1], M := M(a) > 0$ and $\delta := \delta(a) > 0$ such that

$$|f(a+t)-f(a^+)| + |f(a-t)-f(a^-)| \le Mt^\alpha, \quad \forall 0 < t \le \delta$$

then

$$\lim_{N\to\infty} F_N f(a) = \frac{1}{2}[f(a^+)+f(a^-)],$$

where $f(a^+) := \lim_{t\to a^+} f(t)$ and $f(a^-) := \lim_{t\to a^-} f(t)$.

Remark 4.13. Note that the hypothesis in (iii) corresponds to the fact that f behaves like a $C^{0,\alpha}$ function at one single point. It is therefore more general than being piecewise $C^{0,\alpha}$.

Remark 4.14. (iii) is more technical, but we can easily see that $(iii) \Rightarrow (ii)$. Indeed, in this case $f(a) = f(a^+) = f(a^-)$ and

$$\frac{|f(a+t)-f(a)|}{|t|^\alpha} \le \|f\|_{C^{0,\alpha}} =: M.$$

We can also prove that $(iii) \Rightarrow (i)$. If f is (left and right) differentiable (even if it may have a jump discontinuity in a),

$$\begin{aligned} f(a+t) &= f(a^+) + f'(a^+)t + o(t) \quad \forall t > 0 \\ f(a+t) &= f(a^-) + f'(a^-)t + o(t) \quad \forall t < 0 \\ \Rightarrow f(a+t) - f(a^+) &\le \underbrace{(|f'(a^+)|+1)}_{M(a),\alpha=1} |t| \quad \forall |t| \text{ sufficiently small,} \\ f(a-t) - f(a) &\le \underbrace{(|f'(a^-)|+1)}_{M(a),\alpha=1} |t| \quad \forall |t| \text{ sufficiently small.} \end{aligned}$$

Remark 4.15. (i) The theorem is false if f is only continuous.

(ii) One can weaken the continuity assumption in the sense of Hölder and replace it with Dini's criterion, namely: if there exists $\delta \in]0,\pi]$, $a \in \mathbb{R}$ such that

$$\int_0^\delta \frac{|f(a-y)+f(a+y)-2f(a)|}{|y|} dy < +\infty$$

then $\lim_{N\to\infty} F_N f(a) = f(a)$. A Hölder continuous function obviously satisfies this. Indeed

$$|f(a-y)-f(a)| + |f(a+y)-f(a)| \le M|y|^\alpha$$

and thus

$$\int_0^\pi \frac{|f(a-y)+f(a+y)-2f(a)|}{|y|}dy \le M\int_0^\pi |y|^{\alpha-1}dy = M\frac{|\pi|^\alpha}{\alpha} < \infty.$$

Remark 4.16. If $f \in L^1(-\pi,\pi)$ is 2π-periodic, then it's possible for the Fourier series to diverge everywhere (cf. an example due to Kolmogorov). However, if $f \in L^p(-\pi,\pi)$, with $p > 1$, the Fourier series will converge to the function almost everywhere (if $p = 2$ it's the famous result of Carleson, which has been generalized to the case $p > 1$ by Hunt).

To prove Theorem 4.11, we first need an auxiliary lemma:

Lemma 4.12 (Riemann-Lebesgue)**.** *Let* $f \in L^1(\mathbb{R})$ *(the same can be repeated in general dimension* $\mathbb{R}^d$*). Define*

$$\hat f(\xi) := \int_{\mathbb{R}} f(x)e^{-i2\pi\xi\cdot x}dx.$$

Then, $\lim_{|\xi|\to\infty}\hat f(\xi) = 0$.

Proof. This exercise is proposed in Exercise 2 of Series 8. We prove it in dimension $d = 1$.

$$\begin{aligned}\hat f(\xi) &= \int_{\mathbb{R}} f(x-1/(2\xi))\underbrace{e^{-i2\pi\xi(x-1/(2\xi))}}_{1/\xi\text{ periodic}}dx = -\int_{\mathbb{R}} f(x-1/2\xi)e^{-i2\pi\xi x}dx\\ &= \frac{1}{2}\int_{\mathbb{R}}[f(x)-f(x-1/(2\xi))]e^{-i2\pi\xi x}dx.\end{aligned}$$

Now,

$$|\hat f(\xi)| \le \|f(\cdot)-f(\cdot-1/(2\xi))\|_{L^1} \to 0 \quad \text{as } |\xi|\to\infty, \tag{4.18}$$

by continuity of translations in the L^1-norm (Exercise 3 of Series 6). □

Proof of Dirichlet's Theorem. By Remark 4.14, we will just prove (iii) because (i) and (ii) follow automatically. We are going to express $F_N f$ as a suitable sort of convolution and then do typical estimates for convolutions.

Step 1: we define the Dirichlet kernel D_N:

$$F_N f(x) = \sum_{n=-N}^N \hat f(n)e^{i2\pi nx} = \sum_{n=-N}^N\int_0^1 f(y)e^{i2\pi n(x-y)}dy = \int_0^1 f(y)\underbrace{\sum_{n=-N}^N e^{i2\pi n(x-y)}}_{:=D_N(x-y)}dy = f * D_N, \tag{4.19}$$

where

$$\begin{aligned}D_N(y) &= \sum_{n=-N}^N (e^{i2\pi y})^n \underbrace{=}_{n'=n+N} \sum_{n'=0}^{2N}(e^{i2\pi y})^{n'}\\ &= e^{-i2\pi Ny}\frac{e^{i2\pi y(2N+1)}-1}{e^{i2\pi y}-1}\\ &= \frac{e^{i2\pi y(N+1/2)}-e^{-i2\pi y(N+1/2)}}{e^{i2\pi y}-1}e^{i\pi y}\\ &= \frac{e^{i2\pi y(N+1/2)}-e^{-i2\pi y(N+1/2)}}{e^{i\pi y}(e^{i\pi y}-e^{-i\pi y})}e^{i\pi y}\\ &= \frac{\sin(2\pi y(N+1/2))}{\sin(\pi y)}.\end{aligned}$$

Note that

$$\int_0^1 D_N(x)dx = \sum_{n=-N}^{N} \int_0^1 e^{i2\pi n y}dy = \int_0^1 1dx = 1.$$

The Dirichlet kernel has similar properties to the Fejer kernel, but it is not positive. For the sake of simplicity, let us suppose that $f(a^+) = f(a^-) = f(a)$. Now, consider

$$F_N f(a) - f(a) = \int_{-1/2}^{1/2} \frac{\sin(2\pi y(N+1/2))}{\sin(\pi y)}(f(a+y) - f(a))dy. \tag{4.20}$$

Now, define:

$$\varphi_a(y) = \frac{f(a+y) - f(a)}{\sin(\pi y)}.$$

The assumptions in (iii) enable us to show that $\varphi_a \in L^1(-1/2, 1/2)$. Indeed,

$$|\varphi_x(y)| \leq \frac{M|y|^\alpha}{|\sin(\pi y)|}, \quad y \in (-\delta, \delta)$$

and the right-hand side is integrable. Indeed (we see below the importance of the fact that $\alpha > 0$):

$$\int_{-1/2}^{1/2} |\varphi_a(y)|dy \leq \int_{-1/2}^{1/2} \frac{M|y|^\alpha}{|\sin(\pi y)|}dy \leq \int_{-1/2}^{1/2} \frac{M}{2} y^{\alpha-1}dy < +\infty,$$

because $\sin(\pi y) \geq 2y$ on $[0, 1/2]$.

By Lemma 4.12 applied to $\varphi_x|_{[-1/2,1/2]}$ extended to 0 outside $[-1/2, 1/2]$ and with $\xi = N + 1/2$, we have that

$$\int_0^1 \sin(2\pi y(N+1/2))\varphi_x(y)dy \to 0 \quad \text{as } N \to \infty. \tag{4.21}$$

Step 2: without the simplifying assumption, the proof works similarily:

$$F_N f(a) - \frac{1}{2}(f(a^+) + f(a^-)) = \int_{-1/2}^{1/2} \underbrace{\frac{\sin(2\pi y(N+1/2))}{\sin(\pi y)}}_{=:g} \underbrace{\left(f(a+y) - \frac{1}{2}(f(a^+) + f(a^-))\right)}_{=:F} dy.$$

Since for g even:

$$\int_{-1/2}^{1/2} g(y)F(\cdot + y)dy = \int_0^{1/2} g(y)F(\cdot + y)dy + \int_0^{1/2} g(y)F(\cdot - y)dy = \int_0^{1/2} g(y)(F(\cdot + y) + F(\cdot - y))dy.$$

Now,

$$F_N f(x) = \int_0^{1/2} \underbrace{\frac{f(a-y) + f(a+y) - f(a^-) - f(a^+)}{\sin(\pi y)}}_{=:\varphi_a(y)} \sin(2\pi(N+1/2)y)dy.$$

By Lemma 4.12 (again with $\xi = 2\pi(N+1/2)$ and with φ_a extended to 0 outside $[-1/2, 1/2]$), we have that $\int_0^{1/2} \varphi_a(y)\sin(2\pi(N+1/2)y)dy \to 0$ as $N \to \infty$ if $\varphi_a \in L^1(0, 1/2)$. This holds, because:

$$\begin{aligned} \int_0^\delta |\varphi_a(y)|dy &\leq \int_0^\delta \frac{|f(a-y) - f(a^-)|}{2|y|} + \frac{|f(a+y) - f(a^+)|}{2|y|}dy \\ &\leq \int_0^\delta \frac{M|y|^\alpha}{2|y|} + \frac{M|y|^\alpha}{2|y|}dy < +\infty, \end{aligned}$$

because $\sin(\pi y) \geq 2|y|$ for $y \in (0, 1/2)$, and

$$\int_\delta^{1/2} |\varphi_a(y)|dy \leq \int_\delta^{1/2} \frac{|f(a+y)| + |f(a-y)|}{\sin(\pi\delta)} dy + \frac{|f(a^-)| + |f(a^+)|}{\sin(\pi\delta)} < +\infty,$$

because $f \in L^1(\mathbb{R}/\mathbb{Z})$.

□

4.8. Uniform convergence of Fourier Series

Note that with Fourier's Theorem we have only obtained convergence of the Fourier series $\sum_{n=-\infty}^{\infty} \hat{f}(n)e_n$ to f in the L^2 metric. One may ask whether one has convergence in the uniform or pointwise sense as well, but it turns out (perhaps somewhat surprisingly) that the answer is no to both of those questions. However, if one assumes that the function f is not only continuous, but is also continuously differentiable, then one can recover pointwise convergence; if one assumes continuously twice differentiable, then one gets uniform convergence as well. These results are beyond the scope of this text and will not be proven here. However, we will prove one theorem about when one can improve the L^2 convergence to uniform convergence:

Theorem 4.13 (Uniform convergence of Fourier Series). *Let $f \in C^1(\mathbb{R})$ and 1-periodic, then*

$$\sum_{n=-\infty}^{\infty} |\hat{f}(n)| < +\infty \tag{4.22}$$

and

$$F_N f \to f \text{ uniformly on } [0,1]. \tag{4.23}$$

Proof. We prove the two statements of the theorem:

(i) $f \in C^1 \Rightarrow \sum_{n=-\infty}^{\infty} |\hat{f}(n)| < +\infty$.
Thanks to Remark 4.11, we have that

$$\hat{f'}(n) = i2\pi n \hat{f}(n).$$

Applying Plancherel's formula, we then have that

$$+\infty > \|f'\|^2_{L^2(0,1)} = \sum_{n=-\infty}^{\infty} |\hat{f'}(n)|^2 = \sum_{n=-\infty}^{\infty} (2\pi)^2 n^2 |\hat{f}(n)|^2.$$

Hence,

$$\left(\sum_{n=-\infty}^{\infty} |\hat{f}(n)|\right)^2 = \left(\sum_{n=-\infty}^{\infty} \left|\frac{1}{n} n\hat{f}(n)\right|\right)^2 \leq \underbrace{\sum_{n=-\infty}^{\infty} \frac{1}{n^2}}_{<+\infty} \sum_{n=-\infty}^{\infty} n^2 |\hat{f}(n)|^2 < +\infty.$$

(ii) $\sum_{n=-\infty}^{\infty} |\hat{f}(n)| < +\infty \Rightarrow F_n f \to f$ uniformly on $[0,1]$.
By the assumption, we have that $F_N f$ is a Cauchy sequence in L^∞:

$$\|F_N f - F_M f\|_{C^0} = \sup_{x\in[0,1]} \sum_{n=N+1}^{M} |\hat{f}(n)e^{i2\pi nx}| \leq \sum_{|n|\geq N+1} |\hat{f}(n)|.$$

Hence, the sequence has a limit point: $F_N f \to F$ in C^0, and therefore also in L^2. Since $F_N f \to f$ in L^2, we have that $F = f$.

□

Remark 4.17. If the function is regular enough, the Fourier coefficients decay in n. Indeed,

$$f \in C^2 \Rightarrow \hat{f''}(n) = (i2\pi n)\hat{f'}(n) = -(2\pi n)^2 \hat{f}(n).$$

Since $|\hat{f''}(n)| = \left| \int_0^1 f''(x) e^{-i2\pi n x} dx \right| \leq \int_0^1 |f''(x)| dx =: C$, we have that

$$|\hat{f}(n)| \leq \frac{C}{(2\pi n)^2}.$$

THEOREM 4.14 (Bernstein's Theorem). *Let $f \in C^{0,\alpha}$, with $\alpha > 1/2$. Then,*

$$\sum_{n=-\infty}^{\infty} |\hat{f}(n)| < +\infty$$

and

$$F_n f \to f \text{ uniformly on } [0,1].$$

Remark 4.18. We cannot conclude the same for $\alpha = 1/2$. Indeed, there exists an example (due to Hardy-Littlehood) of a function $f \in C^{0,1/2}$ such that $\sum_{n=-\infty}^{\infty} |\hat{f}(n)| = +\infty$, that is

$$f(x) = \sum_{n=1}^{\infty} \frac{e^{in\log n}}{n} e^{inx}.$$

4.9. Fourier Series only in sines or cosines

We now observe that some special properties of functions (real-values, symmetries...) reflect on properties of the Fourier coefficients.

Remark 4.19. $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R} \Rightarrow a_n, b_n \in \mathbb{R}$. What about $\hat{f}(n)$ and $\hat{f}(-n) = \overline{\hat{f}(n)}$?

Remark 4.20. Given $f \in L^2(\mathbb{R})$, $2\pi-$periodic and odd (that means, $f(-x) = -f(x)$), writing the Fourier series of f as

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(nx) + \sum_{n=1}^{\infty} b_n \sin(nx),$$

we have that $a_n = 0 \quad \forall n \in \mathbb{N}$.

If instead $f \in L^2(\mathbb{R})$, $2\pi-$periodic is even (that means, $f(-x) = f(x)$), we have $b_n = 0 \quad \forall n \in \mathbb{N}$.

PROOF OF REMARK 4.20. We consider in the proof only the case f odd; the case with f even can be proved analogously.

$$\begin{aligned} a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)\cos(nx)dx &= \frac{1}{\pi}\left[\int_{-\pi}^{0} f(x)\cos(nx)dx + \int_0^{\pi} f(x)\cos(nx)dx \right] \\ &\overset{\star}{=} \frac{1}{\pi}\left[-\int_0^{\pi} f(y)\cos(ny)dy + \int_0^{\pi} f(x)\cos(nx)dx \right] \\ &= 0, \end{aligned}$$

where $\star$ follows from the change of variables $y = -x$ and from the fact that $f(y) = -f(x)$ being f odd. □

We now prove that we can write the Fourier series of a general function only with sines.

Given $l \geq 0$ and $f : [0,l] \to \mathbb{R}$, we perform the following steps:

(i) we extend f oddly on $[-l, l]$, by defining

$$\tilde{f}(x) = \begin{cases} f(x) & \text{in } [0, l] \\ -f(-x) & \text{in } [-l, 0] \end{cases}$$

(ii) we extend $\tilde{f}$ to a $2l-$periodic function;
(iii) we write the Fourier series of $\tilde{f}$:

$$F_\infty \tilde{f}(x) = \underbrace{\frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\left(\frac{\pi n x}{l}\right)}_{=0,\text{ being } \tilde{f} \text{ odd}} + \sum_{n=1}^{\infty} b_n \sin\left(\frac{\pi n x}{l}\right).$$

The Fourier coefficients b_n can be expressed only in terms of f:

$$\begin{aligned} b_n &= \frac{1}{l}\int_{-l}^{l} \tilde{f}(y) \sin\left(\frac{\pi n y}{l}\right) dy \\ &= \frac{1}{l}\int_{-l}^{0} \tilde{f}(y) \sin\left(\frac{\pi n y}{l}\right) dy + \frac{1}{l}\int_{0}^{l} \tilde{f}(y) \sin\left(\frac{\pi n y}{l}\right) dy \\ &= \frac{2}{l}\int_{0}^{l} f(y) \sin\left(\frac{\pi n y}{l}\right) dy. \end{aligned}$$

Definition. The Fourier series in sines of f is defined as

$$F_\infty^s f := \sum_{n=1}^{\infty} b_n \sin\left(\frac{\pi}{l} n x\right).$$

Corollary 4.15. *Let $f \in L^2(0, l)$; then,*
(i) $F_N^s f \to f$ in $L^2(0, l)$,
(ii) if $f \in C^{0,\alpha}([0, l])$ for some $\alpha > 0$, then

$$F_N^s f \to f \text{ pointwise in } (0, l).$$

Proof. (i) We can apply Fourier Theorem (Theorem 4.9) to $\tilde{f}$, retriving that

$$F_N \tilde{f} \to \tilde{f} \quad \text{in} \quad L^2(-l, l),$$

which leads to the thesis because in $(0, l)$ we have that $F_N \tilde{f} = F_N^s f$ and $\tilde{f} = f$;
(ii) The pointwise convergence follows by applying Dirichlet's theorem to $\tilde{f}$.

□

Remark 4.21. Note that $F_N^s f(0) \to 0$, therefore the second thesis of Corollary 4.15 does not hold at $x = 0$.

Proceeding analogously, we can write the Fourier series of a function only with cosines.

Definition. The Fourier series in cosines of f is defined as

$$F_\infty^c f := \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\left(\frac{\pi}{l} n x\right),$$

where

$$a_n := \frac{2}{l}\int_0^l f(y) \cos\left(\frac{\pi}{l} n y\right) dy.$$

Remark 4.22. As in the case of sines, the Fourier series in cosines corresponds to the Fourier series $F\bar{f}$, where $\bar{f}$ is the even extension of f.

Corollary 4.16. *Let $f \in L^2(0,l)$, then*

(i) $F_N^c f \to f$ in $L^2(0,l)$;
(ii) if $f \in C^{0,\alpha}([0,l])$ for some $\alpha > 0$, then $F_N^c f \to f$ pointwise in $[0,l]$;
(iii) if $f \in C^1$ or $f \in C^{0,\alpha}([0,l])$ with $\alpha > 1/2$, then $F_N^c f \to f$ uniformly.

Remark 4.23. Point (iii) of Corollary 4.16 follows from Bernstein's Theorem, and it cannot be true for the expansion in sines.

For the sines, you can recover the same result in (iii) only if you assume $f(0) = f(l) = 0$. This is due to the fact that the even extension of a C^1 function is not necessarily C^1, because we can create angles: take for example $f(x) = x \in C^1(0,l)$, whose even expansion is $f(x) = |x| \notin C^1(-l,l)$. However, we have that the even extension of f is C^1 if $f'(0) = f'(l) = 0$.

Proof of Corollary 4.16. For (i) and (ii), we proceed as for Corollary 4.15. For (ii), observe that the even extension of a $C^{0,\alpha}$ function is $C^{0,\alpha}$. □

Remark 4.24 (Parseval's Identity for f expressed as a Fourier serie only of sines or cosines)**.** Let $f \in L^2$ be a $2l-$periodic function. By expanding it in Fourier series, we get

$$f = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\left(\frac{n\pi}{l}x\right) + b_n \sin\left(\frac{n\pi}{l}x\right).$$

Then,

$$\frac{1}{l}\int_0^{2l} |f(x)|^2 dx = \frac{a_0^2}{2} + \sum_{n=1}^{\infty} a_n^2 + b_n^2.$$

Indeed, $f(2lx)$ is $1-$periodic, and

$$\begin{aligned} f(2lx) &= \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(2\pi n x) + b_n \sin(2\pi n x) \\ &\overset{\star}{=} \frac{a_0}{2} e^{i\pi 0} + \sum_{n=1}^{\infty} \frac{a_n - i b_n}{2} e^{i2\pi n x} + \frac{a_n + i b_n}{2} e^{-i2\pi n x}, \end{aligned}$$

where $\star$ follows from the fact that

$$\cos(2\pi n x) = \frac{e^{i2\pi n x} + e^{-i2\pi n x}}{2} \text{ and } \sin(2\pi n x) = \frac{e^{i2\pi n x} - e^{-i2\pi n x}}{2i}.$$

By Parseval's identity, we have that

$$\|f(2lx)\|^2_{L^2(0,1)} = \frac{a_0^2}{4} + \sum_{n=1}^{\infty} \left|\frac{a_n - i b_n}{2}\right|^2 + \left|\frac{a_n + i b_n}{2}\right|^2 = \frac{a_0^2}{4} + \sum_{n=1}^{+\infty} \frac{a_n^2}{2} + \frac{b_n^2}{2}$$

and since

$$\|f(2lx)\|^2_{L^2(0,1)} = \int_0^1 f(2lx)^2 dx = \frac{1}{2l}\int_0^{2l} f(x')^2 dx',$$

we retrieve

$$\frac{1}{l}\int_0^{2l} f(x')^2 dx' = \frac{a_0^2}{2} + \sum_{n=1}^{+\infty} a_n^2 + b_n^2.$$

We will now analyze how the decay of the coefficients of the Fourier Series only in sines b_n is affected by the regularity of f. For simplicity, let us suppose $l = \pi$.

Proposition 4.17. *Let* $f \in C^{2k}([0,\pi])$ *with* $f^{(2j)}(0) = f^{(2j)}(\pi) = 0$ *for all* $j = 0, \dots, k-1$. *Then,*

$$b_n = o(n^{-2k}) \text{ as } n \to \infty.$$

Let $g \in C^{2k-1}([0,\pi])$ *with* $g^{(2j)}(0) = g^{(2j)}(\pi) = 0$ *for all* $j = 0, \dots, k-1$. *Then,*

$$b_n = o(n^{-2k+1}) \text{ as } n \to \infty.$$

PROOF. Using integration by parts we have

$$\begin{aligned}
\frac{\pi}{2} b_n &= -\frac{1}{n} f(x)\cos(nx)\Big|_0^\pi + \frac{1}{n}\int_0^\pi f'(x)\cos(nx)dx \\
&= \frac{1}{n^2} f'(x)\sin(nx)\Big|_0^\pi - \frac{1}{n^2}\int_0^\pi f''(x)\sin(nx)dx \\
&= \dots \\
&= \frac{\cos(k\pi)}{n^{2k}} \int_0^\pi f^{2k}(x)\sin(nx)dx.
\end{aligned}$$

The Riemann-Lebesgue lemma tells us that

$$\lim_{n\to\infty} \int_0^\pi f^{2k}(x)\sin(nx)dx = 0$$

which gives the first result.
The second result is proved in the same way. □

The proposition could be stated more concisely as $f \in C^k([0,\pi])$ then $b_n = o(n^{-k})$ as $n \to \infty$. Notice that the result is no longer true without the assumptions $f^{(2j)}(0) = f^{(2j)}(\pi) = 0$ for $j = 0, \dots, k-1$ as can be seen by considering the constant function 1 for which the coefficients are given by

$$b_n = \frac{2}{\pi}\int_0^\pi \sin(nx)dx = \frac{2}{n\pi}(1 - \cos(n\pi)) = \frac{1 - (-1)^n}{n\pi}$$

which is not $o(n^{-1})$.
Also, it is important to talk about the regularity of f and not of $\tilde{f}$ as the odd extension of a function does not inherit a priori its regularity property. More precisely, it is possible to find a C^∞ function on $[0,\pi]$ whose odd extension is not even continuous: consider for example a constant non zero function. Even if the odd extension turns out to be differentiable, it need not be C^2 as can be seen by considering $f(x) = x^2$ which odd extension is $\tilde{f}(x) = x|x|$ which derivative is $\tilde{f}'(x) = 2|x|$ (not C^1).
Similar results can be stated for the case of even extensions but one would need conditions on the derivatives of odd orders to get the decay result.

CHAPTER 5

Fourier Transform

This chapter is inspired by [SS03, Chapter 5] and [Dac, Chapter 18].

The theory of Fourier series applies to periodic functions on $\mathbb{R}$. In this chapter, we develop an analogous theory for functions on the entire real line which are non-periodic. The functions we consider will be suitably "small" at infinity. There are several ways of defining an appropriate notion of "smallness", but it will nevertheless be vital to assume some sort of vanishing at infinity. On the one hand, recall that the Fourier series of a periodic function associates a sequence of numbers, namely the Fourier coefficients, to that function; on the other hand, given a suitable function f on $\mathbb{R}$, the analogous object associated to f will in fact be another function $\hat{f}$ on $\mathbb{R}$ which is called the Fourier transform of f. Since the Fourier transform of a function on $\mathbb{R}$ is again a function on $\mathbb{R}$, one can observe a symmetry between a function and its Fourier transform, whose analogue is not as apparent in the setting of Fourier series.

This tool will also enable us to transform problems involving PDEs, such as the Heat Equation, into ODEs, that are easier to deal with. The workflow is the following: once the explicit solution of the ODE is found in the space of Fourier Transforms, with the Inverse Transform we will be able to recover the solution of the original hard problem.

Definition. [Fourier Transform] Let $f \in L^1(\mathbb{R})$. The Fourier Transform of f is defined as

$$\mathcal{F}f(\xi) = \hat{f}(\xi) := \int_{-\infty}^{\infty} f(y) e^{-2\pi i \xi y} dy$$

Remark 5.1. Sometimes, the Fourier transform is defined without a 2π in the exponential and with a multiplicative factor $1/\sqrt{2\pi}$ in front.

Example 5.1. Consider $f(x) = I_{[-1,1]}(x)$. We can compute

$$\hat{f}(\xi) = \int_{-1}^{1} e^{-2\pi i \xi y} dy = \left[\frac{e^{-2\pi i \xi y}}{-2\pi i \xi} \right]_{-1}^{1} = \frac{1}{\pi \xi} \frac{e^{2\pi i \xi} - e^{-2\pi i \xi}}{2i} = \frac{\sin(2\pi \xi)}{\pi \xi}.$$

Remark 5.2. $|\hat{f}(0)| = \left| \int_{\mathbb{R}} f(x) dx \right| \leq ||f||_{L^1(\mathbb{R})}$.

Lemma 5.1 (Basic properties of the Fourier Transform)**.** *Let $f, g \in L^1(\mathbb{R})$, $a, b \in \mathbb{R}$. Then*

(i) $\hat{f}$ is a continuous function, $\lim_{\alpha \to \infty} |\hat{f}(\alpha)| = 0$ and $\|\hat{f}\|_{L^\infty} \leq \|f\|_{L^1}$
(ii) $\mathcal{F}(af + bg) = a\mathcal{F}f + b\mathcal{F}g$
(iii) If f is k times differentiable and $f^{(l)}$ is in $L^1(\mathbb{R})$ for all $l = 1, \ldots, k$, then

$$\widehat{f^{(k)}}(\xi) = (2\pi i \xi)^k \hat{f}(\xi);$$

(iv) If If $h_l(x) = x^l f(x) \in L^1(\mathbb{R})$ for some $l \in \mathbb{N}$ then $\hat{f}$ is l-times differentiable and

$$\frac{d^k}{(d\xi)^k} \hat{f}(\xi) = (-2\pi i) \hat{h}_k(\xi) \quad \forall k = 1, \ldots, l$$

(v) If $h(x) = f(x + a)$, then $\hat{h}(\xi) = e^{2\pi i \xi a} \hat{f}(\xi)$
(vi) If $h(x) = f(ax)$, then $\hat{h}(\xi) = \frac{1}{a} \hat{f}(\frac{\xi}{a})$

(vii) The multiplication formula holds:

$$\int_{-\infty}^{\infty} \hat{f}(x)g(x)dx = \int_{-\infty}^{\infty} f(x)\hat{g}(x)dx.$$

PROOF. Let us now prove all the properties:

(i) continuity follows from the fact that

$$|\hat{f}(\xi+h) - \hat{f}(\xi)| \le \int_{-\infty}^{\infty} f(y)|e^{-2\pi i\xi h} - 1|e^{-2\pi i\xi y}dy.$$

and since $|e^{-2\pi i\xi h} - 1| \to 0$ as $h \to 0$, we can apply the dominated convergence theorem.

Then, notice that $\lim_{\alpha\to\infty} |\hat{f}(\alpha)| = 0$ is granted again by dominated convergence and $\|\hat{f}\|_{L^\infty} = \|\hat{f}\|_{C^0} \le \|f\|_{L^1}$ because $|e^{-2\pi i\xi y}| \le 1$.

(ii) Linearity follows from the linearity of the integral. Indeed,

$$\begin{aligned}\mathcal{F}(af+bg)(\xi) &= \int_{\mathbb{R}} (af(x)+bg(x))e^{-2\pi i\xi x}\,dx = a\int_{\mathbb{R}} f(x)e^{-2\pi i\xi x}\,dx + b\int_{\mathbb{R}} g(x)e^{-2\pi i\xi x}\,dx \\ &= a\mathcal{F}(f)(\xi) + b\mathcal{F}(g)(\xi)\,.\end{aligned}$$

(iii) Since f is absolutely integrable, $\liminf_{x\to\pm\infty} |f(x)| = 0$. In particular, consider two subsequences $\{x_n\}_{n\in\mathbb{N}}, \{y_n\}_{n\in\mathbb{N}}, x_n \to +\infty, y_n \to -\infty$ such that $f(x_n) \to 0,\ f(y_n) \to 0$. Integrating by parts, we obtain:

$$\int_{y_n}^{x_n} f'(x)e^{-2\pi i\xi x}\,dx = (f(x)e^{-2\pi i\xi x})|_{y_n}^{x_n} + \int_{y_n}^{x_n} 2\pi i\xi f(x)e^{-2\pi i\xi x}\,dx\,.$$

By the choice of $\{x_n\}_{n\in\mathbb{N}}, \{y_n\}_{n\in\mathbb{N}}$ we can make the first term of the right-hand-side vanish, and we can pass to the limit for $n \to +\infty$ in the second one and in the left-hand-side thanks to the dominated convergence theorem, because $f, f' \in L^1(\mathbb{R})$. Therefore, we obtain

$$\left|\int_{\mathbb{R}} f'(x)e^{-2\pi i\xi x}\,dx - 2\pi i\xi\int_{\mathbb{R}} f(x)e^{-2\pi i\xi x}\,dx\right| = 0, \tag{5.1}$$

and we conclude $\mathcal{F}(f')(\xi) = 2\pi i\xi\mathcal{F}(f)(\xi)$.

Now, we prove the general result by induction. Assume it holds for some n and prove it for $n+1$. Let $f \in C^{n+1}(\mathbb{R})$ and $f^{(k)} \in L^1(\mathbb{R})$ for all $k = 1,\dots,n+1$. Then, since the result holds for n,

$$\mathcal{F}(f^{(n+1)})(\xi) = (2\pi i\xi)^n\mathcal{F}(f')(\xi) = (2\pi i\xi)^{n+1}\mathcal{F}(f)(\xi),$$

where the last equality follows from the case $n = 1$ which we already proved.

(iv) We will prove that for every $\xi \in \mathbb{R}$ we have

$$\lim_{\varepsilon\to 0} \frac{\widehat{f}(\xi+\varepsilon) - \widehat{f}(\xi)}{\varepsilon} + 2\pi i\widehat{h}(\xi) = 0.$$

This proves both the differentiability of $\hat{f}$ and the claimed formula for its derivative. Let $\varepsilon > 0$ be arbitrary. We have

$$\frac{\widehat{f}(\xi+\varepsilon) - \widehat{f}(\xi)}{\varepsilon} + 2\pi i\widehat{h}(\xi) = \int_{\mathbb{R}} f(x)e^{-2\pi i\xi x}\left[\frac{e^{-2\pi i\varepsilon x} - 1}{\varepsilon} + 2\pi ix\right]dx.$$

Notice that

$$\left|\frac{e^{-2\pi i\varepsilon x} - 1}{\varepsilon}\right| \le 2\pi|x|$$

and

$$\lim_{\varepsilon\to 0}\left[\frac{e^{-2\pi i\varepsilon x}-1}{\varepsilon}-2\pi i x\right]=0$$

pointwise. In order to be able to apply the dominated convergence theorem, notice that by assumption

$$\left|f(x)e^{-2\pi i\xi x}\left[\frac{e^{-2\pi i\varepsilon x}-1}{\varepsilon}+2\pi i x\right]\right|\le 4\pi|xf(x)|=4\pi|h(x)|\in L^1(\mathbb{R})\,.$$

By the dominated convergence theorem, we get,

$$\lim_{\varepsilon\to 0}\int_{\mathbb{R}} f(x)e^{-2\pi i\xi x}\left[\frac{e^{-2\pi i\varepsilon x}-1}{\varepsilon}+2\pi i x\right]dx=0,$$

which proves the result.

Now, in order to prove the formula for higher order derivatives, we use an induction argument. Assume by induction that the formula holds for some l and prove it for $l+1$. We assume that $h_{l+1}\in L^1(\mathbb{R})$. In order to apply the formula for l, we need to make sure that $h_l\in L^1(\mathbb{R})$. Indeed,

$$\begin{aligned}\int_{\mathbb{R}}|x^l f(x)|\,dx &= \int_{|x|\le 1}|x|^l|f(x)|\,dx+\int_{|x|>1}|x|^l|f(x)|\,dx\\ &\le \int_{|x|\le 1}|f(x)|\,dx+\int_{|x|>1}|x|^{l+1}|f(x)|\,dx<\infty\end{aligned}$$

Thus, using the induction hypothesis and the case $l=1$ that we already proved, we have

$$\begin{aligned}\mathcal{F}(f)^{(l+1)}(\xi)&=\frac{d}{d\xi}[\mathcal{F}(f)^{(l)}(\xi)]=\frac{d}{d\xi}[(-2\pi i)^l\mathcal{F}(h_l)(\xi)]=(-2\pi i)^l\frac{d}{d\xi}[\mathcal{F}(h_l)(\xi)]\\ &=(-2\pi i)^{(l+1)}\mathcal{F}(h_{l+1})(\xi).\end{aligned}$$

(v) Using Proposition 2.23, we obtain

$$\mathcal{F}(h)(\xi)=\int_{\mathbb{R}} f(x+a)e^{-2\pi i\xi x}\,dx\overset{??}{=}\int_{\mathbb{R}} f(x)e^{-2\pi i\xi(x-a)}\,dx=e^{2\pi i\xi a}\mathcal{F}(f)(\xi)\,.$$

(vi) Using Proposition 2.23, we obtain

$$\mathcal{F}(h)(\xi)=\int_{\mathbb{R}} f(ax)e^{-2\pi i\xi x}\,dx\overset{??}{=}\frac{1}{a}\int_{\mathbb{R}} f(x)e^{-2\pi i\xi x/a}\,dx=\frac{1}{a}\mathcal{F}(f)(\xi/a)\,.$$

(vii) First of all, notice that both integrals are finite and well-defined because $f,g\in L^1$ and $\widehat{f},\widehat{g}\in L^\infty$. More precisely, we have

$$\int_{\mathbb{R}}\widehat{f}(x)g(x)\,dx=\int_{\mathbb{R}}\left(\int_{\mathbb{R}} f(y)e^{-2\pi i xy}\,dy\right)g(x)\,dx$$

and

$$\int_{\mathbb{R}} f(y)\widehat{g}(y)\,dy=\int_{\mathbb{R}} f(y)\left(\int_{\mathbb{R}} g(x)e^{-2\pi i xy}\,dx\right)dy.$$

In order to prove the result we will use Fubini's theorem. However, we first need to show that the function

$$(x,y)\mapsto f(y)g(x)e^{-2\pi i xy}$$

is integrable on $\mathbb{R}^2$. Indeed, by Tonelli's theorem

$$\int_{\mathbb{R}^2} |f(y)g(x)e^{-2\pi ixy}|\, d(x,y) = \int_{\mathbb{R}^2} |f(y)||g(x)|\, d(x,y) = \left(\int_{\mathbb{R}} |f(y)|\, dy\right)\left(\int_{\mathbb{R}} |f(y)||g(x)|\, dx\right)$$
$$= \|f\|_{L^1(\mathbb{R})}\|g\|_{L^1(\mathbb{R})} < \infty.$$

Thus, using Fubini's theorem, we obtain

$$\begin{aligned}\int_{-\infty}^{+\infty} \widehat{f}(x)g(x)\, dx &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(y)e^{-2\pi ixy}\, dy\right) g(x)\, dx \\ &= \int_{\mathbb{R}^2} f(y)e^{-2\pi ixy} g(x)\, d(x,y) \\ &= \int_{\mathbb{R}} f(y) \left(\int_{\mathbb{R}} g(x)e^{-2\pi ixy}\, dx\right) dy \\ &= \int_{\mathbb{R}} f(y)\widehat{g}(y)\, dy.\end{aligned}$$

□

Proposition 5.2 (Gaussians are good kernels)**.** *Let* $f(x) = e^{-\pi x^2}$*. Then, we have* $\hat{f} = f$*.*

Proof. Note that $f'(x) = -2\pi x f(x)$, thus

$$\mathcal{F}(f')(\xi) = -2\pi\mathcal{F}(xf(x))(\xi)$$

Now, we can apply (iii) of Theorem 5.1, and obtain

$$2\pi i\xi\mathcal{F}(f)(\xi) = -i\frac{d}{d\xi}\mathcal{F}(f)(\xi).$$

Therefore $\hat{f} = \mathcal{F}(f)$ satisfies the same ODE as f, that is $\hat{f}' = -2\pi\xi\hat{f}$ and since

$$\hat{f}(0) = \int_{\mathbb{R}} f(y)dy \stackrel{\star}{=} 1,$$

we can conclude that $\hat{f} = f$.

The step in $\star$ is due to the fact that

$$\left(\int_{-\infty}^{\infty} e^{-\pi x^2} dx\right)^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\pi(x^2+y^2)} dxdy = \int_0^{2\pi}\int_0^{\infty} e^{-\pi r^2} r dr d\theta$$
$$= (2\pi)\int_0^{\infty} e^{-\pi r^2} r dr = (2\pi)[e^{-\pi r^2}]_0^{\infty} = 1.$$

□

Corollary 5.3. *If* $\delta > 0$*, let* $K_\delta(x) = \delta^{-1/2}e^{-\pi x^2/\delta}$*.*
Then

$$\hat{K}_\delta(\xi) = e^{-\pi\delta\xi^2}.$$

and K_δ *enjoys the following properties:*

(i) $K_\delta \geq 0$*;*
(ii) $\int_{-\infty}^{\infty} K_\delta(x)dx = \int_{-\infty}^{\infty} K_1(x)dx = 1$*;*
(iii) $\forall \eta > 0, \int_{|x|>\eta} K_\delta(x)dx \to 0$ *as* $\delta \to 0$*.*

PROOF. Note that $K_\delta(x) = \delta^{-1/2} K_1(x/\sqrt{\delta})$; we can apply (vi) of Theorem 5.1 to conclude that

$$\hat{K}_\delta(\xi) = e^{-\pi\delta\xi^2}.$$

Let us now prove the three properties:

(i) $K_\delta \geq 0$ follows immediately from the definition of K_δ and the non-negativity of the exponential function;

(ii) with the change of variables $y = x/\sqrt{\delta}$, we have

$$\int_{-\infty}^{\infty} K_\delta(x)dx = \delta^{-1/2} \int_{-\infty}^{\infty} e^{-\pi x^2/\delta} dx = \int_{-\infty}^{\infty} e^{-\pi y^2} dy = 1;$$

(iii) $\int_{|x|>\eta} K_\delta(x) = \int_{|y|>\eta/\sqrt{\delta}} e^{-\pi y^2} dy \to 0$ as $\delta \to 0$ because $\eta/\sqrt{\delta} \to \infty$.

□

Remark 5.3. The statement (iii) of Theorem 5.3 implies that, as $\delta \to 0$, K_δ concentrates at 0 and $\hat{K}_\delta$ gets flatter. This is an example of Heisenberg's uncertainty principle: f and $\hat{f}$ cannot be both essentially localized.

5.1. Fourier Inversion Formula

THEOREM 5.4 (Fourier Inversion Formula). *Let $f \in L^1(\mathbb{R})$ s.t. $|\hat{f}| \in L^1(\mathbb{R})$, then for a.e. $x \in \mathbb{R}$*

$$f(x) = \int_{\mathbb{R}} \hat{f}(\xi) e^{2\pi i\xi y} d\xi \tag{5.2}$$

Remark 5.4. Note that it is natural to suppose $|\hat{f}| \in L^1$ to define the right-hand-side of (5.2); under this assumption, the right-hand-side is continuous , therefore f coincides a.e. with a continuous function.
If $f \in C^0(\mathbb{R})$, the inversion formula holds $\forall x \in \mathbb{R}$.

Now, it is natural to ask when the condition $|\hat{f}| \in L^1$ is satisfied.

We define the Schwartz space $\mathcal{S}(\mathbb{R}) \subset C^\infty(\mathbb{R})$ as the set of all functions $f \in C^\infty(\mathbb{R})$ such that their derivatives are rapidly decreasing, namely

$$\sup_x |x|^k |f^{(l)}(x)| < \infty \quad \forall k, l \in \mathbb{N}.$$

The Schwartz space $\mathcal{S}(\mathbb{R})$ contains all smooth compactly supported functions, and Gaussians (whose derivatives are of the form $P(x)e^{-cx^2}$ with P polynomial).
Moreover, if $f \in \mathcal{S}(\mathbb{R})$, then $\hat{f} \in \mathcal{S}(\mathbb{R})$. We can now prove the following corollary:

Corollary 5.5. *The Fourier Transform in the Schwartz space $\mathcal{F} : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ is bijective.*

PROOF. Let $\mathcal{F}^* f = \int_{-\infty}^{\infty} f(y) e^{2\pi i\xi y} dy$. Then,

$$\mathcal{F}^* \circ \mathcal{F} = \text{Id on } \mathcal{S}(\mathbb{R})$$

Furthermore, since $\mathcal{F}^* f = \mathcal{F}(f(-x))$, therefore we also have that

$$\mathcal{F} \circ \mathcal{F}^* = \text{Id}.$$

□

Let us now prove the Fourier Inversion Formula.

PROOF OF THEOREM 5.4. We will first prove the result in a simpler case, for $f \in C^0 \cap L^\infty$: here, we will prove the Fourier Inversion Formula for all x.
For $x = 0$, we want to show that

$$f(0) = \int_{-\infty}^{\infty} \hat{f}(\xi) d\xi$$

We apply (vii) of Theorem 5.1 to f and $G_\delta = e^{-\pi x^2 \delta}$; note that $\hat{G}_\delta = K_\delta$, that is a good kernel. Hence, we have that:

$$\int_{-\infty}^{\infty} f(x) K_\delta(x) dx = \int_{-\infty}^{\infty} \hat{f}(\xi) G_\delta(\xi) d\xi.$$

We first want to show that

$$\int_{-\infty}^{\infty} \hat{f}(\xi) G_\delta(\xi) d\xi \to \int_{-\infty}^{\infty} \hat{f}(\xi) d\xi \text{ as } \delta \to 0;$$

note that $\hat{f} \in L^1$, $G_\delta \to 1$ pointwise, and $G_\delta \leq 1$, therefore the claim follows by applying dominated convergence on $|\hat{f}|$.

Let us now focus on the other side of the equality to prove; take $\varepsilon > 0$ and consider

$$\begin{aligned}\left| f(0) - \int_{-\infty}^{\infty} f(x) K_\delta(x) dx \right| &\leq \int_{-\infty}^{\infty} |f(0) - f(x)| K_\delta(x) dx \\ &\leq \int_{|x|<\eta} |f(0) - f(x)| K_\delta(x) + \leq \int_{|x|>\eta} |f(0) - f(x)| K_\delta(x) \\ &\leq \int_{|x|<\eta} |f(0) - f(x)| K_\delta(x) + 2||f||_{L^\infty} o(1).\end{aligned}$$

If we fix η such that $|f(0) - f(x)| < \varepsilon \quad \forall |x| < \eta$, we can conclude that

$$\left| f(0) - \int_{-\infty}^{\infty} f(x) K_\delta(x) dx \right| \leq \varepsilon \underbrace{\int_{-\infty}^{\infty} K_\delta(x) dx}_{=1} + 2\|f\|_{L^\infty} o(1),$$

and the proof is concluded, because $o(1)$ goes to 0 in δ at any fixed η.

For general x, instead, recall that for $F(y) = f(x + y)$, then

$$f(x) = F(0) = \int_{-\infty}^{\infty} \hat{F}(\xi) d\xi = \int_{-\infty}^{\infty} \hat{f}(\xi) e^{2\pi i \xi x} d\xi,$$

by (v) of Theorem 5.1. We can now deal with the general case.

Now, we consider $f, \hat{f} \in L^1(\mathbb{R})$: for any x, we write the multiplication formula applied to $F(y) = f(x + y)$ and K_δ; we get

$$\int_{-\infty}^{\infty} F(y) K_\delta(y) dy = \int_{-\infty}^{\infty} \hat{F}(\xi) G_\delta(\xi) d\xi = \int_{-\infty}^{\infty} \hat{f}(\xi) e^{i 2\pi \xi x} G_\delta(\xi) d\xi.$$

Now we let $\delta \to 0$; then the right-hand-side converges pointwise in x:

$$\int_{-\infty}^{\infty} \hat{f}(\xi) e^{i2\pi\xi x} G_\delta(\xi) d\xi \to \int_{-\infty}^{\infty} \hat{f}(\xi) e^{i2\pi\xi x} d\xi.$$

Furthermore, we claim that the left-hand-side converges to $f(x)$ for almost every $x \in \mathbb{R}$:

$$f_\delta(x) = \int_{-\infty}^{\infty} f(x + y) K_\delta(y) dy \to f(x).$$

The claim can be proved by noting that

$$|f_\delta(x) - f(x)| \leq \left| \int_{-\infty}^{\infty} (f(x+y) - f(x)) K_\delta(y) \right| dy.$$

Thus,

$$\begin{aligned}\int_{-\infty}^{\infty} |f_\delta(x) - f(x)| dx &\leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |f(x+y) - f(x)| K_\delta(y) dy dx \\ &\overset{\star}{=} \int_{-\infty}^{\infty} \|f(\cdot + y) - f\|_{L^1} K_\delta(y) dy \\ &\overset{\diamondsuit}{=} \int_{-\infty}^{\infty} \|f(\cdot + \sqrt{\delta} z) - f\|_{L^1} e^{-\pi z^2} dz,\end{aligned}$$

where $\star$ follows from the Fubini-Tonelli theorem and $\diamondsuit$ from the change of variables $y = \sqrt{\delta} z$. Now, recall that $\forall f \in L^p(\mathbb{R})$, the translations are continuous in the L^p norm:

$$\|f(\cdot + y) - f\|_{L^p} \to 0 \text{ as } y \to 0.$$

Thus, the quantity above is dominated and the integral goes to 0.

□

Proposition 5.6. *If $f, g \in \mathcal{S}(\mathbb{R})$ then $f * g \in \mathcal{S}(\mathbb{R})$.*

Proof. To prove that $f * g$ is rapidly decreasing, observe first that for any $\ell \geq 0$ we have $\sup_{x \in \mathbb{R}} |x|^\ell |g(x-y)| \leq A_\ell (1 + |y|)^\ell$, because g is rapidly decreasing (to check this assertion, consider separately the two cases $|x| \leq 2|y|$ and $|x| \geq 2|y|$). From this, we see that

$$\sup_x \left| x^\ell (f * g)(x) \right| \leq A_\ell \int_{-\infty}^{\infty} |f(y)| (1 + |y|)^\ell dy$$

so that $x^\ell (f * g)(x)$ is a bounded function for every $\ell \geq 0$. These estimates carry over to the derivatives of $f * g$, thereby proving that $f * g \in \mathcal{S}(\mathbb{R})$ because, as observed for (3.13),

$$\left(\frac{d}{dx} \right)^k (f * g)(x) = \left(f * \left(\frac{d}{dx} \right)^k g \right)(x) \quad \text{for } k = 1, 2, \ldots$$

This identity is proved first for $k = 1$ by differentiating under the integral defining $f * g$. The interchange of differentiation and integration is justified in this case by the rapid decrease of dg/dx, that enables to apply Theorem 2.17. The identity then follows for every k by iteration. □

5.2. Plancherel Identity

Theorem 5.7 (Plancherel). *Let $f \in L^1 \cap L^2(\mathbb{R})$, then $\hat{f} \in L^2(\mathbb{R})$ and*

$$\|f\|_{L^2} = \|\hat{f}\|_{L^2}. \tag{5.3}$$

Before proving the theorem, let us first prove a preliminary result

Proposition 5.8. *Let $f, g \in L^1(\mathbb{R})$ and consider their convolution*

$$f * g = \int_{-\infty}^{\infty} f(x-y) g(y) dy.$$

Then

(i) $f * g$ is well defined for a.e. and $\|f * g\|_{L^1} \leq \|f\|_{L^1} \|g\|_{L^1}$
(ii) $f * g = g * f$

(iii) $\widehat{f * g} = \hat{f}\hat{g}$

Proof. For (i), we just prove the inequality, without showing measurability:

$$\begin{aligned}\int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} f(x-y)g(y)dy \right| dx &\le \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |f(x-y)g(y)| dy dx \\ &= \int_{-\infty}^{\infty} \|f\|_{L^1} |g(y)| dy \\ &\le \|f\|_{L^1} \|g\|_{L^1}.\end{aligned}$$

For (ii), observe that

$$\int_{-\infty}^{\infty} f(x-y)g(y)dy = \int_{-\infty}^{\infty} g(x-z)f(z)dz$$

with the change of variables $z = x - y$. For (iii), with the change of variables $z = x - y$, we have:

$$\begin{aligned}\mathcal{F}(f * g)(\xi) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x-y)g(y)dy e^{-2\pi i \xi x} dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(z)g(y)e^{-2\pi i \xi z} e^{-2\pi i \xi y} dz dy \\ &= \int_{-\infty}^{\infty} f(z)e^{-2\pi i \xi z} dz \int_{-\infty}^{\infty} g(y)e^{-2\pi i \xi y} dy \\ &= \hat{f}(\xi)\hat{g}(\xi).\end{aligned}$$

□

Remark 5.5. For $f, g \in \mathcal{S}(\mathbb{R})$, we can prove that $\widehat{fg} = \hat{f} * \hat{g}$.

Proof of Theorem 5.7. we divide the proof in two steps.

Step 1: we prove that $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ and $\hat{f} \in L^2(\mathbb{R})$ imply (5.3).
Let $g(x) = \overline{f(-x)}$, so that $\hat{g}(x) = \int_{-\infty}^{+\infty} \overline{f(-y)} e^{-i2\pi\xi y} dy = \overline{\hat{f}(\xi)}$ (note indeed that the sign does not change, because we have one negative sign from the differential and one negative sign from the inversion of extrema of the interval).

Let $h = f * g$, then $\hat{h} = \hat{f}\hat{g} = \hat{f}\overline{\hat{f}} = |\hat{f}|^2$, exploiting Theorem 5.8 and the above observation.

Note that we have

$$\begin{aligned}h(0) &= \int_{-\infty}^{\infty} f(y)g(0-y)dy \\ &= \int_{-\infty}^{\infty} |f|^2(y)dy.\end{aligned}$$

Now, we want to compute $h(0)$ via the Fourier Inverse Formula:

$$h(0) = \int_{-\infty}^{\infty} \hat{h}(\xi)d\xi = \|\hat{f}\|_{L^2}^2.$$

However, note that in principle the Fourier Inverse Formula cannot be applied at $x = 0$, as we only have equality almost everywhere. But actually h is continuous, because

$$\begin{aligned}|h(x+\epsilon)-h(x)| &= \left|\int_{-\infty}^{\infty} f(y)(g(x+\epsilon-y)-g(x-y))dy\right| \\ &\le \int_{-\infty}^{\infty} |f(y)||g(x+\varepsilon-y)-g(x-y)|dy \\ &\le \|f\|_{L^2}\|g(x+\varepsilon-\cdot)-g(x-\cdot)\|_{L^2} \\ &\le \|f\|_{L^2}\|g(\varepsilon-\cdot)-g(\cdot)\|_{L^2} \to 0\end{aligned}$$

as $\varepsilon \to 0$, thanks to the continuity of translations in L^2.

Step 2: we prove that $f \in L^1 \cap L^2 \Rightarrow \hat{f} \in L^2$ and $\|\hat{f}\|_{L^2} \le \|f\|_{L^2}$.

We define $f_\delta := f * K_\delta$ and we want to apply the result in **Step 1** to f_δ and then let $\delta \to 0$. We know that $\hat{f}_\delta = \hat{f}\hat{K}_\delta$; now, we need to verify that f_δ satisfies the right assumptions. We know that $f_\delta \in L^1$ from Theorem 5.8. Furthermore, $\hat{f}_\delta = \hat{f}\hat{K}_\delta$, and since $\hat{f} \in L^\infty(\mathbb{R})$ and $e^{-\pi\delta x^2} \in L^2(\mathbb{R})$ (because Gaussians belong to L^p for all $p > 0$), by the Hölder inequality $\hat{f}_\delta$ is in $L^2(\mathbb{R})$.

Now, we first bound f_δ pointwise: we fix $x \in \mathbb{R}$, and study

$$\begin{aligned}f_\delta^2(x) &= \left[\int_{-\infty}^{\infty} f(y)K_\delta(x-y)dy\right]^2 \\ &\le \int_{-\infty}^{\infty} f^2(y)K_\delta(x-y)dy \underbrace{\int_{-\infty}^{\infty} K_\delta(x-y)dy}_{=1} \\ &= \int_{-\infty}^{\infty} f^2(y)K_\delta(x-y)dy\end{aligned}$$

by Hölder inequality with factors $f(y)\sqrt{K_\delta(x-y)}$ and $\sqrt{K_\delta(x-y)}$ and exponents $p = q = 2$). We now integrate with respect to x, thus, applying Fubini's Theorem,

$$\begin{aligned}\int_{-\infty}^{\infty} f_\delta^2(x)dx &\le \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y)^2 K_\delta(x-y)dydx \\ &= \int_{-\infty}^{\infty} f(y)^2 \underbrace{\int_{-\infty}^{\infty} K_\delta(x-y)dx}_{=1}\, dy \\ &= \int_{-\infty}^{\infty} f(y)^2 dy.\end{aligned}$$

Now, we have that

$$\|f\|_{L^2} \ge \|f_\delta\|_{L^2} = \|\hat{f}_\delta\|_{L^2} = \|\hat{f}e^{-\pi\delta\xi^2}\|_{L^2} = \left(\int_{-\infty}^{\infty} |\hat{f}(\xi)|^2 e^{-2\pi\delta\xi^2} d\xi\right)^{1/2}$$

Notice that $e^{-2\pi\delta\xi^2} \to 1$ pointwise as $\delta \to 0$, and that we can apply monotone convergence to conclude that

$$\left(\int_{-\infty}^{\infty} |\hat{f}|^2 e^{-2\pi\delta\xi^2}\right)^{1/2} \to \|\hat{f}\|_{L^2}.$$

Therefore, $\|f\|_{L^2} \ge \|\hat{f}\|_{L^2}$, and this concludes the proof.

□

CHAPTER 6

Fourier Transforms and PDEs

This chapter is inspired by [Dac, Chapter 19], [SS03, Chapter 5].

Definition (Partial Differential Equation). A Partial Differential Equation (PDE) is an equation whose solution u is such that

$$F(x, u(x), \nabla u(x), \dots, \nabla^k u(x)) = 0$$

where $x \in \Omega \subset \mathbb{R}^d$, $u : \Omega \to \mathbb{R}^N$, $F : \Omega \times \mathbb{R}^N \times \mathbb{R}^{d\times N} \times \dots \times \mathbb{R}^{d\times \dots \times d \times N} \to \mathbb{R}^m$.

Example 6.1. Examples of well-known PDEs are:

(i) Given $u : \mathbb{R} \to \mathbb{R}$, $\partial_{xx} u = 0$, whose solutions are $u = ax^2 \quad a \in \mathbb{R}$;

(ii) The Laplace Equation in $\mathbb{R}^2$ for $u : \mathbb{R}^2 \to \mathbb{R}$:

$$\partial_{xx} u + \partial_{yy} u = 0$$

(iii) The Laplace Equation in $\mathbb{R}^d$ for $u : \mathbb{R}^d \to \mathbb{R}$:

$$\Delta u = \sum_{i=1}^{d} \partial_{x_i x_i} u = 0$$

(iv) The Poisson Equation, non-homogeneous version of the Laplace Equation with a given datum f:

$$\Delta u = f$$

(v) The Heat Equation

$$\partial_t u - \Delta u = 0$$

for which the function u has a $d+1$-dimensional domain, as $x \in \mathbb{R}^d$ and $t \in \mathbb{R}$;

(vi) The Wave Equation

$$\partial_{tt} u - \Delta u = 0$$

(vii) The PDE

$$\nabla u = f$$

(viii) The Burgers Equation

$$\partial_t u - \partial_x (u^2) = 0$$

Remark 6.1. The Laplace, Heat and Wave Equations are linear PDEs, namely, $F(x, \cdot, \cdot, \dots, \cdot)$ is a linear function. In the equations, when there is no given right-hand side, if u, v are solutions, then $\alpha u + \beta v$ is a solution $\forall \alpha, \beta \in \mathbb{R}$. On the other hand, the Burgers Equation is nonlinear.

Typically, PDEs are associated to given conditions describing the behavior at the boundary of the domain or the initial condition. For example, for the Heat and Wave Equations that describe an evolution in time, we prescribe the starting condition by setting

$$u(x, 0) = \phi(x) \text{ for some given function } \phi$$

On the other hand, for the Laplace Equation we prescribe conditions on the function or its derivatives at the domain's edge: we can set Dirichlet boundary conditions

$$u(x) = u_0(x) \text{ on } \partial\Omega$$

or Neumann boundary conditions

$$\partial_\nu u(x) = v_0(x) \text{ on } \partial\Omega$$

We will now focus instead on few of the main PDEs recurrent in the applications and on particular methods to solve them, that are:

- Separation of Variables
- Fourier Transform the equation and solve the ODEs that appear there
- Solution via Fourier Series.

6.1. The heat equation on $\mathbb{R}$

The heat equation has the general form

$$\begin{cases} \partial_t u - \partial_{xx} u = 0 \\ u(x,0) = f(x), \end{cases} \tag{6.1}$$

where $f(x) : \mathbb{R} \to \mathbb{R}$ is given, while $u : [0,\infty) \times \mathbb{R} \to \mathbb{R}$ is to be found.

We formally derive a solution using the following strategy: first of all, we compute the Fourier Transform in x

$$\begin{cases} \partial_t \hat{u}(\xi,t) + 4\pi^2\xi^2 \hat{u}(\xi,t) = 0 \\ \hat{u}(\xi,0) = \hat{f}(\xi). \end{cases}$$

Now notice that, for a fixed ξ, this is an ODE in time for the function $\hat{u}$.
Therefore, we can compute its solution:

$$\partial_t[\ln \hat{u}] = -4\pi^2\xi^2$$
$$\ln \hat{u}(\xi,t) - \ln \hat{u}(\xi,0) = -4\pi^2\xi^2 t$$

Taking the exponential, we find

$$\hat{u}(\xi,t) = \hat{u}(\xi,0)e^{-4\pi^2\xi^2 t}\hat{f}(\xi)e^{-4\pi^2\xi^2 t},$$

and finally, by inverting the Fourier Transform and applying Corollary 5.3 on the Fourier Transform of the exponential, we retrieve

$$u(x,t) = f * \mathcal{F}^{-1}(e^{-4\pi^2 t\xi^2}) = f * \frac{1}{(4\pi t)^{1/2}} e^{-x^2/4t}.$$

Theorem 6.1 (Solution to heat equation). *Let $f \in \mathcal{S}(\mathbb{R})$, and define the Heat Kernel as*

$$H_t(x) \coloneqq \frac{1}{(4\pi t)^{1/2}} e^{-x^2/4t}$$

Let $u = f * H_t$ for $t > 0$; then, u satisfies the following:

(i) $u \in C^2(\mathbb{R})$ for $x \in \mathbb{R}, t > 0$ and solves the heat equation $\partial_t u - \partial_{xx} u = 0$
(ii) $u(x,t) \to f(x)$ uniformly in x as $t \to 0$
(iii) $u(\cdot,t) \to f$ in the L^2 norm as $t \to 0$, namely $\int_{\mathbb{R}} |u(x,t) - f(x)|^2 dx \to 0$ as $t \to 0$.

Remark 6.2. Observe that H_t is a Gaussian for every t fixed and $\int_{\mathbb{R}} H_t(x)dx = 1 \quad \forall t > 0$. Moreover, as $t \to 0$, $H_t \to 0$ a.e.; in particular, $H_t dx$ converges to the Dirac Delta centered at the origin δ_0, which is rigorously expressed as

$$\forall \varepsilon > 0 \quad \int_{[-\varepsilon,\varepsilon]} H_t \to 1 \text{ as } t \to 0. \tag{6.2}$$

Remark 6.3. If $f \in C_c^0$ is non zero and $f \geq 0$, then $\operatorname{supp}(u(\cdot,t)) = \mathbb{R}$ and $\int_{\mathbb{R}} f(x)dx = \int_{\mathbb{R}} u(x,t)dx$.

PROOF OF THEOREM 6.1. Let us prove the points of the theorem:

(i) Take the Fourier transform of u. Notice that this can be done, since $f \in \mathcal{S}(\mathbb{R}) \subset L^1(\mathbb{R})$ and $H_t \in L^1(\mathbb{R})$, so $f * H_t \in L^1(\mathbb{R})$ (see Proposition 5.8 (i)). Then, by Proposition 5.8 (iii), we have:

$$\hat{u} = \hat{f}\widehat{H_t} = \hat{f}(\xi)e^{-4\pi^2\xi^2 t}$$

Then, the Fourier Inversion Formula, whose application is justified by $u, \hat{u} \in L^1(\mathbb{R})$ (Theorem 5.4), gives

$$u(x,t) = \int_{-\infty}^{\infty} \hat{f}(\xi)e^{-4\pi^2\xi^2 t + 2\pi i\xi x}d\xi.$$

If u is differentiable in x, then differentiating under the integral (see Theorem 2.17) gives the meaningful formula

$$\partial_x u(x,t) = \int_{-\infty}^{\infty} \hat{f}(\xi)e^{-4\pi^2\xi^2 t + 2\pi i\xi x}(2\pi i\xi)d\xi.$$

Another way to differentiate under the integral in a rigorous way consists in computing

$$\frac{u(x+h,t) - u(x,t)}{h} = \int_{-\infty}^{\infty} \hat{f}(\xi)e^{-4\pi^2\xi^2 t}e^{2\pi i\xi x}\left(\frac{e^{2\pi i\xi h} - 1}{h}\right)d\xi.$$

Letting $h \to 0$, we have that

$$\frac{e^{2\pi i\xi h} - 1}{h} \to 2\pi i\xi$$

and we can retrieve the same formula by dominated convergence (Theorem 2.14). Indeed, we can take as dominant $\|\hat{f}\|_{L^\infty}(2\pi\xi)e^{-4\pi^2\xi^2 t} \in L^1(\mathbb{R}) \quad \forall t > 0$, because $\hat{f}(\xi)$ is bounded in ξ and $|2\pi i\xi| \leq 2\pi|\xi|$.

The same justification holds for $\partial_{xx}u$ and $\partial_t u$:

$$\partial_{xx}u(x,t) = \int_{-\infty}^{\infty} \hat{f}(\xi)e^{-4\pi^2\xi^2 t}e^{2\pi ix\xi}(2\pi i\xi)^2 d\xi,$$

$$\partial_t u(x,t) = \int_{-\infty}^{\infty} \hat{f}(\xi)e^{-4\pi^2\xi^2 t}e^{2\pi ix\xi}(-4\pi^2\xi^2)d\xi.$$

Therefore, $\partial_t u = \partial_{xx}u \quad \forall t > 0, x \in \mathbb{R}$.

(ii) Then, we want to prove that $u(t,x) - f(x) \to 0$ as $t \to 0$.
Let $\varepsilon > 0$. We have that

$$u(x,t) - f(x) = \int_{-\infty}^{\infty} H_t(y)(f(x-y) - f(x))dy.$$

We claim that for t sufficiently small, the modulus of the integrand is smaller than ε: to do so, fix $R > 0$ such that $|f| \leq \varepsilon/4$ outside $[-R, R]$.
Since f is uniformly continuous in $[-R-1, R+1]$, there exists δ such that

$$|f(x) - f(x-y)| \leq \frac{\varepsilon}{2} \quad \forall |y| < \delta.$$

Then,

$$\begin{aligned}\left|\int_{-\infty}^{\infty} H_t(y)(f(x-y) - f(x))dy\right| &\leq \int_{|y|<\delta} H_t(y)|f(x-y) - f(x)|dy + \int_{|y|>\delta} H_t(y)|f(x-y) - f(x)|dy \\ &\leq \sup_{x\in\mathbb{R}, |y|<\delta}|f(x) - f(x-y)|\int_{|y|<\delta} H_t(y)dy + 2\sup_{x\in\mathbb{R}}|f(x)|\int_{|y|>\delta} H_t(y)dy.\end{aligned}$$

Now, we make the following observations:

- $\sup_{x\in\mathbb{R},|y|<\delta}|f(x)-f(x-y)|\leq \varepsilon/2$ because on $[-R-1,R+1]$ we have uniform continuity, while on $[-R-1,R+1]^C$ we have $|f|\leq\varepsilon/4$ by the choice of R;
- $\int_{|y|<\delta}H_t(y)dy\leq 1$ because $\int_{\mathbb{R}}H_t(x)dx=1\quad\forall t>0$;
- $\int_{|y|>\delta}H_t(y)dy\to 0$ as $t\to 0$ by the properties of good kernels (Corollary 5.3 (iii)).

Therefore, we can conclude that, for t small enough,

$$\left|\int_{-\infty}^{\infty}H_t(y)(f(x-y)-f(x))dy\right|\leq\frac{\varepsilon}{2}+2\sup_{x\in\mathbb{R}}|f(x)|\int_{|y|>\delta}H_t(y)dy\leq\varepsilon$$

(iii) To prove the third point, use Plancherel (Theorem 5.7) to conclude that

$$\begin{aligned}\int_{-\infty}^{\infty}|u(x,t)-f(x)|^2dx&=\int_{-\infty}^{\infty}|\hat{u}(\xi,t)-\hat{f}(\xi)|^2d\xi\\&=\int_{-\infty}^{\infty}|\hat{f}(\xi)||e^{-4\pi^2\xi^2t}-1|d\xi\to 0\end{aligned}$$

as $t\to 0$ by dominated convergence with dominant $2|\hat{f}|^2$.

□

Remark 6.4. If we try to solve Equation 6.1 *backwards* (with $t<0$), the formal computations are the same, but the formula we get for u has problems and there is no analogue for Theorem 6.1.

Remark 6.5. $u(\cdot,t)\in\mathcal{S}(\mathbb{R})$ uniformly in t, namely:

$$\sup_{x\in\mathbb{R},0<t<T}|x|^k\left|\frac{\partial^l}{\partial x^l}u(x,t)\right|<+\infty\quad\forall k,l\geq 0$$

We can prove it for $k=l=0$:

$$|u(t,x)|\leq\int_{|y|\leq|x|/2}|f(x-y)||H_t(y)|dy+\int_{|y|\geq|x|/2}|f(x-y)||H_t(y)|dy.$$

For the first term in the integrand, note that $f\in\mathcal{S}(\mathbb{R})$ implies that $\forall N\in\mathbb{N}$ we have

$$|f(x-y)|\leq\frac{C_N}{1+x^N}\qquad\text{for some } C_N>0 \text{ and } \forall|y|\leq|x|/2.$$

On the other hand, for the second term in the integrand, we have that f is bounded on $\mathbb{R}$ and $H_t(y)\leq(4\pi t)^{-1/2}e^{-|x|^2/16t}$ $\forall|y|\geq|x|/2$. Therefore,

$$|u(x,t)|\leq\frac{C_N}{1+|x|^N}+\frac{C}{\sqrt{t}}e^{-cx^2/t}.$$

Remark 6.6 (What about uniqueness?)**.** Notice that it is sufficient to prove that for $f\equiv 0$, the solution is uniquely equal to 0. Indeed, if f were an initial datum associated to two distinct solutions u and v, then $u-v$ would be a non null solution with null initial datum.

We will only sketch the proof of the following theorem:

THEOREM 6.2. *If* $u:\mathbb{R}\times[0,\infty)\to\mathbb{R}$ *is*

(i) a solution of the heat equation and $u(x,0)=0$*;*
(ii) $u\in C^0(\mathbb{R}\times[0,\infty))\cap C^2(\mathbb{R}\times(0,\infty))$*;*
(iii) $u(\cdot,t)\in\mathcal{S}(\mathbb{R})$ *uniformly in* t*;*

then, $u\equiv 0$.

Sketch of the proof. Introduce the energy

$$E(t) = \int_{-\infty}^{\infty} |u|^2(x,t)dx \geq 0$$

Note that this quantity is decreasing, because

$$\frac{dE}{dt}(t) = \int_{-\infty}^{\infty} 2\partial_t u(x,t)u(x,t)dx = 2\int_{-\infty}^{\infty} \partial_{xx}u(x,t)u(x,t)dx = -2\int_{-\infty}^{\infty} |\partial_x u(x,t)|^2 dx \leq 0$$

where the integration by parts can be justified by considering intervals of the form $[-N, N]$ and then let $N \to \infty$.

Since $u(\cdot, 0) = 0$, $E(0) = 0$. Therefore, $E(t) = 0 \quad \forall t > 0$ and $u \equiv 0$. □

Remark 6.7. $u(x,t) = x/tH_t(x)$ solves Equation 6.1 for $t > 0$ and

$$\lim_{t\to 0} u(x,t) = 0 \quad \forall x \in \mathbb{R}$$

but u is not continuous at 0.

6.2. The heat equation on an interval

Let $L, c > 0, f \in C^{0,\alpha}([0,L])$ such that $f(0) = f(L) = 0$. Consider the solution u to the PDE

$$\begin{cases} \partial_t u = c^2 \partial_{xx} u \quad x \in (0,L), t > 0 \\ u(x,0) = f(x) \\ u(0,t) = u(L,t) = 0 \end{cases} \tag{6.3}$$

First, notice that we may reduce to the case $c = L = 1$; indeed, if u solves the heat equation in this special case, then, $v(x,t) = u(Lx, L^2t/c^2)$ solves the problem

$$\begin{cases} \partial_t v = \partial_{xx} v \quad x \in (0,1), t > 0 \\ v(x,0) = f(Lx) \\ v(0,t) = v(1,t) = 0, \end{cases}$$

because $\partial_t v = L^2/c^2 \partial_t u = L^2 \partial_{xx} u = \partial_{xx} v$. Once we found v, u can be retrieved as

$$u(x,t) = v\left(\frac{x}{L}, \frac{c^2}{L^2}t\right).$$

Let us start constructing the solution via separation of variables: we can look for solutions of the form $v(x,t) = Z(x)W(t)$, for which the heat equation rewrites as

$$\begin{cases} Z(x)W'(t) = Z''(x)W(t) \\ Z(0)W(t) = Z(1)W(t) = 0. \end{cases}$$

Dividing both sides by $V(x)W(t)$ gives

$$\begin{cases} \frac{W'(t)}{W(t)} = \frac{Z''(x)}{Z(x)} = \lambda \\ Z(0)W(t) = Z(1)W(t) = 0. \end{cases}$$

This can be separated in two ODEs which we can easily solve:

$$\begin{cases} W'(t) = \lambda W(t) \\ Z''(x) = \lambda Z(x) \\ Z(0) = Z(1) = 0. \end{cases}$$

By Remark 4.1, for $\lambda = -(n\pi)^2$, the solutions to

$$\begin{cases} Z''(x) = \lambda Z(x) \\ Z(0) = Z(1) = 0 \end{cases}$$

are given by $z_n(x) = \sin(n\pi x)$. Notice that from the theory of ODEs we are restricting to the case $\lambda < 0$, because the solutions for $\lambda > 0$ are exponential and for $\lambda = 0$ are lines. On the other hand, the ODE

$$W'(t) = \lambda W(t)$$

has solutions $W_n(t) = e^{-(n\pi)^2 t}$. Overall, $\forall n \in \mathbb{N}$, we have that

$$v_n(x,t) = A_n \sin(\pi n x) e^{-(n\pi)^2 t}$$

is a solution, for a given constant $A_n \in \mathbb{R}$.

Now, recall that the heat equation is linear: therefore, if u, v are solutions, $\forall \alpha, \beta \in \mathbb{R}$ also $\alpha u + \beta v$ is a solution. Therefore, if we take

$$v(x,t) = \sum_{n=1}^{\infty} a_n \sin(\pi n x) e^{-(n\pi)^2 t}; \tag{6.4}$$

this is formally a solution to Equation 6.3. Notice also that $u(0,t) = u(1,t) = 0$; we want to prescribe that

$$v(x,0) = \sum_{n=1}^{\infty} a_n \sin(\pi n x) = f(x).$$

Let a_n be the coefficients of the Fourier series in sines only of f, namely

$$a_n := 2\int_0^1 f(x)\sin(n\pi x)dx.$$

We now state the theorem which validates the construction of solutions:

THEOREM 6.3. *Let $f \in C^2$, and let v be the one defined in (6.4). Then,*
(i) $v \in C^2((0,1)\times(0,\infty))$ and $\partial_t v = \partial_{xx} v$;
(ii) $\lim_{x\to 0} v(x,t) = \lim_{x\to 1} v(x,t) = 0$;
(iii) $f(x) = \lim_{t\to 0} v(x,t)$ (pointwise or uniformly).

PROOF. Under our hypothesis, $f \in L^1$. Therefore, $\sup_{n\in\mathbb{N}} |a_n| < +\infty$.

(i) Let

$$v_N(x,t) = \sum_{n=1}^{N} a_n \sin(\pi n x) e^{-(\pi n)^2 t} \in C^\infty((0,1)\times(0,\infty))$$

Then, by linearity

$$\partial_t v_N(x,t) = \sum_{n=1}^{N} a_n[-(\pi n)^2]\sin(n\pi x) e^{-(\pi n)^2 t} \in C^\infty((0,1)\times(0,\infty))$$

As $N \to +\infty$, the series converges locally uniformly to $\partial_t v$, becuase it is Cauchy:

$$\left| \sum_{n=1}^{N} a_n[-(\pi n)^2]\sin(n\pi x)e^{-(\pi n)^2 t} - \sum_{n=1}^{M} a_n[-(\pi n)^2]\sin(n\pi x)e^{-(\pi n)^2 t} \right| \le \sum_{n=M+1}^{N} C n^2 e^{-(\pi n)^2 t},$$

and $n^2 e^{-(\pi n)^2 t} \le n^{-2}$ for n large enough.

Similarily, we can prove that $\partial_x v_N \in C^\infty((0,1)\times(0,\infty))$ and it converges locally uniformly to

$$\sum_{n=1}^\infty a_n(\pi n)\cos(n\pi x)e^{-(\pi n)^2 t}$$

We can use the following Proposition 6.4 below, so we get that $v \in C^1$. We can now go on with higher order derivatives, obtaining $v \in C^\infty((0,1)\times(0,+\infty))$ and for m even

$$\partial_t^l \partial_x^m v(t,x) = \pm \sum_{n=1}^\infty a_n (n\pi)^m \sin(n\pi x) e^{-(n\pi)^2 t}(-n^2\pi^2)^l$$

and for m odd

$$\partial_t^l \partial_x^m v(t,x) = \pm \sum_{n=1}^\infty a_n (n\pi)^m \cos(n\pi x) e^{-(n\pi)^2 t}(-n^2\pi^2)^l.$$

Finally, note that v solves $\partial_t v = \partial_{xx} v$, because we compute both sides and use the fact that each piece solves the heat equation.

(ii) Note that $v(\cdot,t) \in C^1((0,1))$ with bounded derivative (possibly depending on t), therefore $v(\cdot,t)$ is continuous and $v(0,t) = 0 = v(1,t)$.

(iii) Since $f \in C^2([0,1])$, we know that $\sum_{n=1}^\infty |a_n| < \infty$ and $f(x) = \sum_{n=1}^\infty a_n \sin(n\pi x)$ for every $x \in [0,1]$. Let $\varepsilon > 0$ and take $N \in \mathbb{N}$ so large that $\sum_{n>N} |a_n| < \varepsilon$. Using the bound $|1 - e^{-s}| \leqslant s$ for $s \geqslant 0$, we get

$$\begin{aligned}|f(x) - v(x,t)| &\leqslant \sum_{n>N} |a_n \sin(n\pi x)| + \sum_{n=1}^N |a_n \sin(n\pi x)(1 - e^{-(n\pi)^2 t})| + \sum_{n>N} |a_n \sin(n\pi x) e^{-(n\pi)^2 t}| \\ &\leqslant 2\varepsilon + t\sum_{n=1}^N (n\pi)^2 |a_n|.\end{aligned}$$

From this we get $\limsup_{\substack{t\to 0\\ t>0}} \sup_{x\in[0,1]} |f(x) - v(x,t)| \leqslant 2\varepsilon$ for every $\varepsilon > 0$ which gives uniform convergence.

□

Proposition 6.4. *Let $\Omega \subseteq \mathbb{R}^d$ open, $\{u_n\} \subset C^1(\Omega;\mathbb{R})$ such that, locally uniformly,*

$$u_n \to u$$
$$\nabla u_n \to v$$

Then, $u \in C^1(\Omega,\mathbb{R})$ and $\nabla u = v$.

Proof.

$$\begin{aligned}u_n(x + he_i) - u_n(x) &= h\int_0^1 \partial_{e_i} u_n(x + she_i)ds \\ &= h\int_0^1 \nabla u_n(x + she_i)\cdot e_i ds \\ &= h\int_0^1 v(x + she_i)\cdot e_i ds\end{aligned}$$

Dividing by h and passing to the limit $h \to 0$,

$$\partial_{e_i} u(x) = v(x)\cdot e_i$$

□

6.3. The Laplace equation in a box

Given $L, M > 0$, we can consider the Laplace equation in the box $(0, L) \times (0, M)$:

$$\begin{cases} \Delta u = 0 \text{ in } (0, L) \times (0, M) \\ u(x, 0) = \alpha(x) \text{ and } u(x, M) = \beta(x) \\ u(0, y) = \gamma(y) \text{ and } u(L, y) = \delta(y). \end{cases} \tag{6.5}$$

Ignoring temporarily the boundary conditions and focussing on the equation, we can look for a solution with the method of separation of variables:

$$u(x, y) = \phi(x)\psi(y)$$

which gives

$$\phi''(x)\psi(y) = -\phi(x)\psi''(y)$$
$$\frac{\phi''(x)}{\phi(x)} = -\frac{\psi''(y)}{\psi(y)}$$

Following Remark 4.1 and proceeding as in section 4.1, we have that the only possibility is that the above functions of different variables as equal to a real constant $\lambda \in \mathbb{R}$. We then get the two ODEs $\phi''(x) = \lambda\phi(x)$ and $\psi''(y) = -\lambda\psi(y)$.

Now, if $\lambda < 0$ we get

$$\phi(x) = \alpha \sin(\sqrt{-\lambda}x) + \beta \cos(\sqrt{-\lambda}x);$$

if $\lambda = 0$, ϕ is an affine function; if $\lambda > 0$ we get

$$\phi(x) = \gamma e^{\sqrt{\lambda}x} + \delta e^{-\sqrt{\lambda}x},$$

that we can write ϕ as

$$\phi(x) = (\gamma + \delta) \cosh(\sqrt{\lambda}x) + (\gamma - \delta) \sinh(\sqrt{\lambda}x).$$

Overall, solutions will be of the form

$$\left(\alpha \sin(\sqrt{-\lambda}x) + \beta \cos(\sqrt{-\lambda}x)\right)\left(\gamma \cosh(\sqrt{-\lambda}y) + \delta \sinh(\sqrt{-\lambda}y)\right), \quad \text{for } \lambda < 0,$$

and

$$\left(\alpha \sinh(\sqrt{\lambda}x) + \beta \cosh(\sqrt{\lambda}x)\right)\left(\gamma \sin(\sqrt{\lambda}y) + \delta \cos(\sqrt{\lambda}y)\right), \quad \text{for } \lambda > 0.$$

We can now split this into two simpler problem, suppose we can solve

$$\begin{cases} \Delta v = 0 \\ v(x, 0) = \alpha(x) \text{ and } v(x, M) = \beta(x) \\ v(0, y) = 0 \text{ and } v(L, y) = 0 \end{cases}$$

as well as

$$\begin{cases} \Delta w = 0 \\ w(x, 0) = 0 \text{ and } w(x, M) = 0 \\ v(0, y) = \gamma(y) \text{ and } v(L, y) = \Delta(y) \end{cases}$$

Then u, the solution of the original problem is $u = v + w$.
Let's solve the first problem by separation of variables, writing $u = \phi(x)\psi(y)$, we get

$$\frac{\phi''(x)}{\phi(x)} = -\frac{\psi''(y)}{\psi(y)} = \lambda$$

and the initial condition implies $\phi(0) = \phi(L) = 0$, thus $\lambda = -(\frac{n\pi}{L})^2$, thus $\phi = \alpha_n \sin(\frac{n\pi}{L}x)$ and $\psi_n(x) = \xi_n \cosh(\frac{n\pi}{L}y) + \eta_n \sinh(\frac{n\pi}{L}y)$.
Thus, the general solution is

$$\sum_n \left[\xi_n \cosh\left(\frac{n\pi}{L}y\right) + \eta_n \sinh\left(\frac{n\pi}{L}y\right)\right] \sin\left(\frac{n\pi}{L}x\right)$$

Now we impose our boundary solutions $\alpha(x) = v(x,0) = \sum \xi_n \sin\left(\frac{n\pi}{L}x\right)$ and so we compute the fourier coefficients of α (in sines), which are given by

$$\xi_n = \frac{2}{L}\int_0^L \alpha(x) \sin\left(\frac{n\pi}{L}x\right) dx$$

Similarly, we find

$$\eta_n = \frac{1}{\sinh\left(\frac{2\pi}{L}M\right)}\left[\frac{2}{L}\int_0^L \beta(x)\sin\left(\frac{n\pi}{L}x\right) dx - \xi_n \cosh\left(\frac{n\pi}{L}M\right)\right].$$

Proposition 6.5. *Let $\alpha, \beta \in L^1(0,L)$ and $(a_n)_{n\in\mathbb{N}}, (b_n)_{n\in\mathbb{N}}$ be their respective Fourier coefficients in sine, namely*

$$a_n = \frac{2}{L}\int_0^L \sin\left(\frac{n\pi}{L}y\right)\alpha(y)dy, \qquad b_n = \frac{2}{L}\int_0^L \sin\left(\frac{n\pi}{L}y\right)\beta(y)dy.$$

Let $w : (0,L)\times(0,M) \to \mathbb{R}$ be defined by

$$w(x,y) = \sum_{n=1}^{+\infty}\left(a_n \frac{\sinh(\frac{n\pi}{L}(M-y))}{\sinh(\frac{n\pi}{L}M)} + b_n \frac{\sinh(\frac{n\pi}{L}y)}{\sinh(\frac{n\pi}{L}M)}\right)\sin\left(\frac{n\pi}{L}x\right)$$

then $w \in C^\infty((0,L)\times(0,M))$ and satisfies the Laplace equation, i.e. $\Delta w = 0$ in $(0,L)\times(0,M)$. Also if $\alpha, \beta \in C^3([0,L])$ then $\lim_{y\to 0} w(x,y) = \alpha(x)$ and $\lim_{y\to M} w(x,y) = \beta(x)$ uniformly in $x \in [0,L]$.

Proof. Without loss of generality (up to a rescaling), we consider $L = 1$. We start with the following observation for $y \in (0,M)$ and $n \in \mathbb{N}$

$$0 \leqslant \frac{\sinh(n\pi(M-y))}{\sinh(n\pi M)} = \frac{e^{n\pi(M-y)} - e^{-n\pi(M-y)}}{e^{n\pi M} - e^{-n\pi M}} = e^{-n\pi y}\left(\frac{1 - e^{-2n\pi(M-y)}}{1 - e^{-2n\pi M}}\right) \leq Ce^{-n\pi y},$$

where $C = \frac{2}{1-e^{-2\pi M}}$.
Similarly, for $y \in (0,M)$ and $n \in \mathbb{N}$ we have

$$\left|\frac{\sinh(n\pi y)}{\sinh(n\pi M)}\right| \leq Ce^{-n\pi(M-y)},$$

$$\left|\frac{\cosh(n\pi(M-y))}{\sinh(n\pi M)}\right| \leq Ce^{-n\pi y},$$

$$\left|\frac{\cosh(n\pi y)}{\sinh(n\pi M)}\right| \leq Ce^{-n\pi(M-y)}.$$

Since $\alpha, \beta \in L^1(0,1)$, we have the following trivial bound for their Fourier coefficients

$$\sup_{n\in\mathbb{N}}|a_n|, \sup_{n\in\mathbb{N}}|b_n| \leq 2(\|\alpha\|_{L^1(0,1)} + \|\beta\|_{L^1(0,1)}) =: D.$$

Using these estimates, we are able to prove that $w \in C^\infty((0,1)\times(0,M))$. We prove that w is C^1 and the higher regularity is proved similarly.

To this end, let $(S_n)_{n\in\mathbb{N}}$ be the partial sums of the series defining w, i.e.

$$S_n(x,y)=\sum_{k=1}^{n}\left(a_k\frac{\sinh(k\pi(M-y))}{\sinh(k\pi)M}+b_k\frac{\sinh(k\pi y)}{\sinh(k\pi M)}\right)\sin(k\pi x).$$

Notice that S_n is C^1 for every n with partial derivatives given by

$$\frac{\partial S_n}{\partial x}(x,y)=\sum_{k=1}^{n}k\pi\left(a_k\frac{\sinh(k\pi(M-y))}{\sinh(k\pi)M}+b_k\frac{\sinh(k\pi y)}{\sinh(k\pi M)}\right)\cos(k\pi x)$$

$$\frac{\partial S_n}{\partial y}(x,y)=\sum_{k=1}^{n}k\pi\left(-a_k\frac{\cosh(k\pi(M-y))}{\sinh(k\pi M)}+b_k\frac{\cosh(k\pi y)}{\sinh(k\pi M)}\right)\sin(k\pi x)$$

For any compact $K\subset(0,L)\times(0,M)$ we have

$$\sum_{k=1}^{n}\sup_{(x,y)\in K}\left|k\pi\left(a_k\frac{\sinh(k\pi(M-y))}{\sinh(k\pi)M}+b_k\frac{\sinh(k\pi y)}{\sinh(k\pi M)}\right)\cos(k\pi x)\right|\leq\sum_{k=1}^{\infty}2D\pi ke^{-k\pi d}<+\infty,$$

where $d=\inf_{(x,y)\in K}\max\{y,M-y\}$.
We get a similar inequality for $\frac{\partial S_n}{\partial y}$ and for the sequence S_n as well. We can thus use the Weierstrass M-test to conclude that S_n converges locally uniformly to w (which is in particular well defined) and the partial derivatives also converge locally uniformly to their respective series (see the partial sums above) and so $w\in C^1((0,1)\times(0,M))$, with partial derivatives given by

$$\frac{\partial w}{\partial x}(x,y)=\sum_{n=1}^{\infty}n\pi\left(a_n\frac{\sinh(n\pi(M-y))}{\sinh(n\pi M)}+b_n\frac{\sinh(n\pi y)}{\sinh(n\pi M)}\right)\cos(n\pi x),$$

$$\frac{\partial w}{\partial y}(x,y)=\sum_{n=1}^{\infty}n\pi\left(-a_n\frac{\cosh(n\pi(M-y))}{\sinh(n\pi M)}+b_n\frac{\cosh(n\pi y)}{\sinh(n\pi M)}\right)\sin(n\pi x).$$

Formulas for higher derivatives can be derived without much difficulty, in particular, we have the following

$$\frac{\partial^2 w}{\partial x^2}(x,y)=\sum_{n=1}^{\infty}-n^2\pi^2\left(a_n\frac{\sinh(n\pi(M-y))}{\sinh(n\pi M)}+b_n\frac{\sinh(n\pi y)}{\sinh(n\pi M)}\right)\sin(n\pi x),$$

$$\frac{\partial^2 w}{\partial y^2}(x,y)=\sum_{n=1}^{\infty}n^2\pi^2\left(a_n\frac{\sinh(n\pi(M-y))}{\sinh(n\pi M)}+b_n\frac{\sinh(n\pi y)}{\sinh(n\pi M)}\right)\sin(n\pi x).$$

From this we are able to see that w is harmonic in $(0,L)\times(0,M)$.
Now assume further that $\alpha,\beta\in C^3([0,L])$ with $\alpha(0)=\alpha(L)=\beta(0)=\beta(L)=0$, we first prove that $\lim_{\substack{y\to0\\y>0}}w(x,y)=\alpha(x)$ uniformly in $x\in[0,L]$. Since $\alpha\in C^3([0,L])$ with $\alpha(0)=\alpha(L)=0$, the Fourier series in sine of α converges uniformly to α and its Fourier coefficients satisfy $|a_n|\leqslant\frac{A}{n^3}$ for some $A>0$.
Let $\varepsilon>0$ and take $N\in\mathbb{N}$ so large that $\sum_{n>N}|a_n|\leqslant\varepsilon$, then for $(x,y)\in(0,L)\times(0,M)$, we have

$$\left|\sum_{n=1}^{\infty}a_n\left(\frac{\sinh(n\pi(M-y))}{\sinh(n\pi M)}-1\right)\sin(n\pi x)\right|\leqslant\sum_{n=1}^{N}\left|a_n\left(\frac{\sinh(n\pi(M-y))}{\sinh(n\pi M)}-1\right)\right|+\varepsilon.$$

This bound implies that

$$\lim_{\substack{y\to0\\y>0}}\sum_{n=1}^{\infty}a_n\frac{\sinh(n\pi(M-y))}{\sinh(n\pi M)}\sin(n\pi x)=\alpha(x)$$

uniformly in $x \in [0, 1]$.
To conclude it suffices to prove that

$$\lim_{\substack{y \to 0 \\ y > 0}} \sum_{n=1}^{\infty} b_n \frac{\sinh(n\pi y)}{\sinh(n\pi M)} \sin(n\pi x) = 0$$

uniformly in $x \in [0, L]$.
To see this, let $\varepsilon > 0$ and take $N \in \mathbb{N}$ so large that $\sum_{n>N} |b_n| \leqslant \varepsilon$ (this is possible since $\beta \in C^3([0, L])$), then we have

$$\sum_{n=1}^{\infty} \left| b_n \frac{\sinh(n\pi y)}{\sinh(n\pi M)} \sin(n\pi x) \right| \leqslant \varepsilon + \sum_{n=1}^{N} \left| b_n \frac{\sinh(n\pi y)}{\sinh(n\pi M)} \right|.$$

The other condition, namely $\lim_{\substack{y \to M \\ y < M}} w(x, y) = \beta(x)$ uniformly in $x \in [0, 1]$ is proved in the same way. $\square$

6.4. The Laplace equation in a disc

We complete in this section, with Fourier analysis at hand, the analysis of the Laplace equation in a disc (4.5), whose formal solution was found in Section 4.1 and more precisely in (4.12)

Proposition 6.6. *Let $f \in L^1(0, 2\pi)$ and consider the Laplace equation on the unit disc (written in polar coordinates) with Dirichlet boundary conditions, i.e.*

$$\begin{cases} r^2 \partial_{rr} v + r \partial_r v + \partial_{\theta\theta} v = 0, & (r, \theta) \in (0, 1) \times (0, 2\pi) \\ v(1, \theta) = f(\theta), & \theta \in [0, 2\pi] \end{cases}$$

Let $(a_n)_{n\in\mathbb{N}}, (b_n)_{n\in\mathbb{N}}$ be the real Fourier coefficients of f, namely

$$a_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(nx) dx, \qquad b_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin(nx) dx.$$

Recall the formal solution in polar coordinates given by

$$v(r, \theta) = \frac{a_0}{2} + \sum_{n=1}^{\infty} r^n (a_n \cos(n\theta) + b_n \sin(n\theta))$$

then $v \in C^\infty((0, 1) \times (0, 2\pi))$.
Also if $f \in C^2([0, 2\pi])$ (with $f(0) = f(2\pi)$) then $\lim_{\substack{r \to 1 \\ r < 1}} v(r, \theta) = f(\theta)$ uniformly in $\theta \in [0, 2\pi]$.

Proof. Since $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ are bounded sequences, the Weierstrass M-test gives local uniform convergence in $(0, 1) \times (0, 2\pi)$ and so v is well defined and continuous on $(0, 1) \times (0, 2\pi)$. For any $k \in \mathbb{N}$, the series $\sum_{n=1}^{\infty} n^k r^n$ converges locally uniformly for $r \in (0, 1)$. Since for any $k, j \in \mathbb{N}$ we have

$$\sum_{n=1}^{\infty} \left| \frac{\partial^{k+j}}{\partial r^k \partial \theta^j} r^n (a_n \cos(n\theta) + b_n \sin(n\theta)) \right| \leqslant C_{k,j} \sum_{n=1}^{\infty} n^{k+j} r^n,$$

for some constant $C_{k,j}$ depending only in k and j. From this we deduce that the series of the derivatives converge locally uniformly on $(0, 1) \times (0, 2\pi)$ and by Proposition 6.4 we prove inductively that $v \in C^\infty((0, 1) \times (0, 2\pi))$ and the derivatives of v are computed by differentiating every term in

the sum.
In particular, we get the following

$$\begin{aligned} r^2\partial_{rr}v(r,\theta) &= \sum_{n=1}^{\infty} n(n-1)r^n(a_n\cos(n\theta)+b_n\sin(n\theta)),\\ r\partial_r v(r,\theta) &= \sum_{n=1}^{\infty} nr^n(a_n\cos(n\theta)+b_n\sin(n\theta)),\\ \partial_{\theta\theta}v(r,\theta) &= \sum_{n=1}^{\infty} -n^2r^n(a_n\cos(n\theta)+b_n\sin(n\theta)), \end{aligned}$$

which proves that v satisfies the Laplace equation on the unit disc.
Now assume that $f \in C^2([0,2\pi])$ with $f(0)=f(2\pi)$, then f is equal to its real Fourier series by Dirichlet's theorem and its Fourier coefficients satisfy $|a_n|,|b_n| \leqslant \frac{A}{n^2}$ for some $A>0$ thanks to Remark 4.17.
Let $\varepsilon>0$ and take $N\in\mathbb{N}$ so large that $\sum_{n>N}|a_n|+|b_n|<\varepsilon$. Then for $r\in(0,1)$ and $\theta\in[0,2\pi]$, we have

$$\begin{aligned} |v(r,\theta)-f(\theta)| &\leqslant \sum_{n=1}^{N}(1-r^n)|a_n\cos(n\theta)+b_n\sin(n\theta)|+2\varepsilon\\ &\leqslant 2\varepsilon+\sum_{n=1}^{N}(1-r^n)(|a_n|+|b_n|) \end{aligned}$$

which concludes the proof. □

6.5. The wave equation

We will now derive the wave equation, that can describe for example the behavior of a vibrating rope, whose position in the vertical direction is denoted by $y=u(x,t)$.

The rope can be modelled as N masses with $x-$coordinate $x_n=nL/N$ and y-coordinate y_n to be determined. Denote by $h\coloneqq L/N$ the distance between consecutive particles, and the mass above x_n as $m_n\coloneqq \rho hL/N$, where ρ is the density of the rope.

Now, make the following assumptions:

- the mass above x_n moves only vertically;
- the mass moves according to Newton's law and the forces that act on it are generated by the neighbors and proportional to $(y_n-y_{n-1})/h$;
- each mass moves by Newton's law;
- forces are generated by neighbors $\sim (y_n-y_{n-1})/h$.

We now will write the equation solved by y_n and by letting $N\to+\infty$ (or equivalently $h\to 0$), find a PDE solved by u.

By applying Newton's law, we get that:

$$\rho h y_n'' = \frac{1}{h}[\underbrace{y_{n+1}-y_n}_{\geq 0 \text{ if } y_{n+1}>y_n} - \underbrace{(y_n-y_{n-1})}_{\geq 0 \text{ if } y_n>y_{n-1}}]$$

Dividing by h, we get:

$$\rho y_n'' = \frac{1}{h^2}[u(x_{n+1},t)-2u(x_n,t)+u(x_{n-1},t)]$$

or, recalling that $x_{n\pm 1} = x_n \pm h$,

$$\rho\partial_{tt}u(x_n,t) = \frac{1}{h^2}[u(x_n+h,t) - 2u(x_n,t) + u(x_n-h,t)]$$

Now, fix x and let $n \to +\infty$:

$$\rho\partial_{tt}u(x_n,t) \to \rho\partial_{tt}u(x,t)$$

and

$$\frac{1}{h^2}[u(x_n+h,t) - 2u(x_n,t) + u(x_n-h,t)] \to \partial_{xx}u(x,t)$$

We then retrieve the wave equation

$$\partial_{tt}u(x,t) = \frac{1}{\rho}\partial_{xx}u(x,t)$$

Why is this called wave equation?

Let $f \in C_c^2(\mathbb{R})$. Then, $f(x \pm ct)$ solves

$$\partial_{tt}u = c^2\partial_{xx}u$$

Indeed, applying the chain rule for derivatives,

$$\begin{aligned}\partial_{tt}u &= f''(x-ct)(-c^2)\\ \partial_{xx}u &= f''(x-ct).\end{aligned}$$

Remark 6.8. One can easily prove that this holds in $\mathbb{R}^d$ as well, for $c \in \mathbb{R}^d$.

Remark 6.9. Let us consider the following rescaling: let $a, b > 0$, and let U be such that

$$U(ax,bx) = u(x,t)$$

Then, U solves

$$\partial_{tt}U = \frac{ca^2}{b^2}\partial_{xx}U$$

In particular, choosing $a = \pi/L$ and $b = \sqrt{c}\pi/L$, we can reduce to $c = 1$ and $L = \pi$, because

$$U(x,t) = u(x/a,t/b) \quad \Rightarrow \quad \partial_{xx}U = \partial_{xx}u(x/a,t/b)a^{-2}, \quad \partial_{tt}U = \partial_{tt}u(x/a,t/b)b^{-2}$$

Hence, $b^2\partial_{tt}U = \partial_{tt}u = c\partial_{xx}u = ca^2\partial_{xx}U$.

6.5.1. The wave equation in a bounded interval.

Let $L, c > 0$, $f, g : [0,L] \to \mathbb{R}$ such that $f(0) = f(L) = 0$, $g(0) = g(L) = 0$. Consider:

$$\begin{cases}\partial_{tt}u = c^2\partial_{xx}u & x \in (0,L), t \in (0,+\infty)\\ u(0,t) = u(L,t) = 0 & \forall t \in (0,+\infty)\\ u(x,0) = f(x) & \forall x \in (0,L)\\ \partial_t u(x,0) = g(x) & \forall x \in (0,L)\end{cases} \tag{6.6}$$

Now, we find all $u : [0,L] \times [0,+\infty) \to \mathbb{R}$ such that $u(x,t) = v(x)w(t)$ by separation of variables, ignoring temporarily the initial condition, and we will write the formal solution to Equation 6.6.

By separation of variables,

$$v'(x)w''(t) = v''(x)w(t)$$

therefore

$$\frac{v''(x)}{v(x)} = \frac{w''(t)}{w(t)} = -\lambda$$

The wave equation is then equivalent to the ODEs

$$\begin{cases} \frac{v''(x)}{v(x)} = -\lambda \\ v(0) = v(L) = 0, \end{cases} \qquad \frac{w''(t)}{c^2 w(t)} = -\lambda$$

By Remark 4.1, the general solution to the ODE for $v(x)$ is

$$v(x) = \alpha \sin(\sqrt{\lambda} x) + \beta \cos(\sqrt{\lambda} x)$$

from the boundary condition, to respect the periodicity, we retrieve

$$\lambda = \left(\frac{n\pi}{L}\right)^2$$

hence

$$v_n(x) = \sin\left(\frac{n\pi}{L} x\right)$$

Since w solves the same equation (but with $-\lambda c^2$), we can define

$$w_n(t) = \alpha_n \cos\left(\frac{n\pi c}{L} t\right) + \beta_n \sin\left(\frac{n\pi c}{L} t\right).$$

A formal solution is given by the (infinite) linear combination of the solutions:

$$u(x,t) = \sum_{n=1}^{\infty} \left[\alpha_n \cos\left(\frac{n\pi c}{L} t\right) + \beta_n \sin\left(\frac{n\pi c}{L} t\right)\right] \sin\left(\frac{n\pi}{L} x\right)$$

Note that there is no guarantee that all solutions are of this form, but we can try to impose the boundary condition and see if we are able to find one:

$$f(x) = u(x,0) = \sum_{n=1}^{\infty} \alpha_n \sin\left(\frac{n\pi}{L} x\right)$$

This is the Fourier expansion only in sines obtained by reflecting $f(x)$ oddly. The unique choice for α_n is therefore:

$$\alpha_n = \frac{2}{L} \int_0^L f(x) \sin\left(\frac{n\pi}{L} x\right) dx =: a_n$$

Similarily, by differentiating term by term, we obtain

$$g(x) = \partial_t u(x,0) = \sum_{n=1}^{+\infty} \frac{n\pi c}{L} \beta_n \sin\left(\frac{n\pi}{L} x\right)$$

hence

$$\beta_n = \frac{2}{n\pi c} \int_0^L g(x) \sin\left(\frac{n\pi}{L} x\right) dx =: \frac{L}{n\pi c} b_n.$$

Proposition 6.7. *Assume that $f \in C^4([0,L])$ and $g \in C^3([0,L])$ are $L-$periodic (note that for g we need one derivative less, because it represents the time derivative of u), and such that $f(0) = f''(0) = f(L) = f''(L) = 0$ and $g(0) = g''(0) = g(L) = g''(L) = 0$. Then, $u \in C^2((0,L) \times [0,+\infty))$ and*

$$f(x) = \lim_{t\to 0} u(x,t), \qquad g(x) = \lim_{t\to 0} \partial_t u(x,t) \quad \textit{uniformly in } x$$

PROOF. By Proposition 4.17 the regularity of f and g implies that

$$|a_n| \le \frac{C}{n^4}, \quad \text{and} \quad |b_n| \le \frac{C}{n^3}$$

Formally, we would have that

$$\partial_x u(x,t) = \sum_{n=1}^{\infty} \left(\frac{n\pi}{L}\right)^2 \left[a_n \cos\left(\frac{n\pi c}{L}t\right) + \frac{L}{n\pi c} b_n \sin\left(\frac{n\pi c}{L}t\right)\right] \cos\left(\frac{n\pi}{L}x\right) \tag{6.7}$$

Note that the series converges absolutely:

$$|\partial_x u| \leq \sum_{n=1}^{\infty} \left(\frac{n\pi}{L}\right)^2 \left(\frac{C}{n^4} + \frac{L}{n\pi c}\frac{C}{n^3}\right) < +\infty$$

To proceed rigorously, we can define

$$u_N(x,t) = \sum_{n=1}^{N} \left[a_n \cos\left(\frac{n\pi c}{L}t\right) + b_n \sin\left(\frac{n\pi c}{L}t\right)\right] \sin\left(\frac{n\pi}{L}x\right),$$

so that

$$\partial_x u_N(x,t) = \sum_{n=1}^{N} \left(\frac{n\pi}{L}\right)^2 \left[a_n \cos\left(\frac{n\pi c}{L}t\right) + \frac{L}{n\pi c} b_n \sin\left(\frac{n\pi c}{L}t\right)\right] \cos\left(\frac{n\pi}{L}x\right).$$

Then, we show that both sequences are Cauchy in C^0; let $M < N$, and consider

$$|\partial_x u_N(x,t) - \partial_x u_M(x,t)| \leq \left|\sum_{n=M+1}^{N} \left(\frac{n\pi}{L}\right)^2 \left[a_n \cos\left(\frac{n\pi c}{L}t\right) + \frac{L}{n\pi c} b_n \sin\left(\frac{n\pi c}{L}t\right)\right] \cos\left(\frac{n\pi}{L}x\right)\right|$$

$$\leq \sum_{n=M+1}^{\infty} \left(\frac{n\pi}{L}\right)^2 \left(\frac{C}{n^4} + \frac{L}{n\pi c}\frac{C}{n^3}\right)$$

which goes to 0 as $M \to \infty$, because the series converges. Hence $\partial_x u_N \to v$ and similarly $u_N \to u$ uniformly. We apply Proposition 6.4 to claim that $v = \partial_x u$ and justify (6.7). Now, observe that the formal partial derivative in t of u is given by

$$\partial_t u(x,t) = \sum_{n=1}^{\infty} \left[b_n \cos\left(\frac{n\pi c}{L}t\right) - \frac{n\pi c}{L} a_n \sin\left(\frac{n\pi c}{L}t\right)\right] \sin\left(\frac{n\pi}{L}x\right).$$

This series converges uniformly, hence $|\partial_t u(\cdot,t)| \leq C$ for all $x \in (0,L)$. Therefore, $u(\cdot,t)$ is Lipschitz and we get $u(0,t) = \lim_{x\to 0} u(x,t) = 0$, $u(L,t) = \lim_{x\to L} u(x,t) = 0$.

Now, we want to prove that $f(x) = \lim_{t\to 0} u(x,t)$ uniformly. Recall that by Proposition 4.17,

$$f(x) = \sum_{n=1}^{\infty} a_n \sin\left(\frac{n\pi}{L}x\right), \qquad \text{with } |a_n| \leq \frac{C}{n^4}.$$

Then

$$|u(x,t) - f(x)| \leq \sum_{n=1}^{\infty} \left|a_n \left(\cos\left(\frac{n\pi}{L}t\right) - 1\right) + \frac{L}{n\pi c} b_n \sin\left(\frac{n\pi}{L}t\right)\right| \underbrace{\left|\sin\left(\frac{n\pi}{L}x\right)\right|}_{\leq 1} \tag{6.8}$$

$$\leq \sum_{n=1}^{\infty} \left[\frac{C}{n^4}\left|\cos\left(\frac{n\pi}{L}t\right) - 1\right| + \frac{LC}{n^4\pi c} \sin\left(\frac{n\pi}{L}t\right)\right] \tag{6.9}$$

$$\leq \sum_{n=1}^{\infty} \frac{Cn^2t^2}{n^4} + \frac{CLnt}{n^4\pi c} \tag{6.10}$$

$$\leq t\sum_{n=1}^{\infty} \left(\frac{Ct}{n^2} + \frac{CL}{n^3\pi c}\right), \tag{6.11}$$

which goes to 0 as $t \to 0$ because the series in (6.11) converges.

To prove $g(x) = \lim_{t\to 0} \partial_t u(x,t)$, we apply again Proposition 4.17 to g; its Fourier Series converges uniformly to g on $[0,L]$. Then, we take $\varepsilon > 0$ and $N \in \mathbb{N}$ so large that

$$\sum_{n>N} |b_n| + \frac{n\pi c}{L}|a_n| < \varepsilon$$

which is possible since $na_n, b_n = O(n^{-3})$ as $n \to \infty$.
From this, for any $(x,t) \in [0,L] \times (0,\infty)$

$$\begin{aligned}\left|\sum_{n=1}^{\infty}\left(b_n \cos\left(\frac{n\pi c}{L}t\right)\sin\left(\frac{n\pi}{L}x\right)\right) - g(x)\right| &= \left|\sum_{n=1}^{\infty} b_n\left(\cos\left(\frac{n\pi c}{L}t\right) - 1\right)\sin\left(\frac{n\pi}{L}x\right)\right| \\ &\leq \sum_{n=1}^{N}\left|b_n\left(\cos\left(\frac{n\pi c}{L}t\right)-1\right)\right| + 2\sum_{n>N}|b_n|.\end{aligned}$$

We also have

$$\left|\sum_{n=1}^{\infty}\frac{n\pi c}{L}a_n \sin\left(\frac{n\pi c}{L}t\right)\sin\left(\frac{n\pi}{L}x\right)\right| \leq \sum_{n=1}^{N}\frac{n\pi c}{L}\left|a_n \sin\left(\frac{n\pi c}{L}t\right)\right| + \sum_{n>N}\frac{n\pi c}{L}|a_n|.$$

From these two inequalities, we deduce that for $(x,t) \in [0,L] \times (0,\infty)$, we have

$$|\partial_t u(x,t) - g(x)| \leq \sum_{n=1}^{N}\left|b_n\left(\cos\left(\frac{n\pi c}{L}t\right)-1\right)\right| + \sum_{n=1}^{N}\frac{n\pi c}{L}\left|a_n \sin\left(\frac{n\pi c}{L}t\right)\right| + 3\varepsilon$$

which in turn implies

$$\limsup_{t\to 0^+}\sup_{x\in[0,L]} |\partial_t u(x,t) - g(x)| \leqslant 3\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, we get the result.

□

Remark 6.10. A similar result to the one on $\partial_x u$ could be obtained even for the second derivative with this method, but not for the third derivative. Another question that we can investigate is whether $u \in C^\infty((0,1)\times(0,L))$. To fix ideas, take $L = 5$. In general, this is not true: for example, take $f \in C^4 \setminus C^5$, $g = \partial_t|_{t=0} f(x-ct) = -cf'(x)$, $c = 1$, and $L = 5$. Then, we choose f such that $\operatorname{supp}(f) \subset (2,3)$, because, as we will argue in Section 6.6, we want to avoid issues at the boundaries of the domain. Note that $u(x,t) = f(x-t)$ is a solution to (6.6) for $t < 2$ and it has the same regularity of f, hence it is not C^∞.

6.6. D'Alembert's approach to the wave equation

Lemma 6.8. *Let $F, G : \mathbb{R} \to \mathbb{R}$ be two C^2 functions then $u : \mathbb{R}^2 \to \mathbb{R}$ defined by*

$$u(x,t) = F(x+t) + G(x-t)$$

is a solution to the wave equation, i.e. $\partial_{tt} u = \partial_{xx} u$.

PROOF. $\partial_{tt}u(x,t) = F''(x+t) + G''(x-t)$ and $\partial_{xx}u(x,t) = F''(x+t) + G''(x-t)$. □

We are now interested in finding F and G compatible with some boundary conditions, more precisely we have the following proposition.

Proposition 6.9 (D'Alembert's formula)**.** *Let $f \in C^2([0,\pi])$ and $g \in C^1([0,\pi])$ satisfying $f(0) = f(\pi) = 0$, $f''(0) = f''(\pi) = 0$ and $g(0) = g(\pi) = 0$, then the solution to following boundary value problem*

$$\begin{cases} \partial_{tt}u(x,t) = \partial_{xx}u(x,t), & (x,t) \in (0,\pi) \times (0,\infty) \\ u(x,0) = f(x), & x \in [0,\pi] \\ u(0,t) = u(\pi,t) = 0, & t \geqslant 0 \\ \partial_t u(x,0) = g(x), & x \in [0,\pi] \end{cases}$$

is given by

$$u(x,t) = \frac{1}{2}(f(x+t) - f(x-t)) + \int_{x-t}^{x+t} g(s)ds.$$

Proof. By extending f and g oddly on $[-\pi,\pi]$ and then 2π-periodically on $\mathbb{R}$ we may assume that $f \in C^2(\mathbb{R})$ and $g \in C^1(\mathbb{R})$ are odd 2π-periodic functions.
In view of the Lemma 6.8, we are looking for the solution of the form $u(t,x) = F(x+t) + G(x-t)$ where F and G are C^2 functions. Our goal is thus to find F and G in terms of the boundary conditions.
F and G need to satisfy the following system

$$\begin{cases} F(x) + G(x) = f(x), & x \in [0,\pi] \\ F'(x) - G'(x) = g(x), & x \in [0,\pi] \end{cases}$$

From this we deduce $F'(x) + G'(x) = f'(x)$ which gives

$$2F'(x) = f'(x) + g(x),$$

from which we get

$$F(x) = \frac{1}{2}\left(f(x) + \int_0^x g(s)ds\right) + c.$$

Since $G(x) = f(x) - F(x)$, we get

$$G(x) = \frac{1}{2}\left(f(x) - \int_0^x g(s)ds\right) - c.$$

Notice that these formulae give $F'(x) - G'(x) = g(x)$.
So the solution to the PDE is

$$u(x,t) = \frac{1}{2}(f(x+t) + f(x-t)) + \frac{1}{2}\int_{x-t}^{x+t} g(s)ds.$$

From this formula, we see that $u \in C^2((0,\pi) \times (0,\infty)) \cap C^0([0,\pi] \times [0,\infty))$ satisfies the wave equation on the interval $[0,\pi]$, $u(x,0) = f(x)$ and $\partial_t u(x,0) = g(x)$ for all $x \in [0,\pi]$.
We also have

$$u(0,t) = \frac{1}{2}(f(t) + f(-t)) + \frac{1}{2}\int_{-t}^{t} g(s)ds = 0$$

since f, g are odd.
To prove that $u(t,\pi) = 0$, notice that $f(\pi+t) + f(\pi-t) = f(\pi+t) + f(-\pi-t) = 0$ since f is 2π-periodic and odd. Similarly $g(\pi-t) = g(-\pi-t) = -g(\pi+t)$, which gives

$$\int_{\pi-t}^{\pi+t} g(s)ds = \int_{-t}^{t} g(\pi+s)ds = 0$$

where the last equality follows from the fact that $s \mapsto g(\pi+s)$ is an odd function. ☐

APPENDIX A

Complements on measure theory

We follow the presentation of [Fol99; Sch15]. For a more in-depth introduction to the topic, we refer to [Fol99, Chapter 1].

A.1. Introduction

The construction of the Lebesgue measure m on $\mathbb{R}^d$ can be seen as a particular instance of Carathéodory's construction of measures which in fact applies to a much more general setting of a measurable space $(\Omega, \mathcal{A})$. It allows to characterise measures μ on $(\Omega, \mathcal{A})$ uniquely in terms of their values on a *suitable* family of sets $\mathcal{G}$ generating the σ-algebra (meaning that $\sigma(\mathcal{G}) = \mathcal{A}$) and, on the other hand, to construct measures only from their values on the particular family $\mathcal{G}$ (such a map will be called a pre-measure). As $\mathcal{G}$ is *potentially* much smaller than the full σ-algebra $\mathcal{A}$, this is a useful tool to construct non-trivial measures, which, in general, is a quite difficult task.

In Appendix A.2 we present the constructive part (usually referred to as "Carathéodory's construction"), that is we show under which conditions on the pre-measure and the family $\mathcal{G}$ there exists an extension to a full measure μ on $(\Omega, \mathcal{A})$. In the subsequent Appendix A.3, we discuss the uniqueness of such extension on the basis of Dynkin's theorem. As an application of this technique, we show in Appendix A.4 how Carathéodory's construction can be used to build, from a given cumulative distribution function $F : \mathbb{R}^d \to [0, 1]$ (see Definition A.4 below), a unique probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ which obeys

$$\mathbb{P}\left(\prod_{i=1}^{d}(-\infty, x_i]\right) = F(x_1, \ldots, x_d) \qquad \text{for all } x = (x_1, \ldots, x_d) \in \mathbb{R}^d .$$

Notation. In these notes, we use the following notational conventions. We will denote

- $(\Omega, \mathcal{A})$ a measurable space, meaning that Ω is any set and $\mathcal{A}$ is a σ-algebra on Ω .
- the Lebesgue measure on $\mathbb{R}^d$ by m ,
- the Borel-σ-algebra on $\mathbb{R}^d$ by $\mathcal{B}(\mathbb{R}^d)$,
- the σ-algebra of all Lebesgue measurable sets by $\mathcal{M}(\mathbb{R}^d)$,
- for a collection of sets $\mathcal{G} \subseteq \mathcal{P}(\Omega)$, we denote by $\sigma(\mathcal{G})$ the smallest σ-algebra containing $\mathcal{G}$, that is

  $$\sigma(\mathcal{G}) \coloneqq \bigcap_{\mathcal{A}\,\sigma\text{-algebra}\,:\mathcal{G}\subseteq\mathcal{A}} \mathcal{A} .$$

  Since any intersection of σ-algebras is a σ-algebra and since for any $\mathcal{G}$, $\mathcal{P}(\Omega)$ is a σ-algebra containing $\mathcal{G}$, it is straightforward to show that $\sigma(\mathcal{G})$ is well-defined (we also refer to [Aru21]). We also say that $\mathcal{G}$ generates $\sigma(\mathcal{G})$.

A.2. Existence

Definition (algebra). Let Ω a set. We call a collection of sets $\mathcal{G} \subseteq \mathcal{P}(\Omega)$ an algebra on Ω if the following conditions hold:

(i) $\Omega \in \mathcal{G}$.
(ii) $A \in \mathcal{G} \implies A^c \in \mathcal{G}$.

(iii) $A, B \in \mathcal{G} \implies A \cup B \in \mathcal{G}$.

Remark A.1 (Properties of an algebra)**.** It follows that from the definition that an algebra is stable under finite unions and also under finite intersections as $A \cap B = (A^c \cup B^c)^c$. Hence the only difference between an algebra and a σ-algebra is that a σ-algebra is also stable under countable unions.

Definition (measure). Let $(\Omega, \mathcal{A})$ be a measurable space. A map $\mu : \mathcal{A} \to [0, +\infty]$ is a measure on $(\Omega, \mathcal{A})$, if it satisfies the following conditions:

(i) $\mu(\varnothing) = 0$.
(ii) $\mu(\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n=1}^{+\infty} \mu(A_n)$ for all countable families of pairwise disjoint sets $\{A_n\}_{n\in\mathbb{N}}$ in $\mathcal{A}$.

The following theorem shows how to construct a full measure μ on a measurable space $(\Omega, \mathcal{A})$ starting only from a pre-measure μ_0 defined on an algebra $\mathcal{G}$ which generates $\mathcal{A}$. We will see that such pre-measures are easier to construct as an algebra $\mathcal{G}$ which generates a σ-algebra $\mathcal{A}$ can be much smaller than $\mathcal{A}$ itself and hence, when compared to a measure, the axioms of a pre-measure have to verified only on a smaller family of sets.

Theorem A.1 (Carathéodory's extension Theorem). *Let $(\Omega, \mathcal{A})$ a measurable space, let $\mathcal{G}$ an algebra on Ω generating $\mathcal{A}$ (i.e. $\sigma(\mathcal{G}) = \mathcal{A}$) and let $\mu_0 : \mathcal{G} \to [0, \infty]$ be a map satisfying*

(i) $\mu_0(\varnothing) = 0$,
(ii) $\mu_0\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=1}^{\infty} \mu_0(A_n)$ *for all countable families of pairwise disjoint sets* $\{A_n\}_{n\in\mathbb{N}}$ *in* $\mathcal{G}$ *such that also* $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

Such a map is called a pre-measure. Then μ_0 extends to a measure on $(\Omega, \mathcal{A})$ in the sense that there exists a measure μ on $(\Omega, \mathcal{A})$ with $\mu(G) = \mu_0(G)$ for all $G \in \mathcal{G}$.

Remark A.2 (Properties of a pre-measure)**.** Any pre-measure μ_0 on an algebra $\mathcal{G}$ is monotone and subadditive on $\mathcal{G}$. These properties are deduced as in the case of a full measure from the axioms (i)– (ii). Indeed,

- If $A, B \in \mathcal{G}$ with $A \subseteq B$, then $B = A \sqcup (B \cap A^c)$ and from (ii) (applied to the family $(A, (B \cap A^c), \varnothing, \varnothing, \dots))$ and (i), we deduce that $\mu_0(B) = \mu_0(A) + \mu_0(B \cap A^c) \geq \mu_0(A)$.
- If $\{A_n\}_{n\in\mathbb{N}}$ is a countable family of not necessarily pairwise disjoint sets in $\mathcal{G}$ such that $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{G}$ (which might not be the case for an algebra, see Remark A.1), then $\mu_0\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n=1}^{\infty} \mu_0(A_j)$. Indeed, $\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} B_n$ with $B_n := A_n \setminus \bigcup_{k=1}^{n-1} A_k = A_n \cap \bigcap_{k=1}^{n-1} A_k^c$ and $B_n \in \mathcal{G}$ are pairwise disjoint. It follows from (ii) that
$$\mu_0\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=1}^{\infty} \mu_0(B_n) \leq \sum_{n=1}^{\infty} \mu_0(A_n),$$
where the last inequality is due to the subadditivity shown in the first bullet point.

Remark A.3 (Analogy with Lebesgue's construction)**.** In the context of the construction of the Lebesgue measure we encountered similar objects:

- $\mathcal{G}$ is the algebra generated by all the (open) boxes $B = \prod_{i=1}^{d} (a_i, b_i)$.
- μ_0 is defined on the open boxes by $\mu_0(B) = \mathrm{vol}(B) = \prod_{i=1}^{d} (b_i - a_i)$ and then on an arbitrary set of $\mathcal{G}$ by noticing that any such set can be written as a finite combination of unions and intersections of open boxes and that you can therefore apply the inclusion-exclusion principle (which comes from disjoint additiviy).

Proof of Theorem A.1. We proceed in three steps.

Step 1: We use μ_0 to construct an outer measure μ^*, that is a map $\mu^* : \mathcal{P}(\Omega) \to [0, \infty]$ satisfying the following conditions:

(i) $\mu^*(\varnothing) = 0$.
(ii) monotonicity: $A \subseteq B \implies \mu^*(A) \le \mu^*(B)$, *for all* $A, B \in \mathcal{P}(\Omega)$.
(iii) countable subadditivity: $\mu^*(\bigcup_{n\in\mathbb{N}} A_n) \le \sum_{n\in\mathbb{N}} \mu^*(A_n)$, *for all countable families* $\{A_n\}_{n\in\mathbb{N}}$ *in* $\mathcal{P}(\Omega)$.

We recall that the Lebesgue outer measure on $\mathcal{P}(\mathbb{R}^d)$ was defined by

$$m^*(A) = \inf \left\{ \sum_{n=1}^{\infty} \mathrm{vol}(B_n) \, : \, \{B_n\}_{n\in\mathbb{N}} \text{ countable covering of } A \text{ with open boxes} \right\}.$$

In our more general setting, we similarly set for $A \in \mathcal{P}(\Omega)$

$$\mu^*(A) \coloneqq \inf \left\{ \sum_{n=1}^{\infty} \mu_0(A_n) \, : \, A_n \in \mathcal{G} \text{ for all } n \in \mathbb{N} \text{ and } A \subseteq \bigcup_{n\in\mathbb{N}} A_n \right\}.$$

Observe that μ^* is well-defined as $\Omega \in \mathcal{G}$ and that μ^* is non-negative. The properties (i) – (iii) are verified exactly as in the Lebesgue case using the monotonicity and subadditivity properties of μ_0 of Remark A.2 (cf. the results on outer measure and [Tao16, Lemma 7.2.5]).

Step 2: We show that μ^* is an extension of μ_0 in the sense that $\mu^*(A) = \mu_0(A)$ for every $A \in \mathcal{G}$.
Let $A \in \mathcal{G}$. The inequality $\mu^*(A) \le \mu_0(A)$ is a consequence of the countable subadditivity (iii) of μ^*, established in Step 1, applied to the family $\{A, \varnothing, \varnothing, \dots\}$. We now show the reverse inequality. Let $\{A_n\}_{n\in\mathbb{N}}$ be a countable covering of A made of sets $A_n \in \mathcal{G}$. It then holds that $A = \bigcup_{n\in\mathbb{N}}(A_n \cap A)$. Since $A_n \cap A \in \mathcal{G}$ by Remark A.1, we deduce by monotonicity and subadditivity of μ_0 on $\mathcal{G}$ (see Remark A.2) that

$$\mu_0(A) \le \sum_{n=1}^{\infty} \mu_0(A_n \cap A) \le \sum_{n=1}^{\infty} \mu_0(A_n).$$

Since $\{A_n\}$ was an arbitrary covering of A by sets contained in $\mathcal{G}$, we deduce that $\mu_0(A) \le \mu^*(A)$.

Step 3: We show that μ^* is σ-additive when restricted to the σ-algebra of all measurable sets $\mathcal{M}_{\mu^*}$ (as defined below) and also that $\mathcal{A} \subseteq \mathcal{M}_{\mu^*}$. It thus follows from Step 1 that $\mu \coloneqq \mu^*|_{\mathcal{A}}$ is a measure on $\mathcal{A}$ which is an extension of μ_0 thanks to Step 2.
We call a set $A \in \mathcal{P}(\Omega)$ μ^*-measurable if

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c) \qquad \forall B \in \mathcal{P}(\Omega), \tag{A.1}$$

and we define $\mathcal{M}_{\mu^*} \coloneqq \{A \in \mathcal{P}(\Omega) \, : \, A \text{ is } \mu^*\text{-measurable}\}$. We now show that

(a) $\mathcal{M}_{\mu^*}$ is a σ-algebra on Ω,
(b) $\mathcal{G} \subseteq \mathcal{M}_{\mu^*}$ and hence it follows also that $\mathcal{A} = \sigma(\mathcal{G}) \subseteq \mathcal{M}_{\mu^*}$,
(c) μ^* is σ-additive on $\mathcal{M}_{\mu^*}$, that is for every countable family of pairwise disjoint sets $A_n \in \mathcal{M}_{\mu^*}$ it holds $\mu^*\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=1}^{\infty} \mu^*(A_n)$.

We observe that in the construction of the Lebesgue measure, the same condition (A.1) led to the notion of Lebesgue measurable sets and it was proved that the collection of Lebesgue measurable sets forms a σ-algebra (cf lectures and lemmas 7.4.4 and 7.4.9 of[Tao16]) and that the Lebesgue outer measure is σ-additive on this σ-algebra (cf. lectures and of [Tao16, Lemma 7.4.8]). Inspecting the proofs, they did not make use of the specific structure of the Lebesgue measure and extend without changes to this general setting showing properties (a) and (c) above.

As for property (b), it is enough to show that for every $A \in \mathcal{G}$, it holds $\mu^*(B) \ge \mu^*(B \cap A) + \mu^*(B \cap A^c)$ for every $B \in \mathcal{P}(\Omega)$. Indeed, the reverse inequality follows from the subaddivity of μ^*

on $\mathcal{P}(\Omega)$ (see Step 1). Fix $\varepsilon > 0$. By the definition of the outer measure, we can find a countable family of sets $B_n \in \mathcal{G}$ such that $\mu^*(B) \geq \sum_{n=1}^{\infty} \mu_0(B_n) - \varepsilon$. By finite additivity of μ_0 on $\mathcal{G}$ applied to every $B_n = (B_n \cap A) \sqcup (B_n \cap A^c)$ (see Remark A.2), we deduce

$$\mu^*(B) \geq \sum_{n=1}^{\infty} \mu_0(B_n \cap A) + \sum_{n=1}^{\infty} \mu_0(B_n \cap A^c) - \varepsilon \geq \mu^*(B \cap A) + \mu^*(B \cap A^c) - \varepsilon\,,$$

where we used in the last inequality that $\{B_n \cap A\}_{n\in\mathbb{N}}$ is a covering of $B \cap A$ by sets in $\mathcal{G}$ and $\{B_n \cap A^c\}_{n\in\mathbb{N}}$ is a covering of $B \cap A^c$ by sets in $\mathcal{G}$. We conclude by the arbitrariness of ε.

□

A.3. Uniqueness

Given an algebra $\mathcal{G}$ and a pre-measure μ_0, Theorem A.1 allows to construct a measure μ on the full measure space $(\Omega, \mathcal{A})$ such that $\mu|_{\mathcal{G}} = \mu_0$. However, such extensions of a pre-measure may be non-unique. For instance, let $\mathcal{G}$ be the algebra generated by the intervals of the form $[a, b)$. For an interval $[a, b)$, we define

$$\mu_0([a,b)) := \begin{cases} +\infty & \text{if } [a,b) \neq \varnothing \\ 0 & \text{otherwise} \end{cases}$$

and, enforcing the condition (ii), it is straightforward to extend μ_0 to a pre-measure on $\mathcal{G}$. Two different extensions of μ_0 on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ are then given by

- $\mu_1(A) = \begin{cases} +\infty & \text{if } A \in \mathcal{B}(R) \text{ and } A \neq \varnothing\,, \\ 0 & \text{otherwise}\,. \end{cases}$
- the counting measure, i.e. $\mu_2(A) := \mathrm{card}(A)$.

The goal of this chapter is therefore to establish conditions on $\mathcal{G}$ and μ_0 which guarantee that there is at most one measure μ on $(\Omega, \mathcal{A})$ with $\mu|_{\mathcal{G}} = \mu_0$. The notion of Dynkin systems will prove useful.

Definition (Dynkin system). A family of sets $\mathcal{D} \subseteq \mathcal{P}(\Omega)$ is called a Dynkin system if the following conditions hold:

(i) $\Omega \in \mathcal{D}$.
(ii) $D \in \mathcal{D} \implies D^c \in \mathcal{D}$.
(iii) If $\{D_n\}_{n\in\mathbb{N}}$ is a countable family of pairwise disjoint sets in $\mathcal{D}$, then $\bigcup_{n\in\mathbb{N}} D_n \in \mathcal{D}$.

Remark A.4 (Properties of Dynkin systems)**.**

- Arbitrary intersections of Dynkin systems are Dynkin systems (show it yourself!). Hence it makes sense to introduce the smallest Dynkin system containing a family of set $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ by setting
$$\delta(\mathcal{F}) := \bigcap_{\mathcal{D} \text{ Dynkin system}: \mathcal{F} \subseteq \mathcal{D}} \mathcal{D}\,.$$
We sometimes also say that $\delta(\mathcal{F})$ is generated by $\mathcal{F}$.
- A Dynkin system $\mathcal{D}$ is a σ-algebra if and only if it is stable under intersections (i.e. $\forall A, B \in \mathcal{D}$ it holds $A \cap B \in \mathcal{D}$). Indeed, it is clear that every σ-algebra is in particular a Dynkin system and conversely, if a Dynkin system is stable under intersections then
$$\bigcup_{n\in\mathbb{N}} D_n = \bigcup_{n\in\mathbb{N}} \left[D_n \setminus \bigcup_{j=1}^{n-1} D_j \right] = \bigcup_{n\in\mathbb{N}} \left[D_n \cap D_1^c \cap \cdots \cap D_{n-1}^c \right]$$
also belongs to $\mathcal{D}$, making it a σ-algebra.

The interest in Dynkin systems is motivated by the following fundamental result which ensures that the smallest Dynkin system equals the σ-algebra generated by a family of sets $\mathcal{H}$, provided that the family is stable under intersections. This is remarkable as there are a priori many more Dynkin systems containing $\mathcal{H}$ than σ-algebras.

THEOREM A.2 (Dynkin). *Let Ω a set, $\mathcal{H} \subseteq \mathcal{P}(\Omega)$ a family of sets which is stable under intersections (i.e. $A, B \in \mathcal{H} \implies A \cap B \in \mathcal{H}$). Then $\sigma(\mathcal{H}) = \delta(\mathcal{H})$.*

PROOF OF THEOREM A.2. Since every σ-algebra is a Dynkin system, we already know that $\sigma(\mathcal{H}) \supseteq \delta(\mathcal{H})$ by minimality of $\delta(\mathcal{H})$. Similarly, to prove the reverse inclusion, it is enough to show that $\delta(\mathcal{H})$ is a σ-algebra by the minimality of $\sigma(\mathcal{H})$.

By Remark A.4, we only have to show that $\delta(\mathcal{H})$ is stable under intersections. Fix $A \in \delta(\mathcal{H})$ and let us define $\mathcal{D}_A = \{E \in \mathcal{P}(\Omega) : E \cap A \in \delta(\mathcal{H})\}$. We claim that $\mathcal{D}_A$ is a Dynkin system. Indeed,

(i) $\Omega \in \mathcal{D}_A$, since $\Omega \cap A = A \in \delta(\mathcal{H})$.
(ii) if $E \in \mathcal{D}_A$, then by definition $E \cap A \in \delta(\mathcal{H})$. Moreover, also $A^c \in \delta(\mathcal{H})$ as $\delta(\mathcal{H})$ is stable under complements and hence, exploiting the stability of $\delta(\mathcal{H})$ under disjoint unions, we deduce that $(E^c \cap A)^c = E \cup A^c = A^c \cup (E \cap A) \in \delta(\mathcal{H})$. But this implies $E^c \cap A \in \delta(\mathcal{H})$ (as $\delta(\mathcal{H})$ is stable under complements) and thus by definition $E^c \in \delta(\mathcal{H})$.
(iii) if $\{E_n\}_{n\in\mathbb{N}} \subseteq \mathcal{D}_A$ is a sequence of pairwise disjoint sets, then $E_n \cap A$ are pairwise disjoint sets in $\delta(\mathcal{H})$ and since $\delta(\mathcal{H})$ is stable under countable disjoint unions, we conclude that $\big(\bigcup_{n\in\mathbb{N}} E_n\big) \cap A = \bigcup_{n\in\mathbb{N}}(E_n \cap A) \in \delta(\mathcal{H})$ and hence by definition $\bigcup_{n\in\mathbb{N}} E_n \in \mathcal{D}_A$.

We furthermore claim that $\mathcal{H} \subseteq \mathcal{D}_A$. Indeed, if $B \in \mathcal{H}$, then $\mathcal{H} \subseteq \mathcal{D}_B$ as $\mathcal{H}$ is stable under intersections. By minimality of $\delta(\mathcal{H})$, we then deduce that in fact $\delta(\mathcal{H}) \subseteq \mathcal{D}_B$ and hence in particular $A \cap B \in \delta(\mathcal{H})$. Since B was an arbitrary set in $\mathcal{H}$, this shows the claimed inclusion.

Using once more the minimality of $\delta(\mathcal{H})$, we infer that $\delta(\mathcal{H}) \subseteq \mathcal{D}_A$, which means, by definition, that $A \cap B \in \delta(\mathcal{H})$ for all $B \in \delta(\mathcal{H})$. Since A was an arbitrary set in $\delta(\mathcal{H})$, this shows that $\delta(\mathcal{H})$ is stable under intersections, thus a σ-algebra. ☐

The relevance of Dynkin's theorem lays in the fact that it is the right tool to study the uniqueness of the extension of pre-measures. We give here one possible uniqueness criterion (see also Remark A.5).

Corollary A.3 (Uniqueness of Carathédory's extension)**.** *Let $(\Omega, \mathcal{A})$ be a measurable space. Let $\mathcal{G}$ be an algebra generating $\mathcal{A}$ and μ_0 be a pre-measure on $\mathcal{G}$ as in Theorem A.1. Under the additional hypothesis that $\mu_0(\Omega) < +\infty$, the extension measure μ on $(\Omega, \mathcal{A})$, constructed in Theorem A.1, is unique on $\mathcal{A}$.*

PROOF OF COROLLARY A.3. Let μ_1, μ_2 be two measures on $(\Omega, \mathcal{A})$ extending μ_0, that is $\mu_1(G) = \mu_2(G) = \mu_0(G)$ for all $G \in \mathcal{G}$. Observe that since $\Omega \in \mathcal{G}$ the additional hypothesis guarantees that

$$\mu_1(\Omega) = \mu_0(\Omega) = \mu_2(\Omega) < +\infty\,. \tag{A.2}$$

In other words, both μ_1 and μ_2 are finite measures and hence by σ-additivity it holds in particular that

$$\mu_i(A^c) = \mu_0(\Omega) - \mu_i(A) \tag{A.3}$$

for all $A \in \mathcal{A}$ and $i = 1, 2$. We now introduce $\mathcal{D} \coloneqq \{A \in \mathcal{A} : \mu_1(A) = \mu_2(A)\}$ and claim that it is a Dynkin system. Indeed,

(i) $\Omega \in \mathcal{D}$ follows from (A.2).
(ii) let $A \in \mathcal{D}$. Using (A.3) we have $\mu_1(A^c) = \mu_0(\Omega) - \mu_1(A) = \mu_0(\Omega) - \mu_2(A) = \mu_2(A^c)$ and thus also $A^c \in \mathcal{D}$.

(iii) if $\{D_n\}_{n\in\mathbb{N}} \subseteq \mathcal{D}$ are pairwise disjoint sets, then by σ-additivity of μ_1 and μ_2 on $\mathcal{A}$ it holds

$$\mu_1\left(\bigcup_{n\in\mathbb{N}} D_n\right) = \sum_{n=1}^{\infty}\mu_1(D_n) = \sum_{n=1}^{\infty}\mu_2(D_n) = \mu_2\left(\bigcup_{n\in\mathbb{N}} D_n\right),$$

showing that $\bigcup_{n\in\mathbb{N}} D_n \in \mathcal{D}$.

By assumption $\mathcal{G} \subseteq \mathcal{D}$ and by minimality, we infer $\delta(\mathcal{G}) \subseteq \mathcal{D}$. Moreover by Theorem A.2 (since $\mathcal{G}$ is stable under intersections, see Remark A.1), we have $\delta(\mathcal{G}) = \sigma(\mathcal{G})$ and, since $\sigma(\mathcal{G}) = \mathcal{A}$, we deduce $\mathcal{A} \subseteq \mathcal{D}$, or in other words, $\mu_1(A) = \mu_2(A)$ for all $A \in \mathcal{A}$, showing that the extension is unique. ☐

Remark A.5 (Uniqueness under a σ-finiteness assumption)**.** The assumption $\mu_0(\Omega) < +\infty$ in Corollary A.3 can be relaxed. Indeed, it is enough that there exists a countable family $\{G_n\}_{n\in\mathbb{N}} \subseteq \mathcal{G}$ such that $\Omega = \bigcup_{n\in\mathbb{N}} G_n$ and for every $n \in \mathbb{N}$ it holds $\mu_0(G_n) < +\infty$. Up to considering $G'_n := G_n \setminus \bigcup_{k=1}^{n-1} G_k$ instead of G_n, we can always assume that G_n are pairwise disjoint.
Observe that the only point in the proof of Corollary A.3 that fails under these relaxed assumptions, is the stability of $\mathcal{D}$ under complements (as (A.3) no longer holds). Instead, we introduce for every $n \in \mathbb{N}$ the collection $\mathcal{D}_n := \{A \in \mathcal{A} : \mu_1(A\cap G_n) = \mu_2(A \cap G_n)\}$. As $\mu_0(G_n) < +\infty$, we have by σ-additivity of μ_1 and μ_2 on $\mathcal{A}$ that $\mu_i(A^c \cap G_n) = \mu_0(E_n) - \mu_i(A\cap G_n)$ for all $A \in \mathcal{A}$ and $i = 1, 2$. With this observation, one proceeds as in the proof of Corollary A.3 to show that $\mathcal{D}_n$ is a Dynkin system and that $\mathcal{A} \subseteq \mathcal{D}_n$. Therefore $\mathcal{A} \subseteq \bigcap_{n\in\mathbb{N}} \mathcal{D}_n$ and thus for $A \in \mathcal{A}$, it holds by σ-additivity

$$\mu_1(A) = \sum_{n\in\mathbb{N}}\mu_1(A\cap G_n) = \sum_{n=1}^{\infty}\mu_2(A\cap G_n) = \mu_2(A),$$

showing the uniqueness of the extension.

Exercise 1 (Uniqueness of Lebesgue measure). *Let μ be a measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ satisfying the following conditions:*

(1) μ is translation-invariant, i.e. $\mu(A + x) = \mu(A)$ for all $A \in \mathcal{B}(\mathbb{R}^d)$ and every $x \in \mathbb{R}^d$.
(2) $\mu([0, 1)^d) =: \lambda < +\infty$.

Prove that $\mu = \lambda m$ on $\mathcal{B}(\mathbb{R}^n)$, where m denotes the n-dimensional Lebesgue measure.

Solution. *One can proceed in three steps.*

Step 1: We show that $\mu(B) = km(B)$ holds for every box $B = \prod_{i=1}^d [a_i, b_i)$ with $a_i, b_i \in \mathbb{Q}$.
Consider first a box $B = \prod_{i=1}^d [0, \frac{p_i}{q_i})$. Observe that $[0, p_i)$ is the union of p_i translates of the interval $[0, 1)$ (being $[k, k + 1)$ for $k = 0, \dots, p_i - 1$). In particular, $\prod_{i=1}^d [0, p_i)$ is the finite union of $\prod_{i=1}^d p_i$ disjoint boxes of the form $\prod_{i=1}^d [k_i, k_i + 1)$ where $k_i = 1, \dots, p_i - 1$ for $i = 1, \dots, d$. By translation-invariance of μ, we have, setting $\bar{k} := (k_1, \dots, k_d)$, $\mu(\prod_{i=1}^d [k_i, k_i + 1)) = \mu(\bar{k} + [0, 1)^d) = \lambda$ and by σ-additivity of μ, $\mu(\prod_{i=1}^d [0, p_i)) = \lambda \prod_{i=1}^n p_i$. With a similar reasoning, we have for any $q_1, \dots, q_d \in \mathbb{N}$ by translation invariance and σ-additivity that $\mu(\prod_{i=1}^d [0, p_i)) = \prod_{i=1}^d q_i\ \mu\big(\prod_{i=1}^d [0, \frac{p_i}{q_i})\big)$. Combining both properties, we deduce that $\mu(B) = \lambda \prod_{i=1}^d \frac{p_i}{q_i} = \lambda m(B)$. By translation invariance, we deduce that this equality holds in fact for all boxes with rational endpoints.

Step 2: Consider $\mathcal{H} := \{B = \prod_{i=1}^d [a_i, b_i) : a_i, b_i \in \mathbb{Q}\}$. We show that $\mathcal{H}$ is stable under intersections and that $\sigma(\mathcal{H}) = \mathcal{B}(\mathbb{R}^d)$.
The stability under intersections is straightforward to verify. It is clear that $\sigma(\mathcal{H}) \subseteq \mathcal{B}(\mathbb{R}^d)$ as $\mathcal{H} \subseteq \mathcal{B}(\mathbb{R}^d)$. To show the reverse inequality, it is by minimality enough to show that $\sigma(\mathcal{H})$ contains all open sets. By definition, every open set $U \subseteq \mathbb{R}^d$ can be written as a countable union of open

boxes with rational endpoints $\prod_{i=1}^d (a_i, b_i)$ *. Since open boxes with rational endpoints can be written as a countable union of half-open boxes* $\prod_{i=1}^d (a_i, b_i) = \bigcup_{k=1}^\infty \prod_{i=1}^d [a_i + \frac{1}{k}, b_i)$ *, they belong to* $\sigma(\mathcal{H})$ *.*

Step 3: The Dynkin argument.
Let $B_k := [-k, k)^d$ *for* $k \geq 1$ *. Let* $\mathcal{D}_k := \{A \in \mathcal{B}(\mathbb{R}^d) : \mu(A \cap B_k) = \lambda m(A \cap B_k)\}$ *. We claim that* $\mathcal{D}_k$ *is a Dynkin system and that* $\mathcal{H} \subseteq \mathcal{D}_k$ *. With the claim, we apply Theorem A.2 and Step 2, to deduce that* $\mathcal{B}(\mathbb{R}^d) = \sigma(\mathcal{H}) = \delta(\mathcal{H}) \subseteq \mathcal{D}_k$ *. Thus* $\mathcal{B}(\mathbb{R}^d) \subseteq \bigcap_{k \geq 1} \mathcal{D}_k$ *and we conclude that for every* $A \in \mathcal{B}(\mathbb{R}^d)$

$$\mu(A) = \lim_{k\to\infty} \mu(A \cap B_k) = \lim_{k\to\infty} \lambda m(A \cap B_k) = \lambda m(A) .$$

Remark A.6. The attentive reader will notice that we prove that the Lebesgue measure on $\mathcal{B}(\mathbb{R}^d)$ is the unique translation invariant-measure with $\mu([0,1]^d) = 1$ only on $\mathcal{B}(\mathbb{R}^d)$, but we do not give any uniqueness statement of the Lebesgue measure on larger σ-algebra of all Lebesgue measurable sets $\mathcal{M}(\mathbb{R}^d)$. The uniqueness on $\mathcal{M}(\mathbb{R}^d)$ is related to the procedure of "completion of a measurable fact" (a measurable space is called complete if the subsets of all null-sets are measurable) and the fact that the completion of $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), m)$ is given by $(\mathbb{R}^d, \mathcal{M}(\mathbb{R}^d), m)$.

A.4. Probability measures from cumulative distribution functions

We recall the definition of a joint cumulative distribution function from Probability (Definition 1.31 of [Aru21]).

Definition. Any function $F : \mathbb{R}^d \to [0, 1]$ is called a joint cumulative distribution function (short "cdf"), if it satisfies the following conditions:

(i) F is non-decreasing in each coordinate.
(ii) $F(x_1, \ldots, x_d) \to 1$ if all of $x_i \to +\infty$.
(iii) $F(x_1, \ldots, x_d) \to 0$ if at least one of $x_i \to -\infty$.
(iv) F is right-continuous, meaning that for every $(x_1^m, \ldots, x_d^m)$ converging to $(x_1, \ldots, x_d)$ such that for all $m \geq 1$ we have $x_i^m \geq x_i$ for all $i = 1, \ldots, d$, it holds that $F(x_1^m, \ldots, x_d^m) \to F(x_1, \ldots, x_d)$.
(v) Let $A := (a_1, b_1] \times \cdots \times (a_d, b_d]$ and $V := \{a_1, b_1\} \times \cdots \times \{a_d, b_d\}$, where $a_i, b_i \in (-\infty, +\infty)$ $\forall i = 1, \ldots, d$ (V is the set of the vertices of the finite rectangle A); if $v \in V$, let $\operatorname{sgn}(v) := (-1)^{\#\text{ of } a_i \text{ in } v}$. Then,

$$\Delta_A F = \sum_{v \in V} \operatorname{sgn}(v) F(v).$$

We will let $\mu(A) = \Delta_A F$, so we must assume

$$\Delta_A F \geq 0 \text{ for all rectangles } A.$$

A fundamental result in Probability shows that there is a one-to-one correspondence between probability measures $\mathbb{P}$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and cumulative distribution functions F. More precisely, we have the following

Theorem A.4. *(Theorem 1.32 of [Aru21])*

(i) Each probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ *gives rise to a cdf* $F : \mathbb{R}^d \to [0, 1]$ *through*

$$F(x_1, \ldots, x_d) := \mathbb{P}\left(\prod_{i=1}^d (-\infty, x_i]\right) . \tag{A.4}$$

(ii) Conversely, given a cdf $F : \mathbb{R}^d \to [0, 1]$*, there exists a unique probability measure* $\mathbb{P}$ *on* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ *such that (A.4) holds for all* $(x_1, \ldots x_d) \in \mathbb{R}^d$ *.*

The item (i) of Theorem A.4 is a straightforward consequence of basic properties of probability measures and the definition (A.4) and has been established in the probability course. The item (ii) of Theorem A.4 instead has been proven only in dimension $d = 1$ in the probability course. The proof took advantage of the fact that F has pseudo-inverse which is a peculiarity of the one-dimensional case and as such, the proof does not extend to higher dimension. We here present an alternative proof, relying on Carathéodory's construction, which extends to all dimensions.

PROOF OF (II) OF THEOREM A.4 VIA CARATHÉODORY. We introduce

$$\mathcal{G}_0 := \left\{ \prod_{i=1}^{d} I_i \text{ where } I_i = (s_i, t_i] \text{ or } I_i = (s_i, \infty) \text{ with } -\infty \leq s_i < t_i < \infty \right\} \cup \{\varnothing\}$$

and $\mathcal{G} := \left\{ \bigcup_{i=1}^{N} G_i \,:\, G_i \in \mathcal{G}_0 \right\}$. One verifies that $\mathcal{G}$ is an algebra (see Exercise 2). Moreover, we claim that $\sigma(\mathcal{G}) = \mathcal{B}(\mathbb{R}^d)$. The inclusion $\sigma(\mathcal{G}) \subseteq \mathcal{B}(\mathbb{R}^d)$ is trivial as $\mathcal{G} \subseteq \mathcal{B}(\mathbb{R}^d)$. For the reverse inclusion, we observe that $\sigma(\mathcal{G})$ also contains all open boxes $B = \prod_{i=1}^{d} (s_i, t_i)$. Since every open set can be written as a countable union of open boxes, this implies that $\sigma(\mathcal{G})$ contains all open sets and hence, since $\mathcal{B}(\mathbb{R}^d)$ is generated by all open sets, we have $\mathcal{B}(\mathbb{R}^d) \subseteq \sigma(\mathcal{G})$.

It now suffices to construct a pre-measure μ_0 on $\mathcal{G}$ (i.e. a map $\mu_0 : \mathcal{G} \to [0, +\infty]$ satisfying (i) and (ii) of Theorem A.1) such that

- $\mu_0(\prod_{i=1}^{d} (-\infty, x_i]) = F(x_1, \ldots, x_d)$ for all $x \in \mathbb{R}^d$,
- $\mu_0(\mathbb{R}^d) = 1$.

Indeed, Theorem A.1 and Corollary A.3 then guarantee the existence of a unique measure μ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ such that $\mu|_{\mathcal{G}} = \mu_0$; this guarantees both the validity of (A.4) as well as $\mu(\mathbb{R}^d) = 1$, making μ a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

In order to present the main idea, we show the construction of the pre-measure μ_0 only for $x = 1$ and leave the case $d \geq 2$ as an exercise (see Exercise 2). We first extend the cdf $F : \mathbb{R} \to [0, 1]$ to a function defined on $\mathbb{R} \cup \{\pm\infty\}$ by setting $F(-\infty) := 0$ and $F(+\infty) := 1$.

In a first step, we define μ_0 on $\mathcal{G}_0$ by setting

$$\begin{aligned} \mu_0((s, t]) &:= F(t) - F(s) \qquad \text{if } -\infty \leq s < t < \infty, \\ \mu_0((s, \infty)) &:= F(+\infty) - F(s) = 1 - F(s). \end{aligned}$$

By definition $\mu(\varnothing) = 0$ and $\mu_0(\mathbb{R}) = F(+\infty) - F(-\infty) = 1$. Moreover, by construction it holds that $\mu_0((s, t]) + \mu_0((t, r]) = \mu_0((s, r])$ and $\mu_0((s, t]) + \mu_0((t, \infty)) = \mu_0((s, \infty))$. Using this two properties it is easy to verify that μ_0 is finitely additive on $\mathcal{G}_0$.

The finite additivity allows us in a second step to define μ_0 on all of $\mathcal{G}$. Indeed, every $G \in \mathcal{G}$ can be written as $G = \bigcup_{i=1}^{N} G_i$ with $G_i \in \mathcal{G}_0$. Up to considering $G_i' = G_i \setminus \bigcup_{k=1}^{i-1} G_k = G_i \cap \bigcap_{k=1}^{i-1} G_k^c$ instead of G_i ($G_i' \in \mathcal{G}_0$ because $\mathcal{G}_0$ is closed under intersections and complements), we can assume w.l.o.g. that the sets G_i are pairwise disjoint. This allows to define $\mu_0(G) := \sum_{i=1}^{N} \mu_0(G_i)$. μ_0 is well-defined (i.e. it is independent of the choice of the family $\{G_i\}_{i=1}^{N}$ thanks to the finite additivity) and by construction, μ_0 is finitely additive on $\mathcal{G}$, meaning that (ii) holds for finite families of pairwise disjoint sets. We are left to establish it for countable families. Let therefore $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{G}$ be a countable family of pairwise disjoint sets such that also $A := \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{G}$. For $k \geq 1$ fixed, we rewrite $A = B_k \cup \bigcup_{i=1}^{k} A_i$ with $B_k := \bigcup_{i \geq k+1} A_i = A \cap \bigcap_{i=1}^{k} A_i^c$. By stability under complements and finite intersections, $B_k \in \mathcal{G}$ and hence by finite additivity of μ_0 on $\mathcal{G}$ we have

$$\mu_0(A) = \mu_0 \left(\bigcup_{i=1}^{k} A_i \right) + \mu_0(B_k) = \sum_{i=1}^{k} \mu_0(A_i) + \mu_0(B_k).$$

We deduce the validity of (ii) by taking $k \to \infty$, if

$$\lim_{k\to\infty} \mu_0(B_k) = 0\,. \tag{A.5}$$

To show (A.5), observe that $B_{k+1} \subset B_k$ is a decreasing sequence of sets with $\bigcap_{k\geq 1} B_k = \varnothing$. Hence, since μ_0 is monotone, $\lim_{k\to\infty}\mu_0(B_k) \in [0,1]$ exists and we assume by contradiction that $\lim_{k\to\infty}\mu_0(B_k) =: c > 0$. We will show that this assumption is absurd by showing that under this assumption, $\bigcap_{k\in\mathbb{N}} B_k$ would contain a non-empty set.

To do so, we first observe that for every interval $I \in \mathcal{G}_0$ and any $\varepsilon > 0$, there exists $I' \in \mathcal{G}_0$ and a compact K such that $I' \subseteq K \subseteq I$ and such that $\mu_0(I') \geq \mu_0(I) - \varepsilon$. Indeed,

- if $I = (s,t]$: then $\mu_0(I) = F(t) - F(s)$ and thanks to the right-continuity of F, there exists $s' \in (s,t)$ such that $F(s') - F(s) \leq \varepsilon$ and hence setting $I' := (s',t]$ and $K := [(s+s')/2, t]$, we have that K is compact, $I' \subseteq K \subseteq I$ and $\mu_0(I') = F(t) - F(s') \geq \mu_0(I) - \varepsilon$.
- if $I = (s,\infty)$: then $\mu_0(I) = 1 - F(s)$. Thanks to the to the right-continuity of F, there exists $s' \in (s,\infty)$ such that $F(s') - F(s) \leq \frac{\varepsilon}{2}$ and since $\lim_{t\to\infty} F(t) = 1$, there exists $t' \in (s',\infty)$ such that $F(t') \geq 1 - \frac{\varepsilon}{2}$. Hence setting $I' := (s',t']$, $K' := [(s+s')/2, t']$, we have that K is compact, $I' \subseteq K \subseteq I$ and $\mu_0(I') = F(t') - F(s') \geq \mu_0(I) - \varepsilon$.

Since every B_k is made out of a disjoint union of intervals $I \in \mathcal{G}_0$, this shows that for every B_k there exists a compact K_k and $B_k' \in \mathcal{G}$ such that $B_k' \subseteq K_k \subseteq B_k$ and $\mu_0(B_k') \geq \mu_0(B_k) - c2^{-(k+1)}$. Now we set $C_k := \bigcap_{j=1}^k K_j$ and we observe that $\{C_k\}_{k\geq 1}$ is a decreasing family of compact sets. We claim that $\bigcap_{k\in\mathbb{N}} C_k \neq \varnothing$ which gives the desired contradiction since $\bigcap_{k\in\mathbb{N}} C_k \subseteq \bigcap_{k\in\mathbb{N}} B_k$ by construction. It is a general fact from topology that the intersection of a decreasing sequence of non-empty compact sets is non-empty[1] and so, it suffices to show that $C_k \neq \varnothing$ for every $k \geq 1$. By construction $\bigcap_{j=1}^k B_j' \subseteq C_k$ and, since $B_k \setminus \bigcup_{j=1}^k (B_j \setminus B_j') \subseteq \bigcap_{j=1}^k B_j'$, we have by finite (sub-)additivity

$$\mu_0\left(\bigcap_{j=1}^k B_j'\right) \geq \mu_0(A_n) - \mu_0\left(\bigcup_{j=1}^k (B_j \setminus B_j')\right) \geq \mu_0(A_n) - c\sum_{j=1}^k 2^{-(j+1)} \geq \frac{c}{2}\,.$$

In particular, $\bigcap_{j=1}^k B_j' \neq \varnothing$ and hence $C_k \neq \varnothing$. □

EXERCISE 2. *Let $F : \mathbb{R}^2 \to [0,1]$ a two-dimensional joint cumulative distribution function. The goal of this exercise is to use Carathéodory's construction to show the existence of a unique probability measure $\mathbb{P}$ on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ such that*

$$\mathbb{P}((-\infty,x_1] \times (-\infty,x_2]) = F(x_1,x_2) \qquad \textit{for all } x = (x_1,x_2) \in \mathbb{R}^2\,. \tag{A.6}$$

(a) As above, we introduce again the collections of sets

$$\mathcal{G}_0 := \left\{ \prod_{i=1}^2 I_i \text{ where } I_i = (s_i,t_i] \text{ or } I_i = (s_i,\infty) \text{ with } -\infty \leq s_i < t_i < \infty \right\} \cup \{\varnothing\}\,,$$

$$\mathcal{G} := \left\{ \bigcup_{i=1}^N G_i \,:\, G_i \in \mathcal{G}_0 \right\}\,.$$

Show that $\mathcal{G}$ is an algebra.

(b) Define a map $\mu_0 : \mathcal{G}_0 \to [0,1]$ such that $\mu_0(\varnothing) = 0$, such that $\mu_0(\mathbb{R}^2) = 1$ and such that μ_0 is additive on finitely many disjoint sets in $\mathcal{G}_0$ (meaning that whenever $A, B \in \mathcal{G}_0$ with $A \cap B = \varnothing$, then it holds that $\mu_0(A \cup B) = \mu_0(A) + \mu_0(B)$).

[1] Indeed, by contradiction, assume that $C_{k+1} \subseteq C_k$ are compact, that $C_k \neq \varnothing$ for all $k \geq 1$ and that $\bigcap_{k\in\mathbb{N}} C_k = \varnothing$. Then the family $O_k = C_k^c$ is an open cover for C_1 (and hence for every C_k with $k \geq 1$). By compactness, we can extract a finite subcover. Since the C_k are decreasing, the family O_k is increasing and the subcover in fact only consist of one open set $O_{\bar{k}}$. This is absurd as $O_{\bar{k}}$ cannot possibly cover $C_{\bar{k}}$ which by assumption is non-empty.

(c) *Extend μ_0 to $\mathcal{G}$ ad verify that μ_0 is a pre-measure on $\mathcal{G}$ (i.e. verify the properties (i) and (ii) of Theorem A.1).*

(d) *Deduce that there exists a unique probability measure $\mathbb{P}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ satisfying* (A.6).

Solution. (a) *Since $\mathcal{G}$ is by definition stable by unions, we only have to show that it is stable under complements and intersections with the following steps:*

(i) *if $G_1, G_2 \in \mathcal{G}_0$ then $G_1 \cap G_2 \in \mathcal{G}_0$:*
We can use the fact that for some sets A, B, C, D we have $(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D)$ to reduce the argument to each dimension and apply the fact that the intersections of two intervals (of one of the forms in the definition of $\mathcal{G}_0$) is either an interval (of one of the same forms) or the empty set. Hence the intersections of sets in $\mathcal{G}_0$ is still in $\mathcal{G}_0$ ✓

(ii) *if $G_1 = \cup_{i=1}^{N} G_{1,i}$, $G_2 = \cup_{j=1}^{M} G_{2,i} \in \mathcal{G}$ (with each $G_{1,i}$ and $G_{2,i}$ in $\mathcal{G}_0$) then $G_1 \cap G_2 \in \mathcal{G}$:*
because

$$G_1 \cap G_2 = \left(\bigcup_{i=1}^{N} G_{1,i}\right) \cap \left(\bigcup_{i=1}^{M} G_{2,i}\right) = \bigcup_{i=1}^{N} \bigcup_{j=1}^{M} \underbrace{G_{1,i} \cap G_{2,j}}_{\in \mathcal{G}_0 \text{ by (i)}} \in \mathcal{G}_0 \checkmark$$

(iii) *if $G \in \mathcal{G}_0$ then $G^c \in \mathcal{G}$:*
For some sets A, B we can't say that $(A \times B)^c = (A^c) \times (B^c)$ so we can't use the argument used for (i). However, as shown in the figure 1, we can actually obtain the complement by adding up to 4 other sets of $\mathcal{G}_0$. This can be easily generalized to the case where $A \times B$ is unbounded. ✓

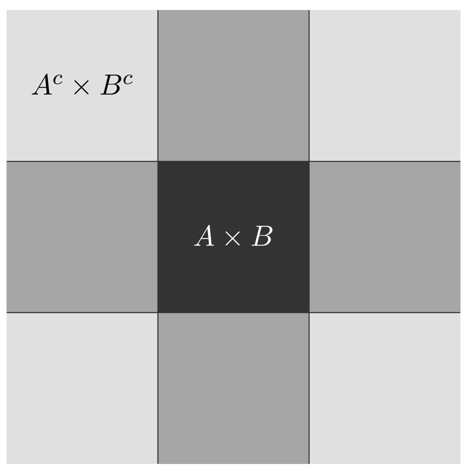


Figure 1. Black: the set $A \times B \in \mathcal{G}_0$ we are considering. Gray: the set $(A^c) \times (B^c)$. Dark gray: the 4 sets we need to add to obtain the complement $(A \times B)^c$

(iv) *if $G = \cup_{i=1}^{N} G_i \in \mathcal{G}$ (with each G_i in $\mathcal{G}_0$) then $G^c \in \mathcal{G}$:*
because

$$\left(\bigcup_{i=1}^{N} G_i\right)^c = \bigcap_{i=1}^{N} \underbrace{G_i^c}_{\in \mathcal{G} \text{ by (iii)}} \underbrace{\in \mathcal{G}}_{\text{by (ii)}} \checkmark$$

(b) *We define $\mu_0(\varnothing) = 0$, and for $a < b$ and $c < d$*

$$\mu_0((a, b] \times (c, d]) = F(b, d) - F(a, d) - F(b, c) + F(a, c).$$

For simplicity, let us just focus on the case where a, b, c, d *are all finite (we can handle the cases* $a = +\infty$*, etc. as done for the one-dimensional case in the notes). To verify finite additivity on* $\mathcal{G}_0$*, we need to show that if* $A, B \in \mathcal{G}_0$ *with* $A \cap B = \varnothing$ *and* $A \cup B \in \mathcal{G}_0$*, then* $\mu_0(A \cup B) = \mu_0(A) + \mu_0(B)$*. Indeed, by our definition of* μ_0*, for* $a_1 < b_1 < c_1$ *and* $a_2 < b_2 < c_2$*, we have*

$$\mu_0((a_1, b_1] \times (a_2, b_2]) + \mu_0((a_1, b_1] \times (b_2, c_2]) = \mu_0((a_1, b_1] \times (a_2, c_2]),$$

and

$$\mu_0((a_1, b_1] \times (a_2, b_2]) + \mu_0((b_1, c_1] \times (a_2, b_2]) = \mu_0((a_1, c_1] \times (a_2, b_2]).$$

Note that μ_0 *satisfies the desired properties because*

- $\mu_0(\varnothing) = 0$
- $\mu_0((-\infty, b] \times (-\infty, d]) = F(b, d)$
- $\mu_0((a, b] \times (-\infty, d]) = F(b, d) - F(a, d)$
- $\mu_0((-\infty, b] \times (c, d]) = F(b, d) - F(b, c)$
- $\mu_0((a, b] \times (c, d]) = F(b, d) - F(a, d) - F(b, c) + F(a, c)$
- $\mu_0((-\infty, b] \times \mathbb{R}) = \lim_{d\to\infty} F(b, d)$
- $\mu_0(\mathbb{R} \times (-\infty, d]) = \lim_{b\to\infty} F(b, d)$
- $\mu_0((a, \infty) \times (c, d)) = F(b, d) - F(a, d) - F(b, c) + F(a, c)$
- $\mu_0((a, b) \times (c, \infty)) = F(b, d) - F(a, d) - F(b, c) + F(a, c)$
- $\mu_0((-\infty, b] \times (c, \infty)) = \mu_0((-\infty, b] \times \mathbb{R}) - \mu_0((-\infty, b] \times (-\infty, c]) = -F(b, c) + \lim_{d\to\infty} F(b, d)$
- $\ldots$
- $\mu_0((a, \infty) \times (c, \infty)) = F(b, d) - F(a, d) - F(b, c) + F(a, c)$

(c) *The first assumption (i) is verified by definition. For the exact same reasons as for the one-dimensional case (from the notes), we can extend* μ_0 *to a premeasure on* $\mathcal{G}$*. The only (slight) difference is we have to prove that for every box* $I \in \mathcal{G}_0$ *and any* $\epsilon > 0$*, there exists* $I' \in \mathcal{G}_0$ *and a compact* K *such that* $I' \subset K \subset I$ *and such that* $\mu_0(I') \geq \mu_0(I) - \epsilon$*. For simplicity, let's just show this is true for the case where* $I = (a, b] \times (c, d]$ *and* a, b, c, d *are finite.*

By the right-continuity of F*, there is an* $a' \in (a, b)$ *and* $c' \in (c, d)$ *so that* $F(a', d) - F(a, d) \leq \epsilon/2$ *and* $F(b', c) - F(b, c) \leq \epsilon/2$*. Set* $I' = (a', b] \times (c', d]$ *and* $K = [(a + a')/2, b] \times [(b + b')/2, d]$*. Then,*

$$\begin{aligned}\mu_0(I) - \mu_0(I') &= (F(b, d) - F(a, d) - F(b, c) + F(a, c)) - (F(b, d) - F(a', d) - F(b, c') + F(a', c')) \\ &= F(a', d) - F(a, d) + F(b, c') - F(b, c) + F(a, c) - F(a', c') \leq \epsilon,\end{aligned}$$

where we used that $F(a, c) - F(a', c') \leq 0$ *since* F *is nondecreasing in each coordinate. By the right-continuity of* F*, there is an* $a' \in (a, b)$ *and* $c' \in (c, d)$ *so that* $F(a'', d) - F(a, d) \leq \epsilon/3$ *for all* $a'' \in (a, a']$ *and* $F(b'', c) - F(b, c) \leq \epsilon/3$ *for all* $b'' \in (b, b']$*. By the right-continuity of* F*, there is a* $t' > 0$ *such that* $F(a + t, c + t) - F(a, c) \leq \epsilon/3$ *for all* $t \in (0, t']$*. Let* $\tilde{a} = \min\{a', a + t'\}$ *and* $\tilde{b} = \min\{b', b + t'\}$*. Set* $I' = (\tilde{a}, b] \times (\tilde{c}, d]$ *and* $K = [(a + \tilde{a})/2, b] \times [(b + \tilde{b})/2, d]$*. Then,*

$$\begin{aligned}\mu_0(I) - \mu_0(I') &= (F(b, d) - F(a, d) - F(b, c) + F(a, c)) - (F(b, d) - F(\tilde{a}, d) - F(b, \tilde{c}) + F(\tilde{a}, \tilde{c})) \\ &= F(\tilde{a}, d) - F(a, d) + F(b, \tilde{c}) - F(b, c) + F(a, c) - F(\tilde{a}, \tilde{c}) \leq \epsilon.\end{aligned}$$

(d) *This follows directly from Theorem 2.4 and Corollary 3.4 (as for the one-dimensional case).*

APPENDIX B

The Laplace Transform

We introduce an integral transform akin to the Fourier transform. We will see that it enjoys properties making it applicable to resolution of ODEs, turning these into algebraic equations.

B.1. Definition

We have seen in Chapter 5 that integral transforms can allow one to rephrase certain problems into a more tractable language. A kernel integral transform of general type takes the following form. Given a function $f : \Omega \subset \mathbb{R} \to \mathbb{C}$ of sufficient regularity, one can define

$$K[f](y) := \int_{\Omega} f(x)K(x,y)dx$$

for some sufficiently regular "kernel" $K : \Omega \times \Omega \to \mathbb{C}$. The case of the Fourier transform takes $\Omega = \mathbb{R}$ and $K(x,y) = e^{-2\pi ixy}$. In a similar manner, the Laplace transform is a kernel integral transform of the following shape.

Definition. [Laplace Transform] Let $f : \mathbb{R}_{\geq 0} \to \mathbb{C}$. Its Laplace transform is

$$\mathcal{L}[f](s) := \int_0^{\infty} e^{-st} f(t)dt$$

defined on the domain $\mathcal{D}(\mathcal{L}[f]) = \{s \in [0, +\infty[: e^{-st}f(t) \in L^1(\mathbb{R})\}$. We will also denote the Laplace transform of f by $F := \mathcal{L}[f]$.

Remark B.1. (1) $\mathcal{L}$ is an integral transform with kernel $K(s,t) = e^{-st}$ and $\Omega = \mathbb{R}_{\geq 0}$;
(2) We take the domain of $\mathcal{L}[f]$ to live within the reals. More generally one usually considers a complex variable s;
(3) As we will see, $\mathcal{L}$ transforms derivatives into multiplications. Thus its principal use will be in solving differential equations, by turning ODEs (of solution f) into algebraic equations solvable in F. This then raises the question of whether one can invert the Laplace transform in a meaningful way.

As is explicited in the definition of the Laplace transform, $\mathcal{L}[f]$ is generally not defined on $\mathbb{R}_{\geq 0}$. In fact $\mathcal{D}(\mathcal{L}[f])$ might be empty. If non-empty however, $\mathcal{D}(\mathcal{L}[f])$ is large in the following sense.

Lemma B.1. *If $s \in \mathcal{D}(\mathcal{L}[f])$, then $\mathbb{R}_{\geq s} \subset \mathcal{D}(\mathcal{L}[f])$.*

Proof. Let $s \in \mathcal{D}(\mathcal{L}[f])$, $s' > s$. Then

$$\int_0^{\infty} |f(t)|e^{-s't}dt = \int_0^{\infty} |f(t)|e^{(s-s')t}e^{-st}dt \leq \int_0^{\infty} |f(t)|e^{-st} < \infty,$$

so that $s' \in \mathcal{D}(\mathcal{L}[f])$. □

Remark B.2. $\mathcal{D}(\mathcal{L}[f])$ is not necessarily closed, as is shown by $f(t) = e^t$, of domain $\mathbb{R}_{>1}$.

We want a nice class of functions for which $\mathcal{L}[f]$ is non-empty. We take Remark B.2 as inspiration.

Definition. Let $a \in \mathbb{R}_{>0}$. The exponential class Exp_a is defined by

$$\mathrm{Exp}_a := \left\{ f \in L^1_{loc}(\mathbb{R}_{\geq 0}) : |f(t)| = O(e^{at}) \text{ as } t \to +\infty \right\}.$$

Lemma B.2. *Let* $f \in \mathrm{Exp}_a$. *Then* $\mathbb{R}_{>a} \subset \mathcal{D}(\mathcal{L}[f])$.

Proof. Let $K > 0, C > 0$ be such that $|f(t)| \le C^{at}$ $\forall t > K$. Then

$$\int_0^\infty |f(t)|e^{-st}dt = \int_0^K |f(t)|e^{-st}dt + \int_K^\infty |f(t)e^{st}$$
$$\le \int_0^K |f(t)|e^{-st}dt + C\int_K^\infty e^{(a-s)t}dt$$

The first integral converges since $f \in L^1_{loc}$, while the second converges iff $s > a$, as required. □

B.2. Properties and Applications

In what follows we do not carry along domain considerations, and assume the unexplicited functions f, g are regular enough for each expression to make sense.

Proposition B.3 (Properties of the Laplace Transform)**.** *The Laplace transform enjoys the following properties:*

(1) (Linearity) $\mathcal{L}[af + bg] = a\mathcal{L}[f] + b\mathcal{L}[g]$ $\forall a, b \in \mathbb{C}$
(2) (Derivatives in t) Derivatives are transformed into products: $\forall n \in \mathbb{N}$,

$$\mathcal{L}[f^{(n)}](s) = s^n\mathcal{L}[f](s) - \sum_{i=0}^{n-1} s^{n-i-1}f^{(i)}(0)$$

(3) (Derivatives in s) Products are transformed into derivatives:

$$\mathcal{L}[t^n f(t)](s) = (-1)^n\frac{d^n}{ds^n}\mathcal{L}[f](s)$$

(4) (Translation) $\mathcal{L}[f](s-a) = \mathcal{L}[e^{at}f(t)](s)$
(5) (Scaling) For $a > 0$, $\mathcal{L}[f(at)](s) = a^{-1}\mathcal{L}[f](sa^{-1})$

Proof. (1) This is an immediate consequence of linearity of integrals.
(2) We first consider $n = 1$. Integrating by parts gives

$$\int_0^\infty f'(t)e^{-st}dt = f(t)e^{-st}|_0^\infty + s\int_0^\infty f(t)e^{st}dt = s\mathcal{L}[f](s) - f(0).$$

One can then conclude by induction:

$$\mathcal{L}[f^{(n)}](s) = s\mathcal{L}[f^{(n-1)}](s) - f^{(n-1)}(0) = s^n\mathcal{L}[f](s) - \sum_{i=0}^{n-1} s^{n-i-1}f^{(i)}(0).$$

(3) We begin with $n = 1$. By Corollary 2.18 we have

$$\frac{d}{ds}\mathcal{L}[f](s) = \frac{d}{ds}\int_0^\infty e^{-st}f(t)dt = \int_0^\infty (-te^{-st})f(t)d = -\mathcal{L}[tf(t)].$$

By induction we then conclude, for

$$\frac{d^n}{ds^n}\mathcal{L}[f](s) = (-1)^{n-1}\frac{d}{ds}\mathcal{L}[t^{n-1}f(t)] = (-1)^n\mathcal{L}[t^n f(t)].$$

(4)

$$\mathcal{L}[f](s-a) = \int_0^\infty f(t)e^{(a-s)t}dt = \int_0^\infty e^{at}f(t)e^{-st}dt = \mathcal{L}[e^{at}f(t)](s).$$

(5)

$$\mathcal{L}[f(at)](s) = \int_0^\infty f(at)e^{-st}dt = a^{-1}\int_0^\infty f(t)e^{-sa^{-1}t}dt = a^{-1}\mathcal{L}[f](sa^{-1})$$

□

Let us now see how these properties, in particular (2), help us to solve ODEs. Suppose given an ODE of the form

$$\begin{cases} G(f, f', ..., f^n) = h & h \text{ "nice", say in } \operatorname{Exp}_a \\ f^{(i)}(0) = a_i & a_i \in \mathbb{C}, i = 0, ..., n-1 \end{cases}$$

If G is compatible enough with $\mathcal{L}$, for example it is linear, or has non constant coefficients given by polynomials of degree $k < n$, we obtain after applying the Laplace transform

$$\tilde{G}(F, F', ..., F^{(k)}, f(0), ..., f^{(n)}(0)) = H$$

with some new set of initial conditions. The hope is that this is easier to solve and the solution F has an easy to find Laplace inverse.

Remark B.3. (1) We will not explicit $\mathcal{L}^{-1}$ or show injectivity of $\mathcal{L}$. However in practice it suffices to find an appropriate candidate for this inverse and check explicitly that it satisfies the initial ODE.

(2) The case discussed above is particular, and can already be solved by standard methods of ODE resolution. However beyond providing us an alternative means of resolution, the Laplace transform also gives a transform to an often physically meaningful domain (in electrical engineering, the s-space is often called the frequency domain for reasons similar to the Fourier transform, while the t-space is the time domain). It also makes exact computer resolution of ODEs computationally simpler.

Before seeing explicit examples of this procedure, we need to have a toolkit of a few know Laplace transforms.

Example B.1. A quick integral computation shows that $\mathcal{L}[1](s) = s^{-1}$, defined on $\mathbb{R}_{>0}$. We can use this with the translation property of $\mathcal{L}$ to immediately obtain $\mathcal{L}[e^{at}](s) = (s-a)^{-1}$. Differentiating in s gives $\mathcal{L}[t^n](s) = n!s^{-(n+1)}$. Finally, we may observe that

$$\mathcal{L}[\cos(bt) + i\sin(bt)] = \frac{1}{s-ib} = \frac{s+ib}{s^2+b^2} \implies \mathcal{L}[\cos(bt)] = \frac{s}{s^2+b^2}, \mathcal{L}[\sin(bt)] = \frac{b}{s^2+b^2}$$

We thus have the following preliminary table of Laplace transforms:

Laplace transforms	
$f(t)$	$\mathcal{L}[f](s)$
C	Cs^{-1}
t^n	$n!s^{-(n+1)}$
e^{at}	$(s-a)^{-1}$
$\cos(bt)$	$s(s^2+b^2)^{-1}$
$\sin(bt)$	$b(s^2+b^2)^{-1}$

Let us now run a few example ODEs

Example B.2. (1)

$$\begin{cases} y'' - 2y' + y = 0 \\ y(0) = 1 \\ y'(0) = 1 \end{cases}$$

Applying $\mathcal{L}$ gives

$$s^2Y - sy(0) - y'(0) - 2sY + 2y(0) + Y = 0$$

which has solution (after applying initial conditions) $Y = \frac{s-2}{(s-1)^2}$. It now remains to find a Laplace inverse. We give two methods. Note that $Y(s+2) = s\frac{1}{(s+1)^2} = s\mathcal{L}[t](s+1)$. Thus

$$Y(s) = (s-2)\mathcal{L}[t](s-1) = (s-2)\mathcal{L}[te^t] = \mathcal{L}[(te^t)' - 2te^t] = \mathcal{L}[e^t - te^t]$$

we thus claim $y(t) = e^t(1-t)$ is a solution of our ODE. This can be verified by plugging it back in (thus all computations above are only formal in nature). Alternatively we could write $Y = \frac{1}{s-1} - \frac{1}{(s-1)^2}$ and run similar computations using properties (2) and (4) of the Laplace transform to find the same answer.

(2)

$$\begin{cases} y'' + y = \sin(2t) \\ y(0) = 0 \\ y'(0) = 1 \end{cases}$$

Then applying $\mathcal{L}$ we obtain

$$Y = \frac{1}{s^2+1}\left(\frac{2}{s^2+4} + 1\right)$$

Note that $(s^2+1)^{-1} = \mathcal{L}[\sin(t)]$. We thus get, after a partial fractions decomposition

$$Y = \frac{2}{3}\left(\frac{1}{1+s^2} - \frac{1}{s^2+4}\right) + \mathcal{L}[\sin(t)] = \mathcal{L}\left[\frac{5}{3}\sin(t) - \frac{1}{3}\sin(2t)\right]$$

and one may again verify that $y = \frac{5}{3}\sin(t) - \frac{1}{3}\sin(2t)$ solves the initial ODE.

Main bibliography

[Aru21] J. Aru. *Probability*. Lecture Notes, 2021.

[Buf22] B. Buffoni. *Analyse Fonctionnelle I*. Lecture Notes, 2022.

[Dac] B. Dacorogna. *Analyse avancée pour mathématiciens*. Lecture Notes.

[Fol99] G. B. Folland. *Real analysis. Modern Techniques and Their Applications*. Second. Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, 1999, pp. xvi+386.

[Sch15] B. Schlein. *Stochastik*. Lecture Notes, 2015.

[SS03] E. M. Stein and R. Shakarchi. *Fourier analysis. An Introduction*. English. Vol. 1. Princeton Lect. Anal. Princeton, NJ: Princeton University Press, 2003.

[SS05] E. M. Stein and R. Shakarchi. *Real analysis. Measure theory, integration, and Hilbert spaces*. Vol. 3. Princeton Lectures in Analysis. Princeton University Press, Princeton, NJ, 2005, pp. xx+402.

[Tao16] T. Tao. *Analysis. II*. Third. Vol. 38. Texts and Readings in Mathematics. Electronic edition of [MR3310023]. Hindustan Book Agency, New Delhi; Springer, New Delhi, 2016, pp. xvii+220.

Additional references

[Can84] G. Cantor. *Über unendliche, lineare Punktmannigfaltigkeiten*. Vol. 2. Teubner-Archiv zur Mathematik [Teubner Archive on Mathematics]. Arbeiten zur Mengenlehre aus den Jahren 1872–1884. [Papers on set theory from the years 1872–1884], Edited and with a foreword and commentary by G. Asser. BSB B. G. Teubner Verlagsgesellschaft, Leipzig, 1984, p. 180.

[Fle94] J. F. Fleron. "A note on the history of the Cantor set and Cantor function". In: *Math. Mag.* 67.2 (1994), pp. 136–140.

[Smi75] H. J. S. Smith. "On the integration of discontinuous functions". English. In: *Proc. Lond. Math. Soc.* 6 (1875), pp. 140–153.

[Vol81] V. Volterra. "Some observation on pointwise discontinuous functions." Italian. In: *Battaglini G.* 19 (1881), pp. 76–87.