

REALA
NDFUN
CTION
ALANA
LYSIS

fubini ⊗ tonelli

edizioni e convoluzioni

REAL AND FUNCTIONAL ANALYSIS

THE DEADLIEST COURSE OF MATHEMATICAL ENGINEERING

fubini ⊗ tonelli
edizioni e convoluzioni

© The authors. Some rights reserved

This work is licensed under CC BY-NC-SA 4.0.

<http://creativecommons.org/licenses/by-nc-sa/4.0/>

In particular, without the authors' permission, it is forbidden to make digital or printed copies in order to sell them.

The L^AT_EX source code for this book is available at

<https://github.com/fubinitonelli/rfa>

fubinitonelli.it - info@fubinitonelli.it

DOCUMENT CREATED ON NOVEMBER 27, 2023

Version 3b9b14f7ef09c0d269b03a678ff681457a99b1fd

DEVELOPED WITH GIT  BY:

GABRIELE GABRIELLI - gabriele@fubinitonelli.it

TEO BUCCI - teo@fubinitonelli.it

GIOELE CERRI - gioele@fubinitonelli.it

BRUNO GUINDANI - bruno@fubinitonelli.it

Preface

This book is a collection of notes for the “Real and Functional Analysis” (RFA, but universally known as ARF) course held in Politecnico di Milano, in the Mathematical Engineering Master’s Degree.

It was written over the course of many years, with contributions by several Fubini-Tonelli authors: Gabriele Gabrielli, Teo Bucci, Gioele Cerri, Bruno Guindani. We would also like to thank some external collaborators who helped in correcting some mistakes: Filippo Cipriani, Gabriele Corbo, Marco Lucchini.

This is also the first collection of lecture notes to be entirely written in English, which allows Fubini-Tonelli to expand in yet another area of knowledge, and hopefully among many new readers. Unfortunately, the language barrier means that we will not be able to write as many bad jokes as we did in the Italian books, but we hope that readers will enjoy the journey nonetheless.

Real and Functional Analysis is a mandatory stepping stone for all future M.Sc. mathematical engineers, and it is notoriously one of the most difficult exams of this degree, perhaps *the* most difficult one. Therefore, we immediately warn the readers: it will *not* be easy, and you *will* spend sleep-deprived weekend nights trying to understand how the heck bidual Banach spaces interact with Ascoli-Arzelà’s theorem. But also, you *will* encounter Fubini-Tonelli’s theorem eventually, so there is really nothing to be worried about in the end.

The authors

Contents

1	Fundamentals of Set Theory and Topology	9
1.1	Set Theory	9
1.1.1	Set conventions and main operations	10
1.1.2	Relations and functions	15
1.1.3	Magnitude and Infinities	22
1.1.4	Main results of the axiomatic theory	31
1.2	Topology	34
1.2.1	Metric spaces	34
1.2.2	Topological structure and sequences on metric spaces	39
1.2.3	Topological spaces	43
1.2.4	Topological structure of topological spaces	45
1.2.5	Limits and continuity for functions	51
1.2.6	Other topological notions	56
1.2.7	Complete spaces	58
1.2.8	Compact spaces	60
1.2.9	The Cantor set and the Vitali function	67
2	Real Analysis	79
2.1	Measure theory	80
2.1.1	Measurable spaces	80
2.1.2	Measurable functions	82
2.1.3	Positive measures	90
2.1.4	Lebesgue measure	98
2.2	Abstract integration	107
2.2.1	Abstract integral	107
2.2.2	Monotone convergence theorem and Fatou's lemma	109
2.2.3	Derivative of a measure: definitions	114
2.3	Lebesgue Integral	118
2.3.1	Integrating real valued functions	118
2.3.2	Dominated convergence theorem	119
2.3.3	The almost everywhere concept	122
2.3.4	Derivative of a measure: properties	130
2.3.5	Metric spaces and convergences	132
2.4	Fundamental theorems of calculus	144
2.4.1	First fundamental theorem of calculus	144
2.4.2	Absolute continuity	146
2.4.3	Second fundamental theorem and bounded variation functions	149

2.4.4	Further results	156
2.5	Integrals on product spaces	163
2.5.1	Product measures	163
2.5.2	Integrals and Fubini–Tonelli’s theorems	167
3	Functional Analysis	171
3.1	Banach spaces	171
3.1.1	Vector spaces	171
3.1.2	Normed vector spaces	174
3.1.3	Banach spaces	182
3.1.4	L^p spaces	191
3.1.5	Convergence and separability in L^p spaces	200
3.2	Linear operators	208
3.2.1	Definition of linear operator and boundedness	208
3.2.2	Isomorphism	214
3.2.3	Baire’s categories	217
3.2.4	Uniform boundedness principle	221
3.2.5	Open mapping, bounded inverse mapping and closed graph theorem	224
3.3	Duality	229
3.3.1	Dual spaces	229
3.3.2	Hahn–Banach theorem and consequences	232
3.3.3	Reflexivity	236
3.3.4	Weak convergence	242
3.3.5	Linear compact operators	249
3.4	Hilbert spaces	258
3.4.1	Hilbert spaces	258
3.4.2	Orthogonality and projection	264
3.4.3	Dual of a Hilbert Space and Riesz’s representation theorem	268
3.4.4	Proof of the Radon–Nikodym Theorem	271
3.4.5	Bilinear Forms and Lax Milgram lemma	274
3.4.6	Bases in Hilbert spaces	281
3.4.7	Linear Bounded Operators in Hilbert Spaces	289
	Appendices	303
	Appendix A Limits	303
	Appendix B Additional proofs	305

1 Fundamentals of Set Theory and Topology

This chapter's aim is to provide some basic tool that are required before studying the Measure Theory. Here we try to figure out what is a *set* and what is *infinity*. Later we will introduce some notions of Topology.

1.1 Set Theory

Sets are of fundamental importance for mathematics, but it may be surprisingly difficult to understand what a set actually *is*. This problem arises due to limitations in language: common language is not a reliable tool to deal with logical or abstract concepts. The entry “set” in the Oxford English Dictionary takes up a large number of pages, which just goes to show that understanding the topic is not an easy task. As if it was not enough, the *mathematical* concept of set presents more difficulties, as explained by Bertrand Russel through the so-called “Barber paradox”:

The barber is the “one who shaves all those, and those only, who do not shave themselves”. The question is, does the barber shave himself?

Another formulation, more formal, of the same paradox is the following: “What is the set of all sets that are not members of themselves?”

The mathematicians' solution for the definition of set is an **axiomatic theory**, which is an abstract theory based on axioms: you don't have to explain something which is arbitrarily defined as true. This theory claims that there exist particular objects, called sets, which satisfy a list of axioms. That list was written in order to avoid paradoxes. Such kind of theory is really difficult to grasp, and was constructed by Ernst Zermelo and Abraham Fraenkel (ZF) at the beginning of the 20th century.¹

Nevertheless, Georg Cantor worked on a *naïve set theory*, which simplifies many concepts from ZF theory in order to be more usable, at expense of rigor. We will follow this approach, which doesn't provide a rigorous definition of set, but it uses only its intuitive notion: this theory defines the set as a “collection” of objects called elements. The Cantor naïve set theory will be our reference framework for set theory.

¹To understand how complex the theory is, just know that the definition of empty set is provided at page 200.

1.1.1 Set conventions and main operations

Basic notions and definitions First of all, now we provide the conventions that are used in this text:

- **sets** are denoted with upper case letters:

$$A, B, X, Y$$

- **elements** of sets are denoted with lower case letter:

$$a, b, x, y$$

- **belonging** of an element to a set is denoted with the symbol “ \in ”: the notation $a \in A$ means that the element “ a ” *belongs to* the set “ A ”;
- **not belonging** of an element to a set is denoted with the symbol “ \notin ”: the notation $a \notin A$ means that the element “ a ” *does not belong to* the set “ A ”.

The following is an example of recursive definition of a set: $A = \{x : x \in A\}$, this namely means that A is the set of all the elements x of A .

Intuitively, two sets A and B are equal if and only if each element of A belongs to B and each element of B belongs to A ; that is, in symbols:

$$A = B \iff (\forall a \in A \implies a \in B) \wedge (\forall b \in B \implies b \in A).$$

A set with only one element, like $\{a\}$, is called **singleton**.

If X has a “reasonable” finite number of elements, say a, b, c, d than we can represent the set listing its elements:

$$X = \{a, b, c, d\} = \{d, a, b, c\}$$

Notice that order does not matter for sets.

We state the following axiom: “there exists a set such that no element belongs to it”, this particular set is named as **emptyset**, and can be defined as follows:

$$\emptyset = \{x : x \neq x\}.$$

Observe that the empty set is unique; this can be proved by the condition on equality between two set that was stated before: if no element belongs

to A (thus A is an empty set) and no element belongs to B (thus B is an empty set), then A is equal to B .

Definition 1.1.1

Sets can also exist as elements of another set, which is sometimes referred to as **collection** or **family** of sets, in particular:

- **inclusion:** $A \subseteq B$

The set A is included in the set B , that is every element which belongs to A also belongs to B :

$$x \in A \implies x \in B.$$

- **strict or proper inclusion:** $A \subsetneq B$

The set A is included in B and some elements of B doesn't belong also to A :

$$(x \in A \implies x \in B) \wedge A \neq B.$$

Then, if A is strictly contained in B , that is $A \subsetneq B$, the set A is a **proper subset** of B .

A set A can be determined, and defined, as the collection of all the elements belonging to a set "universe" X which has a certain property $P(x)$, by writing:

$$A = \{x \in X : P(x) \text{ is true}\}.$$

Main set operations Now we introduce the fundamental operations between sets:

Definition 1.1.2

The following operation are defined for two sets, further some of them are defined in case of family of sets:

- **union:** takes every elements which belong to at least one set:

$$A \cup B := \{x : (x \in A) \vee (x \in B)\};$$

- **intersection:** takes only the elements which belong to both of

sets:

$$A \cap B := \{x : (x \in A) \wedge (x \in B)\};$$

- **difference:** takes only the elements which belong to the first set and do not belong to the second one:

$$A \setminus B = A - B := \{x \in A : x \notin B\};$$

- **complement of a set** with respect to a given universe X such that $A \subset X$:

$$C_X A = A^C := \{x \in X : x \notin A\};$$

- **symmetric difference:** takes only the elements which belong to only one of the two sets:

$$A \Delta B := (A \cup B) \setminus (A \cap B) = \{x : (x \in A \vee x \in B) \wedge \overline{(x \in A \wedge x \in B)}\}.$$

Moreover, if $A \cap B = \emptyset$ then A and B are **disjoint**.

Some operations can be defined also for been applied not only to two but to multiple set. In case of unions:

- **union for infinite sets:**

$$\bigcup_{j=0}^{\infty} A_j := \{x : \exists i \in \mathbb{N} : x \in A_j\};$$

- **union of sets indexed by a set:**

$$\bigcup_{\alpha \in J} A_\alpha := \{x : \exists \alpha \in J : x \in A_\alpha\};$$

- **union of sets belonging to a family of sets:**

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

Similar in case of intersections:

- **intersection for infinite sets:**

$$\bigcap_{j=0}^{\infty} A_j := \{x : \forall i \in \mathbb{N} : x \in A_j\};$$

- **intersection of sets indexed by a set:**

$$\bigcap_{\alpha \in J} A_\alpha := \{x : \forall \alpha \in J : x \in A_\alpha\};$$

- **intersection of sets belonging to a family of sets:**

$$\bigcap_{A \in \mathcal{A}} A := \{x : \forall A \in \mathcal{A} : x \in A\}.$$

Each operation has properties related to it. The most basic ones are for example **commutativity** and **associativity**. In the case of union and intersection of sets as defined above, we have these **distributive** properties:

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C), \quad (A \cap B) \cup C = (A \cup C) \cap (B \cup C),$$

and the **De Morgan's Laws**:

$$(A \cup B)^C = A^C \cap B^C, \quad (A \cap B)^C = A^C \cup B^C.$$

These holds also for multiple sets, for instance:

$$\left(\bigcup_{\alpha \in J} A_\alpha \right)^C = \bigcap_{\alpha \in J} A_\alpha^C, \quad \left(\bigcap_{\alpha \in J} A_\alpha \right)^C = \bigcup_{\alpha \in J} A_\alpha^C.$$

Often a universe set X is used: that is consider the set of all the possible elements. In this case, $\emptyset^C = X$ and $X^C = \emptyset$

Consider now the case in which we need to slice a set into multiple subsets. If every element of the original set will belong to only one subsets than those sets are mutually disjoint, so we have construct a partition of a set:

Definition 1.1.3

Let X be a non-empty set: a **partition** of the set X is a family of subset of X such that they are pairwise disjoint and their union is X itself:

$$X_i \in \mathcal{P}(X), \quad \bigcup_j X_j = X, \quad X_i \cap X_j = \emptyset \quad \forall i \neq j.$$

Power set Now we will introduce two new set operations, which give as a result a different “kind” of set from the initial ones.

Definition 1.1.4

The **power set**² of an another existing set X is the set whose elements are all the possible subsets of X :

$$\mathcal{P}(X) = \{A : A \subseteq X\}$$

It is important to become familiar with the formalism and the correct notation. Given $a \in X$, notice that:

- $a \in \mathcal{P}(X)$ is wrong, because a is an element with respect to X , while the elements of \mathcal{P} are sets with respect to X .
- $\{a\} \in \mathcal{P}(X)$ is correct.

For the same reasons, writing $\{a\} \in X$ is wrong, and the correct notation is $\{a\} \subset X$.

It should be known that the power set of the empty set is the set which contains the empty set, not the empty set itself:

$$\mathcal{P}(\emptyset) = \{\emptyset\}$$

Cartesian product Now consider two non-empty set A and B , we want to combine each element of A with every element of B into ordered pairs:

$$a \in A, b \in B \mapsto (a, b)$$

Notice that the first element of the couple belongs to the first set and similarly for the second element. So we consider the pair as a new element. We can define a set which contain all such elements:

²Italian translation: *insieme delle parti*.

Definition 1.1.5

Let A and B two non-empty sets.

The set which contains all the possible couples (a, b) where $a \in A$ and $b \in B$ is the **cartesian product** of A and B :

$$A \times B := \{ \text{all the ordered pairs } (a, b) : a \in A, b \in B \}.$$

Let's show some properties:

- $(a, b) = (\tilde{a}, \tilde{b})$ if and only if $a = \tilde{a}$ and $b = \tilde{b}$;
- in general the cartesian product is not commutative, that is: $A \times B \neq B \times A$, unless $A = B$;
- the cartesian product is associative: $(A \times B) \times C = A \times (B \times C)$;
- thanks to the associative property we can define the cartesian product for multiple sets: $X_1 \times X_2 \times \dots \times X_N$;
- from the previous point we define the following: $X^N := X \times \dots \times X$, N times.

The real number set \mathbb{R} The set of real number will be the setting in which we will work in next sections. We define a proper symbol for its cartesian product:

$$\mathbb{R}^N = \underbrace{\mathbb{R} \times \mathbb{R} \times \dots \times \mathbb{R}}_{n \text{ times}}.$$

1.1.2 Relations and functions

Now introduce the concept of functions, starting from the more general notion of relation.

Definition 1.1.6

Consider two sets $X, Y \neq \emptyset$.

A **binary relation** r from X to Y is a subset of $X \times Y$.

The **domain** of a relation is the set of all elements of X that belongs to the relation:

$$\text{dom}(r) := \{x \in X : (x, y) \in r \text{ for some } y \in Y\}.$$

The **range** of a relation is the set of all elements of Y that belongs to the relation:

$$\text{rng}(r) := \{y \in Y : (x, y) \in r \text{ for some } x \in X\}.$$

When $(x, y), (x, z) \in r \implies y = z$, then r is a single valued relation; this kind of relation plays an important role in mathematics, they are known also as functions:

Definition 1.1.7

A law (rule) $f : X \rightarrow Y$ is a **function** (or mapping, map, transformation) if it associates to an element of X one and only one element of Y .

If $X = \mathbb{N}$, then this function is called **sequence**.

This definition was formulated by Dirichlet as follows:

y is a function of a variable x defined on the interval $a < x < b$ if to every value of the variable x in this interval there corresponds a definite value of the variable y . Also, it is irrelevant in what way this correspondence is established.³

A function can be applied also to an entire set: consider for example $A \subset X$, then:

$$f(A) := \{y \in Y : (x, y) \in f \text{ for some } x \in \text{dom}(f) \cap A\}.$$

There exist different ways to write functions. If we want to highlight the relation to the elements instead, we will use one of those notation:

$$f : x \mapsto y, \quad y = f(x)$$

which are equivalent to $(x, y) \in f$, where y is the **image** of x and x is the **preimage**⁴ (or counterimage, inverse image) of y .

In case we want to highlight how two set are related through a function f , then we will denote that as:

$$f : X \rightarrow Y$$

In this case the set Y image of the set X , while X is the preimage of Y .

³Cited in I. Kleiner, Evolution of the Function Concept: A Brief Survey, College Mathematics Journal vol. 20 iss. 4, page 291.

⁴Italian translation: *controimmagine*.

Definition 1.1.8

The **inverse function** of a function $f : X \rightarrow Y$ is denoted by f^{-1} and, if $B \subseteq Y$, is defined as:

$$f^{-1}(B) = \{x \in X : x = f^{-1}(y) \text{ for some } y \in B\}.$$

Here we state some properties with respect to union and intersection:

Proposition 1.1.9 (Set function properties)

Consider the sets $A, A_j \subset X$ and $B_j \subset Y$ for $j = 1, 2$.

Then:

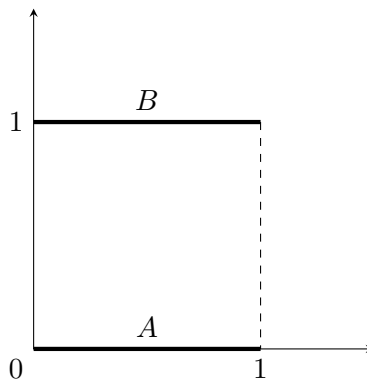
- $f^{-1}(A^C) = (f^{-1}(A))^C$;
- $f(A_1 \cup A_2) = f(A_1) \cup f(A_2)$;
- $f(A_1 \cap A_2) \subseteq f(A_1) \cap f(A_2)$;
- $f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2)$;
- $f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2)$.

Pay attention: the third property is not an equivalence like the other ones, but it states an inclusion. Let us give a counterexample.

Example 1.1.10. Let $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, $f(x, y) \mapsto (x, 0)$ be the function that projects on the x axis.

Let also $A = \{(x, 0) : x \in [0, 1]\}$, and $B = \{(x, 1) : x \in [0, 1]\}$. So we have $A \cap B = \emptyset$, and thus $f(A \cap B) = \emptyset$.

But $f(B) = A = f(A)$, and finally $f(A) \cap f(B) = A$.



Functions can have different behavior with respect to how many times they relate each element of a set to another:

Definition 1.1.11

A function $f : \text{dom}(f) \rightarrow Y$ is:

- an **injection** if $f(x) \neq f(\tilde{x})$ for $x \neq \tilde{x}$;
- a **surjection** if $\text{rng}(f) = Y$ or equivalently $f(\text{dom}(f)) = Y$;
- a **bijection** if it is both surjective and injective.

Now consider the case in which we have a function $f : X \rightarrow Y$ and a function $g : Y \rightarrow Z$. If every image of f is in the domain of g than we can build a new function from X to Z through a composition of f and g :

Definition 1.1.12

Let $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ be two functions.

Then the **composition** of g with f is defined by:

$$g \circ f := \{(x, z) \in X \times Z : \exists y \in Y : (x, y) \in f \wedge (y, z) \in g\}.$$

It is also possible to define a binary relation from a function:

$$r(x, y) = f(x) - y, \quad r = \{(x, y) \in X \times Y : r(x, y) = 0\}.$$

In this case $(x, y) \in r$ if and only if $f(x) - y = 0$, that is $r(x, y) = 0$.

Order relations We have seen that the relations are a generalization of functions, so functions are a sort of special kind of relations. There exists also other notably kind of relations, one of those are used to define an order inside a set:

Definition 1.1.13

Let $X \neq \emptyset$, $r \subseteq X \times X$ is an **order relation** in X if the following properties holds (let $x, y, z \in X$):

- **reflexivity:**

$$(x, x) \in r;$$

- **antisymmetry:**

$$((x, y) \in r \wedge (y, x) \in r) \iff x = y;$$

- **transitivity:**

$$((x, y) \in r \wedge (y, z) \in r) \implies (x, z) \in r.$$

A generic order relation is denoted by \preceq .

A famous order relation is the canonical order in \mathbb{N} or in \mathbb{Q} or in \mathbb{R} : this order is the principle that allow as to say that a number is greater to another. This particular relation is denoted by \leq . For example: $3 \leq 5$ or $7 \geq 2$. Observe that the strict “version” of the symbol ($<$) doesn’t represent an order as it isn’t reflexive.

With the canonical order we can always compare two number and say which is greater. This peculiarity doesn’t belong to order relation in general:

Definition 1.1.14

An order relation $r \subseteq X \times X$ is a **total order** in X if:

$$(x, y) \in r \vee (y, x) \in r \quad \forall x, y \in X,$$

otherwise r is a **partial order**.

If exists a total order relation in a set X , then X is a totally ordered set.

The inclusion between sets (\subset) is, for example, a partial order in $\mathcal{P}(X)$. Also the divisibility between two number in \mathbb{N} , that is $n \preceq m$ if n divides m (m is a multiple of n), is a partial order.

Equivalence relation Like order relation, it exists another kind of binary relation with wide applications:

Definition 1.1.15

Let $X \neq \emptyset$, $r \subseteq X \times X$ is an **equivalence relation** in X if the following properties holds (let $x, y, z \in X$):

- **reflexivity:**

$$(x, x) \in r;$$

- **symmetry:**

$$(x, y) \in r \iff (y, x) \in r;$$

- **transitivity:**

$$((x, y) \in r \wedge (y, z) \in r) \implies (x, z) \in r.$$

Equivalence relations are denoted by the symbol \sim .

Those properties are similar to the ones of the equality symbol ($=$), but while the equality state that two element are the same, an equivalent relation states that two elements are equivalent with respect to a certain property. Here some examples of equivalent relations:

Example 1.1.16. Consider the set of all straight lines in a given planes. Than the parallelism of two lines is an equivalent relation in such set.

Example 1.1.17. Fix $q \in \mathbb{N}$, then

$$x \sim y \iff \exists k \in \mathbb{Z} : x - y = kq$$

is an equivalence relation in \mathbb{Z} .

Example 1.1.18. The relation

$$(a, b) \sim (c, d) \iff a + d = b + c$$

is an equivalence relation in $\mathbb{N} \times \mathbb{N}$.

Example 1.1.19. The relation

$$(a, b) \sim (c, d) \iff ad = bc$$

is an equivalence relation in $\mathbb{Z} \times \mathbb{Z}_0$.

To sum up with all this kind of relations, consider these two proposition $(x, y) \in r$ and $(y, x) \in r$. If r is an order relation then both proposition can be false at the same time. If r is a total order, the at least one of the two can be true. If r is an equivalence relation, then they must be both true or both false.

Equivalence class We said that an equivalence relation states that two elements are “equivalent” with respect to a certain property. Now consider if we need to divide the elements of a set with respect to this kind of equivalence. We would create a partition of our set where in each subset all the elements are equivalent:

Definition 1.1.20

Consider a set $X \neq \emptyset$ and an equivalence relation \sim in $X \times X$.

We define the **equivalence class** of x on X with respect to \sim as follows:

$$[x] := \{y \in X : y \sim x\}.$$

The element x is called the **equivalence class representative** of $[x]$.

We define the **quotient set** of X the collection of all the equivalence classes of X , and we denote as:

$$\frac{X}{\sim} := \{[x] : x \in X\}.$$

Let’s do some considerations from these definition. If an element $y \in X$ is equivalent to $x \in X$, then they are both representative of the same class. Moreover if y belongs to the equivalence class whose x is the representative, again x and y are representative of the same class:

$$x \sim y \implies [x] = [y], \quad y \in [x] \implies [x] = [y].$$

If two equivalence class aren’t disjoint, then they are the same equivalence class. If two equivalence class are disjoint, then their representative are not in the equivalence relation:

$$[x] \cap [y] \neq \emptyset \implies [x] = [y], \quad [x] \cap [y] = \emptyset \implies x \not\sim y$$

Another trivial result is that the union of all the equivalence class is the set X , and the elements of the quotient set are a partition of X .

Recalling the previous example:

Example 1.1.21. Considering $[x]$ as the set of all straight lines parallel to x , the quotient set $\frac{X}{\sim}$ can be identified as the set of all the possible direction of the plane.

Example 1.1.22. Considering $[x]$ as the congruence class of x modulo q , fixing for instance $q = 5$ we have:

$$\mathbb{Z}_5 = \{[0], [1], [2], [3], [4]\}.$$

Example 1.1.23. The set of integer numbers can be defined as follows:

$$\mathbb{Z} := \frac{\mathbb{N} \times \mathbb{N}}{\sim}.$$

Example 1.1.24. Also the correct way to define the rational number set is through a quotient set:

$$\mathbb{Q} := \frac{\mathbb{Z} \times \mathbb{Z}_0}{\sim};$$

indeed, there exist several representations of the same object: $\frac{p'}{q'} = [(p, q)]$.

1.1.3 Magnitude and Infinities

Here we provide a quote from Galileo's "Two New Sciences" in which a paradox caused by the common idea of infinity set is presented. The characters of this dialogue are: *Simplicio*, who is the stereotype of scientist of his period, *Salviati*, who is the voice of Galileo, and *Sagredo*, who acts like a moderator:

Simplicio: Here a difficulty presents itself which appears to me insoluble. Since it is clear that we may have one line greater than another, each containing an infinite number of points, we are forced to admit that, within one and the same class, we may have something greater than infinity, because the infinity of points in the long line is greater than the infinity of points in the short line. This assigning to an infinite quantity a value greater than infinity is quite beyond my comprehension.

Salviati: This is one of the difficulties which arise when we attempt, with our finite minds, to discuss the infinite, assigning to it those properties which we give to the finite and limited; but this I think is wrong, for we cannot speak of infinite quantities as being the one greater or less than or equal to another. To prove this I have in mind an argument which, for the sake of clearness, I shall put in the form of questions to Simplicio who raised this difficulty. I take it for granted that you know which

of the numbers are squares and which are not.

Simplicio: I am quite aware that a squared number is one which results from the multiplication of another number by itself; thus 4, 9, etc., are squared numbers which come from multiplying 2, 3, etc., by themselves.

Salviati: Very well; and you also know that just as the products are called squares so the factors are called sides or roots; while on the other hand those numbers which do not consist of two equal factors are not squares. Therefore if I assert that all numbers, including both squares and non-squares, are more than the squares alone, I shall speak the truth, shall I not?

Simplicio: Most certainly.

Salviati: If I should ask further how many squares there are one might reply truly that there are as many as the corresponding number of roots, since every square has its own root and every root its own square, while no square has more than one root and no root more than one square.

Simplicio: Precisely so.

Salviati: But if I inquire how many roots there are, it cannot be denied that there are as many as the numbers because every number is the root of some square. This being granted, we must say that there are as many squares as there are numbers because they are just as numerous as their roots, and all the numbers are roots. Yet at the outset we said that there are many more numbers than squares, since the larger portion of them are not squares. Not only so, but the proportionate number of squares diminishes as we pass to larger numbers, Thus up to 100 we have 10 squares, that is, the squares constitute $1/10$ part of all the numbers; up to 10000, we find only $1/100$ part to be squares; and up to a million only $1/1000$ part; on the other hand in an infinite number, if one could conceive of such a thing, he would be forced to admit that there are as many squares as there are numbers taken all together.

Sagredo: What then must one conclude under these circumstances?

Salviati: So far as I see we can only infer that the totality of all numbers is infinite, that the number of squares is infinite, and that the number of their roots is infinite; neither is the number

of squares less than the totality of all the numbers, nor the latter greater than the former; and finally the attributes “equal”, “greater”, and “less”, are not applicable to infinite, but only to finite, quantities. When therefore Simplicio introduces several lines of different lengths and asks me how it is possible that the longer ones do not contain more points than the shorter, I answer him that one line does not contain more or less or just as many points as another, but that each line contains an infinite number.

Galileo Galilei, *Two New Sciences*, 1638

Magnitude of sets We introduce some definitions in order to construct a rigorous classification of sets based on their “largeness”.

Definition 1.1.25

We say that sets X and Y have the **same power** or are **equipotent**, and we write “ $A \sim B$ ”, if there exists a bijection $f : X \rightarrow Y$.

So when we can establish a one-to-one correspondence between two sets, these sets have the same power.

Now we pose the problem: how to recognize if a set contains an infinite number of elements? Well, the answer is exactly the definition of infinite set, which was provide by Dedekind, like lots of other results in this area:

Definition 1.1.26 (by Dedekind)

A set X is an **infinite** set if it is equipotent to one or more of its proper subsets. Otherwise X is called **finite**.

Notice that if two set are finite, they are equipotent if and only if they have the same number of elements.

Example 1.1.27. The set \mathbb{N} is infinite, indeed $\mathbb{P} = \{2n : n \in \mathbb{N}\} \subsetneq \mathbb{N}$, and

$$f : \mathbb{N} \rightarrow \mathbb{P}, \quad f : n \mapsto 2n$$

is a bijection.

Example 1.1.28. The set \mathbb{R} is infinite, indeed

$$f(x) = \frac{1}{\pi} \arctan(x) + \frac{1}{2}, \quad x \in \mathbb{R}, \quad f : \mathbb{R} \rightarrow (0, 1)$$

is a bijection.

Countable and uncountable sets Not all infinite sets contains the “same number” of elements: there are some infinite sets that are “larger” than others. Let’s introduce a distinction between infinite sets:

Definition 1.1.29 (by Dedekind)

A set is **countable**⁵ if it is equipotent to \mathbb{N} or to one of its subsets. Otherwise, the set is called **uncountable**.

Then we can say that all finite sets are countable. As we will examine later, countably infinite sets are the “smallest” kind of infinite set. Indeed, one can prove the following:

Proposition 1.1.30

Every infinite set contains a countably infinite subset.

Example 1.1.31. The sets \mathbb{N} , \mathbb{Z} , \mathbb{Q} are countable.

Example 1.1.32. The set of **algebraic numbers** \mathcal{A} is the set whose elements are roots of an algebraic polynomial with rational coefficients. This set is countable; to prove that, notice that countable unions of countable sets are also countable. See [B.0.2 on page 305](#) for further reading.

Observe that $\mathbb{Q} \subsetneq \mathcal{A}$, indeed $\sqrt{2} \in \mathcal{A}$ while $\sqrt{2} \notin \mathbb{Q}$, it is enough to consider the polynomial $P(x) = x^2 - 2$.

Example 1.1.33. The set \mathbb{R} is uncountable. This is proved by contradiction: the proof is known as *Cantor’s argument*.

Example 1.1.34. The set $\mathbb{R} \setminus \mathcal{A}$ is uncountable. The elements of this set, such as π and e , are called **transcendent numbers**.

Cardinal numbers As we understand that exists sets that are bigger than others and sets that are equal between them, we want to find a label that represent how big are the set. To do that we have to extend the notion of natural numbers:

⁵Italian translation: *numerabile*.

Definition 1.1.35

The **cardinality** of a set X , or its **cardinal number**, is the equivalence class of all the sets which are equipotent to X . This is denoted by $m(X)$, where m stands for magnitude.

If X is finite, then $m(X)$ is identified by the numbers of elements in X . As the cardinal numbers are a proper extension of the natural number, of course we are going to see if an ordering is possible also for cardinals.

Now consider two sets, A and B such that either:

- A and B are equipotent: $A \sim B$;
- A is in some sense “smaller” than B , that is:

$$(\exists \tilde{B} \subset B : \tilde{B} \sim A) \wedge (\nexists \tilde{A} \subset A : \tilde{A} \sim B)$$

(that is, it does exist a subset of B that is equipotent to A , but at the same time it does not exist a subset of A that is equipotent to A);

in this case we can introduce a partial order in the set of cardinal number by setting:

$$m(A) \preceq m(B).$$

So we succeed in introducing a partial order, but can we introduce also a total ordering among the cardinals?

Theorem 1.1.36 (Cantor–Bernstein)

These two propositions hold:

- if $(\exists \tilde{B} \subset B : \tilde{B} \sim A) \wedge (\exists \tilde{A} \subset A : \tilde{A} \sim B)$, then $m(A) = m(B)$;
- it never happens that $(\exists \tilde{B} \subset B : \tilde{B} \sim A) \wedge (\nexists \tilde{A} \subset A : \tilde{A} \sim B)$.

The second point is not universal, it depends on which axioms are chosen during the development of the theory. Specifically, to have the second proposition, we assume the *axiom of choice*, which will be discussed later (see definition 1.1.47 on page 31). By using this axiom, we can introduce a total order among the cardinals, as well as sum and multiplication operations:

$$m(A) + m(B) := m(A \cup B), \quad m(A) \cdot m(B) := m(A \times B).$$

Those operation are associative and commutative.

There is not an upper limit to the cardinal numbers, as there is not a “largest set”: this because whatever set do you consider, its power set has a greater cardinality. This was proved by Cantor (see proposition 1.1.38):

Theorem 1.1.37 (Cantor)

For any set X :

$$m(X) \preceq m(\mathcal{P}(X)) \quad \text{and} \quad m(X) \neq m(\mathcal{P}(X)).$$

It can be proven that:

Proposition 1.1.38

Let X be a set. The magnitude of $\mathcal{P}(X)$ is $2^{m(X)}$:

$$m(\mathcal{P}(X)) = 2^{m(X)}.$$

Magnitude of \mathbb{N} We are now going to discuss the magnitude of two important numerical sets: the set of natural numbers and the set of real numbers. The “smallest” kind of infinite belongs to the set \mathbb{N} and has an appropriate name:

Definition 1.1.39

The magnitude of the set of natural numbers is called **aleph-zero**:

$$\aleph_0 := m(\mathbb{N}).$$

As we said before, it does always exists a larger set: it can be proven that the set of real numbers is equipotent to the power set of natural numbers, see proposition 1.1.40.

Magnitude of \mathbb{R} Also the cardinality of the set of real numbers has its own symbol, first acknowledge this concept.

Proposition 1.1.40

The set or real numbers is equipotent to the power set of natural

numbers:

$$\mathbb{R} \sim \mathcal{P}(\mathbb{N}).$$

Proof. We have to prove that $|\mathcal{P}(\mathbb{N})| = |\mathbb{R}|$.

Proof of $|\mathbb{R}| = |(0, 1)|$:

To prove this is sufficient considering the following bijective function and its inverse:

$$f(x) = \frac{1}{\pi} \arctan x + \frac{1}{2} \quad f^{-1}(x) = \tan\left(\pi\left(x - \frac{1}{2}\right)\right).$$

Proof of $|(0, 1)| = |[0, 1]|$:

See proposition [1.1.42 on the facing page](#) for the general case.

Proof of $|[0, 1]| = |\mathcal{P}(\mathbb{N})|$:

Consider $x \in [0, 1]$, its binary expression is the following:

$$z = \sum_{k=1}^{\infty} \alpha_k \frac{1}{2^k}$$

where $\alpha_k \in \{0, 1\} \quad \forall k$.

To determine α_k we divide $[0, 1]$ into 2^k intervals of length $\frac{1}{2^k}$:

$$I_1(k) = \left[0, \frac{1}{2^k}\right), I_2^{(k)} = \left[\frac{1}{2^k}, \frac{2}{2^k}\right) \cdots I_j^{(k)} = \left[\frac{j-1}{2^k}, \frac{j}{2^k}\right) \cdots I_{2^k}^{(k)} = \left[\frac{2^k-1}{2^k}, 1\right)$$

So it exists a unique j such that $x \in I_j^{(k)}$ and we choose $\alpha_k = 0$ if j is odd or $\alpha_k = 1$ if j is even.

Now we define the map:

$$f : [0, 1] \rightarrow \mathcal{P}(\mathbb{N}) \\ x \mapsto A = \{n \in \mathbb{N} : \alpha_n = 1\}$$

and this is a bijection that ends the proof. ■

So the set of real numbers is far more larger than the set of natural numbers, they are not equipotent and so \mathbb{R} is not countable. The following is the definition of its cardinality.

Definition 1.1.41

The magnitude the set of real numbers is called **continuum**:

$$c := m(\mathbb{R}) = 2^{\aleph_0}.$$

Now, due to Cantor theorem, there will be a more “powerful” infinite than the continuum, consider the magnitude of $\mathcal{P}(\mathbb{R})$. An interesting question is about the existence of an “intermediate” infinite between \aleph_0 and the continuum. We will see the answer to this question later.

Proposition 1.1.42

Let $a, b \in \mathbb{R}$, with $a < b$. Then:

$$[a, b] \sim (a, b) \sim [a, b) \sim (a, b].$$

In general, it is not easy to construct a bijection between an open and a closed set. Indeed, it would not be continuous, as continuous functions are the ones which map open sets into open sets.

Ordinal numbers With cardinal number we labeled set according to how “big” they are. Now we want to label set according also to their order.

Definition 1.1.43

Let A, B be two totally ordered sets: $(A, \leq), (B, \preceq)$.

We say that these sets have the same **order** or **type** if there is a bijection $f : A \rightarrow B$ which preserves order, namely:

$$\forall x, y \in A, \quad x \leq y \implies f(x) \preceq f(y).$$

If A and B have the same order, by definition they also have the same power. In general, the converse is not true. For instance, \mathbb{N} can be totally ordered in an uncountable number of ways: if we choose $A = B = \mathbb{N}$ but two different total orders, A and B have obviously the same power, but an appropriate f does not exist.

Definition 1.1.44

Let A be a totally ordered set.

We say that A is **well-ordered** if any non-empty subset of A has a minimum.

For example, \mathbb{N} is well-ordered with respect to its canonical order: we can find a minimum for each of its subsets. Instead, \mathbb{Z} is not well-ordered with respect to its canonical order: for example, any half-line such as $\{z \leq 2 : z \in \mathbb{Z}\}$ has no minimum.

We will now introduce the concept of *ordinal numbers*, as we did with cardinals.

Definition 1.1.45

An **ordinal number** is a equivalence class of well-ordered sets with the same order.

As for cardinals, if a well-ordered set is finite then its ordinal can be identified with the number of its element.

The difference between cardinal numbers and ordinals is the same as the difference between “one, two, three ...” and “first, second, third ...”.

Ordinal numbers of infinite well-ordered sets are called *transfinite*. The transfinite associated to $(\mathbb{N}, \preceq_{\text{canonical}})$ is called ω .

How to define a total order among the ordinals? It's sufficient, among well-ordered set, to compare their cardinal numbers. If two sets A and B are well-ordered, then we can compare $m(A)$ and $m(B)$: then either $m(A) \preceq m(B)$ or $m(B) \preceq m(A)$; now consider A and B as representatives of their respective equivalence classes, and introduce an order for ordinals based on the cardinality of those representatives.

We could introduce the operations of sum and multiplication among ordinals, but unlike cardinals, they would not be commutative.

Having find a total order for ordinals by comparing them with the cardinal and, in a certain sense, using the total order among the cardinal to obtain a total order for the ordinals, we now must remember that this was possible only if we include the *axiom of choice* in our theory (see theorem 1.1.36 on page 26). Indeed there is a theorem that closes the circle, and this theorem need that axiom to be proved:

Theorem 1.1.46 (Zermelo)

Any set can be well ordered.

Extension of \mathbb{R} The set of real numbers can be extended to include the infinities;

$$\mathbb{R}^* := \mathbb{R} \cup \{-\infty, +\infty\};$$

in this case, the conventional extension of the arithmetization is as follows, considering $a \in \mathbb{R}$:

$$a + \infty = +\infty + a = +\infty \quad 0 \cdot +\infty = 0 \quad a \cdot +\infty = +\infty \text{ if } a > 0$$

the multiplication's sign rule holds also in these cases.

1.1.4 Main results of the axiomatic theory

In our introduction we gave a taste of how deep is the theory behind set theory. Now, as this theory plays a role in the fundamentals of analysis we now show some details of those axioms which allow us to reach useful results.

Here we will present the aforementioned axiom of choice.

Definition 1.1.47 (Axiom of choice - AC)

There are different but equivalent formulations for the axiom:

- For any $X \neq \emptyset$ there exists a choice function $f : \mathcal{P}(X) \rightarrow X$, $f : A \mapsto f(A) \in A$ for any $A \neq \emptyset$;
- Let $(E_\alpha)_{\alpha \in A}$ be a family of non-empty sets E_α , indexed by an index set A . Then we can find a family $(x_\alpha)_{\alpha \in A}$ of elements x_α of E_α , indexed by the same set A ; ⁶
- Any cartesian product of a family of non-empty sets is non-empty: let $\chi = \{X_\alpha\}_{\alpha \in J}$ be a family of sets indexed by a set $J \neq \emptyset$. The cartesian product $\prod_{\alpha \in J} X_\alpha$ is defined as the set of all mappings $x : J \rightarrow \cup_{\alpha \in J} X_\alpha$ such that $x(\alpha) \in X_\alpha$ for each $\alpha \in J$. Each mapping x is defined as a choice mapping and $x(\alpha)$ is defined as the α th coordinate of x . Then if $X_\alpha \neq \emptyset$ for each $\alpha \in J$ then $\prod_{\alpha \in J} X_\alpha \neq \emptyset$;
- Any member of a family of non-empty sets has at least one element.

⁶See: T. Tao, An introduction to Measure Theory, page XV, notation, axiom 0.0.4 .

This axiom means that it is always possible to choose one element in each set of a collection of sets through a so-called choice function.

The family is assumed to be non-countable; if it is just countable, a weaker version of the axiom, known as the *countable* axiom of choice (CAC), holds.

One interesting consequence arising from the axiom of choice is the following.

Proposition 1.1.48 (Banach–Tarski paradox)

Given a solid ball in 3-dimensional space, there exists a decomposition of the ball into a finite number of disjoint subsets, which can then be put back together in a different way to yield two identical copies of the original ball.

Despite the existence of this paradox, the axiom of choice is essential to prove even some of the most basic calculus theorems.

Equivalent axioms Other axioms can be take instead axiom of choice, and some of those has been proved to be equivalent: here we present two of those after having introduced some definitions.

Definition 1.1.49

Let X, \preceq be a partially ordered set. His subset $A \subset X$ is called **chain** if it's a totally ordered set.

A chain is **maximal** if it's not properly contained in another chain.

An element b of a partially ordered set (X, \preceq) is an **upper bound** for $A \subset X$ if $a \preceq b$ for all $a \in A$.

A partially ordered set (X, \preceq) is an **inductive set** if every chain has an upper bound.

Proposition 1.1.50 (Hausdorff's maximal principle)

Every chain in a partially ordered set is contained in a maximal chain.

Proposition 1.1.51 (Zorn's lemma)

Any inductive set has a maximal element.

The Zermelo–Fraenkel theory and the axiom of choice Ernst Zermelo demonstrated in 1904 that every set can be well-ordered if we assume the axiom of choice. In fact, AC is independent from ZF axioms: Kurt Gödel proved that the negation of AC isn't deducible from ZF in 1938 while Paul Cohen, in 1963, proved that AC neither can be deduced from ZF.

From now on two (different) theories are possible: the Zermelo–Fraenkel theory with AC (ZFC), in which cardinal numbers has a total order and it is possible to prove a number of important results even if there exist some counter-intuitive results, and the Zermelo–Fraenkel theory without AC. In this last case one can include only the countable axiom of choice and lots of results can be proved anyway.

The continuum hypothesis Remembering the ordinal number theory, we can assign to two particular transfinite ordinal two specific symbols. The least transfinite ordinal is ω and we denote its cardinality with \aleph_0 . The first transfinite uncountable ordinal, that is the set of all countable ordinals, has its cardinality denoted by \aleph_1 . Those set are themselves ordinals, in the sense that they are representatives from their equivalence classes.

Definition 1.1.52 (Continuum hypothesis (CH))

The continuum has the same cardinality of the first transfinite uncountable ordinal, namely:

$$2^{\aleph_0} = \aleph_1$$

So, this states that there is not an “intermediate” magnitude of infinities between the cardinality of \mathbb{N} and the cardinality of \mathbb{R} (see definition [1.1.39 on page 27](#)).

As for the axiom of choice, in 1940 Kurt Gödel proved that the negation of the CH is not a theorem in ZFC while in 1963 Paul Cohen showed that CH is not a theorem in ZFC; so the CH is another optional axiom of the ZF.

From now on, we will use $ZF + AC + CH$ as our reference theory.

1.2 Topology

With the study of topology we can obtain many useful tools to develop our theory. First we will define the notion of metrical space and we will compare it with the notion of topological spaces.

1.2.1 Metric spaces

First we need to deal with the concept of distance: we have to obtain an abstract tool to quantify how close are two elements of the same set:

Definition 1.2.1

Let X be a non-empty set.

A function $d: X \times X \rightarrow [0, +\infty)$ is called **distance** or **metric** on X if $\forall x, y, z \in X$ these properties hold:

- **non-negativity:**

$$d(x, y) \geq 0;$$

- **identity of indiscernibles:**⁷

$$d(x, y) = 0 \iff y = x;$$

- **symmetry:**

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

- **triangle inequality or subadditivity:**

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

Observe that for this definition we had no need of an algebraic structure on X . It is possible to define multiple distances on the same set. Here we present some broadly-used distances.

Example 1.2.2. On a generic non-empty set X we can define the **discrete distance** as follows:

$$d(x, y) = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y \end{cases}.$$

⁷Italian translation: *annullamento*.

Example 1.2.3. On \mathbb{R}^N we can define the **euclidean distance**:

$$d_e(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

which is a particular case of the following ($p = 2$):

$$d_p(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}}.$$

For exercise check that d_e is a distance.⁸

Example 1.2.4. From the previous formula, with particular values of p , other metric can be defined:

$$d_1(x, y) = \sum_{i=1}^n |x_i - y_i|, \quad d_\infty(x, y) = \max_{i=1, \dots, n} |x_i - y_i|.$$

Considering a set with a metric we get a powerful environment on which we can build useful tools:

Definition 1.2.5

If d is a distance on $X \neq \emptyset$, then the couple (X, d) is called **metric space**.

Open balls or spherical neighbourhood are the sets that contain all points which are “close” from a given point. The geometry of those balls depends on the metric considered.

Definition 1.2.6

Let (X, d) a metric space, $x_0 \in X$ and $r > 0$.

The **open ball** or **spherical neighbourhood** of center x_0 and radius r is defined as:

$$B_d(x_0, r) := \{x \in X : d(x, x_0) < r\}$$

⁸Use Cauchy–Schwarz inequality: $\sum_{k=1}^n |a_k b_k| \leq (\sum_{k=1}^n a_k^2)^{\frac{1}{2}} (\sum_{k=1}^n b_k^2)^{\frac{1}{2}}$

Clearly different distances will produce differently shaped balls in \mathbb{R}^N (see figure 1.1):

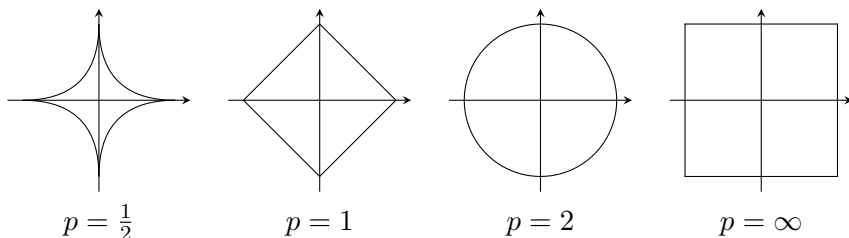


Figure 1.1: balls in different distances on \mathbb{R}^2

Example 1.2.7. Let d be the discrete metric on X . Then we have:

$$B_d(x_0, r) = \begin{cases} \{x_0\} & \text{if } r \leq 1 \\ X & \text{if } r > 1 \end{cases}.$$

On function spaces distances and neighborhoods can be defined, consider $X = C^0([a, b])$ and these two possible distances:

$$d_1(f, g) := \int_a^b |f(x) - g(x)| dx, \quad d_\infty(f, g) := \max_{x \in [a, b]} |f(x) - g(x)|.$$

These two metrics generate very different neighborhoods (see figure 1.2 on the facing page):

- $B_{d_1}(f, r)$ is the set of functions in X subject to $\int_a^b |f - g| < r$;
- $B_{d_\infty}(f, r)$ is instead the set of functions in X whose graphs is contained in a “strip” around f .

Metrics equivalence What differences are there between a distance and another? Some times two distances are very “similar” each other:

Definition 1.2.8

Let $X \neq \emptyset$ and d and δ two different metrics on X .

We say that d and δ are **equivalent** if there exist c_1 and $c_2 > 0$ such

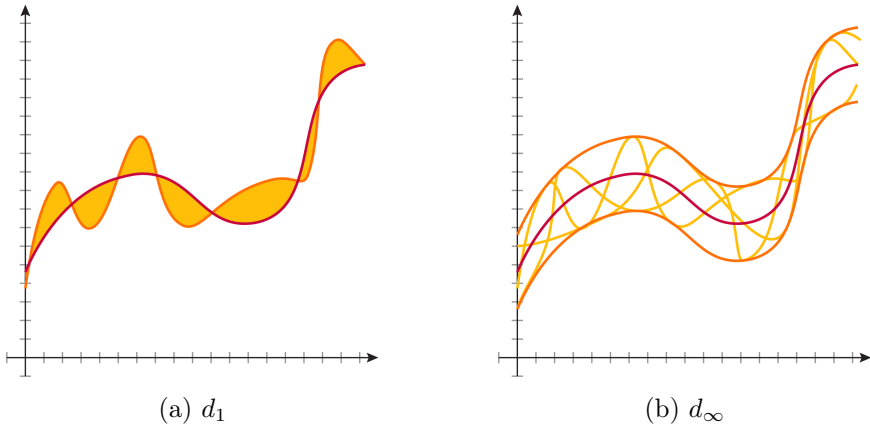


Figure 1.2: spherical neighbourhoods in different distances on \mathcal{C}^0

that:

$$c_1 \delta(x, y) \leq d(x, y) \leq c_2 \delta(x, y) \quad \forall x, y \in X$$

Consider now two equivalent metrics, namely d and δ , observe that:

$$\begin{aligned} y \in B_d(x, r) &\implies d(x, y) < r \implies \delta(x, y) < \frac{r}{c_1} \\ &\implies y \in B_\delta\left(x, \frac{r}{c_1}\right) \quad \forall x \in X \quad \forall r > 0 \end{aligned}$$

In the same way we have: $y \in B_\delta(x, r) \implies y \in B_d(x, r c_2)$: the open balls with respect to d and δ , and of two equivalent metrics in general, are nested one inside another.

Example 1.2.9. Now we prove that d_1, d_e, d_∞ on \mathbb{R}^N are equivalent:

$$\begin{aligned} d_\infty(x, y) &= \max_{k=1, \dots, n} |x_k - y_k| = |x_{\bar{k}} - y_{\bar{k}}| \\ &= \sqrt{(x_{\bar{k}} - y_{\bar{k}})^2} \leq \sqrt{\sum_{k=1}^n (x_k - y_k)^2} = d_e(x, y) \end{aligned}$$

Now remember that if $a_1, \dots, a_n \geq 0$ then $a_1^2 + \dots + a_n^2 \leq (a_1 + \dots + a_n)^2$:

$$d_e(x, y) = \left(\sum_{k=1}^n (x_k - y_k)^2 \right)^{\frac{1}{2}} \leq \left[\left(\sum_{k=1}^n |x_k - y_k| \right)^2 \right]^{\frac{1}{2}} = d_1(x, y)$$

Then we just need to prove that $d_1 \leq c_1 d_\infty$:

$$d_1(x, y) = \sum_{k=1}^n |x_k - y_k| \leq \sum_{k=1}^n \max_{i=1, \dots, n} |x_i - y_i| = n \cdot d_\infty(x, y)$$

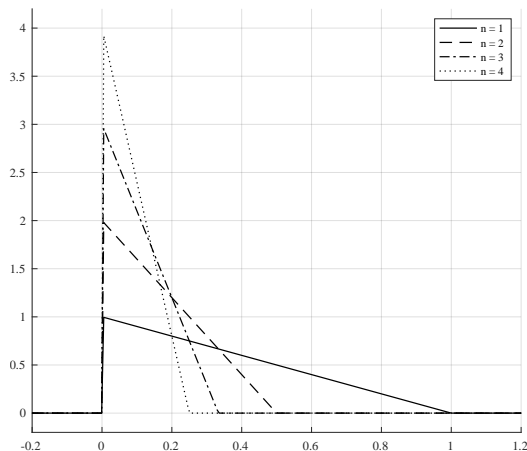
In the end we have $d_\infty(x, y) \leq d_e(x, y) \leq d_1(x, y) \leq c_1 d_\infty(x, y)$, and the thesis is proven.

Example 1.2.10. The distances d_1 and d_∞ on the space $\mathcal{C}^0([a, b])$ are *not* equivalent. Indeed the distance d_1 can be controlled by the distance d_∞ :

$$d_1(f, g) = \int_a^b |f(x) - g(x)| dx \leq \int_a^b d_\infty(f, g) dx = d_\infty(f, g) (b - a)$$

However d_∞ cannot be controlled by d_1 : let us prove that with a counterexample. Take $[a, b] = [0, 1]$ and define f_n as follows:

$$f_n(x) = \begin{cases} n - n^2x & \text{if } x \in [0, \frac{1}{n}] \\ 0 & \text{if } x \in [\frac{1}{n}, 1] \end{cases}$$



Then we have:

$$d_1(f_n, 0) = \int_0^{\frac{1}{n}} (n - n^2x) dx = \left[nx - \frac{n^2x^2}{2} \right]_0^{\frac{1}{n}} = 1 - \frac{1}{2} = \frac{1}{2},$$

and

$$d_\infty(f_n, 0) = n.$$

Hence, there is no $c > 0$ such that $d_\infty(f_n, 0) \leq c \cdot d_1(f_n, 0)$ for any n .

1.2.2 Topological structure and sequences on metric spaces

Thanks to the notion of metric and the definition of balls we can now build a topological structure on metric spaces:

Definition 1.2.11

Let (X, d) be a metric space, $A \subset X$, $x_0 \in X$. We say that:

- x_0 is an **interior point** of A if there exists $r > 0$:

$$B_r(x_0) \subset A;$$

- x_0 is an **exterior point** of A if there exists $r > 0$:

$$B_r(x_0) \subset (X \setminus A) = A^C;$$

- x_0 is a **boundary point** of A if it is neither interior nor exterior;
- x_0 is a **adherent point** of A if for all $r > 0$:

$$B_r(x) \cap A \neq \emptyset$$

so any point of A is an adherence point of A ;

- x_0 is an **accumulation point** (or limit point or cluster point) of A if for all $r > 0$ exists x_n :

$$x_n \in \{B_r(x_0) \cap A\} \setminus \{x_0\};$$

- x_0 is an **isolated point** of A if:

$$x_0 \in A \text{ and } \exists r > 0 : B_r(x_0) \cap A = \{x_0\}.$$

Moreover, the set of each “kind” of point can be identified with a proper name:

Definition 1.2.12

Let (X, d) be a metric space, $A \subset X$, $x_0 \in X$, than:

- the **interior** set of A is defined as:

$$\overset{\circ}{A} = \text{int}(A) := \{x_0 \in A : x_0 \text{ is an interior point of } A\};$$

- the **exterior** set of A is defined as:

$$\text{ext}(A) := \{x_0 \in A : x_0 \text{ is an exterior point of } A\};$$

- the **boundary** of A is defined as:

$$\partial A := \{x_0 \in A : x_0 \text{ is a boundary point of } A\};$$

- the **closure** of A is defined as:

$$\bar{A} := \{x_0 \in A : x_0 \text{ is an adherent point of } A\};$$

- the **derived** set of A is defined as:

$$A' := \{x_0 \in X : x_0 \text{ is an accumulation point of } A\}.$$

This allow us to specify whether a set is open or closed:

Definition 1.2.13

The set A is **open** if $A = \overset{\circ}{A}$; that is every $x \in A$ is an interior point.
The set A is **closed** if $X \setminus A$ (or A^C) is open.

With these new concepts we can extend the notion of disjoint sets (see definition 1.1.2 on page 11):

Definition 1.2.14

We say that two sets, A and B , are **almost disjoint**, or disjoint except for boundaries, if

$$\overset{\circ}{A} \cap \overset{\circ}{B} = \emptyset.$$

Now we highlight some basic properties.

Proposition 1.2.15 (Topological properties of sets)

Let $A \subset X$, then:

- the sets $\overset{\circ}{A}$, ∂A and $\text{ext}(A)$ are a partition of X ;
- A is open $\iff A = \overset{\circ}{A} \iff A \cap \partial A = \emptyset$;
- $\bar{A} = A \cup \partial A = \overset{\circ}{A} \cup \partial A$;

- A is closed $\iff A = \bar{A} \iff \partial A \subseteq A$;
- $A' = \bar{A} \setminus \{\text{isolated points of } A\}$;
- $\overset{\circ}{A}$ is the largest open set containing A ;
- \bar{A} is the smallest closed subset of X containing A .

Proposition 1.2.16 (Topological property for family of sets)

Now, let I be a family of indexes of any cardinality (countable or uncountable) and $\{A_j\}_{j \in I} \subset \mathcal{P}(X)$ and $m \in I$:

- if A_j is a open set $\forall j \in I$, then $\bigcup_{j \in I} A_j$ is open;
- if A_1, \dots, A_m are open sets, then $\bigcap_{j=1}^m A_j$ is open;
- if A_j is a closed set $\forall j \in I$, then $\bigcap_{j \in I} A_j$ is closed;
- if A_1, \dots, A_m are closed sets, then $\bigcup_{j=1}^m A_j$ is closed.

So uncountable union of open set is an open set while uncountable intersection of closed set is a closed set.

Notice that the first two properties are the complementary to the last two due to De Morgan's law.

Sequences in metric spaces Much of the previous definition about distances allow us to study the sequences defined in metric spaces, in particular we now discuss about their convergence. A sequence is, formally, a function which domain is \mathbb{N} (see definition 1.1.7 on page 16). In a less formal language, a sequence can be considered an ordered set of elements which are indexed on \mathbb{N} . We will denote a sequence as follows:

$$\{x_n\}_{n \in \mathbb{N}}.$$

During theory deployment we will be interested in discover the behavior of a given sequence, for example if it converges or not. Here we start crafting tools useful for such goal.

Definition 1.2.17

Let (X, d) be a metric space, $\{x_n\}$ be a sequence in X , $x^* \in X$.

Then the **limit** of the sequence $\{x_n\}$ is x^* , namely $x_n \rightarrow x^*$, as

$n \rightarrow \infty$ if:

$$\forall \varepsilon > 0 \quad \exists \bar{n} = \bar{n}(\varepsilon) : n > \bar{n} \implies d(x_n, x^*) < \varepsilon$$

or, equivalently using open balls:

$$\forall \varepsilon > 0 \quad \exists \bar{n} = \bar{n}(\varepsilon) : n > \bar{n} \implies x_n \in B_\varepsilon(x^*).$$

When such limit does exist, then the sequence is said to be **convergent**.

From this definition comes three remarks: first, the limit is unique and any sub-sequence $\{x_{n_k}\}$ converges to the same limit: to prove this use the triangular inequality and the identity of indiscernibles. Second, considering two different but equivalent metrics the limit is the same, namely let d_1, d_2 be equivalent distances on set X , then $x_n \xrightarrow{d_1} x^* \iff x_n \xrightarrow{d_2} x^*$. Third, there is a relation between a set and the sequences which it contains, in particular we have the following theorem:

Theorem 1.2.18 (theorem closure in metric spaces)

Let (X, d) be a metric space and consider $A \subset X$.

The subset A is closed if and only if, for any sequence $\{x_n\} \subset A$, we have that:

$$x_n \rightarrow x \implies x \in A.$$

This theorem provides a characterization for closedness in metric space. The condition highlighted is known as *sequentially closedness* and in case of metric space is completely equivalent to the closedness (see definition 1.2.55 on page 57, and following results). We will see that this is not true for topological spaces.

Necessary condition (\implies). :

Having that A is closed and any sequence $\{x_n\}$ such that $x_n \rightarrow x$, we have to prove that $x \in A$.

By contradiction assume that $x \in (X \setminus A)$, which is open by definition.

Then, it exist $r > 0$ such that $B_r(x) \subset (X \setminus A)$ and, equivalently, $B_r(x) \cap A = \emptyset$.

By definition 1.2.17 on the preceding page we can choose a point from the sequence which is arbitrarily close d to x , but for all n we have $x_n \notin B_r(x)$, so we have a contradiction.

Sufficient condition (\Leftarrow):

By hypothesis we have that any sequence $\{x_n\}_{n \in \mathbb{N}}$ converges to x which is a point of A . Choose $x \in \bar{A}$; we will prove that $x \in A$.

Consider $B_{\frac{1}{n}}(x)$, by definition of adherence point we can choose a sequence $\{x_n\} \subset A$ such that:

$$\forall n \exists x_n \in \left(A \cap B_{\frac{1}{n}}(x) \right).$$

Notice that $d(x_n, x) < \frac{1}{n} \rightarrow 0$, and so $x_n \rightarrow x$. Having choose $x_n \in A$ we have, by hypothesis, that $x \in A$.

Thus $A = \bar{A}$ and A is closed. ■

Having considered “all the possible sequences included in a set”, the following definition make sense:

Definition 1.2.19

The **sequence closure** of a set X is:

$$\{x \in X : \exists \{x_n\} \in X : x_n \rightarrow x \text{ in } (X, d)\}$$

Bolzano–Weierstrass Here we state a very useful result on which we will return to in next chapters.

Theorem 1.2.20 (Bolzano–Weierstrass)

Consider a metric space given by \mathbb{R}^N coupled with any metric.

Any bounded sequence contains at least one convergent subsequence.

This theorem states that when working on \mathbb{R}^N , having defined any metric, if we have any bounded sequence $\{x_n\}_{n \in \mathbb{N}}$ it is always possible extract a subset of its elements obtaining a subsequence $\{x_{n_h}\}_{h \in \mathbb{N}}$ which is convergent in the metric space. Note the convention of indexes and subindexes.

1.2.3 Topological spaces

Here we try a different approach to deal with “closedness”: alongside metric spaces we define the topological spaces from which we will obtain similar results.

Definition 1.2.21

Let $x \neq \emptyset$ and $\tau \subseteq \mathcal{P}(X)$.

We say that τ is a **topology** if:

- it contains the set itself and the empty set:

$$X \in \tau, \quad \emptyset \in \tau;$$

- it is closed with respect to finite intersections:

$$A, B \in \tau \implies A \cap B \in \tau;$$

- it is closed with respect to uncountable unions:

$$\{A_i\}_{i \in I} \subset \tau \implies \bigcup_{i \in I} A_i \in \tau.$$

In this case, we also say that (X, τ) is a **topological space**.

Moreover, we call **open sets** each set belonging to τ .

As we see, this definition also contains the notion of openness, and here is generalized (see definition 1.2.13 on page 40 for the definition of open set in metric spaces). Anyway, there exists a link between metric spaces and topological spaces:

Definition 1.2.22

Let (X, d) be a metric space.

We define the **topology induced by the distance d** as follows:

$$\tau = \{A \subset X : A \text{ is open with respect to } d\}.$$

We can prove that two metric spaces (X, d_1) and (X, d_2) , with d_1 equivalent to d_2 (see definition 1.2.8 on page 36), produce the same topology; this result is obtainable using the characterization of the closure in metric spaces.

As we define the open set as the set belonging to the topology, closed sets are the sets whose complement are open, namely $A \subset X$ is closed if $X \setminus A$ is open ($X \setminus A \in \tau$).

We can also identify a topological space by considering its closed subsets instead of the open ones. This is possible by the De Morgan's law (see section 1.1.1 on page 13) which allow us to define the *dual properties*. Consider \mathcal{C} as the collection of all the closed subsets of X , then if satisfy the dual properties, which are:

$$X, \emptyset, X \in \mathcal{C}; \quad C, D \in \mathcal{C} \implies C \cup D \in \mathcal{C}; \quad \{D_i\} \text{ family of } \mathcal{C} \implies \bigcap_i D_i \in \mathcal{C};$$

then the family of the complements $\{f^C : f \in \mathcal{C}\}$ is a topology.

Observe that \emptyset and X are *both* open and closed.

Consider now some examples: if $X \neq \emptyset$, $\tau = \mathcal{P}(X)$ is the topology associated with the discrete metric (indeed, all the set are open with respect to this metric), while $\tau = \{\emptyset, X\}$ is called the **trivial topology**; let $X = \{1, \dots, 4\}$: a possible topology of X is given by $\tau = \{\emptyset, X, \{1\}, \{2, 3\}, \{1, 2, 3\}\}$.

1.2.4 Topological structure of topological spaces

As we have done for metric spaces, also for topological spaces we can define a topological structure (see definition 1.2.11 on page 39 for a comparison):

Definition 1.2.23

Let (X, τ) be a topological space, $A \subset X$, $x_0 \in X$. We say that:

- U is an **open neighborhood** of x_0 if:

$$U \in \tau \text{ and } x_0 \in U;$$

- x_0 is an **interior point** of A if:

$$\exists U \in \tau \text{ such that } x_0 \in U \subseteq A;$$

- x_0 is an **exterior point** of A if it is interior to $X \setminus A$, namely:

$$\exists U \in \tau \text{ such that } x_0 \in U \subseteq (X \setminus A);$$

- x_0 is a **boundary point** of A if it is neither interior nor exterior;

- x_0 is an **adherent point** of A if:

$$\forall U \in \tau, \quad x_0 \in U \implies A \cap U \neq \emptyset;$$

- x_0 is an **accumulation point** of A if:

$$\forall U \in \tau, \quad x_0 \in U \implies (A \cap U) \setminus \{x_0\} \neq \emptyset;$$

- x_0 is an **isolated point** of A if:

$$x_0 \in A \text{ and } x_0 \text{ is not an accumulation point.}$$

The notions of interior set ($\overset{\circ}{A}$), exterior set ($\text{ext}(A)$), boundary set (∂A), closure set (\bar{A}) and derived set (A') are the same as for metric spaces, given this topological structure (see definition 1.2.12 on page 39). Also their properties still hold (see propositions 1.2.15 on page 40 and 1.2.16 on page 41).

Base of topological spaces The base of a topological space describes the space in its whole. It is a family of sets for which any set of the topological space can be expressed as an intersection of those.

Definition 1.2.24

Let (X, τ) be a topological space.

We say that $\mathcal{B} \subset \mathcal{P}(X)$ is a **base** for (X, τ) if for every $A \in \tau$ it exists a collection of sets $\{B_i\}_{i \in I} \subset \mathcal{B}$ such that $A = \bigcap_{i \in I} B_i$.

Indexes set I can be either finite, countable or uncountable.

Moreover, let $x \in X$. We say that a family $\mathcal{N}(x) \subset \mathcal{P}(X)$ is a **local base** of (X, τ) at x if for every open neighborhood A of x it exists a set $B \in \mathcal{N}(x)$ such that $x \in B, B \subset A$.

Considering a metric space (X, d) and the topology induced by the metric: then a local base at x is the set of all open balls centered in x : $\mathcal{N}(x) = \{B_r(x) : r > 0\}$. The union of these local bases is a base: $\mathcal{B} = \{B_r(x) : x \in X, r > 0\}$.

Topological subspace Can we build a topology for the subsets of X ?

Definition 1.2.25

Let (X, τ) be a topological space, and let S be a non-empty subset of X .

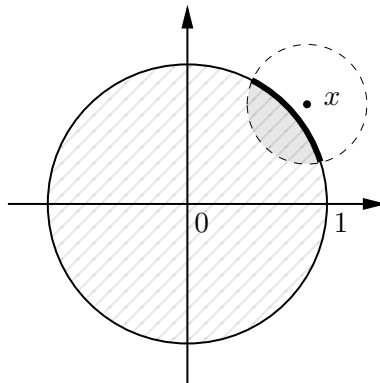
On S it is naturally defined the **topology induced by τ** as

$$\sigma := \{B \subset S : B = S \cap A \text{ where } A \in \tau\}.$$

It's easy to prove that (S, σ) is a topological space.

Moreover if $\mathcal{B} = \{B_n\}_{n \in I}$ is a base for τ , then $\mathcal{B}' = \{B_n \cap S\}_{n \in I}$ is a base for σ .

Example 1.2.26. Let (\mathbb{R}^N, τ) be a topological space with a topology induced by the euclidean metric d_e , and $S = \overline{B_1(0)}$. Then $B_r(x) \cap S$ is open in the induced topology of S even if it is not open as a subset of \mathbb{R}^N .



Separation axioms We have already seen in [1.2.22 on page 44](#) how to build a topological space from a metric space. Under which conditions can we do the opposite, namely construct a distance coherent with the given topology?

Definition 1.2.27

A topological space (X, τ) is called **first countable** if for any $x \in X$ it exists a local countable base $\mathcal{W}(x)$, and **second countable** if it exists a countable base \mathcal{B} for τ so that $|\mathcal{B}| = |\mathbb{N}| = \aleph_0$.

Metric spaces are always first countable with respect to the topology induced by the euclidean distance, but not always second countable: for ex-

ample consider a space made of an uncountable number of isolated points. We can prove that (\mathbb{R}^N, d_e) is second countable: see proposition 1.2.38 on page 51.

Definition 1.2.28

A topological space (X, τ) is said to satisfy the **Tychonoff property**⁹ if:

$$\forall x, y \in X, \exists U, V \in \tau : x \in U, y \notin U, y \in V, x \notin V.$$

Definition 1.2.29

A topological space (X, τ) is called **Hausdorff space** if for all $x, y \in X$, with $x \neq y$, it exists a pair U, V of open sets such that $x \in U, y \in V, U \cap V = \emptyset$.

Proposition 1.2.30

Any metric space is a Hausdorff space.

Proof. Let $x \neq y, \rho := d(x, y)$. Then $x \in U := B_{\rho/4}(x), y \in V := B_{\rho/4}(y)$. By definition of ball, $U \cap V = \emptyset$, so the pair U, V satisfies the requirements of the definition of Hausdorff spaces. ■

An example of a topological space which is not Hausdorff is (X, τ_0) with $\tau_0 = \{\emptyset, X\}$ (see proof of proposition 1.2.34 on the facing page). Another example is the following:

Example 1.2.31. Consider $X \neq \emptyset$ with infinitely many elements, and consider the following topology: $\tau = \{A \subset X : A = \emptyset \vee A^C \text{ is finite}\}$; (X, τ) is a topological space.

Now we show that it is not an Hausdorff space.

Consider $x, y \in X, x \neq y$. Suppose by contradiction that (X, τ) is an Hausdorff space, then exists $U, V \in \tau$ such that $x \in U, y \in V$ and $U \cap V = \emptyset$.

Since $U \in \tau$ then U^C is finite. For the same reason, V^C is finite. So we have:

$$U^C \cup V^C = (X \setminus U) \cup (X \setminus V) = X \setminus (U \cap V) = X \setminus \emptyset = X$$

⁹This property is referenced with “T1”.

But, as X is infinite it can't be a union of two finite set, then we have a contradiction.

Definition 1.2.32

(X, τ) is called **normal space** if, for any pair A, B of closed disjoint sets, it exists a pair U, V of open sets such that $A \subset U, B \subset V, U \cap V = \emptyset$.

Theorem 1.2.33 (Urysohn's metrization theorem)

Let (X, τ) be a second countable topological space.
It is metrizable if and only if it is a Hausdorff, normal space.

This result is said to be *the first non-trivial theorem in topology*.

A comparison between a metric space and the induced topological space In general we have the following result:

Proposition 1.2.34

Consider a metric space (X, d) and the topological space (X, τ) which topology is induced by the distance d , then:

$$\text{Metric space } (X, d) \subsetneq \text{Topological space } (X, \tau).$$

For example consider a topological space whose topology is not induced by any distance, like (X, τ_0) where $X \neq \emptyset$ and $\tau = \emptyset, X$ which is the trivial topology. If X contains at least two elements, then τ_0 is not induced by any distance.

Proof. Suppose by contradiction that exists a distance d on X such that: $\tau_d = \{\text{open sets of } X \text{ according to the metric } d\} = \tau_0 = \{\emptyset, X\}$.

Let $x, y \in X$ such that $x \neq y$, then $d(x, y) = k > 0$.

We can consider $B_d(x, \frac{k}{2})$ and $B_d(y, \frac{k}{2})$.

Both of these balls are non-empty and $B_d(x, \frac{k}{2}) \cap B_d(y, \frac{k}{2}) = \emptyset$.

Since they are open and non-empty, the only non-empty open set in τ_0 is X , so $B_d(x, \frac{k}{2}) = B_d(y, \frac{k}{2}) = X$, so their intersection can't be empty. ■

Sequences in topological spaces As we defined a limit for sequences in metric spaces using their topological structure we can define a notion of limit using the topological structure of the topological spaces:

Definition 1.2.35

Let (X, τ) be a topological space, $\{x_n\}$ be a sequence in X , $x^* \in X$, $U \subset X$ a neighborhood of x .

Then the **limit** of the sequence $\{x_n\}$ is x^* , namely $x_n \rightarrow x^*$, as $n \rightarrow \infty$ if:

$$\forall U \in \tau : x^* \in U \quad \exists \bar{n} = \bar{n}(U) : \forall n > \bar{n} \implies x_n \in U$$

When such limit does exist, then the sequence is said to be **convergent**.

In general, limits defined this way are not unique.

Example 1.2.36. Let $X := \{1, 2, 3, 4\}$, $\tau := \{\emptyset, X, \{1\}, \{2, 3\}, \{1, 2, 3\}\}$, $x_n = 3 \forall n$.

The sequence $\{x_n\}$, by definition, converges to 3 but also to 2: in fact, every open set containing 2 also contains 3.

Proposition 1.2.37

If (X, τ) is a Hausdorff space, then the limit of any sequence is unique.

In general the converse is not true, consider for instance the non-Hausdorff topological space (X, τ_0) with $\tau_0 = \{\emptyset, X\}$: we can show that if $\{x_n\}$ is a sequence in X then it converges to any element $x^* \in X$, indeed the only neighborhood of a point x^* is the whole X .

Topological results on \mathbb{R} The base for the \mathbb{R} canonical topology (the topology induced by the euclidean distance d_E) can be generated by

$$\{(x_0 - \delta, x_0 + \delta)_{x_0, \delta \in \mathbb{R} \delta > 0}\}.$$

A base for the canonical topology of \mathbb{R}^* is generate by

$$\{ \{[-\infty, s)_{s \in \mathbb{R}}\}, \{(x_0 - \delta, x_0 + \delta)_{x_0, \delta \in \mathbb{R} \delta > 0}\}, \{(r, +\infty]_{r \in \mathbb{R}}\} \}.$$

Proposition 1.2.38

The topological spaces (\mathbb{R}, τ_E) and (\mathbb{R}^N, τ_E) , where τ_E is the topology induced by the euclidean distance d_E , are second countable.

This because we can choose as base for the induced topology the family of open balls with rational radii¹⁰ and center with rational coordinates. See definition 1.2.27 on page 47 for second countable.

1.2.5 Limits and continuity for functions

With the tools that we have been developed, now we face the definition of a notion of continuity for functions (recall the definition 1.1.7 on page 16). A general notion for continuous functions is the following:

Definition 1.2.39

A function said to be is continuous if it transforms open sets into open sets.

A formal definition need to specify on which structure we are working, for which we develop the two different cases in the followings; as we are going to see, the definition of limits differs and the definition of continuity is strictly related to the definition of limit. For metric spaces see theorem 1.2.46 on page 54, for topological spaces see theorem 1.2.51 on page 55. We could also equivalently define continuous functions as the ones mapping closed sets in closed sets, but using open sets theory is simpler.

Limits and continuity for functions in metric spaces First, we develop and study the continuity in the context of metric spaces. The following definition is similar to 1.2.17 on page 41 where we defined the limit for sequences, now we extend that definition to functions in general.

Definition 1.2.40

Let (X, d_X) , (Y, d_Y) be metric spaces and consider a function $f : X \rightarrow Y$.

Then the **limit** of $f(x)$ is y_0 , namely $f(x) \rightarrow y_0$, as $x \rightarrow x_0$, with x_0 accumulation point for X , if:

$$\forall \varepsilon > 0 \quad \exists \delta = \delta(x_0, \varepsilon) > 0 : d_X(x, x_0) < \delta \implies d_Y(f(x), y_0) < \varepsilon$$

¹⁰Radii is the plural of radius, as it is a latin word.

or, equivalently using open balls:

$$\forall \varepsilon > 0 \quad \exists \delta = \delta(x_0, \varepsilon) > 0 : f(B_{d_X}(x_0, \delta) \setminus \{x_0\}) \subset B_{d_Y}(y_0, \varepsilon).$$

When such limit does exist, then we say that the function **converges** to y_0 in x_0 , using the following notation:

$$\lim_{x \rightarrow x_0} f(x) = y_0.$$

As for sequences, the limit, if exists, is unique.

Now we can provide the “classical” definition of continuous function:

Definition 1.2.41

Let (X, d_X) , (Y, d_Y) be metric spaces and consider a function $f : X \rightarrow Y$.

We say that f is **continuous** in x_0 if one of the following condition is satisfied:

- x_0 is an accumulation point of X and $f(x) \rightarrow f(x_0)$ as $x \rightarrow x_0$;
- x_0 is an isolated point for X .

We say that f is **continuous in a set** $D \in X$ if it is continuous in every point of D .

A stronger version of continuity is the following:

Definition 1.2.42

Let (X, d_X) , (Y, d_Y) be metric spaces and consider a function $f : X \rightarrow Y$.

We say that f is **uniformly continuous** in $D \subset X$ if:

$$\forall \varepsilon > 0 \quad \exists \delta = \delta(\varepsilon) > 0 \quad \forall x, y \in D : d_X(x, y) < \delta \implies d_Y(f(x), f(y)) < \varepsilon.$$

Notice that in this definition δ does not depend on x , but only on ε .

Proposition 1.2.43

Let (X, d_X) , (Y, d_Y) be metric spaces and consider a function $f : X \rightarrow Y$.

If f is uniformly continuous in D , then f is continuous in D . The

converse isn't true in general.

The following is a particular definition of continuity that occurs when functions converge like sequences in their domain:

Definition 1.2.44

Let (X, d_X) , (Y, d_Y) be metric spaces and consider a function $f : X \rightarrow Y$.

We say that f is **sequentially continuous** in x_0 if:

$$\forall \{x_n\} \subset X \quad x_n \rightarrow x_0 \implies f(x_n) \rightarrow f(x_0).$$

Proposition 1.2.45

Let (X, d_X) , (Y, d_Y) be metric spaces and consider a function $f : X \rightarrow Y$.

The function f is continuous in x_0 if and only if f is sequentially continuous in x_0 .

Proof. If x_0 is an isolated point the thesis is trivial. Otherwise, let x_0 be an accumulation point.

Necessary condition \implies :

Consider a sequence $\{x_n\} \subseteq X$ such that $x_n \rightarrow x_0$ and fix $\varepsilon > 0$.

We gain two consequences, f is continuous at x_0 , namely:

$$\forall \varepsilon > 0 \exists \delta > 0 : 0 < d_X(x_n, x_0) < \delta \implies d_Y(f(x_n), f(x_0)) < \varepsilon$$

and that x_n converges to x_0 , namely:

$$\exists \bar{n} > 0 : \forall n > \bar{n} \implies d_X(x_n, x_0) < \delta.$$

Then if $n > \bar{n}$ we have $d_Y(f(x_n), f(x_0)) < \varepsilon$, that is $f(x_n) \rightarrow f(x_0)$: this prove the implication.

Sufficient condition \longleftarrow :

By contradiction, if f isn't continuous in x_0 then:

$$\exists \bar{\varepsilon} > 0 : \forall \delta > 0 \exists x_\delta \in X : 0 < d_X(x_\delta, x_0) < \delta \quad \text{and} \quad d_Y(f(x_\delta), f(x_0)) \geq \bar{\varepsilon}. \quad (\star)$$

Let $\delta_n = \frac{1}{n}$, with $n \in \mathbb{N}$ and let $x_n = x_{\delta_n}$. Then $\{x_n\} \subseteq X$ and $d_X(x_n, x_0) < \frac{1}{n}$ meaning that $x_n \rightarrow x_0$.

By (\star) we have $d_Y(f(x_n), f(x_0)) \geq \bar{\varepsilon}$, that is $f(x_n)$ doesn't converge to $f(x_0)$ so f is not sequentially continuous. This proves the co-implication. ■

The general definition for continuity 1.2.39 on page 51 can be written as follows in terms of metric spaces:

Theorem 1.2.46 (Characterization of continuity in metric spaces)

Let $(X, d_X), (Y, d_Y)$ be metric spaces and consider a function $f : X \rightarrow Y$.

Then the function f is continuous in X if $\forall A \subset Y$ open set we have that $f^{-1}(A) \subset X$ is an open set.

The continuity is invariant with respect to equivalent distances:

Proposition 1.2.47

Let $(X, d_1), (X, d_2)$ be metric spaces, where d_1 is equivalent to d_2 .

A function $f : (X, d_1) \rightarrow (X, d_1)$ is continuous if and only if $f : (X, d_2) \rightarrow (X, d_2)$ is also continuous.

Proposition 1.2.48

Let $(X, d_X), (X, \tilde{d}_X)$ be metric spaces, such that $\exists C_1 > 0 : d_X \leq C_1 \tilde{d}_X$.

Let also $(Y, d_Y), (Y, \tilde{d}_Y)$ be metric spaces, such that $\exists C_2 > 0 : \tilde{d}_Y \leq C_2 d_Y$.

If $f : (X, d_X) \rightarrow (Y, d_Y)$ is continuous, then $f : (X, \tilde{d}_X) \rightarrow (Y, \tilde{d}_Y)$ is also continuous.

Proof. Let x_0 in X . By hypothesis, we have:

$$\forall \varepsilon > 0 \quad \exists \delta = \delta(x_0, \varepsilon) : d_X(x, x_0) < \delta \implies d_Y(f(x), f(x_0)) < \varepsilon$$

Take $\tilde{\varepsilon} > 0$ and define $\tilde{\delta}(\tilde{\varepsilon}) := \frac{1}{C_1} \cdot \delta\left(\frac{\tilde{\varepsilon}}{C_2}\right)$. Then we have:

$$\begin{aligned} \tilde{d}_X(x, x_0) < \tilde{\delta}(\tilde{\varepsilon}) &\implies d_X(x, x_0) < \delta\left(\frac{\tilde{\varepsilon}}{C_2}\right) \\ &\implies d_Y(f(x), f(x_0)) < \frac{\tilde{\varepsilon}}{C_2} \\ &\implies \tilde{d}_Y(f(x), f(x_0)) < \tilde{\varepsilon} \end{aligned}$$

Because $\tilde{\varepsilon}$ is arbitrary, $f : (X, \tilde{d}_X) \rightarrow (Y, \tilde{d}_Y)$ is continuous by definition. ■

Limits and continuity for functions in topological spaces The notion of continuity function is slightly different.

Definition 1.2.49

Let (X, τ_X) , (Y, τ_Y) be topological spaces, and consider a function $f : X \rightarrow Y$.

Then the **limit** of $f(x)$ is y_0 , namely $f(x) \rightarrow y_0$, as $x \rightarrow x_0$, with x_0 accumulation point for X , if:

$$\forall B \in \tau_Y : y_0 \in B \quad \exists A \in \tau_X : x \in A, f(A \setminus \{x_0\}) \subset B.$$

When such limit does exist, then we say that the function **converges** to y_0 in x_0 , using the following notation:

$$\lim_{x \rightarrow x_0} f(x) = y_0.$$

As for sequences in topological spaces, and as opposed to the metric space definition, the limit can be not unique.

The definitions of continuity and sequential continuity are the same ones already given for metric spaces (see definitions [1.2.41 on page 52](#) and [1.2.44 on page 53](#)), simply with a different notion of convergence. Anyway, there is a difference in their implications:

Proposition 1.2.50

Let (X, τ_X) , (Y, τ_Y) be topological spaces, and consider a function $f : X \rightarrow Y$.

If f is continuous in x_0 , then f is sequentially continuous in x_0 .

The converse is not true in general, as topological spaces cannot be described by sequences. But, if (X, τ_X) is a first countable topological spaces (see [1.2.27 on page 47](#)), then the double implication holds.

As we done in [1.2.46 on the facing page](#), we will now provide another version of the definition of continuity based on open sets (see [1.2.39 on page 51](#)), version adequate to the context of topological spaces:

Theorem 1.2.51 (Characterization of the continuity in topological spaces)

Let (X, τ_X) , (Y, τ_Y) be topological spaces, and consider a function

$f : X \rightarrow Y$.

Then the function f is **continuous** in X if:

$$\forall A \in \tau_Y \quad \text{we have} \quad f^{-1}(A) \in \tau_X.$$

1.2.6 Other topological notions

Now we will examine some particular properties of spaces. First we discuss a remark on topological spaces, then we will present the sequentially closed space and the concepts of density and separability. Further sections will develop the theory with the notions of completeness and compactness.

The following notion allow us to compare two topologies of the same set:

Definition 1.2.52

Let τ_1 and τ_2 be topologies on X .

We say that τ_2 is **weaker** than τ_1 , and τ_1 is **stronger** than τ_2 , if:

$$\tau_2 \subseteq \tau_1.$$

So $\tau = \emptyset$ is the weakest topology, while $\tau = \mathcal{P}(X)$ is the strongest.

Consider the setting of the definition and a continuous function $f : (X, \tau_2) \rightarrow (Y, \tau_y)$: then $f : (X, \tau_1) \rightarrow (Y, \tau_y)$ is continuous as well. This is easy to show as every preimage of an open set of τ_y is an open set belonging to τ_1 .

These definitions are instead about metric spaces only:

Definition 1.2.53

Consider a metric space (X, d) , with $E \subset X$.

The **diameter** of E is

$$\text{diam}(E) := \sup_{x, y \in E} d(x, y).$$

We say that E is **bounded** if $\text{diam}(E) < +\infty$.

We say that E is **totally bounded** if it can be “covered” with finitely

many balls:

$$\forall \varepsilon > 0 \exists \{x_i\}_{i=1}^n : E \subset \bigcup_{i=1}^n B(x_i, \varepsilon).$$

Sequentially closed set This is an alternative definition of closedness for sets.

Definition 1.2.54

We say that $A \subset X$ is **sequentially closed** if for every sequence $\{x_n\} \subset A$, $x_n \rightarrow x^* \implies x^* \in A$.

What is the relation between closed sets and sequentially closed sets? There are any differences between metric space and topological space environment? We summarize all those results in the following; recall theorem closure 1.2.18 on page 42 to get the big picture.

Proposition 1.2.55

Let (X, d_X) be a metric space.

Then $A \in X$ is closed if and only if it is also sequentially closed.

Let (X, τ_X) be a topological space.

If $A \in X$ is closed, then it is also sequentially closed. In general, the converse is not true.

Density and separability Topological spaces can have some particular properties that are very useful. Later in the theory we will discover which benefits separability implies.

Definition 1.2.56

Let (X, τ) be a topological space and $A, B \subset X$, with $A \subset B$.

We say that A is **dense** in B if $\bar{A} = B$.

We say that A is **everywhere dense** if A is dense in X itself.

We say that A is **nowhere dense**¹¹ if $\overset{\circ}{\bar{A}} = \emptyset$ ¹².

Recalling the definition of closure \bar{A} , $\bar{A} = B$ means that if $x \in B$ then every open neighborhood of x intersects A .

¹¹Italian translation: *mai denso*.

¹²This means that a nowhere dense set is a set such that the interior of its closure is empty. In other words, a set is nowhere dense if its closure does not contain any balls.

The same for $\bar{A} = X$: this means that $\forall D \in \tau, D \cap A \neq \emptyset$.

Example 1.2.57. \mathbb{Q} is everywhere dense in \mathbb{R} , id est $\bar{\mathbb{Q}} = \mathbb{R}$. This because $\{(a, b), a < b\}$ is a base for the standard topology of \mathbb{R} , thus $\forall A$ open sets $\subset \mathbb{R} : A \supset (a, b)$, one has $(a, b) \cap \mathbb{Q} \neq \emptyset$ that means $\bar{\mathbb{Q}} = \mathbb{R}$.

Remember that $\overset{\circ}{\mathbb{Q}} = \emptyset$ and $\partial\mathbb{Q} = \mathbb{R}$.

Definition 1.2.58

A topological space (X, τ) is **separable** if exists $A \subset X$ countable such that $\bar{A} = X$.

Example 1.2.59. The topological space (\mathbb{R}^N, τ_e) is separable $\forall n \in \mathbb{N}$.

1.2.7 Complete spaces

Another property for metric spaces is completeness. To define such property we have to consider the sequences that do not *oscillate* nor *diverge*: those sequences are the following:

Definition 1.2.60

Let (X, d) be a metric space.

A sequence $\{x_n\} \subset X$ is a **fundamental sequence** (or Cauchy sequence) if:

$$\forall \varepsilon > 0 \quad \exists \bar{n} = \bar{n}(\varepsilon) : m, n > \bar{n} \implies d(x_m, x_n) < \varepsilon \quad \forall n$$

or, equivalently, considering $m = n + p$:

$$\forall \varepsilon > 0 \quad \exists \bar{n} = \bar{n}(\varepsilon) : m, n > \bar{n} \wedge p \in \mathbb{N} \implies d(x_n, x_{n+p}) < \varepsilon \quad \forall p.$$

From this we can make some considerations. As a converging sequence tends to get closer and closer to a point, we wonder what is the relation between converging and fundamental sequences

Proposition 1.2.61

Let (X, d) be a metric space.

If the sequence $\{x_n\} \subset X$ converges, then it is a fundamental sequence.

The converse is not true.

Moreover we have:

Proposition 1.2.62

Any fundamental sequence is bounded.

Proposition 1.2.63

A fundamental sequence is convergent if and only if exists a convergent subsequence.

The reader should prove that proposition: it’s really simple!

We have to understand why some fundamental sequences does not converge. Sequences can be seen as a subset of elements, and for convergence we need a notion of distance: we have to investigate into the metric space in which sequence are defined.

Definition 1.2.64

A metric space (X, d) is **complete** if every fundamental sequence in X converges to a limit in X .

So, if a space is not complete, the problem can lie in the set or in the distance. Consider the following examples.

Example 1.2.65. The metric space (\mathbb{R}, d_e) is complete, while (\mathbb{Q}, d_e) is not; consider the sequence $(1 + \frac{1}{n})^n$. This sequence is fundamental and for $n \rightarrow \infty$ it converges to $e \notin \mathbb{Q}$.

The metric space $((0, 1), d_e)$ is not complete because $\frac{1}{n} \xrightarrow{n \rightarrow \infty} 0 \notin (0, 1)$.

Example 1.2.66. The metric space (\mathbb{R}, d) with $d(x, y) = |e^x - e^y|$ is not complete. Here the problem lies in the distance.

Consider the sequence $\{x_n\}$ with $x_n = -n$. The sequence is fundamental indeed:

$$d(x_n, x_{n+p}) \xrightarrow{n \rightarrow +\infty} 0 \quad \forall p.$$

However, it is not convergent: if $x_n \xrightarrow{d} x_0$ we expect that $d(x_n, x_0) \rightarrow 0$ but we have

$$d(x_n, x_0) = |e^{-n} - e^{x_0}| \rightarrow e^{x_0} \neq 0 \quad \forall x_0 \in \mathbb{R}.$$

Any metric space (X, d) admits a “completion” (\tilde{X}, d) . This is a complete metric space such that X is dense in \tilde{X} .

insert ref to the exercises, in which we discuss completeness of functional set

The following theorem characterize the subspaces of complete metric spaces.

Proposition 1.2.67

Let (X, d) be a complete metric space and $E \subset X$.

The metric space (E, d) is complete if and only if E is closed.

Proof. Recall that in metric space closure is equivalent to sequentially closure (see proposition 1.2.55 on page 57).

Necessary condition \implies :

Let $\{x_n\}_n \subset E$ such that $x_n \rightarrow x^*$. We have to prove $x^* \in E$.

Since sequence converges, then is fundamental in E , which is, by hypothesis, complete with respect to d .

Then we have $x_n \rightarrow x^* \in E$ so that E is closed.

Sufficient condition \impliedby :

Let $\{x_n\}_n \subset E$ be fundamental in (E, d) , in particular $\{x_n\}_n$ is fundamental in (X, d) .

As (X, d) is complete, it exists x^* such that $x_n \rightarrow x^* \in X$.

But E is closed, so $x^* \in E$ and the thesis is proved. ■

For the sake of a complete discussion, we present the first Cantor’s intersection theorem, which is a characterizations, a necessary and sufficient condition to completeness.

Theorem 1.2.68 (Cantor’s intersection theorem I)

Let (X, d) be a metric space.

It is complete if and only if:

$$\forall \{E_n\}_{n \in \mathbb{N}} \subset \mathcal{P}(X) : E_{n+1} \subset E_n \wedge \text{diam}(E_n) \xrightarrow{n \rightarrow \infty} \emptyset$$

$$\exists x^* \in X : \bigcap_{n \geq 1} E_n = \{x^*\}.$$

1.2.8 Compact spaces

The last property that we will discuss is compactness. It will be discussed for both topological and metric spaces. We now provide some basic definition specific for topological space, even if those make sense for metric spaces also.

Definition 1.2.69

Let (X, τ) be a topological space and $E \subset X$.

A family $\{E_i\}_{i \in I}$ is a **cover** of E if

$$E \subset \bigcup_{i \in I} E_i.$$

An **open cover** is a cover made of open sets.

A **subcover** is a subfamily of a cover which is a cover of E itself.

That is, $\{E_i\}_{i \in J}$, $J \subset I$, is a subcover if $E \subset \bigcup_{i \in J} E_i$.

Definition 1.2.70

Let (X, τ) be a topological space. The set X is **compact** if from each of its open cover, we can select a finite subcover. Namely:

$$\forall \{A_i\}_{i \in I}, A_i \in \tau \forall i \in I : X \subset \bigcup_{i \in I} A_i; \exists J \subset I, J \text{ finite} : X \subset \bigcup_{i \in J} A_i.$$

In case of single set withing a topological spaces, we have the following

Definition 1.2.71

Let (X, τ) be a topological space and $E \subset X$.

We say that E is **compact** if it is compact as topological space with the induced topology¹³, that is every cover of E made of open sets of X admits a finite subcover.

Notice that if τ_1 and τ_2 are topologies on X , with τ_2 weaker than τ_1 , and $E \subset X$ is compact with respect to τ_1 , then E is also compact with respect to τ_2 ; the reverse is false.

Example 1.2.72. The set $E = (0, 1)$ is not compact with respect to d_e . Indeed, consider $E_n = (\frac{1}{n}, 1)$, which is open: $\{E_n\}$ is a cover for E as $(0, 1) \subset \bigcup_{n \in \mathbb{N}} E_n$, but there is no finite subcover.

Example 1.2.73. Consider (X, τ_d) , where τ_d is the discrete topology for which all subset of X are open ($\tau_d = \mathcal{P}(X)$): only finite subsets of X are compact.

¹³An induced topology for a set belonging to a topological space is the smallest topology for which any subset open with respect the topology of the space is open.

Example 1.2.74. Consider (X, τ_0) with $\tau_0 = \{\emptyset, X\}$: any subset of X is compact.

Proposition 1.2.75

Let (X, τ) be a compact topological space and $S \subseteq X$.
If S is closed then S is compact.

Proof. Let $\{E_i\}_{i \in I}$ be a cover of S made of open sets in X .
As if S is closed then $X \setminus S$ is open, $\{E_i\}_{i \in I} \cup (X \setminus S)$ is an open cover of X .

Since (X, τ) is compact, this open cover admits a finite subcover, that is:

$$\exists J \subset I : X \subset (\cup_{j \in J} E_j) \cup (X \setminus S).$$

So $S \subset \cup_{j \in J} E_j$, that is $\{E_j\}_{j \in J}$ is a finite subcover of S , hence the thesis. ■

Notice that the compactness of S does not imply the closedness of S ; for example in (X, τ_0) only X and \emptyset are closed, yet any $A \subset X$ is compact.

Now the reader could prove this proposition.

Proposition 1.2.76

Continuous functions map compact sets into compact sets.

Sequentially compactness We can provide a slightly different definition or compactness, again we provide the definition for topological spaces but it take sense for metric spaces as well.

Definition 1.2.77

We say that (X, τ) is **sequentially compact** if for every sequence in X it exists a convergent subsequence, namely:

$$\forall \{x_n\} \subset X \quad \exists \{x_{n_k}\} \subset X : x_{n_k} \xrightarrow{k \rightarrow \infty} x^* \in X.$$

Compactness and sequential compactness are different notions, and they do not imply each other, but there are some relations. Indeed, for topological spaces the following theorems hold (recall definition of first and second countable spaces [1.2.27 on page 47](#)):

Theorem 1.2.78

Let (X, τ) be a topological space, let $E \subset X$ compact.
If X is first countable, then E is sequentially compact.

Theorem 1.2.79

Let (X, τ) be a topological space, let $E \subset X$ be sequentially compact.
If X is second countable, then E is compact.

Compactness for metric spaces In metric spaces the notion of compactness is the same as for topological spaces; indeed, it is defined upon the notion of openness, which has already been defined for both kind of spaces. Here we see the relation between compactness and boundedness for metric spaces (recall definition of totally bounded set [1.2.53 on page 56](#)).

Theorem 1.2.80 (Characterization of compact spaces)

If (X, d) is a metric space, the following are equivalent:

- the set X is compact;
- the set X is sequentially compact;
- the set X is complete and totally bounded.

Corollary 1.2.81

Let (X, d) be a metric space, with $E \subset X$.
If E is compact then it is closed and bounded.

Proof. First, as E is compact, then it is totally bounded, so it is also bounded: see the previous theorem.

Let us prove that E is closed: it's enough to show that E is sequentially closed.

By the previous theorem, if E is compact then it is also complete.

Take a converging sequence $\{x_n\}_{n \in \mathbb{N}} \subset E$ such that $x_n \xrightarrow{d} x^*$. Is $x^* \in E$? Since $\{x_n\}_{n \in \mathbb{N}}$ is fundamental (it's converging) and E is complete, $\{x_n\}_{n \in \mathbb{N}}$ has a limit in E so $x^* \in E$ and the thesis is reached. ■

The converse in general is not true.

Heine–Borel theorem In case of $X = \mathbb{R}^N$ we have the following important theorem:

Theorem 1.2.82 (Heine–Borel)

Let (\mathbb{R}^N, d_e) and $K \subset \mathbb{R}^N$.

Then K is compact if and only if K is closed and bounded

As in later chapters compact set will play a important role, we denote a compact set with the letter K .

Proof. Necessary condition \implies :

This follows from the previous corollary.

Sufficient condition \impliedby :

As $K \subset \mathbb{R}^N$ is closed and \mathbb{R}^N complete, K is complete as well.

Consider an hypercube C such that $C \supset K$ with l being the length of its side, divide C into m^n small cubes, with $m \in \mathbb{N}$ which have size $\frac{l}{m}$.

Let's call them

$$C_1^{(m)}, \dots, C_m^{(m)}.$$

Their diagonal has length $d_m = \frac{l}{m} \sqrt{n} \xrightarrow{m \rightarrow \infty} 0$.

Now fix $\varepsilon > 0$ and take m large such that $d_m < \varepsilon$. Then:

$$K \subset C = \bigcup_{j=1}^{m^n} C_j^{(m)} \subset \bigcup_{j=1}^{m^n} B_\varepsilon(x_j).$$

So the set is totally bounded and complete, therefore it is also compact, hence the thesis. ■

This *does not work in infinite dimensional spaces* because $m^n \xrightarrow{n \rightarrow \infty} +\infty$ and total boundedness requires a finite amount of balls.

Function defined on compact spaces

Proposition 1.2.83

Let (X, τ_x) and (Y, τ_y) two topological spaces and $E \subset X$ be compact. Let $f : X \rightarrow Y$ be a continuous function.

Then the image $f(E)$ compact.

Proof. Let $\{B_i\}_{i \in I}$ be an open cover of $f(E)$, and $A_i := f^{-1}(B_i)$. Then A_i is open (from the continuity of f) and $E \subset \bigcup_{i \in I} A_i$.

As E is compact, it exists a finite subcover $E \subset \bigcup_{k=1}^n A_{i_k}$. This implies that:

$$f(E) \subset f\left(\bigcup_{c=1}^m A_{i_k}\right) = \bigcup_{k=1}^n B_{i_k}$$

which is a finite subcover of $f(E)$, so it's compact. ■

Theorem 1.2.84 (Weierstrass)

Let (X, τ) topological space, X compact, and $f : X \rightarrow \mathbb{R}$ continuous.

Then F achieves a maximum and a minimum.

Proof. The image $f(X) \in \mathbb{R}$ is compact, so it's closed and bounded. This implies that $\sup f(x) \in f(X)$ and $\inf f(x) \in f(X)$. So there exists a maximum and a minimum. ■

Semicontinuity Now we define a weaker notion of continuity (see definition of limsum and liminf [A on page 303](#)).

Definition 1.2.85

Let (X, τ) a topological space, and let $f : X \rightarrow \mathbb{R}^*$.

We say that f is **lower semicontinuous** in $x_0 \in X$ if:

$$\forall \varepsilon > 0 \quad \exists U \in \tau : \quad x_0 \in U \text{ and } f(x) > f(x_0) - \varepsilon \quad \forall x \in U.$$

Equivalently in a metric space (X, d) , f is lower semicontinuous if:

$$\liminf_{x \rightarrow x_0} f(x) \geq f(x_0).$$

We say that f is **upper semicontinuous** in $x_0 \in X$ if:

$$\forall \varepsilon > 0 \quad \exists U \in \tau : \quad x_0 \in U \text{ and } f(x) < f(x_0) + \varepsilon \quad \forall x \in U.$$

Equivalently in a metric space (X, d) , f is upper semicontinuous if:

$$\limsup_{x \rightarrow x_0} f(x) \leq f(x_0).$$

Observe that f is continuous if and only if it is both lower semicontinuous and upper semicontinuous; as it happens with the existence of the limit.

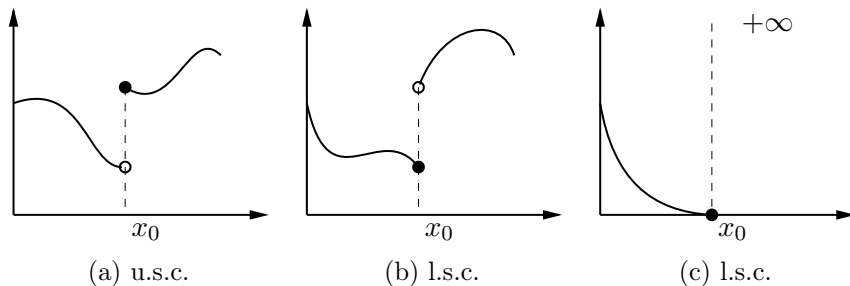


Figure 1.3: Examples of lower and upper semicontinuous functions.

Theorem 1.2.86

Let $f : (X, \tau) \rightarrow \mathbb{R}$ a lower semicontinuous function defined on a topological space, with X compact.
Then f achieves its minimum.

Let $f : (X, \tau) \rightarrow \mathbb{R}$ an upper semicontinuous function defined on a topological space, with X compact.
Then f achieves its maximum.

Now some results to complete the treatise.

Theorem 1.2.87 (Cantor's intersection theorem II)

Let (X, d) be a topological space, $\{K_n\}_{n \in \mathbb{N}}$ be a sequence of compact sets with $K_{n+1} \subset K_n$ and $K_n \neq \emptyset \quad \forall n \in \mathbb{N}$.

Then

$$\bigcap_{n \in \mathbb{N}} K_n \neq \emptyset.$$

Theorem 1.2.88 (Heine–Cantor)

Let (X, d) be a metric space, X a compact set and $f : X \rightarrow \mathbb{R}$ a continuous.

Then f is uniformly continuous.

1.2.9 The Cantor set and the Vitali function

Proves could require the use of counterexample to reach faster the result. Here we present two great result which can be useful to better understand the next results. First we will present the Cantor set which has quite unique characteristics in terms of its measure, then we will present a function, the Vitali–Cantor, which has interesting properties in terms of continuity and differentiation.

To fully understand this section the reader may read the measure chapter before, and study these results only when there is a need.

The Cantor set This set has a very weird definition which leads to really useful properties, especially when they are combined in the same set.

Definition 1.2.89

The **Cantor set** T is defined as follows.

Take the interval $[0, 1]$ and remove the open middle portion of size $\frac{1}{3}$. We call C_1 the remaining set:

$$C_1 := \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right].$$

Now, from each interval of C_1 , we remove an open middle portion of size $\frac{1}{9}$. Thus we obtain

$$C_2 := \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right].$$

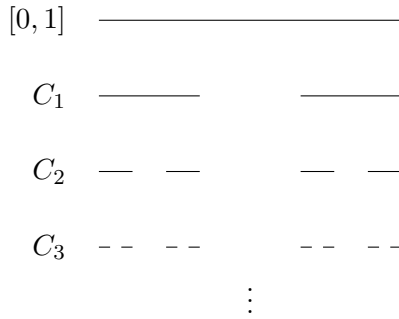
Then we iterate this process: for every k we define:

$$C_k := \bigcup_{n=1}^{2^k} I_n^{(k)},$$

where $I_n^{(k)}$ is a closed interval obtained from C_{k-1} by erasing the open middle portion of size $\left(\frac{1}{3}\right)^k$ from each I_m^{k-1} .

We define the Cantor set as:

$$T := \bigcap_{k=1}^{\infty} C_k.$$



Construction of the Cantor set

The Cantor set T has the following properties:

Proposition 1.2.90

The Cantor set is non-empty, namely: $T \neq \emptyset$, and, in particular, $0, 1 \in T$.

Indeed, we do not remove the extrema of each interval I_m^{k-1} : those extrema $0, \frac{1}{9}, \frac{2}{9}, \frac{1}{3}, \frac{2}{3}, \dots$ are contained in T .

Proposition 1.2.91

The Cantor set is closed.

Proof. To prove that T is closed, notice that the intervals C_k are closed for any k as it is a finite union of closed intervals.

Then, as T is a countable intersection of closed set it is closed itself. ■

Proposition 1.2.92

The Cantor set is compact.

Proof. As T is a subset of \mathbb{R} , to achieve this result we have to show that T is closed and bounded.

It is bounded as its diameter is 1, and it is closed as we proved the previous proposition. ■

Proposition 1.2.93

The Cantor set is a Lebesgue-measurable set.

To understand what this means, see definition 2.1.40 on page 101.

Proof. As T is closed, it is Borel measurable as well, so it is Lebesgue measurable (see proposition 2.1.43 on page 102). ■

Proposition 1.2.94

The length of the Cantor set is zero, namely $|T| = 0$.

Proof. We have $T \in C_k$ for all k , and thus:

$$|T| \leq |C_k| \leq \left| \bigcup_{n=1}^{2^k} I_n^k \right| = \sum_{n=1}^{2^k} |I_n^k| = \sum_{n=1}^{2^k} \left(\frac{1}{3}\right)^k = \left(\frac{2}{3}\right)^k \xrightarrow{k \rightarrow \infty} 0.$$

So it has measure zero, and it does not contain any interval. ■

Later, with the *generalized Cantor set*, we will see that it is not necessary to have measure zero, in order to contain no intervals.

Proposition 1.2.95

The Cantor set T is a **perfect set**, namely every point $x \in T$ is a limit point for T .

Equivalently we say that T does not have any isolated points.

Proof. We have to prove that for any $x \in T$ and for all $\varepsilon > 0$ we have:

$$\{T \cap (x - \varepsilon, x + \varepsilon)\} \setminus \{x\} \neq \emptyset$$

Taking $x \in T$, by its definition we have that $x \in C_k$ for any $k \in \mathbb{N}$.

Then for all $k \in \mathbb{N}$ there exists $n_k \in \{1, \dots, 2^k\}$ such that $x \in I_{n_k}^k = [a_k, b_k]$.

Recall that $|I_{n_k}^k| = \frac{1}{3^k}$.

Fix $\varepsilon > 0$ and take k sufficiently large so that:

$$|I_{n_k}^k| = (b_k - a_k) = \frac{1}{3^k} < \varepsilon.$$

Then $(x - \varepsilon, x + \varepsilon) \supset [a_k, b_k]$. Since both $a_k, b_k \in T$ we have:

$$a_k, b_k \in (x - \varepsilon, x + \varepsilon) \cap T \quad \forall k : \quad \frac{1}{3^k} < \varepsilon.$$

This completes the proof. ■

Proposition 1.2.96

The Cantor set does not contain any open interval, namely:

$$\overset{\circ}{T} = \emptyset.$$

This prove is left to the reader, observe that each $|I_n^k|$ tends to 0 as $k \rightarrow \infty$.

Proposition 1.2.97

The Cantor set is uncountable.

Proof. Rename each interval with L, R as follows:



As $C_k = \cup_{n=1}^{2^k} I_n^k$, at each subsequent step every interval I_n^k is divided into two sub-intervals, namely R and L.

Let $x \in T = \cap_{k=1}^{\infty} C_k$. At step one x can be either in L or in R. If we take $x \in L$, at step two x can be $x \in LL$ or $x \in LR$.

We continue tracing down the intervals where x lies in each C_k , thus getting an unique, infinite sequence of L and R.

Then we define a map $f : T \rightarrow E = \{\{s_k\}_{k \in \mathbb{N}} : s_k = L \text{ or } R \quad \forall k\}$.

We could prove that E is uncountable and f is bijective: we would have $|T| = |E|$ that prove our thesis.

The set E is uncountable:

By contradiction, suppose that E is countable; then we could assign a natural number to every element of E :

$$E = \{x^{(n)} = \{s_k^{(n)}\}_{k \in \mathbb{N}} : n \in \mathbb{N}\}.$$

We construct a new sequence $x = \{t_n^{(n)}\}_{n \in \mathbb{N}}$ with

$$t_n := \begin{cases} L & \text{if } s_n^{(n)} = R \\ R & \text{if } s_n^{(n)} = L \end{cases}$$

Clearly $x \in E$, since it is made of R and L, but $t_n \neq s_n^{(n)}$ and thus $x \neq x^{(n)}$ for all n . Then E is uncountable.

The function f is injective:

Let $x, y \in T$ such that $f(x) = f(y)$.

By construction for any k we have that x and y belong to the same I_n^k .

As

$$|I_n^k| = \frac{1}{3^k}$$

and

$$|x - y| \leq |I_n^k| = \frac{1}{3^k} \quad \forall k,$$

we gain that $x = y$ and f is injective.

The function f is surjective:

Let $\{s_k\}_k \in E$. We have to prove that there exists $x \in T$ such that $f(x) = \{s_k\}_k$.

Fix $\{s_k\}_k$ and let $I_{n_k}^k$ be the interval identified by $\{s_k\}_k$ at each step k .

In each k we can take

$$y_k \in I_{n_k}^k \cap T, \text{ with, by construction, } I_{n_{k+1}}^{k+1} \subset I_{n_k}^k.$$

We can assert that $\{y_k\}$ is a fundamental sequence, indeed we have:

$$d_E(y_k, y_{k+1}) < |I_{n_k}^k| = \frac{1}{3^k} \xrightarrow{k \rightarrow \infty} 0.$$

Moreover, $([0, 1], d_E)$ is complete, indeed it is a closed subset of a complete metric space (\mathbb{R}, d_E) , then $\{y_k\}$ is converging to an $x \in [0, 1]$.

As T is compact, it is closed and sequentially closed, so $x \in T$.

We have still to prove that $f(x) = \{s_k\}_k$.

By contradiction suppose that there exists a sequence $\{t_k\}_k$ and a certain \bar{k} for which $t_{\bar{k}} \neq s_{\bar{k}}$, and suppose that $f(x) = \{t_k\}_k$.

Obviously $\{t_k\}_k \neq \{s_k\}_k$, indeed we have that for all $k > \bar{k}$ we have $t_k \neq s_k$. So $x \in I_{m_1}^k$ (the interval represented by $\{t_k\}_k$), but $y_{\bar{k}}, y_{\bar{k}+1}, \dots$ belongs to $I_{m_2}^k$ (represented by $\{s_k\}_k$).

Then we have:

$$d_E(x, y_{\bar{k}+p}) > \frac{1}{3^k} \quad \forall p \text{ but } y_k \rightarrow x \text{ as } d(x, y_{k+p}) \xrightarrow{p \rightarrow \infty} 0.$$

So we have a contradiction, then f is bijective and this ends the proof. ■

Proposition 1.2.98

The Cantor set has the same cardinality of \mathbb{R} and $[0, 1]$.

The generalized Cantor set Similarly as before, it is construct by removing at each step not a third of the previous interval but a smaller, arbitrary quantity. What we will obtain is a non-zero-measure set which still have useful characteristics.

Definition 1.2.99

The **generalized Cantor set** T is defined as follows.

Fix $\varepsilon \in (0, 1)$ and take $C_0^\varepsilon = I_0^{\varepsilon,0} = [0, 1]$.

We build C_1^ε by erasing the open middle portion of size $\frac{\varepsilon}{3}$ from $I_0^{\varepsilon,0}$, and C_2^ε by erasing the open middle portion of size $\frac{\varepsilon^2}{9}$ from $I_0^{\varepsilon,1}$ and $I_1^{\varepsilon,1}$, the intervals composing C_1^ε .

Now iterate the process: C_k^ε is constructed removing open middle portions of size $(\frac{\varepsilon}{3})^k$ from each interval $I_n^{\varepsilon,k-1}$ of C_{k-1}^ε .

The **generalized Cantor set** is defined as:

$$T_\varepsilon = \bigcap_{k=1}^{\infty} C_k^\varepsilon.$$

The generalized Cantor set has many properties analogues to the Cantor set, in particular holds the following propositions: it is non-empty (1.2.90 on page 68), closed (1.2.91 on page 68), compact (1.2.92 on page 68), perfect (1.2.95 on page 69), does not contains any interval (1.2.96 on page 70) and is uncountable (1.2.97 on page 70) with the same cardinality of \mathbb{R} (1.2.98). Anyway its measure is not zero, indeed it is easy to compute $|T_\varepsilon|$:

$$|T_\varepsilon| = 1 - \frac{\varepsilon}{3} - 2\frac{\varepsilon^2}{9} - \dots = 1 - \sum_{k=1}^{+\infty} 2^{k-1} \left(\frac{\varepsilon}{3}\right)^k = 1 - \frac{1}{2} \sum_{k=1}^{+\infty} \left(\frac{2\varepsilon}{3}\right)^k = \frac{3(1-\varepsilon)}{3-2\varepsilon} > 0.$$

You can still prove that T_ε does not contain any interval: start by computing $|I_{n,\varepsilon}^{(k)}| \rightarrow 0$.

The characteristic function of T_ε , $\mathbb{1}_{T_\varepsilon}$ is Lebesgue-integrable:

$$\int_0^1 \mathbb{1}_{T_\varepsilon} d\mu = |T_\varepsilon|.$$

Observe that it is not Riemann-integrable; you reader should prove it: the interior of T_ε is \emptyset , its measure is zero, the lower Riemann sum converges to 0, but $|T_\varepsilon| > 0$, and thus the upper Riemann sum converges to $|T_\varepsilon|$.

The Vitali–Cantor function This function, also known as *the devil’s staircase*, present many properties in terms of continuity and derivability which make it a good counterexample in many situation. Its definition does not coincide with its construction.

Definition 1.2.100

The **Vitali–Cantor function**, is a function $f : [0, 1] \rightarrow [0, 1]$ such that:

- $f(0) = 0$ and $f(1) = 1$;
- f is continuous and monotone non-decreasing;
- f is derivable almost everywhere in $[0, 1]$, with $f' = 0$ almost everywhere in $[0, 1]$.

Such function can be obtained through a limit of the following series of function:

$$\begin{aligned}
 f_0(x) &= x \\
 f_1(x) &= \begin{cases} \frac{3}{2}x & \text{if } x \in [0, \frac{1}{3}] \\ \frac{1}{2} & \text{if } x \in (\frac{1}{3}, \frac{2}{3}) \\ \frac{3}{2}(x - \frac{1}{3}) & \text{if } x \in [\frac{2}{3}, 1] \end{cases} \\
 &= \begin{cases} \frac{1}{2} & \text{if } x \in [0, 1] \setminus C_1 \\ \text{linear} & \text{if } x \in C_1 \end{cases} \\
 &= \int_0^x g_1(t) dt \quad \text{with} \quad g_1(t) = \frac{3}{2} \mathbf{1}_{C_1}(t) \\
 f_2(x) &= \begin{cases} \frac{1}{2} & \text{if } x \in [\frac{1}{3}, \frac{2}{3}] \\ \frac{1}{4} & \text{if } x \in [\frac{1}{9}, \frac{2}{9}] \\ \frac{3}{4} & \text{if } x \in [\frac{7}{9}, \frac{8}{9}] \\ \text{linear} & \text{if } x \in C_2 \end{cases}
 \end{aligned}$$

$$\begin{aligned}
&= \int_0^x g_2(t) dt && \text{with } g_2(t) = \left(\frac{3}{2}\right)^2 \mathbb{1}_{C_2}(t) \\
&\vdots \\
f_k(x) &= \int_0^x g_k(t) dt && \text{with } g_k(t) = \left(\frac{3}{2}\right)^k \mathbb{1}_{C_k}(t)
\end{aligned}$$

The limit of the sequence $\{f_k\}$ is the Vitali–Cantor function, we will prove this through the following propositions. Let’s see its graph:

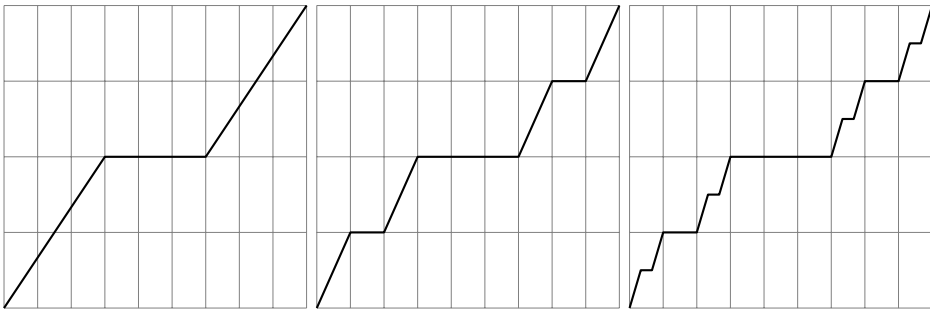


Figure 1.4: Construction of the Vitali–Cantor function.

as we can see, it can be considered as a fractal.

Proposition 1.2.101

For any k we have that $f_k(0) = 0$ and $f_k(1) = 1$.

Proof. Considering the properties of the defined integral, as the integral domain is reduced to a point, we have $f_k(0) = 0$. The other result can be computed:

$$f_k(1) = \int_0^1 \left(\frac{3}{2}\right)^k \mathbb{1}_{C_k}(t) dt = \left(\frac{3}{2}\right)^k |C_k| = 1.$$

■

Proposition 1.2.102

For any k we have that f_k is monotonic increasing and Lipschitz-continuous.

Proof. Observe that for any $x < y$ we have:

$$|f_k(y) - f_k(x)| = \left| \int_x^y g_k(t) dt \right| = \int_x^y g_k(t) dt \leq \left(\frac{3}{2}\right)^k |y - x|.$$

This fits the Lipschitz-continuity (see definition 2.4.7 on page 147) and shows that it is also bounded. ■

Proposition 1.2.103

The sequence is piece-wise constant.

In particular if $x \in C_k^C$, then $f_k(x) = f_{k+m}(x)$ for all $m \in \mathbb{N}$, and precisely there exists $N_x \in \mathbb{N}$ such that:

$$f_k(x) = f_{k+m}(x) = \frac{N_x}{2^k} \quad \forall m \in \mathbb{N}.$$

Proof. First, we prove the equality. If $x \notin C_k = \bigcup_{n=1}^{2^k} I_n^k$, then there exist N_x intervals I_n^k such that $I_n^k \subset [0, x]$, and $(2^k - N_x)$ intervals I_n^k such that $I_n^k \subset [x, 1]$, thus

$$f_k(x) = \int_0^x \left(\frac{3}{2}\right)^k \mathbb{1}_{C_k}(t) dt = \left(\frac{3}{2}\right)^k N_x |I_n^k| = \left(\frac{3}{2}\right)^k N_x \left(\frac{1}{3}\right)^k = \frac{N_x}{2^k}$$

Now we prove the value. If $x \notin C_k$, then $x \notin C_{k+1}$: since each interval I_n^k produces two intervals I_n^{k+1} , when we have N_x intervals $I_n^k \subset [0, x]$, then we also have $2N_x$ intervals $I_n^{k+1} \subset [0, x]$. Therefore:

$$f_{k+1}(x) = \left(\frac{3}{2}\right)^{k+1} \cdot 2N_x \cdot |I_n^{k+1}| = \frac{N_x}{2^k}.$$

Thus $f_k(x) = f_{k+1}(x)$, and by induction $f_k(x) = f_{k+m}(x) \quad \forall m \in \mathbb{N}$. ■

Proposition 1.2.104

Each complementary of a set C_k can be rewritten as an union of open sets, namely:

$$C_k^C = \bigcup J_n^{(k)},$$

where $J_n^{(k)}$ are open.

Moreover, if $x \neq y$ and x, y are contained in the same $J_n^{(k)}$, then $f_k(x) = f_k(y)$.

Proof. Considering the previous proposition, this follows as $x, y \notin C_k$, and $N_x = N_y$. ■

Proposition 1.2.105

The sequence $\{f_k\}$ is fundamental with respect to d_∞ . Moreover, the limit f exists, $f(0) = 0$, $f(1) = 1$ and f is monotonic increasing.

Proof. Take $x \in C_k^C$. Then $f_{k+1}(x) = f_k(x)$.
By continuity of f_k and f_{k+1} we have:

$$f_{k+1}(x) = f_k(x) \quad \forall x \in \overline{C_k^C}.$$

Take now $x \in (\overline{C_k^C})^C = \overset{\circ}{C}_k$ instead. Then $x \in I_n^{\circ k} = (a_k, b_k)$: we have $f_{k+1}(a_k) = f_k(a_k)$ and $f_{k+1}(b_k) = f_k(b_k)$.

Using the triangular inequality and the monotonicity of f :

$$\begin{aligned} |f_{k+1}(x) - f_k(x)| &= |f_{k+1}(x) - f_{k+1}(a_k) + f_{k+1}(a_k) - f_k(x)| \\ &\leq |f_{k+1}(x) - f_{k+1}(a_k)| + |f_k(x) - f_k(a_k)| \\ &\leq (f_{k+1}(b_k) - f_{k+1}(a_k)) + (f_k(b_k) - f_k(a_k)) \\ &= 2 \cdot (f_k(b_k) - f_k(a_k)) \\ &= 2 \int_{a_k}^{b_k} g_k(t) dt \\ &= 2 \int_{a_k}^{b_k} \left(\frac{3}{2}\right)^k dt \\ &= 2 \cdot \left(\frac{3}{2}\right)^k \cdot \frac{1}{3^k} \\ &= \frac{1}{2^{k-1}} \end{aligned}$$

And thus:

$$d_\infty(f_{k+1}, f_k) \leq \frac{1}{2^k - 1} \xrightarrow{k \rightarrow +\infty} 0.$$

Hence $\{f_k\}$ is a fundamental sequence in $(C^0([0, 1]), d_\infty)$, which is complete, and thus there exists

$$f \in C^0([0, 1]) : f_k \xrightarrow{d_\infty} f.$$

Thus $f(0) = 0$, $f(1) = 1$ and f is monotonic increasing. ■

Proposition 1.2.106

The function f is differentiable almost everywhere; where that is possible we have $f'(x) = 0$.

Proof. Let $x \notin T$. Then there exists $k \in \mathbb{N}$ such that $x \notin C_k$. But C_k^C is open, and hence there exists $\varepsilon > 0$ such that $(x - \varepsilon, x + \varepsilon) \in C_k^C$. Therefore, by what we proved in the previous points ([1.2.103 on page 75](#) and [1.2.104 on page 75](#)), we have that

$$f_{k+m}(y) = \frac{N_x}{2^k} \quad \forall y \in (x - \varepsilon, x + \varepsilon) \quad \forall m \in \mathbb{N}.$$

As $m \rightarrow +\infty$, we obtain

$$f(y) = \frac{N_x}{2^k} \quad \forall y \in (x - \varepsilon, x + \varepsilon).$$

This means that f is constant in a neighborhood of x , and thus it is differentiable in $x \in T^C$, with $f'(x) = 0$. Observe that the function f is not differentiable only in T : but as $|T| = 0$ we have that f is differentiable almost everywhere. ■

2 Real Analysis

Now that we have introduced some fundamental notions about set theory and topology, we can begin our study of Real Analysis, which is a branch of mathematical analysis that studies real numbers and the most common operations we can do with them: sequences, series, functions, integration.

This raises some meaningful questions: how are they defined from a rigorous point of view? Which are their properties? Are there some operations that cannot be done for certain reasons? Which characterizations can we find to connect these structures?

In this part we will first introduce measure theory, which as abstract as it may seem, it is essential to define the powerful Lebesgue integral later on. Measure theory also does some *dotting the i's and crossing the t's* on something that the reader may have seen many times in previous courses: probability. Some concepts that maybe did not add up before, now will make sense. We will also explore in a meaningful way the concept of non-measurable sets, which is way more non-trivial than measurable sets.

Then, starting from a formal definition of integration according to Lebesgue, we will see some relevant theorems and properties, useful both in practical and theoretic fields.

We will see the two fundamental theorems of calculus, something that the reader may have also seen in a simplified form in previous courses. To this end, we will explore new definitions of continuity.

At last, we will give some definitions and useful theorems to deal with measures defined in product spaces.

2.1 Measure theory

Here we try to find a way to measure sets. Considering for instance a non-empty set Ω , how can we define a set function $\mu : \mathcal{P}(\Omega) \rightarrow [0, +\infty]$ which is interpretable as a **measure** of the set Ω ? Are there some properties required?

Two reasonable properties for such function are:

$$\mu(\emptyset) = 0, \quad \mu(A \cup B) = \mu(A) + \mu(B) \quad \text{if} \quad A \cap B = \emptyset.$$

Does a function with those properties exist?

2.1.1 Measurable spaces

Our first step is to define a good environment in which we can have all the tools we need: let's define on which kind of sets we can search for such measure.

Definition 2.1.1

Consider a non-empty set Ω . The family of subsets $\mathfrak{M} \subseteq \mathcal{P}(\Omega)$ is a **σ -algebra** of Ω if:

- the **empty set is included**:

$$\emptyset \in \mathfrak{M};$$

- it is **closed with respect to complements**:

$$E \in \mathfrak{M} \implies E^C \in \mathfrak{M};$$

- it is **closed with respect to countable unions**:

$$\{E_i\}_{i \in \mathbb{N}} \subseteq \mathfrak{M} \implies \bigcup_{i \in \mathbb{N}} E_i \in \mathfrak{M}.$$

It can be easily shown from the definition, that also Ω itself belongs to the σ -algebra and, by using the De Morgan's laws, that \mathfrak{M} is closed with respect to countable intersections:

$$\forall \{E_j\}_{j \in \mathbb{N}} \in \mathfrak{M} \quad \bigcap_{j \in \mathbb{N}} E_j = \bigcap_{j \in \mathbb{N}} (E_j^C)^C = \left(\bigcup_{j \in \mathbb{N}} E_j^C \right)^C \in \mathfrak{M}.$$

Similarly it can be proved that the difference of sets belonging to the σ -algebra itself belongs to the σ -algebra.

The power set $\mathcal{P}(\Omega)$ is known to be the trivial σ -algebra.

This allow us to find the ideal environment for a measure:

Definition 2.1.2

Let $\Omega \neq \emptyset$ and $\mathfrak{M} \subseteq \mathcal{P}(\Omega)$ be a σ -algebra.

Then (Ω, \mathfrak{M}) is a **measurable space**.

It is easy to show that the intersection of σ -algebras is still a σ -algebra.

To many purposes is useful to consider the “smallest” σ -algebra containing a given family of subsets of Ω . The following definition formalize this concept.

Definition 2.1.3

Let $\mathcal{E} \in \mathcal{P}(\Omega)$.

The **σ -algebra generated by \mathcal{E}** , written as $\langle \mathcal{E} \rangle$, is defined as the intersection of all the σ -algebras containing \mathcal{E} .

It's easy to prove that this definition is well-posed, to do that, show that the intersection of those σ -algebras itself is a σ -algebra. Furthermore, it's easy to show that such σ -algebra is actually the smallest.

Exercise 2.1.4. Let $\Omega \neq \emptyset$.

Consider $\mathcal{E} = \{E_i\}_{i=1,\dots,N}$ as a partition of Ω : prove that $\langle \mathcal{E} \rangle$ has 2^N elements.

Recalling the concepts of open sets and topological spaces (see definition [1.2.21 on page 44](#)), here is presented a special σ -algebra which has been defined with a special name due to its relevance.

Definition 2.1.5

Let (X, τ) be a topological space with $X \neq \emptyset$.

We call **Borel σ -algebra**, denoted by $\mathcal{B}(X)$, the σ -algebra generated by open sets.

All the set that it contains are called **Borel sets**.

Here some general category of sets which belongs to the Borel σ -algebra.

Definition 2.1.6

We call F_σ **sets** every countable union of closed sets.

We call G_δ **sets** every countable intersection of open sets. ¹⁴

Open and closed sets, F_σ sets and countable intersections of F_σ sets, G_δ sets and countable unions of G_δ sets all belongs to the Borel σ -algebra.

2.1.2 Measurable functions

We have provide a definition for measurable spaces, but, yes, we still don't know how to measure them. Before that we'll discuss the concept of measurability for functions. The first steps will not involves any kind of function, but we will focus on an easily manageable case. It will be easy to extend those results to the general case.

Definition 2.1.7

Let (X, \mathfrak{M}) be a measurable space, and (Y, τ) be a topological space.

We say that $f : X \rightarrow Y$ is a **measurable function** if the preimage of any open set is measurable, namely:

$$f^{-1}(A) \in \mathfrak{M} \quad \text{for all } A \in \tau.$$

Observe that a function defined on two topological spaces, namely $f : (X, \tau_X) \rightarrow (Y, \tau_Y)$, if it maps open sets into open sets then it is \mathcal{B} -measurable with respect to the measurable space $(X, \mathcal{B}(X))$. Indeed, for every open set $A \in \tau$, $f^{-1}(A)$ is open in X by definition, and thus $f^{-1}(A) \in \mathcal{B}(X)$.

Theorem 2.1.8 (Composition of a measurable function with a continuous one)

Let $f : (\Omega, \mathfrak{M}) \rightarrow (X, \tau_X)$ be a measurable function and $g : (X, \tau_X) \rightarrow (Y, \tau_Y)$ be a continuous function.

Then $g \circ f : (\Omega, \mathfrak{M}) \rightarrow (Y, \tau_Y)$ is measurable.

As we said in the previous remark, g is not only continuous but also $\mathcal{B}(X)$ -measurable. Here we prove this case, which is more general with respect to the theorem.

¹⁴Felix Hausdorff (1868 - 1942) adopted this convection; F is for the French word "fermé" which means closed, while σ is for "somme" which means sum (union) in the same language; on the other hand, G is for the German word "Gebiet" meaning area or neighbourhood, indicating open set, and δ is for "Durchschnitt" which means intersection.

Proof. Let $A \subset Y$ be an open set, then $g^{-1}(A) \in \mathcal{B}(X)$. Now let $B \subset X$ be open, and $f^{-1}(B) \in \mathfrak{M}$.

As $\mathcal{B}(X)$ is generated by all the open sets in X and is closed with respect to union and complement, $f^{-1}(g^{-1}(A)) \in \mathfrak{M}$ and this means that $g \circ f$ is \mathfrak{M} -measurable (in case $A = \emptyset$ we have $g \circ f(A) = \emptyset \in \mathfrak{M}$). ■

If f and g are just measurable, then $g \circ f$ is not necessarily measurable, even if f is continuous.

Measurable functions in \mathbb{R} There are some important results that holds when working in \mathbb{R} . First, the previous theorem can be extended to the multi dimensional case.

Theorem 2.1.9

Let $(u, v) : (\Omega, \mathfrak{M}) \rightarrow \mathbb{R}^2$ be \mathfrak{M} -measurable.

Moreover, let $\Phi : \mathbb{R}^2 \rightarrow (X, \tau)$ be continuous.

Then $h = \Phi(u, v) : (\Omega, \mathfrak{M}) \rightarrow (X, \tau)$ is \mathfrak{M} -measurable.

Proof. Let $A \subseteq X$ be open. Then $\Phi^{-1}(A) \subseteq \mathbb{R}^2$ is open.

Thanks to theorem 2.1.10, we have that $\Phi^{-1}(A) = \bigcup_{i \in \mathbb{N}} I_i \times J_i$ where $I_i, J_i \subset \mathbb{R}$ are intervals.

Finally, observe that $h^{-1}(I_i \times J_i) = u^{-1}(I_i) \cap v^{-1}(J_i) \in \mathfrak{M}$. ■

This theorem has several implications: if $u, v : (\Omega, \mathfrak{M}) \rightarrow \mathbb{R}^2$ are measurable, than $u \pm v, uv, |u|, u^+, u_-$ and other elementary operation with these functions returns \mathfrak{M} -measurable functions.

The following proposition shows a topological property of the set \mathbb{R} , it will turn useful in many fields.

Theorem 2.1.10

Every open set $\Omega \in \mathbb{R}$ can be written in a unique way as countable union of mutually disjoint intervals, namely:

$$\Omega = \bigcup_{j=1}^{\infty} I_j \quad I_j = (a_j, b_j).$$

Every open set $\Omega \in \mathbb{R}^N$, with $n > 1$, can be written in a unique way as countable union of n intervals, namely:

$$\Omega = \bigcup_{j=1}^{\infty} I_j \quad I_j = (a_j^1, b_j^1) \times (a_j^2, b_j^2) \times \cdots \times (a_j^n, b_j^n).$$

Proof. Step 1. The intervals are well defined:

First, consider $x \in \Omega$: we need to build the largest open interval $I_x \subset \Omega$ containing x .

We define $I_x = (a_x, b_x)$ where:

$$a_x := \inf\{a < x : (a, x) \subset \Omega\} \quad \text{and} \quad b_x := \sup\{b > x : (x, b) \subset \Omega\};$$

notice that $I_x \neq \emptyset$, $a_x \neq x$ and $b_x \neq x$ since Ω is open and $x \in \Omega$.

Certainly we have that $a_x \in [-\infty, x)$ and $b_x \in (x, +\infty]$.

Moreover, we show that $I_x \in \Omega$; let $w \in I_x$, without loss of generality we assume that $x < w < b_x$. Then by the definition of b_x , it exists $b > w$ such that $(x, b) \in \Omega$: then $w \in \Omega$.

Step 2. The set is the union of the intervals:

Consider the collection of open intervals $\{I_x\}_{x \in \Omega}$. Since each $x \in \Omega$ and $x \in I_x \subseteq \Omega$, then

$$\Omega = \bigcup_{x \in \Omega} I_x.$$

Here we have two problems: first, the family $\{I_x\}_{x \in \Omega}$ may not be disjoint; second the same family can be uncountable.

Step 3. The family of intervals is disjoint:

To solve the first problem we have to prove that if $x, y \in \Omega$, then either $I_x = I_y$ or $I_x \cap I_y = \emptyset$.

If $I_x \cap I_y \neq \emptyset$ then $I_x \cup I_y$ is still an open interval, and it contains both x and y , and it is a subset of Ω .

Since I_x , as we proved before, is the largest open interval in Ω containing x , we have:

$$I_x \supseteq I_x \cap I_y \supseteq I_x \implies I_x = I_x \cup I_y.$$

Similarly we can show that $I_x = I_y$.

Step 4. The family of intervals is countable:

Since for any $x \in \Omega$ there exists a rational number $q \in I_x \subseteq \Omega$, and we have

that $I_x \cap I_q \neq \emptyset$, hence $I_x = I_q$.

This means that:

$$\Omega = \bigcup_{x \in \Omega} I_x = \bigcup_{q \in \mathbb{Q} \cap \Omega} I_q$$

as \mathbb{Q} is countable, then $\Omega \cap \mathbb{Q}$ is countable as well, hence the thesis. ■

This proof can be extended to \mathbb{R}^N case.

Proposition 2.1.11

Every open set $\Omega \subset \mathbb{R}^N$ can be written as a countable union of closed, disjoint cubes, namely:

$$\Omega = \bigcup_{j=1}^{\infty} R_j$$

where

$$R_j = [a_j^1, b_j^1] \times [a_j^2, b_j^2] \times \cdots \times [a_j^n, b_j^n]$$

with

$$a_j^i < b_j^i \quad \forall i \text{ and } \dot{R}_i \cap \dot{R}_k = \emptyset \quad k \neq i.$$

Observe that R_j must be almost disjoint.

This last results is about the Borel σ -algebras of \mathbb{R} and \mathbb{R}^* ; they don't need the entire topology to be generated.

Theorem 2.1.12

The Borel σ -algebra of \mathbb{R} and \mathbb{R}^* are defined as follows:

$$\mathcal{B}(\mathbb{R}) = \langle \{(a, b)\}_{a, b \in \mathbb{R}} \rangle \quad \mathcal{B}(\mathbb{R}^*) = \langle \{(a, +\infty)\}_{a \in \mathbb{R}} \rangle.$$

Proof. Proof of the generator sets of $\mathcal{B}(\mathbb{R})$:

This first statement follows from theorem [2.1.10 on page 83](#).

Proof of the generator sets of $\mathcal{B}(\mathbb{R}^)$:*

For the second one, let us consider the set $(a, +\infty]$, which is open $\forall a \in \mathbb{R}$.

So we have that:

$$\mathcal{B}(\mathbb{R}^*) \supseteq \langle \{(a, +\infty]_{a \in \mathbb{R}}\} \rangle =: \mathfrak{M}.$$

Now we want to prove that $\mathcal{B}(\mathbb{R}^*) \subseteq \mathfrak{M}$, by checking that the sets $[-\infty, a]$, (a, b) , $(b, +\infty]$ are contained in \mathfrak{M} for all a, b :

$$\begin{aligned} [-\infty, a] &= (a, +\infty]^c \in \mathfrak{M} \\ [-\infty, a] &= \bigcup_{n=1}^{\infty} \left(-\infty, a - \frac{1}{n} \right) \in \mathfrak{M} \\ (a, b) &= [-\infty, b) \cap (a, +\infty] \in \mathfrak{M} \end{aligned}$$

The sets $[-\infty, a]$, (a, b) , $(b, +\infty]$ form a base for the topology of \mathbb{R}^* , and thus $\mathfrak{M} \supseteq \mathcal{B}(\mathbb{R}^*)$.

That proves that $\mathcal{B}(\mathbb{R}^*) \subseteq \mathfrak{M}$. ■

Indeed, those structures are generated as follows:

$$\mathcal{B}(\mathbb{R}) = \langle \{(a, b)\}_{a, b \in \mathbb{R}} \rangle \quad \mathcal{B}(\mathbb{R}^*) = \langle \{(a, +\infty]\}_{a \in \mathbb{R}} \rangle$$

See proposition [2.1.12 on the preceding page](#) for the proof of this result.

Measurability for sequence of functions Now we pose the following problem: having a sequence of measurable function which converges to a limit function, is that function measurable? Before look at these results, recall the notions of *liminf* and *limsup* explained in the appendix.

Theorem 2.1.13

Let $f_n : (\Omega, \mathfrak{M}) \rightarrow \mathbb{R}^*$ be \mathfrak{M} -measurable functions for all $n \in \mathbb{N}$. Then $\sup f_n$, $\inf f_n$, $\limsup f_n$, $\liminf f_n$ are all \mathfrak{M} -measurable.

Proof. Step 1. Inferior and superior:

Consider $g(x) := \sup f_n(x)$:

$$\begin{aligned} g^{-1}((a, +\infty]) &= \{x \in \mathbb{R}^* \mid \sup f_n(x) > a\} \\ &= \{x \in \mathbb{R}^* \mid \exists n : f_n(x) > a\} \\ &= \{x \in \mathbb{R}^* \mid \exists n : x \in f_n^{-1}((a, +\infty])\} \\ &= \bigcup_{n \in \mathbb{N}} f_n^{-1}((a, +\infty]) \\ &\in \mathfrak{M}. \end{aligned}$$

We need to prove the measurability only on $(a, +\infty]$, the remaining comes by properties of σ -algebra and topology; indeed, all the generators of the topology τ_C^* have a counterimage in \mathfrak{M} .

Consider now $g(x) := \inf f_n(x)$. Arguing as before we get: $g^{-1}([-\infty, a)) = \bigcup_{n \in \mathbb{N}} f_n^{-1}([-\infty, a)) \in \mathfrak{M}$.

Step 2. Liminf and limsup:

Then consider $g(x) := \limsup f_n(x) = \inf_{n \geq 0} (\sup_{k \geq n} f_k(x))$. In the previous lines we have shown that $\sup_{k \geq 0} f_k$ is measurable, and so is the function $h_n = \sup_{k \geq n} f_k$, for any $n \in \mathbb{N}$; the infimum of such sequence, which is $\inf h_n (\sup_{k \geq n} f_k) = g(x)$, is also measurable as we shown before.

We can argue in a similar way for $\liminf f_n(x)$ and conclude the proof. ■

Simple functions Let consider a simple kind of functions for which is easy to find a measure. Those function will be the prototype of functions in general and though them we will extend their measure to all the measurable functions.

Definition 2.1.14

Let (Ω, \mathfrak{M}) be a measurable space ($\Omega \neq \emptyset$).

We say that $s : \Omega \rightarrow \mathbb{R}$ is a **simple function** if it has the following form:

$$s(t) = \sum_{n=1}^N a_n \mathbb{1}_{E_n}(t)$$

where $a_j \in \mathbb{R}$ and $\{E_j\}_{j=1}^N$ is a partition of Ω made of measurable sets.

The set function $\mathbb{1}_E$ the indicator function.

Example 2.1.15. The function: $s(t) = \mathbb{1}_{[-\infty, 2)}(t) + 3\mathbb{1}_{[2, 3]}(t) - \mathbb{1}_{(3, +\infty)}(t)$ is a simple function.

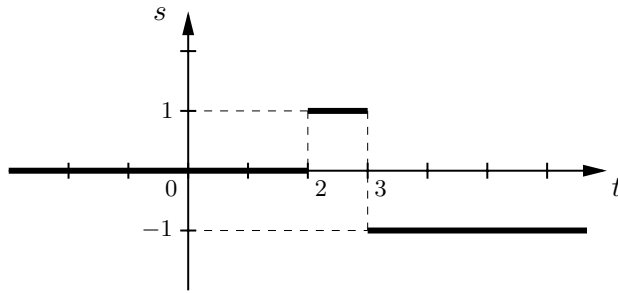


Figure 2.1: Plot of the function of example 2.1.15 on the preceding page.

Theorem 2.1.16

Let (Ω, \mathfrak{M}) be a measurable space, with $\Omega \neq \emptyset$.

For any function $f : (\Omega, \mathfrak{M}) \rightarrow [0, +\infty]$, \mathfrak{M} -measurable, there exists a sequence of simple functions

$$\{s_n\}_{n \in \mathbb{N}} : (\Omega, \mathfrak{M}) \rightarrow [0, +\infty]$$

such that:

- the sequence is monotone:

$$s_n(t) \leq s_{n+1}(t) \quad \forall t \in \Omega \quad \forall n \in \mathbb{N};$$

- the sequence is dominated:

$$0 \leq s_n(t) \leq f(t) \quad \forall t \in \Omega \quad \forall n \in \mathbb{N};$$

- the sequence converges point-wise:

$$\lim_{n \rightarrow +\infty} s_n(t) = f(t) \quad \forall t \in \Omega.$$

Moreover, if the function is bounded, namely there exists $M > 0$ such that $f(t) < M$ for any $t \in \Omega$, the sequence converges uniformly to f in Ω .

Notice that $[0, +\infty]$ is a topological space with the topology induced by (\mathbb{R}^*, τ_C^*) .

The following is a constructive proof in which the sequence object of the theorem is built.

Proof. Let $n \in \mathbb{N}_0$, split $[0, n]$ in $k = n2^n$ intervals $[a_j, b_j]$ of length 2^{-n} .

Take $E_0 = f^{-1}([n, +\infty])$ and $E_j = f^{-1}([a_j, b_j])$, which are \mathfrak{M} -measurable.

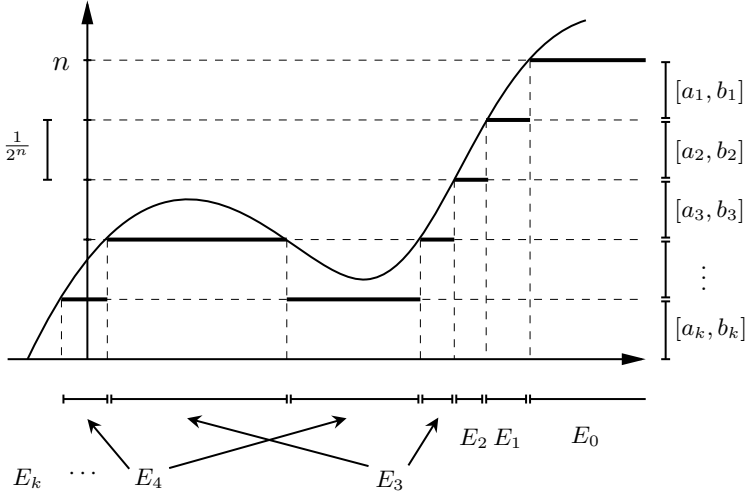


Figure 2.2: Construction of approximating simple functions.

Define now:

$$s_n(t) := n\mathbb{1}_{E_0}(t) + \sum_{j=1}^{n2^n} a_j \mathbb{1}_{E_j}(t).$$

By definition $s_n(t)$ is a simple function for any $n \in \mathbb{N}$, with $s_n(t) = a_j$ if $f(t) \in [a_j, b_j]$, and $s_n(t) = n$ if $f(t) \in [n, +\infty]$.

Bullet 1:

It can be easily proved that $s_n(t) \leq s_{n+1}(t)$ for all $t \in \Omega$ and for all $n \in \mathbb{N}_0$, since at every step the function either increases or stays put for every t .

Bullet 2:

Notice the following:

$$f(t) - s_n(t) = f(t) - n\mathbb{1}_{E_0}(t) - \sum_{j=1}^{n2^n} a_i \mathbb{1}_{E_i}(t) \geq 0 \quad \forall t \in \Omega,$$

indeed, if $t \in E_j = f^{-1}([a_j, b_j))$, we have $f(t) - s_n(t) = f(t) - \frac{j-1}{2^n} = f(t) - a_j \geq 0$.

The case $t \in E_0$ is trivial; we have $f(t) \geq s_n(t)$ for all $t \in \Omega$ and all $n \in \mathbb{N}$.

Bullet 3:

If $t \in \Omega$, whether $f(t) \geq n$ which implies $s_n(t) = n$ for all $n \in \mathbb{N}$, or $f(t) < n$ which implies that exists $\bar{n} = \bar{n}(t)$ such that

$$0 \leq f(t) - s_n(t) \leq \frac{j}{2^n} - \frac{j-1}{2^n} = \frac{1}{2^n} \quad \forall n \geq \bar{n}$$

and

$$\lim_{n \rightarrow \infty} s_n(t) = f(t) \quad \forall t \in \Omega.$$

Final result:

Finally notice that if f is bounded, then \bar{n} does not depend on t anymore, hence the converge is uniform on Ω . ■

2.1.3 Positive measures

Here we introduce a first yet abstract notion of measure: it will not satisfy all of ours expectations but it is the first draft from which we will build an actual measure.

Definition 2.1.17

Let Ω be a non-empty set and \mathfrak{M} one of its σ -algebras. We say that a set function $\mu : \mathfrak{M} \rightarrow [0, +\infty]$ is a **positive measure** if:

1. the function μ is **countably additive**, namely for all sequence of disjoint sets $\{E_j\}_{j \in \mathbb{N}} \subset \mathfrak{M}$ we have:

$$\mu \left(\bigcup_{j \in \mathbb{N}} E_j \right) = \sum_{j \in \mathbb{N}} \mu(E_j);$$

2. it exists at least a set $E \in \mathfrak{M}$ which has a finite measure, namely:

$$\mu(E) < \infty.$$

If μ is a positive measure $(\Omega, \mathfrak{M}, \mu)$ is called **measure space**.

Notice that in the countable additivity equality we may have infinity.

If the measure satisfy some specific property it may gain a special name, for example:

- a measure μ is **finite** if $\mu(\Omega) < \infty$;
- a finite measure μ is a **probability measure** if $\mu(\Omega) = 1$;
- a measure μ is **σ -finite** if there exists $\{E_n\}_{n \in \mathbb{N}} \subset \mathfrak{M}$ such that $\cup_{n \in \mathbb{N}} E_n = \Omega$ and $\mu(E_n) < +\infty$;
- a positive measure on $\mathcal{B}(\Omega)$ is called **Borel measure**.

The following are examples of positive measures:

Example 2.1.18. Take $(\Omega, \mathcal{P}(\Omega))$.

The **counting measure** μ_C is defined as follows:

$$\mu_C(E) := \begin{cases} n & \text{if } m(E) = n \\ +\infty & \text{if } m(E) \geq \aleph_0 \end{cases} \quad \forall E \in \mathfrak{M}$$

where $m(E)$ is the magnitude of E .

Example 2.1.19. Take $(\Omega, \mathcal{P}(\Omega))$.

The **Dirac measure** (or Dirac mass) δ_t is defined as follows:

$$\delta_t(E) := \begin{cases} 1 & \text{if } t \in E \\ 0 & \text{if } t \notin E \end{cases} \quad \forall E \in \mathfrak{M}.$$

Observe that δ_t is also a probability measure.

Example 2.1.20. Take $\Omega = \{x_n\}_{n \in \mathbb{N}}$, $\mathfrak{M} = \mathcal{P}(\Omega)$.

Let $\{p_n\}_{n \in \mathbb{N}} \subset \mathbb{R}^+$ such that $\sum_{n \in \mathbb{N}} p_n = 1$; $\mu(E) = \sum_{x_n \in E} p_n$ is a probability measure.

Main properties The following are seven basic properties of the positive measure. The setting is the measure space $(\Omega, \mathfrak{M}, \mu)$.

Proposition 2.1.21

Any positive measure of the empty set is zero:

$$\mu(\emptyset) = 0.$$

Proof. Take $E \in \mathfrak{M}$ such that $\mu(E) < +\infty$, and let $E_0 = E$, $E_n = \emptyset$ with $n \in \mathbb{N}_0$. Then we apply countable additivity:

$$\mu(\cup_{j \in \mathbb{N}} E_j) = \sum_{j \in \mathbb{N}} \mu(E_j) \implies \mu(E) = \mu(E) + \sum_{n \in \mathbb{N}_0} \mu(\emptyset).$$

Because $\mu(E)$ has finite measure, it has to be $\mu(\emptyset) = 0$. ■

Proposition 2.1.22

Any positive measure is **finitely additive**, namely:

$$\mu\left(\bigcup_{j=0}^n E_j\right) = \sum_{j=0}^n \mu(E_j).$$

Proof. Take a family of disjoint sets $\{E_0, \dots, E_n\} \subset \mathfrak{M}$, and set $E_m = \emptyset$ for every $m > n$. Using countable additivity, we have:

$$\mu\left(\bigcup_{j=0}^n E_j\right) = \mu\left(\bigcup_{j \in \mathbb{N}} E_j\right) = \sum_{j \in \mathbb{N}} \mu(E_j) = \sum_{j=0}^n \mu(E_j)$$

and thus μ is also finitely additive. ■

Proposition 2.1.23

Any positive measure is **monotone increasing** with respect to the partial order given by inclusion, namely:

$$\mu(E) \leq \mu(F) \quad \text{for all } E, F \in \mathfrak{M} \text{ such that } E \subseteq F.$$

Proof. As E and $(F \setminus E)$ are disjoint, we have:

$$\mu(F) = \mu(E \cup (F \setminus E)) = \mu(E) + \mu(F \setminus E) \geq \mu(E).$$

Proposition 2.1.24

Let μ be a positive measure and let $E, F \in \mathfrak{M}$ such that $E \subset F$ and $\mu(E) < +\infty$.

Then:

$$\mu(F \setminus E) = \mu(F) - \mu(E).$$

Proof. See the previous proof. ■

Proposition 2.1.25

Let μ be a positive measure and let $\{E_n\}_{n \in \mathbb{N}} \subset \mathfrak{M}$ be an ascending sequence, namely $E_n \subset E_{n+1}$.

Then:

$$\mu(E_n) \rightarrow \mu\left(\bigcup_{n \in \mathbb{N}} E_n\right).$$

Proof. We build a sequence $\{F_n\}_{n \in \mathbb{N}}$ taking $F_n = E_n \setminus E_{n-1}$: so we have $F_0 = E_0$, $F_1 = E_1 \setminus E_0$, and so on. $\{F_n\}$ is a family of disjoint sets, with

$$\bigcup_{j \in \mathbb{N}} F_j = \bigcup_{n \in \mathbb{N}} E_n = E \quad \text{and} \quad \bigcup_{j=0}^n F_j = E_n.$$

Thus we have:

$$\begin{aligned} \mu(E_n) &= \mu\left(\bigcup_{j=1}^n F_j\right) \\ &= \sum_{j=0}^n \mu(F_j) \xrightarrow{n \rightarrow +\infty} \sum_{j \in \mathbb{N}} \mu(F_j) = \mu\left(\bigcup_{j \in \mathbb{N}} F_j\right) = \mu\left(\bigcup_{n \in \mathbb{N}} E_n\right). \end{aligned}$$
■

Proposition 2.1.26

Let μ be a positive measure and let $\{E_n\}_{n \in \mathbb{N}} \subset \mathfrak{M}$ be a descending sequence, namely $E_n \supset E_{n+1}$.

If $\mu(E_0) < +\infty$, then:

$$\mu(E_n) \rightarrow \mu\left(\bigcap_{n \in \mathbb{N}} E_n\right).$$

This does not hold when $\mu(E_0) = +\infty$. For example, take the measure space $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu_C)$ and the sequence $\{E_n\}_{n \in \mathbb{N}}$, with

$$E_n = \{n^* : n^* \in \mathbb{N}, n^* \geq n\} = \{n, n+1, \dots\}$$

. Then we have $E_0 \supset E_1 \supset E_2 \supset \dots$ and $\mu(E_n) = +\infty$ for any n , but $E = \bigcap_{n \in \mathbb{N}} E_n = \emptyset$.

The reader should try to do this proof before seeing the one provided.

Proof. We construct a sequence of ascending sets in order to take advantage of the previous proposition.

$$F_n = E_0 \setminus E_n \quad \forall n \in \mathbb{N} \quad \text{so that} \quad F_n \subset F_{n+1}.$$

Moreover we have:

$$\mu(F_n) = \mu(E_0) - \mu(E_n) \quad \bigcup_{n \in \mathbb{N}} F_n = E_0 \setminus \bigcap_{n \in \mathbb{N}} E_n,$$

Using the previous property:

$$\mu(F_n) \rightarrow \mu\left(\bigcup_{n \in \mathbb{N}} F_n\right)$$

Thus we have

$$\begin{aligned} \mu(E_0 \setminus E_n) &\rightarrow \mu\left(E_0 \setminus \bigcap_{n \in \mathbb{N}} E_n\right) \\ \mu(E_0) - \mu(E_n) &\rightarrow \mu(E_0) - \mu\left(\bigcap_{n \in \mathbb{N}} E_n\right) \\ \mu(E_n) &\rightarrow \mu\left(\bigcap_{n \in \mathbb{N}} E_n\right) \end{aligned}$$

as n goes to ∞ . ■

Proposition 2.1.27

Any positive measure μ is **countably subadditive**, namely:

$$\mu\left(\bigcup_{j \in \mathbb{N}} E_j\right) \leq \sum_{j \in \mathbb{N}} \mu(E_j)$$

for any sequence $\{E_j\}_{j \in \mathbb{N}} \subset \mathfrak{M}$, where the sets are not necessarily disjoint.

Proof. We start by proving that $\mu(E \cup F) \leq \mu(E) + \mu(F)$, with $E, F \in \mathfrak{M}$. We can apply finite additivity and monotonicity on the two disjoint sets $G_1 = E$, $G_2 = F \setminus E$:

$$\mu(E \cup F) = \mu(G_1 \cup G_2) \stackrel{FA}{=} \mu(G_1) + \mu(G_2) \stackrel{M}{\leq} \mu(E) + \mu(F).$$

Now we generalize this result to finite unions. Let $G_0 = E_0$, $G_n = E_n \setminus \left\{ \bigcup_{j=0}^{n-1} E_j \right\}$ for all $n > 0$; we can easily see that the sets $\{G_j\}$ are disjoint, and thus:

$$\mu\left(\bigcup_{j=0}^n E_j\right) = \mu\left(\bigcup_{j=0}^n G_j\right) \stackrel{FA}{=} \sum_{j=0}^n \mu(G_j) \stackrel{M}{\leq} \sum_{j=0}^n \mu(E_j).$$

The proof is complete letting $n \rightarrow +\infty$. ■

Exercise 2.1.28. Consider a set function μ that is countably subadditive and finitely additive, and prove it is also countably additive.

At last, the characterization:

Proposition 2.1.29

Let (Ω, \mathfrak{M}) be a measurable space.

A set function $\mu : \mathfrak{M} \rightarrow [0, +\infty]$ is a positive measure if and only if it is finitely additive and countably sub-additive.

Proof. The proof of the \implies part is trivial: countable additivity implies sub-additivity and by choosing a suitable collection of sets one easily gets finite additivity.

To prove the converse we argue as follows. Since $\mu \not\equiv +\infty$, there exists $E \in \mathfrak{M}$ such that $\mu(E) < +\infty$. Thus on account of finite additivity:

$$\mu(E) = \mu(E \cup \emptyset) = \mu(E) + \mu(\emptyset),$$

which implies $\mu(\emptyset) = 0$. Moreover it's easy to see that $E \subset F$ implies $\mu(E) \leq \mu(F)$, also for finite additivity:

$$\mu(F) = \mu((F \setminus E) \cup E) = \mu(F \setminus E) + \mu(E) \geq \mu(E).$$

Consider now $\{E_n\} \subset \mathfrak{M}$ a family of countable disjoint sets. For any $N \in \mathbb{N} \setminus \{0\}$ using again finite additivity and the property we just showed:

$$\mu \left(\bigcup_{n=1}^{\infty} E_n \right) \geq \mu \left(\bigcup_{n=1}^N E_n \right) = \sum_{n=1}^N \mu(E_n)$$

then taking the limit as $N \rightarrow \infty$ we get:

$$\mu \left(\bigcup_{n=1}^{\infty} E_n \right) \geq \sum_{n=1}^{\infty} \mu(E_n)$$

on the other hand we have the other inequality from the countable sub-additivity, so we have and equality (i.e. countable additivity). ■

Uniqueness of a measure The following theorem is well known in probability for proving the uniqueness of the probability¹⁵.

Theorem 2.1.30 (Dynkin's lemma)

Let $\Omega \neq \emptyset$, $\mathcal{E} \subset \mathcal{P}(\Omega)$, such that \mathcal{E} is closed with respect to finite intersection and let a sequence $\{E_n\} \subset \mathcal{E}$ exists such that $\Omega = \bigcup_n E_n$.

Let μ and ν be two measures on $\mathfrak{M} = \langle \mathcal{E} \rangle$.

If $\mu(E_n) < +\infty$, $\nu(E_n) < +\infty$ for all n , and $\mu|_{\mathcal{E}} \equiv \nu|_{\mathcal{E}}$ ¹⁶, then $\mu \equiv \nu$.

Example 2.1.31. Take $\Omega = \mathbb{R}$, $\mathfrak{M} = \mathcal{B}(\mathbb{R}) = \langle \{(a, b) : a, b \in \mathbb{R}\} \rangle$.

Suppose that it exists a Borel measure μ such that $\mu((a, b)) = b - a$: in this case μ is unique. This example can be easily generalized to multidimensional case $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$.

¹⁵In Italian that theorem is known as *teorema delle classi monotone*. See F. Bernardi, G. Cerri, A. Di Nardo, G. Gabrielli, B. Guindani, S. Polito, A. Wussler, *Appunti di Probabilità - Edizione L^p* , pages 32-34, section 3.1.1, theorem 3.4 .

¹⁶That is $\mu(E) = \nu(E) \forall E \in \mathcal{E}$

Complete measure This is another property that we will see is required in a number of further results.

Definition 2.1.32

Let $(\Omega, \mathfrak{M}, \mu)$ be a measure space.

We say that μ is **complete** if, for every $A \in \mathfrak{M}$ such that $\mu(A) = 0$ and for every $E \subset A$, E is \mathfrak{M} -measurable.

In plain language, a measure is complete if all the subsets of a zero-measure set are measurable.

Theorem 2.1.33 (Completion of a measure space)

Let $(\Omega, \mathfrak{M}, \mu)$ be a measure space.

Let also $\mathfrak{M}^* = \{P \subset \Omega \mid \exists E, F \in \mathfrak{M} : E \subset P \subset F \text{ and } \mu(F \setminus E) = 0\}$, μ^* such that $\mu^*(P) = \mu(E)$ for every $P \in \mathfrak{M}^*$.

Then \mathfrak{M}^* is a σ -algebra, $\mathfrak{M}^* \supset \mathfrak{M}$, and μ^* is a complete measure.

The measure space $(\Omega, \mathfrak{M}^*, \mu^*)$ is called **completion** of $(\Omega, \mathfrak{M}, \mu)$.

Proof. Step 1, \mathfrak{M}^ is a σ -algebra:*

We prove the consistency of the definition of \mathfrak{M}^* with the definition of σ -algebra.

- To check that $\emptyset \in \mathfrak{M}^*$, notice that $\emptyset \in \mathfrak{M}$; so we can simply take $P = E = F = \emptyset$.
- To check that \mathfrak{M}^* is closed under complementation, take $P \in \mathfrak{M}^*$, we have $E \subset P \subset F$ with $\mu(F \setminus E) = 0$, so $E^C \supset P^C \supset F^C$ with $\mu(E^C \setminus F^C) = 0$, and thus $P^C \in \mathfrak{M}^*$.
- To check that \mathfrak{M}^* is closed under countable union, let $\{P_n\}_{n \in \mathbb{N}} \subset \mathfrak{M}^*$, then $E_n \subset P_n \subset F_n$ with $\mu(F_n \setminus E_n) = 0$ and $n \in \mathbb{N}$. Taking $E = \bigcup_{n \in \mathbb{N}} E_n \in \mathfrak{M}$, $P := \bigcup_{n \in \mathbb{N}} P_n$, and $F = \bigcup_{n \in \mathbb{N}} F_n \in \mathfrak{M}$, we have $E \subset P \subset F$. Also:

$$\mu(F \setminus E) = \mu \left(\bigcup_{n \in \mathbb{N}} (F_n \setminus E_n) \right) \leq \sum_{n \in \mathbb{N}} \mu(F_n \setminus E_n) = 0$$

and thus $\bigcup_{n \in \mathbb{N}} P_n \in \mathfrak{M}^*$.

Step 2, μ^ is a well-defined set function:*

Namely, we have to show that the value of $\mu^*(P) = \mu(E)$ does not change

when choosing a different E .

Let $P, E_1, F_1, E_2, F_2 \in \mathfrak{M}$ such that $E_i \subset P \subset F_i$ and $\mu(F_i \setminus E_i) = 0$.

Then we have $E_1 \setminus E_2 \subset P \setminus E_2 \subset F_2 \setminus E_2$, and thus $\mu(E_1 \setminus E_2) \leq \mu(F_2 \setminus E_2) = 0$.

Because $E_1 = (E_1 \setminus E_2) \cup (E_1 \cap E_2)$, we have $\mu(E_1) = \mu(E_1 \cap E_2)$. Similarly we can prove that $\mu(E_2) = \mu(E_2 \cap E_1)$, and finally that $\mu(E_1) = \mu(E_2)$.

Step 3, μ^ is a measure:*

Finally we check that μ^* is indeed a measure. First of all, $\mu^*(\emptyset) = 0$. Then, consider $\{P_n\}_{n \in \mathbb{N}} \subset \mathfrak{M}^*$ such that $P_i \cap P_j = \emptyset$ when $i \neq j$. For every $n \in \mathbb{N}$, $E_i \subset P_i \subset F_i$, and thus $E_i \cap E_j = \emptyset$ when $i \neq j$. Therefore:

$$\mu^* \left(\bigcup_{n \in \mathbb{N}} P_n \right) = \mu \left(\bigcup_{n \in \mathbb{N}} E_n \right) = \sum_{n \in \mathbb{N}} \mu(E_n) = \sum_{n \in \mathbb{N}} \mu^*(P_n).$$

■

2.1.4 Lebesgue measure

Our goal is to build a complete measure space on \mathbb{R}^N . In particular we want a measure defined for all the sets of the Borel σ -algebra, so we have to define such measure on his σ -algebra $\mathfrak{M}(\mathbb{R}^N) \supset \mathcal{B}(\mathbb{R}^N)$.

First attempt In order to have a spendable definition, we need that our measure $\tilde{\lambda}$ satisfies three properties: first, the measure of rectangles is the product of the euclidean measure of its sides, namely:

$$\lambda(R) = \prod_{j=1}^n (b_j - a_j) \text{ for any } R = (a_1, b_1) \times \cdots \times (a_n, b_n);$$

second, the measure has to be invariant with respect to any translation, namely:

$$\lambda(E + x) = \lambda(E) \quad \forall x \in \mathbb{R}^N \quad \forall E \in \mathfrak{M}(\mathbb{R}^N);$$

and last, measure isn't *concentrated* somewhere, namely:

$$\nexists E \in \mathfrak{M}(\mathbb{R}^N) \text{ such that } \lambda(E) > 0 \text{ and } \forall F \subsetneq E : \lambda(F) = 0.$$

If such kind of measure does exist, by Dynkin's Lemma (2.1.30 on page 96) we know that it is unique.

Another question is this, assuming that such measure exists, can we have that $\mathfrak{M}(\mathbb{R}^N) = \mathcal{P}(\mathbb{R}^N)$, that is every subset of \mathbb{R}^N is measurable? Well, the answer is no, and the explanation requires a little effort.

Definition 2.1.34

Let $(\Omega, \mathfrak{M}, \mu)$ be a measure space.

A set A is an **atom** if it is a countable set, has $\mu(A) > 0$, and it does not contain any measurable subset E such that $\mu(E) > 0$.

Example 2.1.35. Consider $(\Omega, \mathfrak{M}, \mu_t)$, where μ_t is the Dirac measure. Then $\{t\}$ is an atom.

Taking account of the previous definition, due to its third property, our measure is surely **non-atomic**. The following theorem is proven through the axiom of choice and the continuum hypothesis:

Theorem 2.1.36 (Ulam)

The only non-atomic measure on $\mathcal{P}(\mathbb{R})$ is the **zero measure** $\mu(E) = 0$ for all $E \in \mathcal{P}(\mathbb{R})$.

This means that the σ -algebra on which our measure is built must be strictly smaller than $\mathcal{P}(\mathbb{R}^N)$ and this answers the question. So there are some subsets of \mathbb{R}^N that can't be measured by a measure with those sole properties.

Outer measure For the sake of simplicity from now on we will deal with \mathbb{R} instead of \mathbb{R}^N . It is easy to take the following results to the multidimensional case in which, as theory guarantee, they hold as well. To work around the problem our first step is the following, simpler, definition of something that is not a positive measure.

Definition 2.1.37

We define the **outer measure** $\lambda^* : \mathcal{P}(\mathbb{R}) \rightarrow [0, +\infty]$ as follows:

$$\lambda^*(E) := \inf_{\{I_n\}_{n \in \mathbb{N}}} \sum_{n \in \mathbb{N}} \ell(I_n)$$

where I_n is an open interval for any n , $\ell(I_n)$ is its length, and $\bigcup_{n \in \mathbb{N}} I_n \supset E$.

Notice that it is defined on the power set of \mathbb{R} .

Proposition 2.1.38 (Main properties of the outer measure)

Let λ^* be the outer measure, then:

- the outer measure is monotone increasing, namely:

$$E \subset F \implies \lambda^*(E) \leq \lambda^*(F);$$

- the outer measure is countably subadditive.
- for any countable $E \in \mathcal{P}(\mathbb{R})$, $\lambda^*(E) = 0$;

Proof of the countably subadditivity. Consider $\{E_n\}_{n \in \mathbb{N}} \subset \mathcal{P}(\mathbb{R})$ and fix $\varepsilon > 0$. Then, by definition, for all $n \in \mathbb{N}$ there exists a sequence $\{I_j^n\}_{j \in \mathbb{N}}$ such that:

$$\bigcup_{j \in \mathbb{N}} I_j^n \supset E_n \quad \text{and} \quad \sum_{j \in \mathbb{N}} \ell(I_j^n) \leq \lambda^*(E_n) + \frac{\varepsilon}{2^{n+1}}.$$

Now set $E := \bigcup_{n \in \mathbb{N}} E_n$, then we have $E \subset \bigcup_{n \in \mathbb{N}} \bigcup_{j \in \mathbb{N}} I_j^n$; we can observe that:

$$\lambda^*(E) \leq \lambda^* \left(\bigcup_{n \in \mathbb{N}} \bigcup_{j \in \mathbb{N}} I_j^n \right) \leq \sum_{n \in \mathbb{N}} \sum_{j \in \mathbb{N}} \ell(I_j^n) \leq \sum_{n \in \mathbb{N}} \left(\lambda^*(E_n) + \frac{\varepsilon}{2^{n+1}} \right) = \sum_{n \in \mathbb{N}} \lambda^*(E_n) + \varepsilon.$$

Since ε is arbitrary, this gives us sub-additivity. ■

Proof of measure of countable sets. Let us prove that $\lambda^*({x_0}) = 0$, indeed:

$$\lambda^*({x_0}) = \inf_{\varepsilon > 0} (x_0 + \varepsilon - (x_0 - \varepsilon)) = \inf_{\varepsilon > 0} 2\varepsilon = 0$$

Then using countable subadditivity we deduce that any countable set has outer measure zero. ■

Carathéodory condition Outer measure is not a positive measure as it is not finitely additive. If it was, it would be a non-atomic measure defined on $\mathcal{P}(\mathbb{R})$, this contradicts Ulam's theorem (2.1.36 on the preceding page). To better understand this issue, consider $E \in \mathcal{P}(\mathbb{R})$ such that $\lambda^*(E) > 0$, pick $x \in E$ and set $F = E \setminus \{x\}$. Notice that

$$\lambda^*(E) = \lambda^*(F \cup \{x\}) \leq \lambda^*(F) + \lambda^*({x}) = \lambda^*(F).$$

Moreover, by monotonicity $\lambda(F)^* \leq \lambda(E)^*$ so $\lambda^*(E) = \lambda^*(F)$. This shown that the outer measure is not non-atomic.

Anyway our λ^* has the properties we are looking for: $\lambda^*((a, b)) = b - a \forall a, b \in \mathbb{R}$ and it is invariant with respect to any translation; a willing reader can try to prove these properties.

All of those remarks are giving us an hint: maybe to make λ^* a measure we only have to reduce the σ -algebra. So we could move our struggle from the definition of a the measure to a restriction of the σ -algebra. Indeed, by doing so, we can restore the finite additivity. This reasoning is completed in the following results.

Definition 2.1.39

A set $E \in \mathcal{P}(\mathbb{R})$ satisfies the **Carathéodory condition** if:

$$\lambda^*(T) = \lambda^*(T \cap E) + \lambda^*(T \cap E^C) \quad \text{for any } T \in \mathcal{P}(\mathbb{R}).$$

Actually, “ \geq ” is enough in the condition because “ \leq ” is already given by subadditivity.

Definition 2.1.40

The family of the sets that respect the Carathéodory condition is a σ -algebra and is called **Lebesgue σ -algebra**:

$$\mathcal{L}(\mathbb{R}) := \{E \subset \mathcal{P}(\mathbb{R}) : E \text{ satisfy the Carathéodory condition}\}.$$

Its elements are called **Lebesgue sets**¹⁷.

The Lebesgue measure fits also for “smaller” sets: consider any set $\Omega \in \mathcal{L}(\mathbb{R})$. We can define $\mathcal{L}(\Omega)$ by selecting only the subset which satisfy the Carathéodory condition, and the Lebesgue measure on Ω is simply the restriction of the Lebesgue measure.

Definition 2.1.41

The restriction of λ^* to $\mathcal{L}(\mathbb{R})$ is called **Lebesgue measure**.

To conclude the discussion we come back to the main doubt: is the Lebesgue measure actually a measure? Yes, and the following result confirm the conclusion of our quest.

¹⁷For the proof that \mathcal{L} is a σ -algebra and many other results, see: H. L. Royden, Real Analysis, pages 251-253, section 12.1, theorem 1 .

Theorem 2.1.42

The Lebesgue measure is finitely additive.

Proof. Consider two disjoint sets $A, B \in \mathcal{L}(\mathbb{R})$, and let $T = A \cup B$. We have

$$\lambda^*(A \cup B) = \lambda^*(T) \stackrel{CC}{=} \lambda^*(T \cap A) + \lambda^*(T \cap A^C) = \lambda^*(A) + \lambda^*(B).$$

This shows the thesis as it is trivial to extend this argument to any finite number of set. ■

Properties and relevant facts about the Lebesgue measure Here some proposition deduced by previous results to complete our discussion.

Proposition 2.1.43

The Lebesgue σ -algebra contains the Borel σ -algebra:

$$\mathcal{L}(\mathbb{R}) \supset \mathcal{B}(\mathbb{R}).$$

Proof. Step 0, the goal:

To complete this proof, thanks to the generative properties of σ -algebras, is sufficient to prove that $(a, +\infty) \in \mathcal{L}(\mathbb{R})$, satisfying the Carathéodory condition.

Step 1, infinite sets:

Now consider all the subset T of \mathbb{R} . If $\lambda^*(T) = +\infty$, then it is easy to see that $\lambda^*(T) \geq \lambda^*(T \cap (a, +\infty)) + \lambda^*(T \cap (a, +\infty)^C)$.

Step 2, finite sets:

Consider now T such that $\lambda^*(T) < +\infty$ and fix $\varepsilon > 0$; then there exists a sequence of set $\{I_n\}_{n \in \mathbb{N}}$ such that:

$$T \subset \bigcup_n I_n \quad \text{and} \quad \sum_n \ell(I_n) \leq \lambda^*(T) + \varepsilon;$$

by setting $I_n^1 := I_n \cap (a, +\infty)$ and $I_n^2 := I_n \cap (-\infty, a + \frac{\varepsilon}{2^n})$ we have that:

$$\ell(I_n) + \frac{\varepsilon}{2^n} \geq \ell(I_n^1) + \ell(I_n^2).$$

Consider now the following two set:

$$T_1 := T \cap (a, +\infty) \quad \text{and} \quad T_2 := T \cap (-\infty, a].$$

By the previous definition we have that $\bigcup_n I_n^1 \supset T_1$ and $\bigcup_n I_n^2 \supset T_2$. Thus, by monotonicity, $\lambda^*(T_1) \leq \sum_n \ell(I_n^1)$ and $\lambda^*(T_2) \leq \sum_n \ell(I_n^2)$.

Summing up, we have:

$$\begin{aligned} \lambda^*(T_1) + \lambda^*(T_2) &\leq \sum_n \ell(I_n^1) + \sum_n \ell(I_n^2) \\ &\leq \sum_n \ell(I_n) + \sum_n \frac{\varepsilon}{2^n} \\ &\leq \lambda^*(T) + 2\varepsilon. \end{aligned}$$

Taking $\varepsilon \rightarrow 0$ we have that $\lambda^*(T_1) + \lambda^*(T_2) \leq \lambda^*(T)$, and the Carathéodory condition holds for $(a, +\infty)$. ■

The inclusion is strict, indeed it can be proven that the cardinality of $\mathcal{B}(\mathbb{R})$ is 2^{\aleph_0} . However, we have that the **Cantor set** (see definition 1.2.89 on page 67), which belongs to $\mathcal{B}(\mathbb{R})$ since in the construction we use operations which do not bring us outside the sigma algebra, is uncountable (\aleph_1) and L-measurable (zero measure). By completeness of the Lebesgue measure, all its subsets are measurable, hence $\mathcal{P}(\mathcal{C}) \in \mathcal{L}(\mathbb{R})$, but its cardinality is at least $2^{\aleph_1} = 2^{2^{\aleph_0}}$, hence $\mathcal{P}(\mathcal{C})$ cannot be contained in $\mathcal{B}(\mathbb{R})$.

Proposition 2.1.44

The Lebesgue measure space $(\mathbb{R}, \mathcal{L}(\mathbb{R}), \lambda^*)$ is complete.

Proof. Take E, N such that $E \subset N$ and $\lambda(N) = 0$. We need to prove that E is measurable.

First of all, for all $T \in \mathcal{P}(\mathbb{R})$, we have that $T \cap E \subset T \cap N \subset N$, and thus $\lambda^*(T \cap E) = 0$ by monotonicity.

We also have that $T \cap E^C \subset T$, and so $\lambda^*(T \cap E^C) \leq \lambda^*(T)$.

Summing up, $\lambda^*(T) \geq \lambda^*(T \cap E) + \lambda^*(T \cap E^C)$, E respects the Carathéodory condition, and thus it is measurable. ■

In particular, $(\mathbb{R}, \mathcal{L}(\mathbb{R}), \lambda^*)$ is the completion of $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda^*)$. Indeed, the restriction of the measure to the Borel σ -algebra, namely $\lambda|_{\mathcal{B}(\mathbb{R})}$, is a Borel measure, but it is not complete.

Let us clarify the relationship between Lebesgue and Borel sets (recall definition 2.1.6 on page 82).

Proposition 2.1.45

The following statements are equivalent:

1. $E \in \mathcal{L}(\mathbb{R})$;
2. for all $\varepsilon > 0$, there exists A open set such that $A \supset E$ and $\lambda(A \setminus E) < \varepsilon$;
3. for all $\varepsilon > 0$, there exists C closed set such that $C \subset E$ and $\lambda(E \setminus C) < \varepsilon$;
4. there exists a G_δ set $G \supset E$ such that $\lambda(G \setminus E) = 0$;
5. there exists a F_σ set $F \subset E$ such that $\lambda(E \setminus F) = 0$.

Thus every Lebesgue set E can be written as $E = F \cup (E \setminus F)$, where F is a F_σ set, and thus $F \in \mathcal{B}(\mathbb{R})$, $E \setminus F \in \mathcal{L}(\mathbb{R})$ and $\lambda(E \setminus F) = 0$.

Proof. We will prove that 1 implies 2 which implies 4 which implies 1.

Step 1, 1 implies 2:

Start by considering $\lambda(E) < +\infty$. Then, for all $\varepsilon > 0$ there exists a sequence $\{I_n\}_{n \in \mathbb{N}} \subset \mathbb{R}$ such that:

$$\bigcup_{n \in \mathbb{N}} I_n \subset E, \quad \sum_{n \in \mathbb{N}} \lambda(I_n) \leq \lambda(E) + \varepsilon.$$

Set $A := \bigcup_{n \in \mathbb{N}} I_n$, then $A \supset E$ and $\lambda(A \setminus E) < \varepsilon$. As $\lambda(E) = +\infty$, by setting $E_n = E \cap [-n, n]$ we have $\bigcup_{n \in \mathbb{N}} E_n = E$ and $\lambda(E_n) < +\infty$.

From the previous point we know that there exists an open set A_n such that:

$$\lambda(A_n \setminus E_n) \leq \frac{\varepsilon}{2^{n+1}}.$$

Set $A := \bigcup_{n \in \mathbb{N}} A_n \supset E$ and observe that:

$$\lambda(A \setminus E) \leq \sum_{n \in \mathbb{N}} \lambda(A_n \setminus E_n) \leq \sum_{n \in \mathbb{N}} \frac{\varepsilon}{2^{n+1}} = \varepsilon.$$

This proves the implication.

Step 2, 2 implies 4:

Observe that for all $n \in \mathbb{N}_0$, there exists an open set $A_n \subset E$ such that:

$$\lambda(A_n \setminus E) < \frac{1}{n}.$$

Set $G := \bigcap_{n \in \mathbb{N}_0} A_n$, then

$$\lambda(G \setminus E) \leq \lambda(A_n \setminus E) < \frac{1}{n} \quad \forall n \in \mathbb{N}_0.$$

Let $n \rightarrow +\infty$ to get $\lambda(G \setminus E) = 0$. This concludes this second proof.

Step 3, 4 implies 1:

As we can write $E = G \setminus (G \setminus E)$, we have $E \in \mathcal{L}(\mathbb{R})$. This because $G \in \mathcal{B}(\mathbb{R})$ and $(G \setminus E) \in \mathcal{L}(\mathbb{R})$; in particular, we know that its measure is zero.

Step ∞ :

You should prove that 1 implies 3 which implies 5 which implies 1. To do this use the complements! ■

Finally, observe that $\mathcal{L}(\mathbb{R}) \not\subseteq \mathcal{P}(\mathbb{R})$ due to Ulam's theorem.

We can extend our considerations about Lebesgue measure to \mathbb{R}^N and define the measure space $(\mathbb{R}^N, \mathcal{L}(\mathbb{R}^N), \lambda)$. In the construction of λ^* we can simply consider multidimensional intervals and the euclidean n -volume:

$$I = (a_1, b_1) \times \cdots \times (a_N, b_N), \quad V(I) = \prod_{i=1}^N (b_i - a_i).$$

Vitali's theorem This theorem conclude our discussion on the Lebesgue theory. It requires the axiom of choice!

Theorem 2.1.46 (Vitali)

Any set $E \in \mathcal{L}(\mathbb{R})$ with $\lambda(E) > 0$ contains a subset which is not Lebesgue-measurable.

Proof. Without loss of generality, we will prove the theorem for the case $E = [0, 1]$.

In $E = [0, 1]$, consider the equivalence relation $a \sim b \iff a - b \in \mathbb{Q}$ and take the quotient set $\frac{E}{\sim}$. Obviously, one of the equivalence classes in this quotient set contains any and all the rational numbers in E , because the difference of two rationals is rational, so they are equivalent. And in the other classes there are essentially an irrational and the rational translations. Using the axiom of choice, let us pick an element from each of the equivalence classes, and define $\mathcal{V} \subset E$ as the set composed of these picked elements. The set \mathcal{V} is called **Vitali set**.

Then we have:

$$(\mathcal{V} + r) \cap (\mathcal{V} + s) = \emptyset \quad \forall r, s \in \mathbb{Q} : r \neq s,$$

where $(\mathcal{V} + a)$ denotes the set created by translating the elements of \mathcal{V} by a .

Otherwise, there would exist $a, b \in \mathcal{V}$, $a \neq b$ (namely $a - b \notin \mathbb{Q}$) being representative of different classes, such that $a + r = b + s$, that is, $a - b = s - r \in \mathbb{Q}$. Contradiction.

Again by contradiction, suppose that \mathcal{V} is λ -measurable. As $\mathbb{R} = \bigcup_{r \in \mathbb{Q}} (\mathcal{V} + r)$, which is a countable union of disjoint sets, we have:

$$\begin{aligned} +\infty &= \lambda(\mathbb{R}) \\ &= \lambda \left(\bigcup_{r \in \mathbb{Q}} (\mathcal{V} + r) \right) \\ &= \sum_{r \in \mathbb{Q}} \lambda(\mathcal{V} + r) && (\sigma\text{-additivity}) \\ &= \sum_{r \in \mathbb{Q}} \lambda(\mathcal{V}). && (\text{properties of } \lambda) \end{aligned}$$

If $\lambda(\mathcal{V}) = 0$ we get a contradiction and the proof is finished, let's proceed then with $\lambda(\mathcal{V}) > 0$.

Now set $F := \bigcup_{r \in \mathbb{Q} \cap E} (\mathcal{V} + r) \subset [0, 2]$. Via σ -additivity, one has:

$$\lambda(F) = \sum_{r \in \mathbb{Q} \cap E} \lambda(\mathcal{V} + r) = \sum_{r \in \mathbb{Q} \cap E} \lambda(\mathcal{V}) = +\infty.$$

But this is impossible, because $\lambda(F) \leq \lambda([0, 2]) = 2$ by monotonicity. Therefore, \mathcal{V} cannot be λ -measurable. Hence we get a contradiction and the proof is finished. ■

For further reading, try to think (or Google) what is the outer measure of a Vitali set.

2.2 Abstract integration

Measures are about sets, but how about functions? Abstract and theoretical concepts that we explored up to now can be employed to produce a notion of integral.

Riemann integral, which is studied in lower calculus courses, take with him many issues and in many cases it does not provide the answers we are looking for. In order to develop a more powerful integral the first steps here are to define a very general notion of integral, the abstract integral, which can be used only on a very basic kind of function, called simple functions, and study some immediate yet useful result.

Before find the solution of our problem in the next section, we will discuss also some fundamental result of derivatives.

2.2.1 Abstract integral

The setting is very general; we will not fix neither the measure space nor the measure, for instance we could use the Lebesgue measure as the counting measure. From now on, we only require that $(\Omega, \mathfrak{M}, \mu)$ is a complete measure space.

Definition 2.2.1

We define the **abstract integral** of a simple function s as follows:

$$\int_E s \, d\mu := \sum_{j=1}^N a_j \mu(E \cap E_j).$$

Simple functions are defined in definition [2.1.14 on page 87](#). Notice that $\int_E s \, d\mu = \int_\Omega s \mathbf{1}_E \, d\mu$.

Given a simple function and the notion of integral it's possible to define a new measure.

Proposition 2.2.2

The set function $\nu(E) = \int_E s \, d\mu$, defined on any $E \in \mathfrak{M}$, is a measure on \mathfrak{M} .

You can prove this proposition... do it! This result has a great importance as we will see after.

First thing to improve the notion of abstract integral is to expand the range of function that can be measured. The improvement we made in the next definition consist of extend the notion to any positive function; this thanks to combining the abstract integral with the notion of limit.

Definition 2.2.3

For any measurable function $f : \Omega \rightarrow [0, +\infty]$, we define its **Lebesgue abstract integral for positive functions**:

$$\int_E f \, d\mu := \sup_{0 \leq s \leq f} \int_E s \, d\mu \in [0, +\infty]$$

where s is a simple function.

To guarantee the measurability of sums $(f + g)$ and products (fg) of measurable functions we implicitly used an arithmetization similar to the one on \mathbb{R}^* ; the following rules holds for $a \geq 0$:

$$\begin{aligned} a + \infty &= +\infty + a = +\infty; \\ a \cdot (+\infty) &= (+\infty) \cdot a = +\infty; \\ 0 \cdot (+\infty) &= (+\infty) \cdot 0 = +\infty. \end{aligned}$$

This is formally a partial arithmetization on $[0, +\infty]$.

This integral is a much more general notion compared to the Riemann integral on \mathbb{R} . With Riemann integrals, we had very strict criteria for integrability; now almost any function has a Lebesgue integral. Moreover, with Lebesgue we do not have to build a supplementary notion of improper integral.

The Riemann integral is built starting from intervals (which have an easily computable size) on Ω , the domain of integration. With the Lebesgue integral, one is splitting the codomain into intervals instead. Since their preimages may not be intervals, one needs a measure to define the size of such preimages. For the Riemann integral, the Peano measure for intervals is implicitly used.

Proposition 2.2.4 (Basic properties of the Lebesgue integral)

The following properties holds:

1. **monotonicity with respect to functions:**

if $f \leq g$, then $\int_E f d\mu \leq \int_E g d\mu \quad \forall E \in \mathfrak{M}$;

2. **monotonicity with respect to sets:**

if $E, F \in \mathfrak{M}$ and $E \subseteq F$, then $\int_E f d\mu \leq \int_F f d\mu$;

3. **homogeneity:**

for any $\alpha \geq 0$, $\int_E \alpha f d\mu = \alpha \int_E f d\mu$;

4. **annihilation with respect to null functions**

if $f = 0$ in E , then $\int_E f d\mu = 0$;

5. **annihilation with respect to zero-measure sets**

if $\mu(E) = 0$, then $\int_E f d\mu = 0$.

The proofs can be easily achieved from foretold definitions.

2.2.2 Monotone convergence theorem and Fatou's lemma

There are three important results on exchange between limits and integral of sequences of functions: the Levi's monotone convergence theorem, the Fatou's lemma and the dominated convergence theorem. This section is devoted to study the first two. The third one will be discussed in the section [2.3.2 on page 119](#).

Monotone convergence theorem The following is a very useful theorem by an Italian mathematician, Beppo Levi. It provides a way to *exchange limits* between integrals and sequences of functions.

Theorem 2.2.5 (Beppo Levi's monotone convergence theorem)

Let $f_n : \Omega \rightarrow [0, +\infty]$ be a measurable and monotone non-decreasing sequence of function; that is for any $t \in \Omega$, $f_{n+1}(t) \geq f_n(t)$, for all $n \in \mathbb{N}$.

Then $f(t) := \lim_{n \rightarrow \infty} f_n(t)$ is measurable and:

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

Notice that, by setting $f = \lim_{n \rightarrow +\infty} f_n$, we know that f is measurable, as it coincides with the supremum (see theorem [2.1.13 on page 86](#) and the following proof).

Proof. Step 0, initial setting:

Consider the sequence of real numbers $\{\int_{\Omega} f_n d\mu\}_{n \in \mathbb{N}}$, as it is monotone non-decreasing, its limit exists and it's the supremum:

$$\alpha = \lim_{n \rightarrow +\infty} \int_{\Omega} f_n d\mu = \sup_{n \in \mathbb{N}} \int_{\Omega} f_n d\mu \in [0, +\infty].$$

Step 1, proof of $\int_{\Omega} f d\mu \geq \alpha$:

Since $f_n(t) \leq f(t)$, we have:

$$\alpha = \sup_{n \in \mathbb{N}} \int_{\Omega} f_n d\mu \leq \sup_{n \in \mathbb{N}} \int_{\Omega} f d\mu = \int_{\Omega} f d\mu.$$

Step 2, proof of $\int_{\Omega} f d\mu \leq \alpha$:

Fix $c \in (0, 1)$ and s , an arbitrary simple function such that $0 \leq s \leq f$.

Now set

$$E_n := \{t \in \Omega : f_n(t) \geq cs(t)\},$$

where $0 \leq cs(t) \leq f_n(t) \leq f(t)$ for $t \in E_n$.

Recalling that f_n is a non-decreasing sequence, it is easy to see that E_n is ascending, that is $E_1 \subseteq E_2 \subseteq \dots$.

Now, take any $t \in \Omega$. If $f(t) = 0$, then $s(t) = 0$ as well and $t \in E_{n_0}$ for any n_0 .

If instead $f(t) > 0$ (f is non-negative), since $f(t) = \sup_{n \in \mathbb{N}} f_n(t)$, there is room for one more value between $cs(t)$ and $f(t)$, that is:

$$\exists n_0 \in \mathbb{N} : f(t) \geq f_{n_0}(t) \geq cs(t).$$

Thus $t \in E_{n_0}$. Recall that both the sequence of sets and the sequence of integrand functions are monotonous.

Notice an important step here, c cannot be equal 1. Indeed, if $c = 1$ think of a situation when the limit function $f(t)$ is flat at some point, then the simple function coincides with f in that portion. But if this is the case, we can't find an n_0 such that f_n fits in between, unless it reaches the limit in *finite* steps.¹⁸ Since every t is part of a certain E_n , it follows that $\Omega = \bigcup_{n \in \mathbb{N}} E_n$. Observe that if c were greater than or equal to 1 and also $f(t) = s(t)$ for some t , then $t \notin E_n$ for any n .

¹⁸But this hardly ever happens.

Finally:

$$\alpha \geq \int_{\Omega} f_n d\mu \geq \int_{E_n} f_n d\mu \geq c \int_{E_n} s(t) d\mu = c\nu(E_n)$$

and $c\nu(E_n) \xrightarrow{n \rightarrow \infty} c\nu(\Omega)$, therefore $\alpha \geq c\nu(\Omega) = c \int_{\Omega} s d\mu$ for any $c \in (0, 1)$ and for any simple function s such that $0 \leq s \leq f$.

Taking the superior limit on both sides we have:

$$\alpha \geq c \sup_{0 \leq s \leq f} \int_{\Omega} s d\mu = c \int_{\Omega} f d\mu \rightarrow \int_{\Omega} f d\mu \quad \text{as } c \rightarrow 1^-.$$

Thus we have $\alpha \geq \int_{\Omega} f d\mu$. ■

Observe that, if we have a sequence of simple functions $\{s_n\}_{n \in \mathbb{N}}$ such that $s_n(t) \uparrow f(t)$ for all $t \in \Omega$, then, thanks to this theorem, we have: $\int_{\Omega} f d\mu = \lim_{n \rightarrow \infty} \int_{\Omega} s_n d\mu$.

Beppo Levi's theorem is a very useful tool in many proof and it has some interesting consequences.

Proposition 2.2.6 (Additivity of the abstract integral)

Let $(\Omega, \mathfrak{M}, \mu)$ be measurable space, $f, g : \Omega \rightarrow [0, +\infty]$ measurable.

Then:

$$\int_E (f + g) d\mu = \int_E f d\mu + \int_E g d\mu$$

Proof. Consider a sequence of simple functions $\{s_n^1\}$, which is monotonically increasing, namely $s_{n+1}^1 \geq s_n^1 \in \Omega$, and converges point-wise to a function f , namely

$$s_n^1(t) \uparrow f(t) \quad \forall t \in \Omega.$$

Consider also another monotonic sequence of simple function $\{s_n^2\}$ which converges point-wise to g .

Define the sequence of the sum: $s_n = s_n^1 + s_n^2$; it monotonically converges point-wise to the sum of f and g , namely

$$s_n \uparrow f + g.$$

You, reader, should prove, from definition 2.2.1 on page 107, that the integral for simple functions is additive¹⁹:

$$\int_E (s_n^1 + s_n^2) d\mu = \int_E s_n^1 d\mu + \int_E s_n^2 d\mu = \int_\Omega s_n^1 \mathbf{1}_E d\mu + \int_\Omega s_n^2 \mathbf{1}_E d\mu$$

Now we can use monotone convergence theorem to exchange the limit with the integral obtaining the thesis:

$$\begin{aligned} \int_E (f + g) d\mu &= \int_E \lim_{n \rightarrow +\infty} s_n d\mu \\ &= \lim_{n \rightarrow +\infty} \int_E (s_n^1 + s_n^2) d\mu \\ &= \lim_{n \rightarrow +\infty} \int_\Omega (s_n^1 + s_n^2) \mathbf{1}_E d\mu \\ &= \lim_{n \rightarrow +\infty} \int_\Omega s_n^1 \mathbf{1}_E + \int_\Omega s_n^2 \mathbf{1}_E d\mu \\ &= \int_E f d\mu + \int_E g d\mu. \end{aligned}$$

■

This theorem is a version of monotone convergence which holds for series of functions:

Corollary 2.2.7 (Monotone convergence theorem for series)

Let $(\Omega, \mathfrak{M}, \mu)$ be a measure space, and functions $f_n : \Omega \rightarrow [0, +\infty]$ measurable for all $n \in \mathbb{N}$.

Then $\sum_{n \in \mathbb{N}} f_n$ converges point-wise to a measurable function f , and we have:

$$\int_\Omega f d\mu = \int_\Omega \sum_{n \in \mathbb{N}} f_n d\mu = \sum_{n \in \mathbb{N}} \int_\Omega f_n d\mu.$$

Proof. We have a series of positive functions whose sequence of partial sum is monotone: set $F_k(t) = \sum_{i=0}^k f_i(t)$.

The sequence of simple functions $\{s_k\}$ is measurable for any $n \in \mathbb{N}$ and is monotone: $s_{k+1} \geq s_k$ in Ω ; than it converges point-wise to a measurable

¹⁹To prove this, define the integral on Ω by applying the indicative function to s .

function f .

Using the finite additivity we have:

$$\int_{\Omega} F_k d\mu = \sum_{i=0}^k \int_{\Omega} f_i d\mu.$$

Apply the standard monotone convergence theorem to gain the thesis:

$$\int_{\Omega} f d\mu = \int_{\Omega} \lim F_n d\mu = \lim \int_{\Omega} F_k d\mu = \sum_{i=0}^{\infty} \int_{\Omega} f_i d\mu.$$

■

Consider the measure space $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu_c)$, where μ_c is the counting measure. On this space, the non-negative measurable functions are actually non-negative sequences of real numbers: $f(n) = a_n$ with $n \in \mathbb{N}$.

In this case the corollary implies that we can exchange series:

$$\sum_{m \in \mathbb{N}} \sum_{n \in \mathbb{N}} a_{mn} = \sum_{n \in \mathbb{N}} \sum_{m \in \mathbb{N}} a_{mn} \quad a_{mn} \geq 0 \quad \forall m, n.$$

Fatou's Lemma The following simple but powerful result, proved for the first time by the french mathematician Pierre Fatou, is more general than the previous, as it does not require the monotonicity of the sequence, but it is restricted to the inferior limit.

Theorem 2.2.8 (Fatou's lemma)

Let $f_n : (\Omega, \mathfrak{M}, \mu) \rightarrow [0, +\infty]$ be measurable for any $n \in \mathbb{N}$. Then:

$$\int_{\Omega} \liminf_{n \rightarrow +\infty} f_n d\mu \leq \liminf_{n \rightarrow +\infty} \int_{\Omega} f_n d\mu.$$

Sometimes the strict inequality may hold.

Example 2.2.9. Take $(\mathbb{R}, \mathcal{L}(\mathbb{R}), \lambda)$, consider:

$$f_n(t) = \begin{cases} \frac{1}{n} & |t| \leq n \\ 0 & |t| > n \end{cases} \quad \forall n \in \mathbb{N}_0.$$

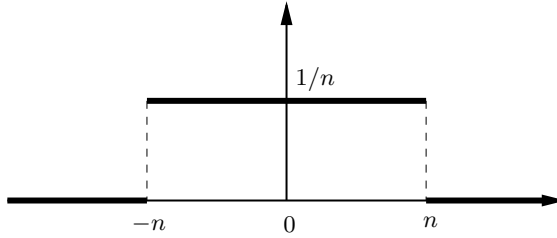


Figure 2.3: Plot of the function of example 2.2.9 on the preceding page.

Then the $\liminf_{n \rightarrow +\infty} f_n(t) = 0$ for all $t \in \mathbb{R}$ but $\int_{\mathbb{R}} f_n d\lambda = 2 > 0 \quad \forall n \in \mathbb{N}_0$.

Proof. Set $g_n(t) = \inf_{j \geq n} f_j(t)$. So g_n is measurable and $g_{n+1} \geq g_n$ in Ω . By definition of inferior limit we have:

$$\sup_{n \in \mathbb{N}} g_n = \liminf_{n \rightarrow \infty} f_n.$$

Using the monotone convergence theorem we get:

$$\lim_{n \rightarrow +\infty} \int_{\Omega} g_n d\mu = \int_{\Omega} \lim_{n \rightarrow +\infty} g_n d\mu = \int_{\Omega} \liminf_{n \rightarrow +\infty} f_n d\mu$$

By integral's monotonicity we obtain:

$$\int_{\Omega} g_n d\mu \leq \int_{\Omega} f_n d\mu;$$

and considering the previous point we have:

$$\lim_{n \rightarrow +\infty} \int_{\Omega} g_n d\mu \leq \liminf_{n \rightarrow +\infty} \int_{\Omega} f_n d\mu$$

which prove the thesis. ■

2.2.3 Derivative of a measure: definitions

Measures defined through measurable functions Here we see how a positive measurable function can provide an alternative measure.

Theorem 2.2.10

Let $\phi : \Omega \rightarrow [0, +\infty]$ be a measurable function. Consider:

$$\nu(E) := \int_E \phi \, d\mu \quad \text{for all } E \in \mathfrak{M}.$$

Then, for any measurable function $f : \Omega \rightarrow [0, +\infty]$, $\nu(E)$ is a measure on \mathfrak{M} , and we have:

$$\int_{\Omega} f \, d\nu = \int_{\Omega} f\phi \, d\mu.$$

Proof. Step 1, $\nu(E)$ is a measure on \mathfrak{M} :

It is easy to see that $\nu(\emptyset) = 0 < +\infty$.

Consider a sequence of mutually disjoint set $\{E_j\}_{j \in \mathbb{N}} \subset \mathfrak{M}$ with $E = \bigcup_{j \in \mathbb{N}} E_j$; using the monotone convergence for series (see corollary 2.2.7 on page 112), we have:

$$\begin{aligned} \nu(E) &= \int_E \phi \, d\mu = \int_{\Omega} \phi \mathbf{1}_E \, d\mu = \int_{\Omega} \sum_{j \in \mathbb{N}} \phi \mathbf{1}_{E_j} \, d\mu \\ &= \sum_{j \in \mathbb{N}} \int_{\Omega} \phi \mathbf{1}_{E_j} \, d\mu = \sum_{j \in \mathbb{N}} \int_{E_j} \phi \, d\mu = \sum_{j \in \mathbb{N}} \nu(E_j); \end{aligned}$$

thus ν is countably additive.

Step 2, exchange of measures:

Now let us focus on the second thesis of the theorem. It is enough to prove it when $f = \mathbf{1}_E$ with $E \in \mathfrak{M}$; then it can be easily extended to simple function and finally to positive functions by its approximation property.

We have:

$$\int_{\Omega} \mathbf{1}_E \, d\nu = \int_E d\nu = \nu(E) = \int_E \phi \, d\mu = \int_{\Omega} \phi \mathbf{1}_E \, d\mu;$$

that conclude the proof. ■

In the first part of this proof we proved the following.

Corollary 2.2.11

Let ϕ be a positive and measurable function, and consider a se-

quence of disjoint sets $\{E_n\}_{n \in \mathbb{N}}$. Then:

$$\int_{\bigcup_{n \in \mathbb{N}} E_n} \phi \, d\mu = \sum_{n \in \mathbb{N}} \int_{E_n} \phi \, d\mu.$$

Radon–Nikodym derivative Suppose to have two measure, and we need to represent one in terms of the other. Possibly through an integral. This is possible by the followings.

Definition 2.2.12

Consider a complete measure space $(\Omega, \mathfrak{M}, \mu)$ and a non-negative measurable function $\phi : \Omega \rightarrow [0, +\infty]$. If ν is such that:

$$\nu(E) = \int_E \phi \, d\mu \quad \text{for all } E.$$

Then ϕ is called **Radon–Nikodym derivative** of ν with respect to μ and it is denoted by $\frac{d\nu}{d\mu}$.

The integrand function is called “derivative” due to the analogy with the first fundamental theorem of calculus (see theorem 2.4.4 on page 145) for which we have $F(x) = \int_a^x F'(t) \, dt$.

For any measurable function $f : \Omega \rightarrow [0, +\infty]$ we can write:

$$\int_E f \, d\nu = \int_E f \phi \, d\mu = \int_E f \frac{d\nu}{d\mu} \, d\mu$$

Since we give the definition of Radon–Nikodym derivative, we ask ourselves if such function exists positive and measurable. We start observing that whenever the measure μ is zero on a set E , $\nu(E)$ must be zero too (otherwise we would have a null denominator and a non-null denominator). Indeed, if $\int_E f \, d\nu = \int_E f \frac{d\nu}{d\mu} \, d\mu$, then we require:

$$\mu(E) = 0 \implies \nu(E) = 0 \quad \text{for all } E \in \mathfrak{M}.$$

This motivates the following definition.

Definition 2.2.13

We say that a measure ν is **absolutely continuous** with respect to

μ , and we write $\nu \ll \mu$, if:

$$\mu(E) = 0 \implies \nu(E) = 0 \quad \text{for any } E \in \mathfrak{M}.$$

In plain language, a measure ν is absolutely continuous with respect to μ if on all the sets for which μ is zero, ν is also zero.

This property holds when a measure can “control” the zero of another measure, it acquires a great meaning when considered with the previous definition: the measure μ controls the measure ν which was defined through μ .

However, the condition $\nu \ll \mu$ is not sufficient alone for the existence of $\frac{d\nu}{d\mu}$; a sufficient condition for its existence is stated by the following theorem.

Theorem 2.2.14 (Radon–Nikodym)

Let $(\Omega, \mathfrak{M}, \mu)$ be a complete measure space, and ν, μ two measures on (Ω, \mathfrak{M}) .

If μ is σ -finite and $\nu \ll \mu$, then the Radon–Nikodym derivative $\frac{d\nu}{d\mu}$ exists.

See definition [2.1.17 on page 90](#) for σ -finite. The theorem will be proved later on section [3.4.4 on page 271](#).

If μ is not σ -finite, then the existence of $\frac{d\nu}{d\mu}$ is not guaranteed. Indeed, consider the measure space $([0, 1], \mathcal{L}([0, 1]), \mu_c)$ where μ is the counting measure and let $\nu = \lambda \ll \mu_c$. You can prove that there is no measure $\phi : [0, 1] \rightarrow [0, +\infty]$ such that $\lambda(E) = \int_E \phi \, d\mu_c \quad \forall E \in \mathcal{L}([0, 1])$.

We will see that this is a powerful generalization of the second fundamental theorem of calculus (see theorem [2.4.11 on page 150](#)).

2.3 Lebesgue Integral

Till now we worked only with positive valued functions, but they are not suitable for many application. The work of Henry Lebesgue (1875 - 1941) was not limited to those function, but it covered the case of a generic real valued function. Our aim now is to define the well-know integral named after him.

2.3.1 Integrating real valued functions

The first step is to establish what functions are integrable. Indeed, we already have a notion of Lebesgue integral, even it is limited to real valued functions we can use it to distinguish between functions. Any absolute value of a real valued function is a positive valued function; so the notion we already have can be used to define a space of integrable functions.

Definition 2.3.1

Let $(\Omega, \mathfrak{M}, \mu)$ be a measurable space.

We define the space of **Lebesgue-integrable functions** as follows:

$$\mathcal{L}_1(\Omega, \mathfrak{M}, \mu) := \{f : \Omega \rightarrow \mathbb{R} : f \text{ measurable and such that } \int_{\Omega} |f| d\mu < +\infty\}.$$

Notice that need the hypothesis of measurability for this definition to have sense; take for instance the space $([0, 1], \mathcal{L}([0, 1]), \lambda)$, and consider the Vitali set $\mathcal{V} \subset [0, 1]$. Define f as follows:

$$f(t) := \begin{cases} 1 & \text{if } t \in \mathcal{V} \\ -1 & \text{if } t \in \mathcal{V}^c \end{cases}.$$

While f is not measurable, $|f|$ is.

Here we can define an integral for real valued functions.

Definition 2.3.2

Let $f \in \mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$. We define the **Lebesgue abstract integral** of f as follows:

$$\int_{\Omega} f d\mu := \int_{\Omega} f^+ d\mu - \int_{\Omega} f^- d\mu$$

where $f^+ = \max\{f, 0\}$ and $f^- = -\min\{f, 0\}$.

In this case f is called **Lebesgue integrable function**.

Notice that $0 \leq f^+ \leq |f|$ and $0 \leq f^- \leq |f|$ are both finite, thus also $\int_{\Omega} f d\mu$ is finite, indeed we have just defined the Lebesgue *abstract* integral.

For a complete discussion we present the following proposition. The notion of vector spaces will be discussed later (see definition 3.1.1 on page 171).

Proposition 2.3.3

The set $\mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$ is a vector space on \mathbb{R} with respect to the canonical operations $f + g$, $c \cdot f$ ($c \in \mathbb{R}$). Indeed, the following inequalities holds:

$$\left| \int_{\Omega} f d\mu \right| \leq \int_{\Omega} |f| d\mu, \quad \int_{\Omega} |f + g| d\mu \leq \int_{\Omega} |f| d\mu + \int_{\Omega} |g| d\mu.$$

This last property, which is, in some sense, a generalization of triangular inequality, allow us to build an algebraic structure on this set.

Integrating on \mathbb{R} 's intervals, the sign of the integral depends on its orientation. Here we define the actual standard.

Definition 2.3.4

Consider $(\mathbb{R}, \mathcal{L}(\mathbb{R}), \lambda)$, for any interval $(a, b) \subset \mathbb{R}$ we set this rule to change the orientation of the interval:

$$\int_a^b f d\lambda := \begin{cases} \int_a^b f d\lambda & \text{if } a < b \\ 0 & \text{if } a = b \\ -\int_b^a f d\lambda & \text{if } a > b. \end{cases}$$

Finally, a remark about series and the counting measure. Consider the measure space $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu_c)$: the series $\{a_n\}_{n \in \mathbb{N}}$ is Lebesgue integrable if and only if $\sum_{n \in \mathbb{N}} |a_n| < +\infty$.

2.3.2 Dominated convergence theorem

Having defined a general yet powerful notion of integral, our focus now is on building many tools that allow us to work on it. Some results are already been proved, the following result extend the monotone convergence (see

2.2.5 on page 109) and the Fatou's lemma (see 2.2.8 on page 113) to real valued functions.

Theorem 2.3.5 (Lebesgue's dominate convergence theorem)

Let $(\Omega, \mathfrak{M}, \mu)$ measure space, $f_n : \Omega \rightarrow \mathbb{R}$ measurable for all $n \in \mathbb{N}$ such that:

- it point-wise converges $f_n(t) \rightarrow f(t)$ as $n \rightarrow +\infty$ for any $t \in \Omega$;
- exists a dominating function, namely $g : \Omega \rightarrow \mathbb{R}$, which is Lebesgue-integrable and such that $|f_n(t)| \leq g(t)$ for all $t \in \Omega$ and for all $n \in \mathbb{N}$.

Then $f_n, f \in \mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$, and we have:

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

Observe that the thesis implies the following:

$$\lim_{n \rightarrow +\infty} \int_{\Omega} f_n d\mu = \int_{\Omega} f d\mu;$$

indeed:

$$\left| \int_{\Omega} f_n d\mu - \int_{\Omega} f d\mu \right| \leq \int_{\Omega} |f_n - f| d\mu \rightarrow 0 \quad \text{as } n \rightarrow +\infty.$$

This shows how powerful and general is this theorem. In general the issue emerging when using it is to find a proper dominating function.

Proof. Using the fact that $|f_n(t)| \leq g(t)$, it is easy to check from the definition that $f_n \in \mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$ for any $n \in \mathbb{N}$.

In addition, we know $f \in \mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$, as $|f(t)| = \lim_{n \rightarrow \infty} |f_n(t)|$ for all $n \in \mathbb{N}$; this because:

$$| |f(t)| - |f_n(t)| | \leq |f(t) - f_n(t)| \rightarrow 0.$$

To prove the limit consider $\phi_n = 2g - |f_n - f|$.

We have that ϕ_n are measurable, non-negative and converging: $\phi_n(t) \rightarrow 2g(t)$ for any $t \in \Omega$.

Owing to Fatou's lemma we have:

$$0 \leq \int_{\Omega} 2g d\mu = \int_{\Omega} \lim_{n \rightarrow +\infty} \phi_n d\mu$$

$$\begin{aligned} &\leq \liminf_{n \rightarrow +\infty} \int_{\Omega} \phi_n \, d\mu \\ &= \int_{\Omega} 2g \, d\mu + \liminf_{n \rightarrow +\infty} \int_{\Omega} (-|f_n - f|) \, d\mu. \end{aligned}$$

So we get:

$$0 \leq \liminf_{n \rightarrow +\infty} \int_{\Omega} -|f_n - f| \, d\mu$$

which implies

$$\limsup_{n \rightarrow +\infty} \int_{\Omega} |f_n - f| \, d\mu \leq 0$$

and from this we can deduce the thesis:

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

■

Case of series of functions The theorem can be formulated also for series of function as follows. To deeply understand this results remember that a series can be seen as the sequence of partial sum.

Theorem 2.3.6 (Dominate convergence theorem for series)

Consider a sequence of functions $\{f_n\}_{n \in \mathbb{N}} \subset \mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$ for any $n \in \mathbb{N}$ such that the series $\sum_{n \in \mathbb{N}} f_n(t)$ converges point-wise $\forall t \in \Omega$, namely:

$$\sum_{n \in \mathbb{N}} \int_{\Omega} |f_n| \, d\mu < +\infty.$$

If exists a function $g \in \mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$ such that:

$$\left| \sum_{j=0}^n f_j(t) \right| \leq g(t) \quad \forall n \in \mathbb{N} \quad \forall t \in \Omega,$$

then $\sum_{n \in \mathbb{N}} f_n$ converges point-wise in Ω to a function $f \in \mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$ and we have:

$$\int_{\Omega} f \, d\mu = \int_{\Omega} \sum_{n \in \mathbb{N}} f_n \, d\mu = \sum_{n \in \mathbb{N}} \int_{\Omega} f_n \, d\mu$$

Proof. Consider $\sum_{n \in \mathbb{N}} |f_n|$ and observe that, by Beppo Levi's theorem:

$$\int_{\Omega} \left(\sum_{n \in \mathbb{N}} |f_n| \right) d\mu = \sum_{n \in \mathbb{N}} \int_{\Omega} |f_n| d\mu < +\infty.$$

Then $\sum_{n \in \mathbb{N}} |f_n|$ converges a.e. in Ω to a $\tilde{f} \in \mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$. Thus we also have that $\sum_{n \in \mathbb{N}} f_n$ (absolutely) converges to some f a.e. in Ω , and moreover:

$$\left| \sum_{n=0}^N f_n(t) \right| \leq \sum_{n=0}^N |f_n(t)| \leq \sum_{n=0}^{+\infty} |f_n(t)| = \tilde{f}(t)$$

For almost any $t \in \Omega$ and for all N .

Thus we can apply dominated convergence to

$$F_N(t) := \sum_{n=0}^N f_n(t) \rightarrow f(t).$$

We have that $f \in \mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$, and:

$$\int_{\Omega} f d\mu = \lim_{N \rightarrow \infty} \int_{\Omega} F_N d\mu = \lim_{N \rightarrow \infty} \sum_{n=0}^N \int_{\Omega} f_n d\mu = \sum_{n \in \mathbb{N}} \int_{\Omega} f_n d\mu.$$

■

Notice that if $\int_{\Omega} \sum_{n \in \mathbb{N}} |f_n| d\mu < +\infty$ then $\sum_{n \in \mathbb{N}} |f_n|$ is finite a.e. in Ω and $\sum f_n$ converges a.e. in Ω .

2.3.3 The almost everywhere concept

By construction, Lebesgue integral is not conditioned by single points. In particular a countable set of points does not affect the result. At this point we have to ask ourself if this concept holds also for other properties. Suppose for example that a function satisfy a property except for a countable set of points, can we still apply the theorems that requires such property? The goal of this section is to extend all the previous results of abstract integration to functions defined up to a zero-measure set.

Definition 2.3.7

A certain property $P(t)$, with $t \in \Omega$ holds **almost everywhere (a.e.)** in Ω if:

$$\mu(\{ t \in \Omega : \neg P(t) \}) = 0.$$

Example 2.3.8. The properties $P(t) = \{\sin(t) \neq 0\}$ and $Q(t) = \{\mathbf{1}_{\mathbb{Q}} = 0\}$ ²⁰ hold a.e. in \mathbb{R} .

Extension of the measurability The first property that we extend to this context from the general one is measurability:

Definition 2.3.9

Let $(\Omega, \mathfrak{M}, \mu)$ be a complete measure space and (X, τ) a topological space.

A function $f : (\Omega, \mathfrak{M}, \mu) \rightarrow (X, \tau)$ is **measurable almost everywhere** in Ω if both

there exists $\Omega_0 \subset \Omega$ such that $\mu(\Omega_0^C) = 0$

and

$f^{-1}(A) \cap \Omega_0 \in \mathfrak{M}$ for all open set $A \subset X$.

This says that we require that the preimage of the function, intersected with a set Ω_0 whose complement has zero measure, is still measurable: in this case the function is said to be measurable almost everywhere. In this way the function can be defined (or not) in Ω_0^C in any manner without affecting its measurability.

It is easy to see that all the results proved so far, in particular Beppo Levi's, Fatou's and Lebesgue's, can be reformulated for functions defined almost everywhere and they still hold.

Essentially boundedness In this context it is useful to give another definition of boundedness.

Definition 2.3.10

A function $f : (\Omega, \mathfrak{M}, \mu) \rightarrow \mathbb{R}$ is **essentially bounded** if exists $M \geq 0$

²⁰The function $\mathbf{1}_{\mathbb{Q}}$, which is the indicator function on the rational number set, is called Dirichlet function.

such that:

$$\mu(\{t \in \Omega : |f(t)| > M\}) = 0.$$

The concept of almost everywhere allow us to redefine the supremum and the infimum of functions as well.

Definition 2.3.11

Let $f : \Omega \rightarrow \mathbb{R}$ be a measurable function.

Then we define its **essential supremum** in Ω as follows:

$$\operatorname{ess\,sup}_{t \in \Omega} f := \inf\{M \geq 0 : \mu(\{t \in \Omega : |f(t)| > M\}) = 0\},$$

and the **essential infimum** in Ω as follows:

$$\operatorname{ess\,inf}_{t \in \Omega} f := \sup\{M \geq 0 : \mu(\{t \in \Omega : |f(t)| < M\}) = 0\}.$$

Example 2.3.12. Consider for example the Dirichlet function $f(t) = \mathbf{1}_{\mathbb{Q}}(t)$ with $t \in \mathbb{R}$. It is bounded and

$$\sup_{\mathbb{R}} f = \max_{\mathbb{R}} f = 1;$$

it is essentially bounded as well, however:

$$\min\{M \geq 0; \lambda(t \in \mathbb{R} : \mathbf{1}_{\mathbb{Q}}(t) > 1)\} = 0,$$

which means

$$\operatorname{ess\,sup}_{\mathbb{R}} f = \operatorname{ess\,inf}_{\mathbb{R}} f = 0 \neq \sup_{\mathbb{R}} f = 1.$$

Example 2.3.13. Consider now the following function:

$$f(t) = \begin{cases} te^{t^2} & t \in \mathbb{Q} \\ \sin(t) & t \in \mathbb{Q}^c; \end{cases}$$

we have:

$$\operatorname{ess\,sup}_{\mathbb{R}} f = 1, \quad \operatorname{ess\,inf}_{\mathbb{R}} f = -1, \quad \sup_{\mathbb{R}} f = +\infty, \quad \inf_{\mathbb{R}} f = -\infty.$$

Extension of the continuity Now we can discuss how also continuity can be defined almost everywhere:

Definition 2.3.14

Let $(\Omega, \mathcal{L}(\Omega), \lambda)$ be a complete measure space where $\Omega \subset \mathbb{R}^N$ is an open set, and (X, τ) be a topological space.

A function $f : (\Omega, \mathfrak{M}, \mu) \rightarrow (X, \tau)$ is **continuous almost everywhere** in Ω if both

there exists $\Omega_0 \subset \Omega$ such that $\mu(\Omega_0^C) = 0$

and

$f^{-1}(A) \cap \Omega_0$ is open for all open set $A \subset X$.

Equivalently a function is continuous a.e. if the measure of its discontinuity points is zero.

Observe that if a function is continuous a.e. then it's Lebesgue measurable a.e..

Proposition 2.3.15

Consider two continuous functions $f, g : \mathbb{R} \rightarrow \mathbb{R}$.

If they coincide a.e. namely

$$f = g \text{ a.e. with respect to } \lambda,$$

then they coincide at any point, namely:

$$f(t) = g(t) \text{ for any } t \in \mathbb{R}.$$

Proof. By contradiction, if there exists $x_0 \in \mathbb{R}$ such that $f(x_0) \neq g(x_0)$ then by continuity there exists also a δ such that $f(x) \neq g(x)$ for all $x \in (x_0 - \delta, x_0 + \delta)$.

But then $\lambda((x_0 - \delta, x_0 + \delta)) = 2\delta > 0$: so there exists an interval of positive measure where the two functions are different, so we have a contradiction. ■

Example 2.3.16. Consider again the Dirichlet function $f(t) = \mathbb{1}_{\mathbb{Q}}(t)$; it is nowhere continuous (it is not continuous a.e. in \mathbb{R}).

However $f = 0$ a.e., so f is equal a.e. to a continuous function. This while the Heaviside function $H = \mathbb{1}_{[0, +\infty)}$ is continuous a.e. but it is not equal a.e. to any continuous function. You can prove this. The same consideration

can be done for the following:

$$f(x) := \begin{cases} \frac{x}{|x|} & x \neq 0 \\ 0 & x = 0 \end{cases}.$$

Example 2.3.17. Consider also the function we seen right before:

$$f(t) = \begin{cases} te^{t^2} & t \in \mathbb{Q} \\ \sin(t) & t \in \mathbb{Q}^c; \end{cases}$$

we see that it is not bounded but it is essentially bounded. It is continuous in $t = 0$; think whether it is continuous elsewhere!

Example 2.3.18. At last consider the function $f : \mathbb{R} \rightarrow \mathbb{R}^*$ defined in this way:

$$f(t) := \begin{cases} \arctan(t) & t \in \mathbb{Q}^c \\ +\infty & t \in \mathbb{Q} \cup [0, +\infty) \\ -\infty & t \in \mathbb{Q} \cup (-\infty, 0]; \end{cases}$$

this function is equal a.e. to a continuous function but it's nowhere continuous, it isn't bounded while is essentially bounded.

Same result with

$$f(t) := \begin{cases} \sin(t) & t \in \mathbb{R} \setminus \mathbb{Q} \\ -\infty & t \in \mathbb{Q}_- \\ +\infty & t \in \mathbb{Q}_+. \end{cases}$$

Proposition 2.3.19

Consider a function $f : (\Omega, \mathcal{L}(\Omega), \lambda) \rightarrow \mathbb{R}$, with $\Omega \subset \mathbb{R}^N$.

As f is Lebesgue measurable, there exists a function $g : \Omega \rightarrow \mathbb{R}$ which is Borel measurable and such that $f = g$ a.e. in Ω .

Extension of convergence theorems The three convergence theorems, which are the Beppo Levi's or monotone convergence theorem (see theorem 2.2.5 on page 109), the Fatou's lemma (2.2.8 on page 113) and the dominated convergence theorem (2.3.2 on page 119), can be trivially reformulated for measurable a.e. functions.

In addition we can prove another theorem for convergence of series of function:

Theorem 2.3.20

Consider a sequence of functions $\{f_n\}_{n \in \mathbb{N}} \subset \mathcal{L}^1(\Omega, \mathfrak{M}, \mu) \forall n \in \mathbb{N}$ such that the series $\sum_{n \in \mathbb{N}} f_n(t)$ converges point-wise $\forall t \in \Omega$, namely:

$$\sum_{n \in \mathbb{N}} \int_{\Omega} |f_n| d\mu < +\infty.$$

Then $\sum_{n \in \mathbb{N}} f_n$ converges a.e in Ω to a function $f \in \mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$ and we have:

$$\int_{\Omega} f d\mu = \int_{\Omega} \sum_{n \in \mathbb{N}} f_n d\mu = \sum_{n \in \mathbb{N}} \int_{\Omega} f_n d\mu.$$

Proof. The series $\sum_{n \in \mathbb{N}} |f_n|$ converges to a non-negative function g and, by Beppo Levi's theorem:

$$\int_{\Omega} \left(\sum_{n \in \mathbb{N}} |f_n| \right) d\mu = \sum_{n \in \mathbb{N}} \int_{\Omega} |f_n| d\mu < +\infty.$$

Then we say that $\sum_{n \in \mathbb{N}} |f_n|$ converges a.e. in Ω to $g \in \mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$; we have that $\sum_{n \in \mathbb{N}} f_n$ absolutely converges to some f a.e. in Ω , and moreover:

$$\left| \sum_{n=0}^N f_n(t) \right| \leq \sum_{n=0}^N |f_n(t)| \leq \sum_{n=0}^{+\infty} |f_n(t)| = g(t)$$

for almost any $t \in \Omega$ and for all N .

Thus we can apply dominated convergence to $\sum_{n=0}^N f_n(t) \rightarrow f(t)$.

We have that $f \in \mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$, and:

$$\int_{\Omega} f d\mu = \lim_{N \rightarrow \infty} \int_{\Omega} \sum_{n=0}^N f_n(t) d\mu = \lim_{N \rightarrow \infty} \sum_{n=0}^N \int_{\Omega} f_n d\mu = \sum_{n \in \mathbb{N}} \int_{\Omega} f_n d\mu.$$

■

Notice that if $\int_{\Omega} \sum_{n \in \mathbb{N}} |f_n| d\mu < +\infty$ then $\sum_{n \in \mathbb{N}} |f_n|$ is finite a.e. in Ω and $\sum f_n$ converges a.e. in Ω .

Further results Here we introduce some other results that can be proven at this state of the theory. First we talk about a useful inequality widely used in probability.

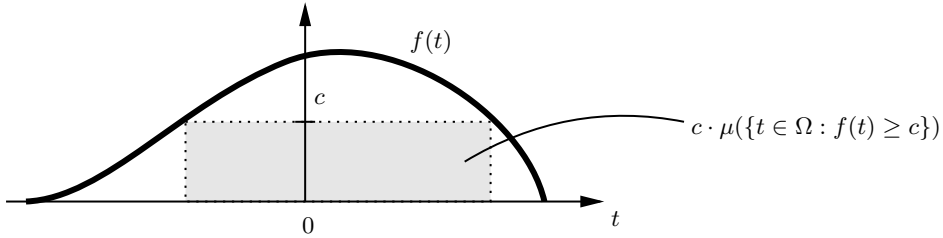
Proposition 2.3.21 (Chebyshev's inequality)

Let $f \in \mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$ such that $f \geq 0$ a.e. in Ω , $c > 0$.

Then the following inequality holds:

$$\mu(\{t \in \Omega : f(t) \geq c\}) \leq \frac{1}{c} \int_{\Omega} f \, d\mu.$$

This inequality is trivial if we think to its geometrical meaning.



Proof. Consider the following chain of inequalities:

$$\int_{\Omega} f \, d\mu \geq \int_{\{t \in \Omega : f(t) \geq c\}} f \, d\mu \geq c \cdot \mu(\{t \in \Omega : f(t) \geq c\}).$$

■

Here another useful tool:

Proposition 2.3.22

Let $(\Omega, \mathfrak{M}, \mu)$ be a σ -finite measure space.

Consider two measurable functions $f, g : \Omega \rightarrow [0, +\infty]$ such that:

$$\int_E f \, d\mu \leq \int_E g \, d\mu \quad \text{for all } E \in \mathfrak{M},$$

then $f \leq g$ almost everywhere in Ω .

Lastly we present a simple yet useful result:

Theorem 2.3.23 (Vanishing lemma)

Let $f \in \mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$ such that $f \geq 0$ a.e. in Ω .

If $\int_{\Omega} f \, d\mu = 0$ then $f = 0$ a.e. in Ω .

Proof. Take the following set for each $n \in \mathbb{N}_0$:

$$E_n = \left\{ t \in \Omega : f(t) \geq \frac{1}{n} \right\} \in \mathfrak{M}.$$

Then:

$$0 = \int_{\Omega} f \, d\mu \geq \int_{E_n} f \, d\mu \geq \frac{1}{n} \mu(E_n) \quad \forall n \in \mathbb{N}_0.$$

Thus $\mu(E_n) = 0$ for all $n \in \mathbb{N}_0$.

Now suppose there exists $t \in \Omega$ such that $f(t) > 0$.

Then it exists $n_0 \in \mathbb{N}$ such that $f(t) \geq \frac{1}{n_0}$, and so

$$t \in E_{n_0} \subset E \text{ where } E = \bigcup_{n \in \mathbb{N}_0} E_n.$$

Thus $\{t \in \Omega : f(t) > 0\} \subseteq E$, so:

$$\mu(E) \leq \sum_{n \in \mathbb{N}} \mu(E_n) = 0.$$

■

A brief final remark: consider $\mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$ and define $d(f, g) := \int_{\Omega} |f, g| \, d\mu$. Notice that such function is not a metric: indeed, by the vanishing lemma, $d(f, g) = 0 \implies f = g$ not everywhere, but *almost* everywhere $\in \Omega$.

To solve the problem, we consider the following equivalence relation:

$$f \sim g \iff f = g \text{ a.e. in } \Omega$$

Take the quotient set $X_1 := \frac{\mathcal{L}^1(\Omega, \mathfrak{M}, \mu)}{\sim}$. Then d is a metric in X_1 .

2.3.4 Derivative of a measure: properties

Now we are ready to proceed in theory development about derivatives. With the definition of almost everywhere we can talk about uniqueness and some properties.

Proposition 2.3.24

If the hypothesis of the Radon–Nikodym theorem (2.2.14 on page 117) hold, then the Radon–Nikodym derivative is unique almost everywhere.

The reader can easily prove this result using proposition 2.3.22 on page 128. Do it now before see the following proof!

Proof. Consider two Radon–Nikodym derivatives for $\frac{d\nu}{d\mu}$: two positive and measurable function ϕ_1 and ϕ_2 .

Then we have:

$$\int_E \phi_1 d\mu = \int_E \phi_2 d\mu \quad \text{for all } E \in \mathfrak{M}.$$

From the previously referenced proposition we have both $\phi_1 \geq \phi_2$ and $\phi_2 \geq \phi_1$ a.e. in Ω , hence the thesis. ■

Moreover, it holds the following.

Proposition 2.3.25

If the Radon–Nikodym derivative ϕ exists and $\mu(\Omega) < +\infty$, then $\phi \in L^1(\Omega, \mathfrak{M}, \mu)$.

Basic properties Here we present some properties of the Radon–Nikodym derivative.

Proposition 2.3.26 (Change of measure)

For any measurable positive function $f : \Omega \rightarrow [0, \infty]$ we have:

$$\int_{\Omega} f d\nu = \int_{\Omega} f \frac{d\nu}{d\mu} d\mu.$$

Proposition 2.3.27 (Linearity of the derivative)

For all $c_1, c_2 \geq 0$ we have:

$$\frac{d(c_1\nu_1 + c_2\nu_2)}{d\nu} = c_1 \frac{d\nu_1}{d\mu} + c_2 \frac{d\nu_2}{d\mu}.$$

Proposition 2.3.28 (Chain rule)

Consider three measures λ, ν and μ such that $\lambda \ll \nu \ll \mu$.

Then we have:

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

Proof. For every $E \in \mathfrak{M}$ observe that:

$$\lambda(E) = \int_E \frac{d\lambda}{d\nu} d\nu = \int_E \frac{d\lambda}{d\nu} \frac{d\nu}{d\mu} d\mu;$$

moreover, as $\lambda \ll \mu$, we have that:

$$\lambda(E) = \int_E \frac{d\lambda}{d\mu} d\mu,$$

and then, on account of 2.3.22 on page 128:

$$\frac{d\lambda}{d\mu} = \frac{d\lambda}{d\nu} \frac{d\nu}{d\mu} \quad \text{a.e..}$$

■

Proposition 2.3.29 (Inverse derivative)

If $\nu \ll \mu$ and $\mu \ll \nu$ then we have:

$$\frac{d\nu}{d\mu} = \left(\frac{d\mu}{d\nu} \right)^{-1}.$$

Proof. For every $E \in \mathfrak{M}$ we have that:

$$\mu(E) = \int_E d\mu = \int_E \frac{d\mu}{d\nu} d\nu = \int_E \frac{d\mu}{d\nu} \frac{d\nu}{d\mu} d\mu$$

thus, always on account of 2.3.22 on page 128:

$$1 = \frac{d\mu}{d\nu} \frac{d\nu}{d\mu} \quad \text{a.e..}$$

■

2.3.5 Metric spaces and convergences

In this part we will see how the tools we developed can be deployed in practice. First we extend the notion of Lebesgue integrability by using the concept of almost everywhere. Then we see how to handle different kinds of limits of series of functions.

The space L^1 As we discovered that many properties can be useful even when they hold almost everywhere, now we can find a space wider than \mathcal{L}_1 but whose functions have the same properties.

Definition 2.3.30

Consider the space $\mathcal{L}_1(\Omega, \mathfrak{M}, \mu)$ and define on it the following equivalence relation:

$$f \sim g \quad \text{if } f = g \text{ a.e. in } \Omega;$$

we define the **space** L_1 by the quotient set:

$$L_1(\Omega, \mathfrak{M}, \mu) = \frac{\mathcal{L}_1(\Omega, \mathfrak{M}, \mu)}{\sim}.$$

In plain language, this set contains all the functions which are equal almost everywhere to a function in \mathcal{L}_1 . For the definition of quotient set see [1.1.20 on page 21](#).

Such set is a vector space on \mathbb{R} with respect to the canonical operations sum $[f] + [g] = [f + g]$ and homogeneity $[cf] = c[f]$, $c \in \mathbb{R}$.

It is also a metric space with respect to the following distance:

$$d_1 := \int_{\Omega} |f - g| d\mu \quad \text{for all } [f], [g] \in L_1(\Omega, \mathfrak{M}, \mu),$$

and is proved that L^1 with d_1 is a complete metric space.

However, \mathcal{L}_1 isn't a metric space with respect to the distance d_1 as $d_1(f, g) = 0$ implies only that $f = g$ a.e. in Ω .

As L^1 is a metric space we can develop a proper notion of convergence on it.

Convergence of series of functions There are four notions of convergence in a measure space: we will introduce the definition and we analyze the relations between them.

Definition 2.3.31

Let $(\Omega, \mathfrak{M}, \mu)$ be a complete measure space, and $\{f_n\}_{n \in \mathbb{N}}$ be a sequence of measurable functions such that $f_n : \Omega \rightarrow \mathbb{R}$:

- we say that $f_n(t)$ **converges point-wise almost everywhere** to $f(t)$ as $n \rightarrow \infty$, and we write “ $f(t)_n \xrightarrow{\text{a.e.}} f(t)$ ”, if we have:

$$\mu(\{t \in \Omega : \lim_{n \rightarrow +\infty} f_n(t) \neq f(t)\}) = 0;$$

- we say that $f_n(t)$ **converges uniformly almost everywhere** to $f(t)$ as $n \rightarrow \infty$, and we write “ $f(t)_n \xrightarrow{\text{u.a.e.}} f(t)$ ”, if we have:

$$\lim_{n \rightarrow +\infty} \text{ess sup}_{\Omega} |f_n - f| = 0;$$

- we say that $f_n(t)$ **converges in mean** to $f(t)$ as $n \rightarrow \infty$, and we write “ $f(t)_n \xrightarrow{L^1} f(t)$ ”, if $\{f_n\}_{n \in \mathbb{N}} \subset L^1(\Omega, \mathfrak{M}, \mu)$ and:

$$\lim_{n \rightarrow +\infty} \int_{\Omega} |f_n - f| d\mu = 0;$$

- we say that $f_n(t)$ **converges in measure** to $f(t)$ as $n \rightarrow \infty$, and we write “ $f(t)_n \xrightarrow{\mu} f(t)$ ”, if we have:

$$\forall \varepsilon > 0 \quad \lim_{n \rightarrow +\infty} \mu(\{t \in \Omega : |f_n(t) - f(t)| > \varepsilon\}) = 0.$$

To memorize those definition notice that there are some elements that are common; the following observations can help to understand and memorize:

- the $\lim_{n \rightarrow \infty}$ occurs in all definitions;
- all definitions states that some quantity is equal to zero;
- all definitions involves the set Ω ;
- all definitions involves the measure μ , the first and the second through a measure of a set, the third through the integral, and the second through the ess sup, which is defined in accordance to a measure.
- all definitions involves $f_n - f$, which, in different ways, tends to zero.

Also the first contains this subtraction implicitly;

- notice that the third definition requires also that the functions are integrable a.e.;
- the first definition consist of a measure of the set in which a limit does not occur, viceversa the third is the limit of the measure of a set.

It's clear that this definitions are not fully equivalent each other, but they are somehow correlated. Here we try to investigate their relations by declaring and proving some propositions.

Some observation before starting the comparisons. Notice that the main vantage of uniform convergence over point-wise is that with uniform convergence we have not any dependence form the points of the domain t .

Notice that the integral in the mean convergence contains the distance d_1 (the module): that is a metric convergence.

The convergence in measure is widely used in probability.

Uniform a.e. convergence implies point-wise a.e. convergence This result is stated in the following proposition:

Proposition 2.3.32

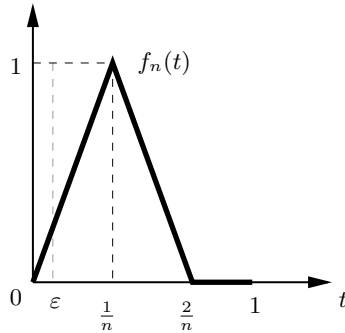
If a series of function converges uniformly almost everywhere, then such series also converges point-wise almost everywhere.

The proof is very easy. Do it!

Point-wise a.e. convergence does NOT imply uniform a.e. convergence

Consider the following series of functions:

$$f_n(t) = \begin{cases} nt & \text{if } 0 \leq t \leq \frac{1}{n} \\ 2 - nt & \text{if } \frac{1}{n} \leq t \leq \frac{2}{n} \\ 0 & \text{if } \frac{2}{n} \leq t \leq 1 \end{cases}$$



In this case $f_n(t) \rightarrow 0$ for all $t \in [0, 1]$, and thus $f_n \xrightarrow{\text{a.e.}} 0$. However, $\text{ess sup } |f_n| = 1$ and so it does not converge to 0, so $f_n \not\xrightarrow{\text{u.a.e.}} f = 0$.

Notice that to obtain also uniform convergence we should consider the interval $[\varepsilon, 1]$, with $\varepsilon > 0$.

In general this little trick holds, indeed we have:

Proposition 2.3.33 (Severini–Egorov)

If $\mu(\Omega) < +\infty$ and $f_n \xrightarrow{\text{a.e.}} f$, then^a

for all $\varepsilon > 0$ there exists $E \subset \mathfrak{M}$

such that

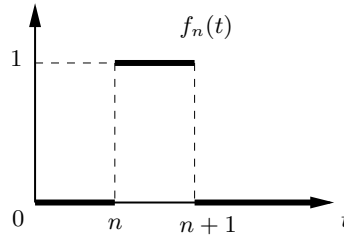
$$\mu(E^C) < \varepsilon \text{ and } f_n \xrightarrow{\text{u.a.e.}} f \text{ in } E.$$

^aFor further discussion, see: R. L. Wheeden, A. Zygmund, Measure and Integral: An Introduction to Real Analysis (2015), page 245, theorem 10.14.

This theorem provides us a condition for which point-wise a.e. convergence implies uniform a.e. convergence, restoring the inverse implication.

If $\mu(\Omega) = +\infty$, the theorem might not hold, as shown in the following counterexample. On $(\mathbb{R}, \mathcal{L}(\mathbb{R}), \lambda)$ take:

$$f_n(t) = \begin{cases} 1 & \text{if } t \in [n, n + 1] \\ 0 & \text{elsewhere.} \end{cases}$$



See that $f_n \xrightarrow{\text{a.e.}} 0$. However, there exist $\varepsilon > 0$ and $E \subset \mathbb{R}$ such that $\lambda(E) < \varepsilon$ and:

$$\sup_{\mathbb{R} \setminus E} |f_n| = 1 \not\rightarrow 0 \text{ as } n \rightarrow +\infty.$$

Point-wise a.e. convergence does NOT imply convergence in mean

Consider the following counterexample.

Take $\Omega = [0, 1]$ and:

$$f_n(t) = \begin{cases} n^2 & \text{if } 0 \leq t < \frac{1}{n} \\ 0 & \text{if } \frac{1}{n} \leq t \leq 1. \end{cases}$$

Then f_n converges a.e. to 0, but $\int_0^1 f_n \, d\lambda = n$ and its limit is $+\infty$, thus f_n does not converge in mean to 0.

Also this implication can be restored if we add more strict hypotheses. If a series of functions respects the requirements of the dominated convergence theorem (see [2.3.2 on page 119](#)) then, if the series converges point-wisely a.e. we have that it converges in mean as well.

Uniform a.e. convergence does NOT imply convergence in mean This can be deduced from the previous results, otherwise it would be a contradiction.

For a counterexample consider the series function $f_n = \frac{1}{n} \mathbb{1}_{(0,n)}(t)$, with $n \in \mathbb{N}_0$.

Then f_n converges a.e. to 0, but $\int_0^{+\infty} f_n \, d\lambda = 1$ for all $n \in \mathbb{N}_0$. Since uniform a.e. convergence implies point-wise a.e. convergence and, with the hypothesis of dominated convergence, it implies the convergence in mean, this convergence can be restored if the hypothesis of that theorem are fulfilled.

Convergence in mean does NOT imply point-wise a.e. convergence In this case, consider the counterexample of the *typewriter sequence*.

Take $\Omega = [0, 1]$ and $\mu = \lambda$, the Lebesgue's measure. For $k \in \mathbb{N}$ consider $m \in \{0, 1, 2, \dots, 2^k - 1\}$, calculate $n = 2^k + m$ and define:

$$E_n = \left[\frac{m}{2^k}, \frac{m+1}{2^k} \right].$$

We have, for each k , a partition of $[0, 1]$, where E_n fill the interval from left to right as a typewriter moves each line down and from left to right in each line. For instance we have the following sets:

$k = 0$	$m \in \{0\}$	$n = 1$	$E_1 = [0, 1]$
		$n = 2$	$E_2 = \left[0, \frac{1}{2}\right]$
		$n = 3$	$E_3 = \left[\frac{1}{2}, 1\right]$
		$n = 4$	$E_4 = \left[0, \frac{1}{4}\right]$
		$n = 5$	$E_5 = \left[\frac{1}{4}, \frac{1}{2}\right]$
		$n = 6$	$E_6 = \left[\frac{1}{2}, \frac{3}{4}\right]$
		$n = 7$	$E_7 = \left[\frac{3}{4}, 1\right]$
$k = 3$	$m \in \{0, 1, 2, 3, 4, 5, 6, 7\}$	$n = 8$	$E_8 = \left[0, \frac{1}{8}\right]$

...

Take $f_n = \mathbb{1}_{E_n}$, as in figure 2.4 on the following page.

Then

$$\int_0^1 \mathbb{1}_{E_n} d\lambda = \frac{1}{2^k} = \frac{1}{n-m} \rightarrow 0 \text{ as } n \rightarrow +\infty$$

and thus f_n converges in mean to 0.

However, $\lim_{n \rightarrow \infty} \mathbb{1}_{E_n}(t)$ does not exist for any $t \in [0, 1]$, and thus f_n does not converge point-wise to any function.

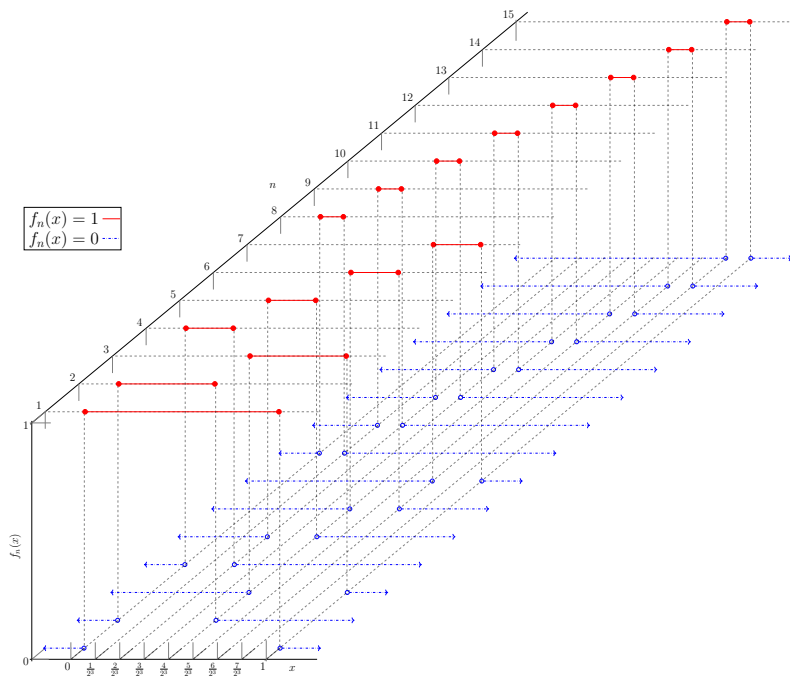


Figure 2.4: The typewriter sequence.

We have the following proposition, as we can restore the implication for only a subsequence.

Proposition 2.3.34

If $f_n \xrightarrow{L^1} f$, then there exists a sub-sequence $\{f_{n_h}\}$ such that $f_{n_h} \xrightarrow{\text{a.e.}} f$.

This is a trivial corollary of the proposition 2.3.37 on page 141 that we will see.

Convergence in mean does NOT imply uniform a.e. convergence Otherwise we would have a contradiction. To restore the convergence, some additional hypothesis are required.

Convergence in measure does NOT imply point-wise a.e. convergence The counterexample of the *typewriter sequence* fits again. We know that $\mathbb{1}_{E_n}$ does not converge point-wisely to any f , but it is easy to check that $\mathbb{1}_{E_n} \xrightarrow{\mu} 0$. Indeed:

$$\lambda(\{t \in [0, 1] : \mathbb{1}_{E_n}(t) > \delta\}) < \frac{1}{2^k} \quad \forall \delta > 0.$$

Nonetheless, the following result holds.

Proposition 2.3.35

If $f_n \xrightarrow{\mu} f$, then there exists a sub-sequence $\{f_{n_h}\}$ such that $f_{n_h} \xrightarrow{\text{a.e.}} f$.

Proof. Let $\delta_n > 0$ such that $\delta_n \rightarrow 0$, and $\varepsilon_n > 0$ be such that $\sum_{n \in \mathbb{N}} \varepsilon_n < +\infty$.

Consider a sub-sequence $\{n_h\}$, with $n_h > n_{h-1}$, such that:

$$E_h = \{t \in \Omega : |f_{n_h}(t) - f(t)| > \delta_{n_h}\} \text{ and } \mu(E_h) < \varepsilon_{n_h}.$$

Set $E = \bigcap_{h \in \mathbb{N}} E_h$.

Since $\mu(E_h) \leq \sum_{h \in \mathbb{N}} \varepsilon_{n_h} < +\infty$ and $\{E_h\}$ is monotone decreasing, we have that $\mu(E_h) \downarrow \mu(E)$.

Moreover, $\mu(E_h) \leq \sum_{j=n}^{\infty} \varepsilon_{n_j} \rightarrow 0$ as $h \rightarrow \infty$, and thus $\mu(E) = 0$.

Take now $t \in E^C$. Then $\exists k \in \mathbb{N}$ such that $t \notin E_k$, that is, $|f_{n_k}(t) - f(t)| \leq \delta_{n_k}$.

Thus $f_{n_h}(t) \rightarrow f(t) \quad \forall h \geq k \quad \forall t \in E^C$. ■

Point-wise a.e. convergence does NOT imply convergence in measure
 Nevertheless, we have the following result.

Proposition 2.3.36

If a series of function converges point-wisely almost everywhere, then such series also converges in measure if $\mu(\Omega) < +\infty$.

Proof. Let $E \in \mathfrak{M}$ such that $\mu(E) = 0$ and:

$$f_n(t) \not\rightarrow f(t) \text{ if } t \in E, \quad f_n(t) \rightarrow f(t) \text{ if } t \in E^C.$$

Fix $\delta > 0$ and consider:

$$A_k(\delta) = \{t \in \mathbb{R} : |f_k(t) - f(t)| > \delta\}, \quad B_n(\delta) = \bigcup_{k=n}^{\infty} A_k(\delta), \quad C(\delta) = \bigcap_{n \in \mathbb{N}} B_n(\delta).$$

See that B_n is decreasing, namely $B_1 \supset B_2 \supset \dots \supset B_n$, and $\mu(B_1) < +\infty$. Then $\mu(B_n(\delta)) \rightarrow \mu(C(\delta))$.

Notice that $C(\delta) \subset E$. Indeed:

$$\begin{aligned} t \in C(\delta) &\implies f \in B_n(\delta) \quad \forall n \in \mathbb{N} \\ &\implies \forall n \exists k \geq n : t \in A_k \\ &\implies \forall n \exists k \geq n : |f_k(t) - f(t)| > \delta \\ &\implies f_n(t) \not\rightarrow f(t) \implies t \in E \end{aligned}$$

Then $\mu(C(\delta)) = 0$, and thus $\mu(B_n(\delta)) \rightarrow 0$.

However $A_n(\delta) \subset B_n(\delta)$, thus also $\mu(A_n(\delta)) \rightarrow 0$ as $n \rightarrow \infty$, that is $f_n \xrightarrow{\mu} f$. ■

If the hypothesis is not satisfied the thesis is not guaranteed, indeed consider for example that $\mu(\Omega) = +\infty$ and set:

$$f_n(t) = \begin{cases} 1 & \text{if } t \geq n \\ 0 & \text{if } t < n \end{cases} \quad \forall n \in \mathbb{N}.$$

Then $f_n \xrightarrow{\text{a.e.}} 0$, but $\lambda(\{t \in \mathbb{R} : f_n(t) > \frac{1}{2}\}) = +\infty \quad \forall n \in \mathbb{N}$, and thus $f_n \not\xrightarrow{\lambda} 0$. Notice that if $\{f_n\}$ would converge in measure then it would necessarily converge to 0.

Convergence in mean implies convergence in measure For this last case, the following result holds.

Proposition 2.3.37

If a series of function converges in mean, then such series also converges in measure.

Proof. Fix $\delta > 0$, and set $E(\delta) := \{t \in \Omega : |f_n(t) - f(t)| > \delta\}$. We want to prove that $\mu(E(\delta)) \rightarrow 0$ as $n \rightarrow +\infty$. We have, by monotonicity:

$$\int_{\Omega} |f_n - f| d\mu \geq \int_{E(\delta)} |f_n - f| d\mu \geq \delta \mu(E).$$

Since $\int_{\Omega} |f_n - f| d\mu \rightarrow 0$ as $n \rightarrow +\infty$, then also $\delta \mu(E) \rightarrow 0$. ■

But the inverse is not true, as shown here.

Convergence in measure does NOT imply convergence in mean indeed, consider the measure space $(\mathbb{R}, \mathcal{L}(\mathbb{R}), \lambda)$ and take $f_n(t) := n \mathbb{1}_{[0, \frac{1}{n}]}(t)$.

Then $\lambda(\{t \in \mathbb{R} : f_n(t) > \delta\}) \leq \frac{1}{n} \rightarrow 0$ as $n \rightarrow \infty$, and thus $f_n \xrightarrow{\mu} 0$.

However $\int_0^1 f_n(t) d\lambda = 1$ as $n \rightarrow \infty$, and thus $f_n \not\xrightarrow{\mu} 0$.

Summary of the relations between kinds of convergence we stated the following relations between convergences:

1. **uniform a.e. convergence implies point-wise a.e. convergence**, see theorem 2.3.32 on page 134, whose proof is left to the reader;
2. **point-wise a.e. convergence does NOT imply uniform a.e. convergence**, as shown with a counterexample, but the implication can be restored as specified in Severini–Egorov theorem (2.3.33 on page 135);
3. **point-wise a.e. convergence does NOT imply convergence in mean**, as shown with a counterexample, but the implication can be restored if the hypothesis of dominated convergence theorem (2.3.2 on page 119) are fulfilled;
4. **uniform a.e. convergence does NOT imply convergence in mean**, as shown with a counterexample, but the implication can be restored if the hypothesis of dominated convergence theorem (2.3.2 on page 119) are fulfilled;

5. **convergence in mean does NOT imply point-wise a.e. convergence**, as shown with the typewriter counterexample, but the implication can be restored for a sub-sequence, see proposition [2.3.35 on page 139](#);
6. **convergence in mean does NOT imply uniform a.e. convergence** otherwise we would have a contradiction, but the implication can be restored with the tools introduced for the previous cases;
7. **convergence in measure does NOT imply point-wise a.e. convergence**, as shown with the typewriter counterexample, but the implication can be restored for a sub-sequence, see proposition [2.3.34 on page 139](#);
8. **point-wise a.e. convergence does NOT imply convergence in measure**, unless the domain has finite measure, see proposition [2.3.36 on page 140](#);
9. **convergence in mean implies convergence in measure**, see theorem [2.3.37 on the previous page](#);
10. **convergence in measure does NOT imply convergence in mean**, as shown with a counterexample.

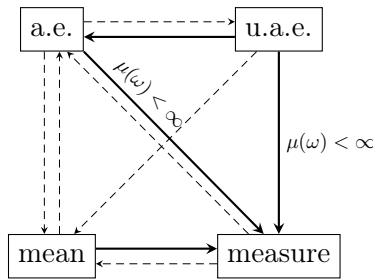


Figure 2.5: Summary of the implication of convergences. Solid: implied (potentially under some condition specified along the line); dashed: not implied.

To conclude our discussion we present the following result about convergence in measure.

Proposition 2.3.38

If $f_n \xrightarrow{\mu} f$ and $f_n \xrightarrow{\mu} g$, then $f = g$ a.e.

The space L^1 belongs to the family of functional space called “ L^p spaces”. In particular the distance $d_\infty(f, g) = \text{ess sup } |f - g|$ is a metric related to the space L^∞ , as (L^∞, d_∞) is a metric space. A wider discussion occurs in section [3.1.4 on page 191](#).

Metric convergence Some convergence can be induced by a metric, in such case they are said to be **metric convergences**.

The *uniform a.e. convergence* is induced by the d_∞ metric; indeed, we have:

$$d_\infty(f_n, f) \rightarrow 0 \iff f_n \xrightarrow{\text{u.a.e.}} f.$$

The *convergence in mean* is a metric convergence too, as it is induced by d_1 :

$$d_1(f_n, f) \rightarrow 0 \iff f_n \xrightarrow{L^1} f.$$

The *point-wise a.e. convergence* is not induced by any metric, so it is not a metric convergence. Some exceptions could occur, in particular when the measure μ is positive only on a countable set.

The case of *measure convergence* is a bit more complicated. We have to operate on a proper space and define a proper metric.

Consider the measure space $(\Omega, \mathfrak{M}, \mu)$ where μ is finite, and the following space:

$$\mathcal{F} := \{f : \Omega \rightarrow \mathbb{R} \text{ measurable}\};$$

then consider related quotient space $\mathcal{U} := \frac{\mathcal{F}}{\sim}$ defined with the following equivalence relation:

$$f \sim g \iff f = g \quad \text{a.e.},$$

that includes all the functions which are equal a.e. to a measurable one. Now define the following distance:

$$d_\mu(f, g) := \int_\Omega \frac{|f - g|}{1 + |f - g|} d\mu.$$

This is a metric in \mathcal{U} and (\mathcal{U}, d_μ) is a complete metric space.

The measure convergence is a metric convergence with the distance d_μ , namely:

$$d_\mu(f_n, f) \rightarrow 0 \iff f_n \xrightarrow{\mu} f \quad \text{as } n \rightarrow +\infty.$$

2.4 Fundamental theorems of calculus

This chapters contains the most important results of the real analysis; here we focus on calculus of one-valued functions, then, in the next chapter, we will consider also multi-valued functions. Some of those results are well-known also in lower calculus courses, but here we will approach the problem from a more technical point of view.

2.4.1 First fundamental theorem of calculus

We try to define a generic goal. Consider the measure space $([a, b], \mathcal{L}([a, b]), \lambda)$, take $f \in L^1([a, b], \mathcal{L}([a, b]), \lambda)$ and set

$$F(x) = \int_a^x f(t) d\lambda,$$

which properties does F have?

Lebesgue points First, our focus is on discontinuity points. The first step is to define a new notion for “continuous points”; on those we are confident that there are no issues with integration and differentiation.

Definition 2.4.1

A point $x \in [a, b]$ is a **Lebesgue point** for a function f if there exists a representative \tilde{f} of f ($\tilde{f} = f$ a.e.) such that:

$$\lim_{h \rightarrow 0} \frac{1}{h} \int_x^{x+h} |\tilde{f}(t) - \tilde{f}(x)| dt = 0,$$

where $h \rightarrow 0^+$ if $x = a$ or $h \rightarrow 0^-$ if $x = b$.

The integral uses the λ measure, the integral variable t has been written in dt for clarity.

Notice that \tilde{f} is a representative of the equivalence class $[f]$, given by almost-everywhere equality; it's typical of \mathcal{L}^1 spaces.

Lebesgue points do not present discontinuity and are a sort of “continuous points”; for instance, jump points are not a Lebesgue point; check it by considering:

$$f(t) = \begin{cases} \frac{t}{|t|} & \text{if } t \in [-1, 1] \setminus \{0\} \\ 0 & \text{if } t = 0. \end{cases}$$

Having checked that jumps are not Lebesgue points, one can wonder if a continuity point is a Lebesgue point. We have the following result.

Theorem 2.4.2

If x_0 is a continuity point for f , then x_0 is a Lebesgue point for f .

Proof. By definition of continuity:

$$\forall \varepsilon > 0 \quad \exists \delta : |x - x_0| < \delta \implies |f(x) - f(x_0)| < \varepsilon,$$

we evaluate the quantity:

$$\left| \frac{1}{h} \int_{x_0}^{x_0+h} |f(x) - f(x_0)| dt \right| \leq \frac{1}{|h|} \varepsilon |\lambda(x_0 + h - x_0)| = \frac{|h|}{|h|} \varepsilon = \varepsilon, \quad \forall \varepsilon > 0.$$

■

Theorem 2.4.3

Let $f \in L^1([a, b], \mathcal{L}([a, b]), \lambda)$.

Then almost any point $x \in [a, b]$ is a Lebesgue point of f . ²¹

The theorem We are able to state a first description of the relation between differentiation and integration.

Theorem 2.4.4 (First fundamental theorem of calculus)

Let $f \in L^1([a, b], \mathcal{L}([a, b]), \lambda)$.

The integral function $F(x) = \int_a^x f(t)$ is differentiable a.e. and $F' = f$ a.e. in $[a, b]$.

Proof. Let $x \in [a, b]$ be a Lebesgue point of f and $h \neq 0$ such that $x + h \in [a, b]$. Notice that we can rewrite the value of a function in the point as follow:

$$f(x) = \lim_{h \rightarrow 0} \frac{1}{h} \int_x^{x+h} f(t) dt.$$

Consider now the incremental quotient of the integral function F ; we need it equal to function f . So we write:

$$\frac{F(x+h) - F(x)}{h} - f(x) = \frac{1}{h} \int_x^{x+h} (f(t) - f(x)) dt,$$

²¹For further discussion and a proof, see: W. Rudin, Real and Complex Analysis, 1987, page 138, theorem 7.6.

then, by triangular inequality (see 2.3.3 on page 119), we have:

$$\left| \frac{F(x+h) - F(x)}{h} - f(x) \right| \leq \frac{1}{|h|} \int_x^{x+h} |f(t) - f(x)| dt.$$

Taking the limit for $h \rightarrow 0$ or $h \rightarrow 0^+$, if $x = a$, or $h \rightarrow 0^-$, if $x = b$, we have:

$$\frac{1}{|h|} \int_x^{x+h} |f(t) - f(x)| dt \rightarrow 0.$$

We have proved the theorem for any Lebesgue point, through the previous theorem (2.4.3 on the preceding page) we know that the set of non-Lebesgue points has measure zero; so the thesis is proven. ■

But this theorem opens another question: is F continuous in $[a, b]$?

2.4.2 Absolute continuity

Which functions can be written as integral functions of their derivatives? In order to answer this question we must first to analyze the properties of F , and find a property that is stronger than simple continuity.

Definition 2.4.5

A function $f : [a, b] \rightarrow \mathbb{R}$ is **absolutely continuous (AC)** in $[a, b]$ if for every $\varepsilon > 0$ it exists $\delta > 0$ such that:

$$\sum_{n=1}^N |f(b_n) - f(a_n)| < \varepsilon,$$

for all $N \in \mathbb{N}$, where the intervals $\{(a_n, b_n)\}_{n=1}^N$ are mutually disjoint and such that:

$$\bigcup_{n=1}^N (a_n, b_n) \subseteq [a, b] \text{ and } \sum_{n=1}^N (b_n - a_n) < \delta.$$

The definition still holds if $\{(a_n, b_n)\}_{n \in \mathbb{N}}$ are countable. This notion is more general than uniform continuity or continuity (1.2.42 on page 52 and 1.2.39 on page 51 respectively), indeed the following holds:

Proposition 2.4.6

If f is absolutely continuous in $[a, b]$, then it is also uniformly continuous in $[a, b]$.

The converse is false in general, consider this function as a counterexample:

$$f(x) = \begin{cases} x \sin \frac{1}{x} & \text{if } x \in [-1, 1] \setminus \{0\} \\ 0 & \text{if } x = 0. \end{cases}$$

We see that f is continuous in $[a, b]$, and thus f is also uniformly continuous (see theorem 1.2.88 on page 66).

However, f is not AC in $[a, b]$ by the definition.

Lipschitz continuity There exists another property of continuity:

Definition 2.4.7

Let $f : [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}$. We say that f is **Lipschitz continuous** if exists $k > 0$ such that:

$$|f(x) - f(y)| \leq k|x - y| \quad \forall x, y \in [a, b].$$

Observe that the derivative of any Lipschitz continuous function is bounded.

Proposition 2.4.8

If f is Lipschitz continuous in $[a, b]$, then it is also absolutely continuous in $[a, b]$.

Again, the converse implication is not guaranteed: consider $[a, b] = [0, 1]$ and $f(x) = \sqrt{x}$. While f is uniformly and absolutely continuous, it is not Lipschitz continuous as $\sup_{[0,1]} f'(x) = +\infty$.

Summary of the continuity notions for functions

- The usual (strong) definition of continuity, a function $f : [a, b] \rightarrow \mathbb{R}$ is continuous in x_0 if $\forall \varepsilon > 0, \exists \delta > 0 : |x - x_0| < \delta \implies |f(x) - f(x_0)| < \varepsilon$.
- Continuity almost everywhere, see 2.3.14 on page 125.
- Lipschitz continuity, see 2.4.7.
- Absolute continuity, see 2.4.5 on the facing page.

complete this with all the definitions of continuity

Absolute continuity of the integral

Theorem 2.4.9

Let $f \in L^1(\Omega, \mathfrak{M}, \mu)$. Then:

$$\forall \varepsilon > 0 \quad \exists \delta > 0 : \quad \int_E |f| \, d\mu < \varepsilon \quad \forall E \in \mathfrak{M} : \mu(E) < \delta.$$

Proof. Let $\varepsilon > 0$. Suppose $f \geq 0$ for simplicity (otherwise is possible to use the same proof for f^- and f^+).

There exists a sequence of simple functions $\{s_n\}_{n \in \mathbb{N}}$ bounded and measurable, such that $0 \leq s_n \leq f$ and $s_n \xrightarrow{L^1} f(t)$ in Ω ; so (see monotone convergence 2.2.5 on page 109 or dominated convergence 2.3.2 on page 119) there exists also a measurable and bounded function g such that

$$\int_E |f - g| \, d\mu \leq \frac{\varepsilon}{2}.$$

Set $M = \text{ess sup}_{t \in \Omega} |g(t)| > 0$ and $\delta = \frac{\varepsilon}{2M}$. Then:

$$\int_E |f| \, d\mu \leq \int_E |f - g| \, d\mu + \int_E |g| \, d\mu < \frac{\varepsilon}{2} + M\mu(E) < \frac{\varepsilon}{2} + M\delta = \frac{\varepsilon}{2} + M \frac{\varepsilon}{2M} = \varepsilon$$

with $\mu(E) < \delta$. ■

This corollary applies the previous theorem to the case of the integral function:

Corollary 2.4.10

Let $f \in L^1([a, b], \mathcal{L}[a, b], \lambda)$, and set:

$$F(x) = \int_a^x f(t) \, d\lambda.$$

Then the integral function F is absolutely continuous in $[a, b]$.

Proof. Fix $\varepsilon > 0$ and let $\delta > 0$ given by the previous theorem 2.4.9.

Let also $N \in \mathbb{N}$, and $\{(a_n, b_n)\}_{n=1:N} \subset [a, b]$ be a collection of mutually disjoint intervals such that $\sum_{n=1}^N (b_n - a_n) < \delta$ and $E = \bigcup_{n=1}^N (a_n, b_n)$ with

$\lambda(E) < \delta$.

Then we have:

$$\sum_{n=1}^N |F(b_n) - F(a_n)| = \sum_{n=1}^N \left| \int_{a_n}^{b_n} f(t) \, d\lambda \right| \leq \sum_{n=1}^N \int_{a_n}^{b_n} |f(t)| \, d\lambda = \int_E |f(t)| \, d\lambda < \varepsilon.$$

Thus, by the definition, F is absolutely continuous in $[a, b]$. ■

For example take $f(t) = \frac{1}{2\sqrt{t}} \in \mathcal{L}^1([0, 1], \mathcal{L}([0, 1]), \lambda)$; then we have:

$$\int_0^x \frac{1}{2\sqrt{t}} \, d\lambda = \sqrt{x}.$$

Thanks to the previous theorem we have that \sqrt{x} is absolutely continuous in $[0, 1]$. However we already know that it is not Lipschitz-continuous.

The space of absolute continuous functions It is easy to see that the set of the function which are absolute continuous on a closed interval is closed with respect to sum and scalar product. Indeed, if f and g are absolutely continuous in $[a, b]$, then $\alpha f + \beta g$ is absolutely continuous too. So the space

$$AC([a, b]) := \{f : [a, b] \rightarrow \mathbb{R} : f \text{ AC in } \Omega\},$$

is a vector space on \mathbb{R} with respect to the canonical operations.

2.4.3 Second fundamental theorem and bounded variation functions

We know that if $f \in L^1([a, b], \mathcal{L}[a, b], \lambda)$, then $F(x) = \int_0^x f(t) \, dt$ is such that $F \in AC([a, b])$ and $F' = f$ a.e. in $[a, b]$.

How can we find, given the functions $F : [a, b] \rightarrow \mathbb{R}$, a complete characterization²² of the following **calculus formula**?

$$F(x) = F(a) + \int_a^x F'(t) \, dt \quad \forall x \in [a, b]$$

These are some necessary conditions:

- $F' \in L^1([a, b], \mathcal{L}[a, b], \lambda)$;

²²A full characterization means a condition which is necessary and sufficient.

- $F \in AC([a, b])$.

Now we try to figure out if those conditions can be also sufficient. But first consider these two examples.

Look at this function:

$$F(x) = \begin{cases} x^2 \sin(\frac{1}{x^2}) & x \in [-1, 1] \setminus \{0\} \\ 0 & x = 0. \end{cases}$$

The function F is differentiable in $[-1, 1]$ but $F' \notin L^1([-1, 0])$.

Consider:

$$F(x) = \begin{cases} -1 & x \in [0, 1] \\ 1 & x \in [-1, 0) \end{cases}$$

It is differentiable in $[1, 1]$ and $F' = 0 \in L^1([-1, 1])$, but the calculus formula doesn't hold.

The characterization we are looking for is given by the following theorem:

Theorem 2.4.11 (Second fundamental theorem of calculus)

A function is absolutely continuous if and only if it is differentiable a.e. in $[a, b]$, its derivative belongs to L^1 and for such function holds the calculus formula. Namely:

$$F \in AC([a, b]) \iff \begin{cases} F \text{ is differentiable a.e. in } [a, b]; \\ F' \in L^1([a, b], \mathcal{L}[a, b], \lambda); \\ F(x) = F(a) + \int_a^x F'(x) \, d\lambda \quad \forall x \in [a, b]. \end{cases}$$

The *sufficient condition* (\implies) part is already proven.

Bounded variation functions To provide also a demonstration of the necessary part we have to introduce a new class of function.

Definition 2.4.12

Given a function $f : [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}$, and $x, y \in [a, b]$ such that

$x < y$, the **total variation** of f in $[x, y]$ is defined as follows:

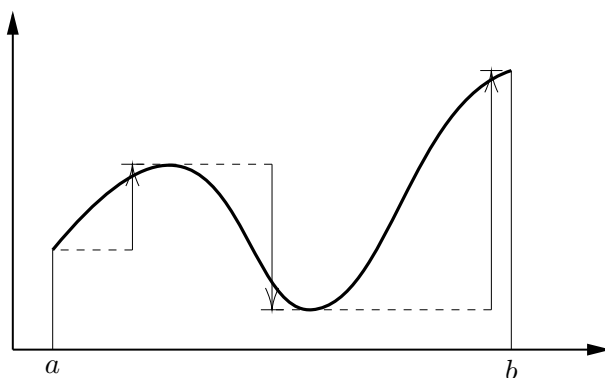
$$V_x^y(f) := \sup_{\{t_i\}} \sum_{n=1}^N |f(t_n) - f(t_{n-1})|,$$

where $x < t_0 < t_1 < \dots < t_N < y$.

The function f is said to be a **bounded variation function** in $[a, b]$ if $V_a^b(f) < +\infty$.

Notice that a function f is not a bounded variation function if and only if there exists $x \in (a, b]$ such that $V_a^x(f) = +\infty$: intuitively, the total variation quantifies jumps and oscillations of the function.

A sufficient condition for bounded variation is the monotonicity on an interval. Indeed, $V_a^b = f(b) - f(a)$ is non-decreasing or $V_a^b = f(a) - f(b)$ is non-increasing.



Let us consider some properties.

Proposition 2.4.13

Let f be a bounded variation function in $[a, b]$.²³ Then:

$$V_a^b(f) = V_a^c(f) + V_c^b(f) \quad \forall c \in [a, b].$$

Moreover the operator $x \mapsto V_a^x(f)$ is non decreasing, *i.e.*:

$$V_a^y(f) - V_a^x(f) = V_x^y(f) \geq 0 \quad \forall x, y : y > x.$$

Observe that the total variation is *absolute homogeneous*:

$$\forall c \in \mathbb{R} \quad V_x^y(cf) = |c| V_x^y(f).$$

The space of bounded variation functions and the relation with absolute continuous ones It is easy to see that the set of the bounded variation function on a closed interval is closed with respect to sum and scalar product:

$$BV([a, b]) := \{G : [a, b] \rightarrow \mathbb{R} : V_a^b(G) < \infty\}.$$

Indeed, if f and g are bounded variation in $[a, b]$ and $\alpha \in \mathbb{R}$, then:

$$V_a^b(\alpha f) = |\alpha| V_a^b(f) \text{ and } V_a^b(f + g) = V_a^b(f) + V_a^b(g).$$

Remember that this point was made also for absolute continuous functions: the spaces of functions with these properties are vector spaces. Also there is a strong relation between these two properties:

Proposition 2.4.14

If a function f is absolute continuous in $[a, b]$ then it is also bounded variation in $[a, b]$ and $x \mapsto V_a^x(f)$ is absolute continuous in $[a, b]$.²⁴
Namely:

$$f \in AC([a, b]) \implies \begin{cases} f \in BV([a, b]) \\ V_a^x(f) \in AC([a, b]). \end{cases}$$

Relations with monotonicity and continuity This new class of function is related to many well-known properties.

In general, uniform continuity does not imply bounded variation; consider the following example:

$$f(x) = \begin{cases} x \cos \frac{1}{x} & x \in (0, 1] \\ 0 & x = 0. \end{cases}$$

²³For further discussion, see: A. N. Kolmogorov, S. V. Fomin, *Introductory Real Analysis*, 1975, page 329, theorem 2.

²⁴For further discussion, see: A. N. Kolmogorov, S. V. Fomin, *Introductory Real Analysis*, 1975, page 337, theorem 2.

This function is uniformly continuous in $[0, 1]$ (see Heine–Cantor’s theorem 1.2.88 on page 66), but it is not bounded variation in $[0, 1]$: prove it using the definition on a suitable partition of such interval. Note also that the length of the graph of f is not finite.

Remember that a sufficient condition for bounded variation is the monotonicity on an interval. Moreover, we have the following proposition:

Proposition 2.4.15

Every bounded variation function can be written as a difference of two non-decreasing functions.

Proof. Observe that $f(x) = V_a^x(f) - [V_a^x(f) - f(x)]$.

Taking $v(x) = V_a^x(f) - f(x)$, by definition of total variation we have:

$$v(y) - v(x) = V_x^y(f) - (f(y) - f(x)) \geq 0 \quad \text{with } y > x$$

since the total variation can be equal to the jump between x and y (if f is monotone) or greater²⁵; and thus $v(x)$ is non-decreasing. ■

Due to this fact we understand that every bounded variation function has at most a countable set of jump discontinuities; so those function are Lebesgue-integrable! Namely if $f \in BV([a, b])$ then $f \in L^1([a, b])$.

Theorem 2.4.16

Any monotone function in $[a, b]$ has finite derivative a.e. in $[a, b]$.²⁶

Corollary 2.4.17

Any $f \in BV([a, b])$ has finite derivative a.e. in $[a, b]$. Moreover $f' \in L^1([a, b])$.

Moreover we have:

Proposition 2.4.18

Let $f : [a, b] \rightarrow \mathbb{R}$ be a non-decreasing function. Then its derivative

²⁵Think, for instance, if it jumps in between.

²⁶For further discussion, see: A. N. Kolmogorov, S. V. Fomin, *Introductory Real Analysis*, 1975, page 321, theorem 6.

$f' \in L^1([a, b])$ and we have:

$$\int_a^b f'(t) dt \leq f(b) - f(a).$$

Proof. Set:

$$\Phi_n(t) = n \left[f \left(t + \frac{1}{n} \right) - f(t) \right]$$

with $n \in \mathbb{N}_0$. We extend f after b , considering $f(t) = f(b) \forall t \in [b, b + \frac{1}{n}]$.

As $f \in L^1([a, b])$ also $\Phi_n \in L^1([a, b]) \forall n \in \mathbb{N}_0$; and $f'(t) = \lim_{n \rightarrow \infty} \Phi_n(t)$ for almost any $t \in [a, b]$ (convergence almost everywhere).

Then we have:

$$\begin{aligned} \int_a^b \Phi_n(t) dt &= n \int_a^b \left[f \left(t + \frac{1}{n} \right) - f(t) \right] dt && (\tau := t + \frac{1}{n}) \\ &= n \left[\int_{a+\frac{1}{n}}^{b+\frac{1}{n}} f(\tau) dt - \int_a^b f(t) dt \right] \\ &= n \left[\int_b^{b+\frac{1}{n}} f(t) dt - \int_a^{a+\frac{1}{n}} f(t) dt \right] && (\text{monotonicity}) \\ &\leq n f(b) \frac{1}{n} - n f(a) \frac{1}{n} = f(b) - f(a). \end{aligned}$$

We now apply Fatou's lemma, and thus we have:

$$\begin{aligned} \int_a^b f'(t) dt &= \int_a^b \lim_{n \rightarrow \infty} \Phi_n(t) dt \\ &\leq \liminf_{n \rightarrow \infty} \int_a^b \Phi_n(t) dt \\ &\leq \int_a^b \Phi_n(t) dt \\ &\leq f(b) - f(a). \end{aligned}$$

Thus $\int_a^b f'(t) dt \leq f(b) - f(a)$ and $f' \in L^1([a, b])$. ■

Observe that the inequality can be strict. Indeed, take:

$$f(x) = \begin{cases} 0 & x \in [0, \frac{1}{2}) \\ 1 & x \in [\frac{1}{2}, 1]; \end{cases}$$

in this case we get: $\int_0^1 f'(t)dt = 0 < 1 = f(1) - f(0)$.

The following last theorem provides a sort of generalization of the characterization of the constant function.²⁷

Theorem 2.4.19

Let $G \in AC([a, b])$ be non-decreasing and such that $G' = 0$ a.e. in $[a, b]$.

Then G is constant in $[a, b]$.

The proof of this result is very hard.

Proof of the second fundamental theorem Now we prove theorem 2.4.11 on page 150. The *sufficient condition* part is already proven. We need to prove that if $F \in AC([a, b])$ then F is differentiable a.e. in $[a, b]$, its derivative $F' \in L^1([a, b])$ and the calculus formula holds.

Proof. Necessary condition \Leftarrow :

Consider a function $F \in AC([a, b])$, we have that (see proposition 2.4.15 on page 153 and 2.4.14 on page 152):

$$F = F_1 - F_2 = V_a^x(F) - [V_a^x(F) - F];$$

where F_1, f_2 are $AC([a, b])$ and non-decreasing.

Thus we can suppose that F is non-decreasing without loss of generality, and prove the theorem in the general case by linearity.

Then, thanks to Corollary 2.4.17 on page 153 we have that F is differentiable a.e. in $[a, b]$ and $F' \in L^1([a, b])$.

Now it remains to prove the calculus formula.

Set:

$$\Phi(x) = F(x) - \int_a^x F'(t) dt \quad \forall x \in [a, b],$$

²⁷For further discussion see: A. N. Kolmogorov, S. V. Fomin, Introductory Real Analysis, page 339

it is $AC([a, b])$. Moreover, since F is non-decreasing, thanks to 2.4.18 on page 153 we have that for all $x_1, x_2 \in [a, b]$ such that $x_1 \leq x_2$:

$$\Phi(x_1) - \Phi(x_2) = F(x_1) - F(x_2) - \int_{x_1}^{x_2} F'(t) dt \geq 0,$$

so also Φ is non-decreasing.

Using the first fundamental theorem of calculus (2.4.4 on page 145), we have that $\Phi' = 0$ a.e. in $[a, b]$. Then (2.4.19 on the previous page) Φ is constant everywhere in $[a, b]$, namely $\Phi(x) = c$ for any $x \in [a, b]$. Take $x = a$, then $F(a) = c$, and thus:

$$F(x) = F(a) + \int_a^x F'(t) dt.$$

■

Notice that the first theorem was proven with the Lebesgue's point and it was quite trivial. Now we stated that absolute continuity is necessary for the calculus formula.

This is a complete characterization of AC functions: being L^1 is not enough for a function to be reconstructed.

This theorem is the largest generalization of the calculus formula currently existing.

2.4.4 Further results

A general version of the change of variable With this new knowledge we can improve our tools:

Theorem 2.4.20

Consider a function $g : [a, b] \rightarrow [c, d] \in AC([a, b])$, such that it is strictly monotone.

Then, for any $f : [c, d] \rightarrow \infty \in L^1([c, d])$ then:

$$\int_c^d f(t) dt = \int_a^b f(g(\tau)) |g'(\tau)| dt = \int_a^b f(g(\tau)) d\mu$$

where $\frac{d\mu}{d\tau} = |g'|$, the Radon–Nikodym derivative.

Notice that the fact that the composition of function is still Lebesgue-integrable is a part of the theorem: the compound function $(f \circ g)|g'|$ is measurable and integrable as a consequence of the assumptions, indeed we have:

$$f \in L^1([c, d]) \iff (f \circ g)|g'| \in L^1([a, b]).$$

The theorem also holds for f just measurable and non-negative. In this case, f is integrable in $[c, d]$ if and only if $(f \circ G)|G'|$ is integrable in $[a, b]$.

A sufficient condition for absolute continuity The following result holds²⁸:

Theorem 2.4.21

Let $f : [a, b] \in \mathbb{R} \rightarrow \mathbb{R}$. If f is differentiable everywhere in $[a, b]$ and $f' \in L^1([a, b])$, then $f \in AC([a, b])$.

Consider for example the following function:

$$f(x) = \begin{cases} x^{\frac{3}{2}} \sin \frac{1}{x} & \text{if } x \in (0, 1] \\ 0 & \text{if } x = 0. \end{cases}$$

Its derivative exists everywhere:

$$f'(x) = \begin{cases} \frac{3}{2}\sqrt{x} \sin \frac{1}{x} - \frac{1}{\sqrt{x}} \cos \frac{1}{x} & \text{if } x \in (0, 1] \\ 0 & \text{if } x = 0; \end{cases}$$

In 0 we computed separately: $\frac{f(h)-f(0)}{h} = \frac{h^{\frac{3}{2}} \sin \frac{1}{h}}{h} \rightarrow 0$ for $h \rightarrow 0^+$.

So f' is not continuous, but it is integrable in $([0, 1])$, and thus $f \in AC([0, 1])$.

Observer also that f is not lipschitz continuous.

What if we have that f is continuous, f' exists a.e., and f' is integrable in $[a, b]$? In this case f might not be $AC([a, b])$: consider for example the Vitali–Cantor function (see definition 1.2.100 on page 73).

Comparison between Lebesgue and Riemann integral There are plenty of definitions for integrals: now we make a comparison between the two integral that we already know. Consider, in this paragraph, the space $(\mathbb{R}, \mathcal{L}(\mathbb{R}), \lambda)$.

²⁸For further discussion, see: W. Rudin, Real and Complex Analysis, 1987, page 149, theorem 7.21.

Theorem 2.4.22

Let $f : (a, b) \subset \mathbb{R} \rightarrow \mathbb{R}$ bounded.

Then f is Riemann-integrable in (a, b) if and only if f is continuous a.e. in (a, b) .

29

Notice for example that $\mathbf{1}_{\mathbb{Q} \cap (a, b)}$ is not Riemann integrable, because it is nowhere continuous. But $\mathbf{1}_{\mathbb{Q} \cap (a, b)} = 0$ a.e. in (a, b) , it is Lebesgue-integrable, and $\int_a^b \mathbf{1}_{\mathbb{Q} \cap (a, b)} d\lambda = 0$.

More in general, if $f : (a, b) \rightarrow \mathbb{R}$ is Lebesgue-measurable and bounded, then f is Lebesgue integrable: indeed, f is measurable.

Theorem 2.4.23

Let $f : (a, b) \subset \mathbb{R} \rightarrow \mathbb{R}$ be bounded and continuous a.e. in (a, b) .

Then the Riemann integral and the Lebesgue integral coincide,³⁰ namely:

$$\underbrace{\int_a^b f(t) dt}_{\text{Riemann}} = \underbrace{\int_a^b f(t) d\lambda}_{\text{Lebesgue}}.$$

Consider for example the generalized Cantor set T_ε of Lebesgue measure $3 \frac{1-\varepsilon}{3-2\varepsilon}$. Then $\mathbf{1}_{T_\varepsilon}$ is bounded but it is not Riemann integrable since it's discontinuities at each point. Indeed, the interior of T_ε is empty and $\mathbf{1}_{T_\varepsilon}$ has no limit as $x \rightarrow x_0$ for any $x_0 \in T_\varepsilon$. Moreover $\mathbf{1}_{T_\varepsilon}$ is Lebesgue integrable in $[0, 1]$ but it is not equal almost everywhere to any Riemann integrable function.

Let us now focus on improper Riemann integrals:

Theorem 2.4.24

Let $f : (a, b) \subseteq \mathbb{R} \rightarrow \mathbb{R}$, with $a, b \in \mathbb{R}^* : a < b$.

If f is Riemann integrable in (a, b) in the improper sense and it changes its sign at most a finite number of times, then f is Lebesgue

²⁹For further discussion, see: W. P. Ziemer, Modern Real Analysis, 2017, page 153, theorem 6.19.

³⁰For the proof, see: A. N. Kolmogorov, S. V. Fomin, Introductory Real Analysis, 1975, pages 309-310, theorem 4.

integrable in (a, b) and the two integrals coincide.

If f changes sign an infinite number of times then the Riemann improper integral can exist but f might not be Lebesgue measurable. Consider the following function:

$$f(t) = \frac{\sin t}{t} \quad t \in (0, +\infty).$$

Then f is Riemann-integrable as $\lim_{a \rightarrow +\infty} \int_0^a f(t) dt$ exists. However, it is not integrable, indeed:

$$\int_0^{+\infty} \left| \frac{\sin t}{t} \right| d\lambda = \int_0^{+\infty} \frac{1}{t} d\lambda = +\infty.$$

Proof. Suppose $f > 0$. We can find two sequences, $\{a_n\}_{n \in \mathbb{N}}$ and $\{b_n\}_{n \in \mathbb{N}}$, such that $a_n < b_n$, $a_n \downarrow a$, $b_n \uparrow b$, and $(a, b) = \bigcup_{n \in \mathbb{N}} (a_n, b_n)$. Suppose f is bounded on $(a_n, b_n) \forall n \in \mathbb{N}$.

Therefore f is Riemann integrable on each (a_n, b_n) and we have:

$$\underbrace{\int_{a_n}^{b_n} f(t) dt}_{\text{Riemann}} = \underbrace{\int_{a_n}^{b_n} f(t) d\lambda}_{\text{Lebesgue}} \quad \forall n \in \mathbb{N}.$$

Set $F_n = \bigcup_{j=0}^n (a_j, b_j)$ and $f_n = f \mathbf{1}_{F_n}$. Then $f_n \uparrow f$ and, using monotone convergence, we have:

$$\lim_{n \rightarrow \infty} \int_a^b f_n(t) d\lambda = \lim_{n \rightarrow \infty} \int_{a_n}^{b_n} f(t) d\lambda = \int_a^b f(t) d\lambda.$$

By definition of Riemann improper integral, we have:

$$\int_a^b f(t) dt = \lim_{n \rightarrow \infty} \int_{a_n}^{b_n} f(t) dt.$$

Putting all together, we get the thesis:

$$\lim_{n \rightarrow \infty} \int_a^b f_n(t) d\lambda = \lim_{n \rightarrow \infty} \int_{a_n}^{b_n} f(t) dt.$$

■

Consider now some examples: the function

$$f(t) = \begin{cases} \sin(t) & \text{if } t \in \mathbb{R} \setminus \mathbb{Z} \\ +\infty & \text{if } t \in \mathbb{Z}^+ \\ -\infty & \text{if } t \in \mathbb{Z}_- \end{cases}$$

is essentially bounded and continuous a.e. in \mathbb{R} , and $f(t) = \sin(t)$ a.e. in \mathbb{R} .

The function

$$f(t) = \begin{cases} \sin(t) & \text{if } t \in [0, 1] \cap (\mathbb{R} \setminus \mathbb{Q}) \\ +\infty & \text{if } t \in [0, 1] \cap \mathbb{Q} \end{cases}$$

is nowhere continuous, bounded, and measurable. Then that f is Lebesgue integrable, hence $\int_0^1 f(t) d\lambda$ exists finite.

Indeed, $f(t) = \sin(t)$ a.e. in $[0, 1]$, thus:

$$\int_0^1 f(t) d\lambda = \underbrace{\int_0^1 \sin t d\lambda}_{\text{Lebesgue}} = \underbrace{\int_0^1 \sin t dt}_{\text{Riemann}}.$$

Lebesgue decomposition of bounded variation functions Any bounded variation function can be written as a sum of three proper function. Let's understand what kind of function they are.

Definition 2.4.25

Let $\{x_n\}_{n \in \mathbb{N}}$ and $\{x'_n\}_{n \in \mathbb{N}}$ be two sequences of points in $[a, b]$.

Let also $\{h_n\}_{n \in \mathbb{N}}, \{h'_n\}_{n \in \mathbb{N}} \subset \mathbb{R}$ such that:

$$\sum_{n \in \mathbb{N}} |h_n| < +\infty, \quad \sum_{n \in \mathbb{N}} |h'_n| < +\infty.$$

A function f is called **jump function** if it can be written as:

$$f(x) = \sum_{\substack{n \in \mathbb{N}: \\ x_n < x}} h_n + \sum_{\substack{n \in \mathbb{N}: \\ x'_n \leq x}} h'_n.$$

Notice that the first sum generates a function continuous from the left, the other one from the right.

Jump functions are step function if the sequences $\{x_n\}$ and $\{x'_n\}$ are strictly increasing, but jump functions can be more general.

There is another more explanatory form for writing a jump function $f : [a, b] \rightarrow \mathbb{R}$ where $f(a) = 0$. In such case we would have:

$$f(x) = \sum_{n \in \mathbb{N}} g_n(x) + \sum_{n \in \mathbb{N}} g'_n(x),$$

where:

$$g_n(x) = \begin{cases} 0 & a \leq x \leq x_n \\ h_n & x_n < x \leq b \end{cases} \quad g'_n(x) = \begin{cases} 0 & a \leq x < x'_n \\ h'_n & x_n \leq x \leq b. \end{cases}$$

We can provide some example of jump functions. For instance we can define a jump function on $[0, 1]$ by setting $x_n = \frac{1}{n}$ and $h_n = \frac{(-1)^n}{n^{\frac{3}{2}}}$ with $n \in \mathbb{N}_0$. In that case we have $f(\frac{1}{100}) = \sum_{n=101}^{+\infty} \frac{(-1)^n}{n^{\frac{3}{2}}}$ or $f(\frac{1}{\sqrt{71}}) = \sum_{n=8}^{+\infty} \frac{(-1)^n}{n^{\frac{3}{2}}}$.

Here an example of a jump function which is not a step function; let $\{x_n\}_{x \in \mathbb{N}}$ be an enumeration of \mathbb{Q} , and set $h_n = \frac{1}{2^n}$. Thus our function is:

$$f(x) = \sum_{x_n < x} \frac{1}{2^n}.$$

It can be proven that f is discontinuous at any $x \in \mathbb{Q}$ and continuous at any $x \in \mathbb{R} \setminus \mathbb{Q}$.³¹

Observe that any jump function $f : [a, b] \rightarrow \mathbb{R}$ belongs to $BV([a, b])$,³² therefore it is differentiable a.e. in $[a, b]$ and its derivative is zero a.e.. Moreover, as they are bounded, they are also Lebesgue-integrable.

There exists non-constant functions $f : [a, b] \rightarrow \mathbb{R}$ which are continuous, non-decreasing with zero derivative a.e.; such functions are BV but not AC since the Calculus formula doesn't hold. A counterexample is again the Vitali–Cantor function (see definition 1.2.100 on page 73).

Thinking to such counterexample, we can define another kind of function.

³¹For further discussion, see: A. N. Kolmogorov, S. V. Fomin, *Introductory Real Analysis*, 1975, pages 315-327.

³²What is its total variation?

Definition 2.4.26

A non constant function $f : [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}$ is a **singular function** (or Cantor-like function) if:

- f is continuous in $[a, b]$;
- f is non-decreasing;
- $\exists f'$ a.e. and $f' = 0$ a.e. in $[a, b]$.

We see lot of kinds of functions: they are somehow related by the following result.

Theorem 2.4.27 (Lebesgue decomposition of bounded variation functions)

Let $f \in BV([a, b])$. Then there exist three functions:

- a function $\phi_1 \in AC([a, b])$;
- a singular function ϕ_2 ;
- a jump function ϕ_3

such that

$$f = \phi_1 + \phi_2 + \phi_3 \text{ in } [a, b].$$

33

Notice that, if $f \in BV([a, b])$, then $f' = \phi_1'$ a.e. in $[a, b]$. Thus, as we already know, we cannot in general reconstruct a bounded variation function by integrating its derivative: the calculus formula only reconstructs the AC function of the Lebesgue decomposition, see that if $f' = \phi_1'$ a.e. in $[a, b]$ then:

$$\int_a^x f'(t) dt = \phi_1(x) - \phi_1(a).$$

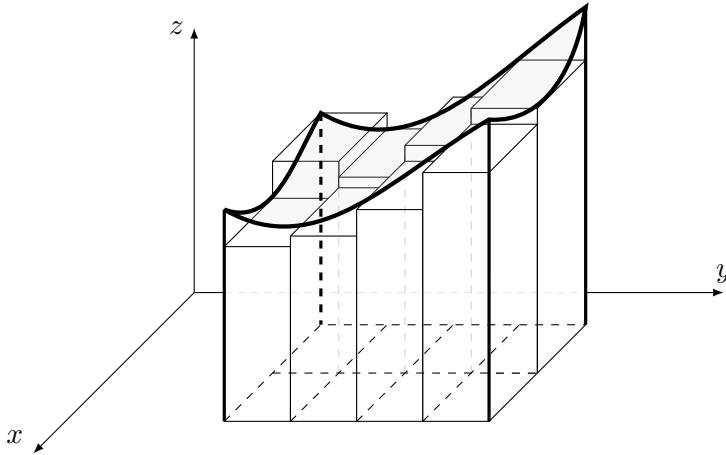
³³For further discussion, see: G. Leoni, A First Course in Sobolev Spaces, 2017, page 104, theorem 3.89.

2.5 Integrals on product spaces

2.5.1 Product measures

In this section we discuss on how to measure a function which domain is not mono-dimensional. We know yet how the Riemann integration has solved this problem:

part of this can be taken as intro of the subsection.



The main idea is to calculate the integral through what we know in one dimension, using the techniques for the integration of one variable:

$$\iint_{\Omega} f(x, y) dx dy = \int_a^b \left(\int_{\Omega_x} f(x, y) dy \right) dx = \int_c^d \left(\int_{\Omega_y} f(x, y) dx \right) dy$$

Now we need to extend this to the Lebesgue and abstract integral. This technique should be extended to proceed for a very general set. Our goal in this part is to extend the “reduction” we have just explained to the Lebesgue integral or, better, to the abstract integral.

Product of σ -algebras Our first step is to build a proper σ -algebra. Let (X, \mathfrak{M}) , (Y, \mathfrak{N}) measure spaces, with $X, Y \neq \emptyset$.

How can we construct a product σ -algebra $\mathfrak{M} \otimes \mathfrak{N}$ such that the “restrictions” of its elements with respect to X (or Y) are measurable set in \mathfrak{N} (or \mathfrak{M})?

Definition 2.5.1

A set $\mathcal{R} \in \mathcal{P}(X \times Y)$ is a **measurable rectangle** if:

$$\mathcal{R} = A \times B$$

where $A \in \mathfrak{M}$ and $B \in \mathfrak{N}$.

Definition 2.5.2

Let \mathfrak{M} and \mathfrak{N} be two σ -algebra. We define the **product σ -algebra** as

$$\mathfrak{M} \otimes \mathfrak{N} := \langle \{A \times B : A \in \mathfrak{M}, B \in \mathfrak{N}\} \rangle.$$

This is the σ -algebra generated by all the measurable rectangles.

This is the smallest σ -algebra containing all the measurable rectangles. At this point you should be able to prove that $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$

The symbol \otimes is known as tensor product but we will not study in deep this operation.

Definition 2.5.3

Let $E \subset X \times Y$. Then:

$E_x := \{y \in Y : (x, y) \in E\} \subset Y$ is called the **x -section** of E ;

$E_y := \{x \in X : (x, y) \in E\} \subset X$ is called the **y -section** of E .

Notice that, if we have a rectangle $E = A \times B$, then we have $E_x = B$ for all $x \in A$ and $E_y = A$ for all $y \in B$.

Through the last definition we can provide a proper characterization of the product σ -algebra $\mathfrak{M} \otimes \mathfrak{N}$:

Proposition 2.5.4 (characterization of the product σ -algebra)

Consider a generic rectangle $E \subset X \times Y$.

We can define $\mathfrak{M} \otimes \mathfrak{N}$ as follows:

$$\begin{aligned} \mathfrak{M} \otimes \mathfrak{N} &= \{E \in \mathfrak{M} \otimes \mathfrak{N} \text{ such that } E_x \in \mathfrak{N} \quad \forall x \in X\} \\ &= \{E \in \mathfrak{M} \otimes \mathfrak{N} \text{ such that } E_y \in \mathfrak{M} \quad \forall y \in Y\} \end{aligned}$$

Proof. Set

$$\mathcal{C} := \{E \in \mathfrak{M} \otimes \mathfrak{N} : E_x \in \mathfrak{N} \quad \forall x \in X\}.$$

Obviously $\mathcal{C} \subseteq \mathfrak{M} \otimes \mathfrak{N}$, and any measurable rectangle is in \mathcal{C} . To prove the opposite relation, we just need to prove that \mathcal{C} is a σ -algebra as $\mathfrak{M} \otimes \mathfrak{N}$ is the generated minimal σ -algebra.

First, it is easy to see that any measurable rectangle belongs to \mathcal{C} , namely for any $A \in \mathfrak{M}$ and $B \in \mathfrak{N}$, $A \times B \in \mathcal{C}$. In particular, $X \times Y \in \mathcal{C}$.

Then, notice that if $E \in \mathcal{C}$ then $(E^C)_x = (E_x)^C \in \mathfrak{M}$, which in turn implies $E^C \in \mathcal{C}$.

Finally, consider $E = \bigcup_{n \in \mathbb{N}} E_n$ with $E_n \in \mathcal{C}$. Then:

$$E_x = \left(\bigcup_{n \in \mathbb{N}} E_n \right)_X = \bigcup_{n \in \mathbb{N}} (E_n)_x$$

The same goes for y-sections. ■

There may exist a set E such that $E_x \in \mathfrak{N}$ for all $x \in X$ and $E_y \in \mathfrak{M}$ for all $y \in Y$ but $E \notin \mathfrak{M} \otimes \mathfrak{N}$.³⁴

Measurable functions Now that we have a proper structure we can talk about measurability.

Definition 2.5.5

Let $f : X \times Y \rightarrow \mathbb{R}$, or $[0, +\infty]$, be defined everywhere.

Then we define the **x-section** f^x and the **y-section** f^y of f by setting:

$$f^x : Y \rightarrow \mathbb{R}, \quad f^x(y) := f(x, y) \quad \forall x \in X \text{ fixed};$$

$$f^y : X \rightarrow \mathbb{R}, \quad f^y(x) := f(x, y) \quad \forall y \in Y \text{ fixed}.$$

(Pay attention: those symbols are not partial derivatives.)

Notice that it's superfluous now to say "everywhere" because there is still no measure.

Proposition 2.5.6

³⁴For further discussion, see: E. Hewitt, K. Stromberg, Real and Abstract Analysis, page 393, exercise 21.26.

If $f : (X \times Y, \mathfrak{M} \otimes \mathfrak{N}) \rightarrow \mathbb{R}$, or $[0, +\infty]$ is $(\mathfrak{M} \otimes \mathfrak{N})$ -measurable, then:

$$f^x \text{ is } \mathfrak{N}\text{-measurable} \quad \forall x \in X;$$

$$f^y \text{ is } \mathfrak{M}\text{-measurable} \quad \forall y \in Y.$$

Proof. Let $A \subseteq f(X \times Y)$ be open. Then:

$$E = f^{-1}(A) = \{(x, y) \in X \times Y : f(x, y) \in A\} \in \mathfrak{M} \otimes \mathfrak{N}.$$

Therefore for any A we have (see characterization of product σ -algebra, proposition 2.5.4 on page 164):

$$E_x = (f^x)^{-1}(A) = (f^x)^{-1}(A) \in \mathfrak{N} \quad \forall x \in X,$$

$$E_y = (f^y)^{-1}(A) = (f^y)^{-1}(A) \in \mathfrak{M} \quad \forall y \in Y.$$

This means that f^x and f^y are measurable with respect to the corresponding σ -algebras. ■

Product measure Now our setting is quite ready: the last theoretical concept that has to be extended is the measure itself. Consider the measure spaces (X, \mathfrak{M}, μ) , (Y, \mathfrak{N}, ν) . We would like to build a new measure space with $X \times Y$ and $\mathfrak{M} \otimes \mathfrak{N}$. Which *measure* could be suitable?

First, take into account the following result.

Proposition 2.5.7

Let μ, ν two σ -finite measures, and $E \in \mathfrak{M} \otimes \mathfrak{N}$. Then:

- $x \rightarrow \nu(E_x)$ is \mathfrak{M} -measurable;
- $y \mapsto \mu(E_y)$ is \mathfrak{N} -measurable;
- $\int_x \nu(E_x) d\mu = \int_y \mu(E_y) d\nu$.

35

The previous proposition suggests the following.

³⁵To see the proof: A. N. Kolmogorov, S. V. Fomin, Introductory Real Analysis, 1975, pages 356-359, theorem 3.

Definition 2.5.8

Consider the set function $\mu \otimes \nu : \mathfrak{M} \otimes \mathfrak{N} \rightarrow [0, +\infty]$ defined as follows:

$$(\mu \otimes \nu)(E) := \int_X \nu(E_x) d\mu = \int_Y \mu(E_y) d\nu \quad \forall E \in \mathfrak{M} \otimes \mathfrak{N}.$$

Then $\mu \otimes \nu$ is a σ -finite measure and is called **product measure** of μ and ν .

Observe that for any measurable rectangle

$$R = A \times B \in \mathfrak{M} \otimes \mathfrak{N}$$

we have

$$(\mu \otimes \nu)(R) = \mu(A) \cdot \nu(B).$$

In general, $\mu \otimes \nu$ is not complete even if μ and ν are so (see definition 2.1.32 on page 97).

Example 2.5.9. Take the sets $X = Y = [0, 1]$ with σ -algebras $\mathfrak{M} = \mathfrak{N} = \mathcal{L}([0, 1])$ and measure $\mu = \nu = \lambda$.

Then let $A \in \mathcal{L}([0, 1])$ such that $\lambda(A) = 0$, and $B \in \mathcal{L}([0, 1])$ such that $\lambda(B) > 0$. So we have $(\lambda \otimes \lambda)(A \times B) = 0$.

Now consider the Vitali set $U \subset B$: $A \times U \subset A \times B$ so $A \times U \notin \mathcal{L}([0, 1]) \otimes \mathcal{L}([0, 1])$.

So $A \times U$ is zero-measure subset of the σ -algebra but it is not measurable then, by definition the measure $\lambda \otimes \lambda$ is not complete.

2.5.2 Integrals and Fubini–Tonelli’s theorems

We have build all the ingredients to define the abstract integral on a multi-dimensional space. Consider the measurable space $(X \times Y, \mathfrak{M} \otimes \mathfrak{N}, \mu \otimes \nu)$. In case $E \in \mathfrak{M} \otimes \mathfrak{N}$ the *reduction formula* for characteristic functions holds:

$$\int_{X \times Y} \mathbb{1}_E d(\mu \otimes \nu) = (\mu \otimes \nu)(E) = \int_X \nu(E_x) d\mu = \int_Y \mu(E_y) d\nu$$

This formula can be generalized: the following result shows that it can be extended to measurable functions. This results is made of two theorem which are different assumptions but the same final result, the *iterated integral formula*.

As painful as it may be to state and comment this two theorems separately for an association named “Fubini⊗Tonelli”, this course requires to at least understand where each one comes from and why they work so well together.³⁶

Theorem 2.5.10 (Tonelli)

Consider the measure space $(X \times Y, \mathfrak{M} \otimes \mathfrak{N}, \mu \otimes \nu)$, and let f be a function $(\mathfrak{M} \otimes \mathfrak{N})$ -measurable and defined everywhere in $X \times Y$. If the function is non-negative, namely $0 \leq f \leq +\infty$, then we have the **iterated integral formula**:

$$\int_{X \times Y} f \, d(\mu \otimes \nu) = \int_X \left(\int_Y f^x \, d\nu \right) \, d\mu = \int_Y \left(\int_X f^y \, d\mu \right) \, d\nu.$$

Observe that the first part of the iterated integral formula is actually an abstract integral, as it is defined on a measure.

For this next theorem we require also completeness for both measures.

Theorem 2.5.11 (Fubini)

Consider the measure space $(X \times Y, \mathfrak{M} \otimes \mathfrak{N}, \mu \otimes \nu)$, and let f be a function $(\mathfrak{M} \otimes \mathfrak{N})$ -measurable and defined everywhere in $X \times Y$. If f is such that $\int_{X \times Y} |f| \, d(\mu \otimes \nu) < +\infty$, then we have:

$$\begin{aligned} f^x &\in L^1(Y, \mathfrak{N}, \nu) && \text{for almost any } x \in X; \\ f^y &\in L^1(X, \mathfrak{M}, \mu) && \text{for almost any } y \in Y. \end{aligned}$$

Moreover:

$$\begin{aligned} x &\mapsto \int_Y f^x \, d\nu \in L^1(X, \mathfrak{M}, \mu); \\ y &\mapsto \int_X f^y \, d\mu \in L^1(Y, \mathfrak{N}, \nu). \end{aligned}$$

Finally, the iterated integral formula holds.

Corollary 2.5.12

Let μ, ν σ -finite and complete respectively on $\mathfrak{M}, \mathfrak{N}$.

³⁶Somehow like the pieces of Exodia.

Consider the measure space $(X \times Y, \mathfrak{M} \otimes \mathfrak{N}, \mu \otimes \nu)$, and let f be a function $(\mathfrak{M} \otimes \mathfrak{N})$ -measurable and defined everywhere in $X \times Y$. If, in addition, we have that:

$$\int_X \left(\int_Y |f^x| d\nu \right) d\mu = \int_Y \left(\int_X |f^y| d\mu \right) d\nu.$$

Then we have:

$$\int_{X \times Y} |f| d(\mu \otimes \nu) < +\infty$$

and the iterated integral formula holds.

Proof of the corollary. Consider $|f|$: using Tonelli's theorem, we deduce that:

$$\int_{X \times Y} |f| d(\mu \otimes \nu) < +\infty,$$

so we can apply Fubini's theorem and obtain the iterated integral formula. ■

Thanks to Fubini's theorem if we have $\sum_{i,j \in \mathbb{N}} |a_{ij}| < +\infty$, then:

$$\sum_{i \in \mathbb{N}} \left(\sum_{j \in \mathbb{N}} a_{ij} \right) = \sum_{j \in \mathbb{N}} \left(\sum_{i \in \mathbb{N}} a_{ij} \right),$$

and the double series converges.

Set $X = Y = \mathbb{N}$, $\mathfrak{M} = \mathfrak{N} = \mathcal{P}(\mathbb{N})$, $\mu = \nu = \mu_c$, and the measure $\mu_c \otimes \mu_c$ is defined with the convention $0 \cdot \infty = 0$.

Completion of a product measure Consider two complete measure space (X, \mathcal{M}, μ) and (Y, \mathcal{N}, ν) with σ -finite measure. Then $(X \times Y, \mathfrak{M} \otimes \mathfrak{N}, \mu \otimes \nu)$ is not complete, but we can consider its completion:

$$(X \times Y, (\mathfrak{M} \otimes \mathfrak{N})^*, (\mu \otimes \nu)^*).$$

Then Fubini–Tonelli's theorem can be relaxed:

Theorem 2.5.13

If $(X \times Y, (\mathfrak{M} \otimes \mathfrak{N})^*, (\mu \otimes \nu)^*)$ is the completion of $(X \times Y, \mathfrak{M} \otimes \mathfrak{N}, \mu \otimes \nu)$ and $f : X \times Y \rightarrow \mathbb{R}$, or $[0, \infty]$, is $(\mathfrak{M} \otimes \mathfrak{N})^*$ -measurable, then Fubini–Tonelli's theorem hold but the inner integrals and the

iterated integral formula are defined almost everywhere.

If we case of the Lebesgue measure we have:

Proposition 2.5.14

The completion of

$$(\mathbb{R}^N \times \mathbb{R}^M, \mathcal{L}(\mathbb{R}^N) \otimes \mathcal{L}(\mathbb{R}^M), \lambda_N \otimes \lambda_M)$$

is

$$(\mathbb{R}^{N+M}, \mathcal{L}(\mathbb{R}^{N+M}), \lambda_{N+M}).$$

Observe that $(\mathbb{R} \times \mathbb{R}, \mathcal{L}(\mathbb{R}) \otimes \mathcal{L}(\mathbb{R}), \lambda \otimes \lambda)$ is not complete.

Take $A \in \mathcal{L}(\mathbb{R})$ such that $\lambda(A) = 0$ and $B \in \mathcal{L}(\mathbb{R})$ such that $\lambda(B) > 0$.

Consider then a Vitali set $\mathcal{V} \subset B$. The rectangle $\mathcal{R} = A \times B$ is such that $(\lambda \otimes \lambda)(\mathcal{R}) = 0$ but $A \times \mathcal{V} \subset \mathcal{R}$ and $(A \times \mathcal{V})_x = \mathcal{V} \notin \mathcal{L}(\mathbb{R})$ so that $A \times \mathcal{V} \notin \mathcal{L}(\mathbb{R}) \otimes \mathcal{L}(\mathbb{R})$.

3 Functional Analysis

Real analysis is somehow a generalization of basic calculus notions. We have seen functions, integrals and in the end we have seen the second theorem of calculus which is a generalization of the Calculus Formula.

At the beginning of this book we introduced topological spaces which are geometrical notions. From now on, these two parts will meet with functional analysis.

This is the field of mathematics in which relations between functional spaces are studied. If in real analysis we worked in finite dimensional space, now we extend our scope to infinite dimensional spaces, which elements are functions.

First we study such spaces which are, as we see, a generalization of euclidean spaces, and in a second step we will study operator between those spaces. We will focus on linear operators.

3.1 Banach spaces

3.1.1 Vector spaces

We are now entering an infinite-dimensional world.

Now we will introduce vector spaces on \mathbb{R} . In some contexts other fields are considered, like \mathbb{C} for quantum mechanics applications.

Definition 3.1.1

Consider the set $V \neq \emptyset$.

A tuple $(V, +, \cdot, \mathbb{R})$ is a **vector space** on \mathbb{R} if the *sum of vectors* law

$$u + v$$

follows these properties:

- **commutativity:**

$$u + v = v + u;$$

- **associativity:**

$$u + (v + w) = (u + v) + w;$$

- **existence of null element**, called zero vector:

$$u + 0 = u;$$

- **existence the inverse element:**

$$u + (-u) = 0;$$

and the *multiplication of a vector by a scalar* law

$$\alpha v \text{ with } \alpha \in \mathbb{R}$$

follows these properties:

- **distributivity with respect to the scalar product:**

$$\alpha(\beta v) = (\alpha\beta)v;$$

- **existence of the identity element:**

$$1u = u;$$

- **distributivity with respect to the sum in \mathbb{R} :**

$$(\alpha + \beta)v = (\alpha v) + (\beta v);$$

- **distributivity with respect to the sum in V :**

$$\alpha(v + u) = (\alpha v) + (\alpha u).$$

The elements of V are called **vectors**.

Generalizing the two operation we have the so called *finite linear combination of vectors*:

$$\sum_{j=1}^N \alpha_j v_j.$$

Numerical set \mathbb{R}^N , with the foretold structure, is a vector space for any n .

A vector space is simply an algebraic structure, there is no topology yet so there is no order.

Definition 3.1.2

Consider a vector space $V \subseteq \mathbb{R}$.

A subset of vectors $A \subset V$ is **linearly independent** if, for any $\{v_1, \dots, v_N\} \subset A$ and for any $\alpha_1, \dots, \alpha_N \in \mathbb{R}$, we have:

$$\sum_{i=1}^N \alpha_i v_i = 0 \implies \alpha_1 = \dots = \alpha_N = 0.$$

Basis and dimension Can we write for each element of the space as a linear combination of a special restricted subset of other elements? Yes, that subset of elements is a basis for the vector space, and for each vector space we may have more than one basis.

Definition 3.1.3

Consider a vector space $V \subseteq \mathbb{R}$.

A subset of vectors $A \subset V$ is an **algebraic** or **Hamel basis** if it is linearly independent and maximal, that is there are no elements which can be added to it without losing linear independence.

So, if A is an algebraic basis of V , then every element of V can be written as a finite linear combination of elements of A .

Theorem 3.1.4 (fundamental theorem of linear algebra)

Every vector space on \mathbb{R} has at least one algebraic basis.

To prove this theorem if V has at least two elements, in the Zermelo–Fraenkel theory, we need the axiom of choice. To be more precise, Andreas Blass proved in 1984 that this theorem is equivalent to the axiom of choice.

Proposition 3.1.5

Let V a vector space on \mathbb{R} .

Then all its algebraic basis has the same cardinality.

Definition 3.1.6

The **dimension** of a vector space is the cardinality of one of its

algebraic basis.

Example 3.1.7. For example consider the space of the algebraic polynomials:

$$\mathcal{P}(x) = \left\{ f(x) = \sum_{j=0}^n a_j x^j : n \in \mathbb{N}; x, a_0, \dots, a_n \in \mathbb{R} \right\}.$$

Since the two canonical operations $P + Q$ and αP can be defined, then \mathcal{P} is a vector space on \mathbb{R} . We have that $\dim \mathcal{P} = \aleph_0$ and one of his algebraic basis is $\{1, x, x^2, x^3, \dots, x^n, \dots\}$.

Example 3.1.8. Consider also the space of continuous functions:

$$\mathcal{C}(\mathbb{R}) = \{f : \mathbb{R} \rightarrow \mathbb{R} : f \text{ continuous in } \mathbb{R}\}.$$

We can define the two canonical operations over $\mathcal{C}(\mathbb{R})$: indeed, $\alpha f + \beta g \in \mathcal{C}(\mathbb{R})$ for all $\alpha, \beta \in \mathbb{R}$ and for all $f, g \in \mathcal{C}(\mathbb{R})$.

Then $\mathcal{C}(\mathbb{R})$ is a vector space on \mathbb{R} and $\dim \mathcal{C}(\mathbb{R}) \geq \aleph_0$ since $\mathcal{P}(x)$, from the previous point, is a subspace of $\mathcal{C}(\mathbb{R})$.

Definition 3.1.9

A set W is a **subspace** of the vector space V if

$$W \subset V$$

and it is closed with respect to vector sum and scalar product, namely

$$\alpha v + \beta u \in W$$

for all $\alpha, \beta \in \mathbb{R}$ and for all $v, u \in W$.

3.1.2 Normed vector spaces

In the section assigned to topology (see section [1.2 on page 34](#)) we have extensively used the notion of distance between two elements. How to adapt that notion to “label” each element of a set, and in particular of a vector space, capturing its “largeness”? Here we define the norm, which is a function that assign a real positive number to each element of the vector space. As we will see later, norm and distance are strictly related to each other, in particular the norm can be seen as the distance of the element from the zero element.

Definition 3.1.10

Let X be a vector space.

A function $\|\cdot\|_X : X \rightarrow [0, +\infty)$ is called **norm** in X if the following properties hold:

- **non-negativity:**

$$\|x\| \geq 0 \quad \forall x \in X;$$

- **annihilation:**

$$\|x\| = 0 \iff x = 0;$$

- **homogeneity with respect to scalar multiplication:**

$$\|\alpha x\| = |\alpha| \|x\| \quad \forall \alpha \in \mathbb{R} \quad \forall x \in X;$$

- **triangular inequality:**

$$\|x + y\| \leq \|x\| + \|y\| \quad \forall x, y \in X.$$

The tuple $(X, \|\cdot\|_X)$ is called **normed vector space**.

In the setting of Euclidean spaces, the norm is interpreted as the vector length; here we have a generalization.

Notice that any normed vector space $(X, \|\cdot\|_X)$ is always a metric space with respect to the metric induced by the norm. This can be done by defining the distance as follows:

$$d(x, y) = \|x - y\| \quad \forall x, y \in X.$$

In particular, it is also a topological space with respect to the inducted topology.

The backward implication is not always true; indeed, a metric space (X, d) is not necessarily a normed vector space, for example consider a ball in \mathbb{R}^3 with the euclidean distance; it is not closet with respect to the sum.

Notice that a metric is necessarily induced by a norm unless it is translation invariant and homogeneous.

Proposition 3.1.11

The function $x \rightarrow \|x\|$ is continuous with respect to the metric. Indeed, the following formula holds:

$$| \|x\| - \|y\| | \leq \|x - y\|.$$

This is a sort of inverse triangular inequality: it allows to control the difference of norms. Prove it!

Proposition 3.1.12

Any norm on \mathbb{R}^N with canonical operations has the following form:

$$\|u\| = c|u| \quad \text{for some } c > 0.$$

Now consider some examples of norm.

Example 3.1.13. Consider $x \in X = \mathbb{R}^N$, with $n \geq 2$, let $p \in [1, +\infty)$. The following are norms:

$$\|x\|_p := \left(\sum_{j=1}^n |x_j|^p \right)^{1/p} \quad \|x\|_\infty := \max\{|x_1|, \dots, |x_n|\}.$$

Example 3.1.14. The space of all continuous function

$$\mathcal{C}([a, b]) := \{f : [a, b] \rightarrow \mathbb{R} \text{ continuous in } [a, b]\}$$

is a vector space. An example of a norm in $\mathcal{C}([a, b])$ is

$$\|f\|_\infty := \max_{t \in [a, b]} |f(t)|.$$

Example 3.1.15. There exist examples of norms for other functional spaces: a norm for $L^1(\Omega, \mathfrak{M}, \mu)$ is

$$\|f\|_1 := \int_{\Omega} |f(t)| d\mu;$$

while a norm for $L^\infty(\Omega, \mathfrak{M}, \mu)$ is

$$\|f\|_\infty := \operatorname{ess\,sup}_{\Omega} |f(t)|.$$

Example 3.1.16. The set of converging series

$$l^p := \left\{ \{x_n\}_{n \in \mathbb{N}} \subset \mathbb{R} : \sum_{u \in \mathbb{N}} |x_u|^p < +\infty \right\}$$

is a vector space with respect to the canonical operations (you should prove it), consider $X = \{x_n\}_{n \in \mathbb{N}} \in l^p$. Then

$$\|x\|_p := \left(\sum_{n \in \mathbb{N}} |x_n|^p \right)^{\frac{1}{p}}$$

is a norm in l^p .

Example 3.1.17. Take now

$$l^\infty := \{ \{x_n\}_{n \in \mathbb{N}} \subset \mathbb{R} : \{x_n\}_{n \in \mathbb{N}} \text{ bounded} \};$$

this is a normed vector space as well and a norm in l^∞ is

$$\|x\|_\infty = \sup_{n \in \mathbb{N}} |x_n|.$$

Sequences As we have done at the beginning with metric spaces, having define a notion of distance allow us to discuss about sequences and their convergence in those new spaces.

Definition 3.1.18

Let $(X, \|\cdot\|)$ be a normed vector space.

We say that $\{x_n\}_{n \in \mathbb{N}} \subset X$ **converges** to an element $x \in X$ if:

$$\forall \varepsilon > 0 \exists n_0 = n_0(\varepsilon) \in \mathbb{N} : d(x_n, x) = \|x_n - x\| < \varepsilon \quad \forall n > n_0.$$

In this case we write $\lim_{n \rightarrow \infty} x_n = x$ or, equivalently $x_n \rightarrow x$ as $n \rightarrow \infty$.

Observe that $x_n \rightarrow x$ if and only if $\|x_n - x\| \rightarrow 0$ as $n \rightarrow \infty$.

Notice that we have (see proposition [3.1.11 on the preceding page](#)):

$$x_n \rightarrow x \implies \|x_n\| \rightarrow \|x\| \quad \text{as } n \rightarrow +\infty.$$

Remember that a converging sequence is fundamental but the converse is not true, think to $(\mathbb{Q}, \|\cdot\|)$.

Definition 3.1.19

Let $(X, \|\cdot\|)$ be a normed vector space.

$\{x_n\}_{n \in \mathbb{N}}$ is a **fundamental sequence** (or Cauchy sequence) in X if:

$$\forall \varepsilon > 0 \exists n_0 = n_0(\varepsilon) \in \mathbb{N} : \|x_n - x_m\| < \varepsilon \quad \forall n, m > n_0.$$

A converging sequence is always fundamental as well, but the converse does not hold in general. For example consider the $\|f\|_1$ in $\mathcal{C}([a, b])$.

Definition 3.1.20

Let $(X, \|\cdot\|)$ be a normed vector space,

We say that $E \subset X$ is **bounded** if exists $M > 0$ such that:

$$E \subseteq \{x \in X : \|x\| < M\} = B_M(0);$$

where $B_M(0)$ is a **ball** centered in 0 of radius M .

As any sequences still remains a set of elements, this definition fit with sequences; indeed, it's easy to see that any fundamental sequence is bounded. As we will see, this will be still valid for series as well. Observe also that if $a_n \rightarrow a$ in \mathbb{R} and $x_n \rightarrow x$ in $(X, \|\cdot\|)$ then $a_n x_n \rightarrow ax$ in X .

The following holds for both metric and topological spaces:

Proposition 3.1.21

Let $A \subset X$, where X is a normed vector space.

Then $x^* \in X$ is a cluster point³⁷ for A if and only if:

$$\exists \{x_n\} \subset A : x_n \neq x^* \quad \forall n \in \mathbb{N} \quad \text{and} \quad x_n \rightarrow x^*.$$

Keep in mind that all the topological notion on those two kind of space depends only on the notion of sequence or neighborhoods respectively.

Series Having a vector space structure and a topology we can introduce something new which can't be explained with the theory of metric spaces only.

³⁷Recalling definition 1.2.11 on page 39 or 1.2.23 on page 45, cluster points are also known as accumulation points. Those points are the ones which has at least one other point of the same set in any of their neighborhood.

Definition 3.1.22

Let $(X, \|\cdot\|)$ be a normed vector space.

Consider a sequence $\{x_n\}_{n \in \mathbb{N}} \subset X$: the sequence of partial sums

$$S_n := \sum_{j=0}^n x_j,$$

is called **series** of the elements x_n and is denoted by

$$\sum_{n \in \mathbb{N}} x_n.$$

We say that such a series **converges** to some $x \in X$ in $(X, \|\cdot\|)$ if we have

$$\|S_n - x\| \rightarrow 0 \text{ as } n \rightarrow \infty.$$

It's important to remember that a series is defined upon a sequence, it is a new sequence of which the element n is the sum from element 0 to element n of the given series. Now let's dive into this new concept.

Proposition 3.1.23 (generalized triangular inequality)

If $\sum_{n \in \mathbb{N}} x_n$ converges in $(X, \|\cdot\|)$, then we have:

$$\left\| \sum_{n \in \mathbb{N}} x_n \right\| \leq \sum_{n \in \mathbb{N}} \|x_n\|.$$

Proof. Step 0:

If $\sum_{n \in \mathbb{N}} \|x_n\| = +\infty$, the inequality holds.

Step 1:

We have to prove that the sequence of absolute values is fundamental; consider $m > n$ and define $S_m = \sum_{i=0}^m x_i$ and $S_n = \sum_{i=0}^n x_j$:

$$\|S_m - S_n\| = \left\| \sum_{i=n+1}^m x_i \right\| \leq \sum_{i=n+1}^m \|x_i\| \xrightarrow{n, m \rightarrow +\infty} 0,$$

where the inequality is true due to the triangular inequality, and the limit is because $\|S_n - x\| \rightarrow 0$ with x finite. As the sequence of absolute values converges, it is fundamental.

Step 2:

Observe that $S_n \rightarrow \sum_{i \in \mathbb{N}} x_i$ as $n \rightarrow \infty$, and thanks to the continuity of the norm, under the same conditions, we have that

$$\|S_n\| \rightarrow \left\| \sum_{i \in \mathbb{N}} x_i \right\|.$$

Step 3:

Again, using the triangular inequality, we have that:

$$\left\| \sum_{i=1}^n x_i \right\| \leq \sum_{i=1}^n \|x_i\| \quad \forall n \in \mathbb{N};$$

both the members of the inequality converges when $n \rightarrow \infty$ and we have the thesis. ■

Separability Here we remember some topological notions.

Definition 3.1.24

Let (X, τ) be a topological space.

We say that $E \subset X$ is **dense** in X if $\bar{E} = X$

Definition 3.1.25

A topological space (X, τ) is **separable** if X contains a countable dense set.

Example 3.1.26. The space of \mathbb{R}^N with the standard metric topology, like $(\mathbb{R}^N, \|\cdot\|_p)$, is separable $p \in [1, +\infty]$, with $E = \mathbb{Q}^n$.

Moreover, $(\mathcal{C}([a, b]), \|\cdot\|_\infty)$ is separable, as we can see from the following result.

At last, a relevant result about continuous function approximation.

Theorem 3.1.27 (Stone–Weierstrass)

Let $\mathcal{P}([a, b])$ be the subspace of algebraic polynomials.

Then $\mathcal{P}([a, b])$ is dense in $(\mathcal{C}([a, b]), \|\cdot\|_\infty)$.

This means that any continuous function can be approximated as precisely as desired by a polynomial.

Norms equivalence As we have done with metrics also for norm we have a notion of equivalence (see definition 1.2.8 on page 36).

Definition 3.1.28

Let $(X, \|\cdot\|_{\clubsuit}), (X, \|\cdot\|_{\spadesuit})$ be normed vector spaces.

We say that the norms $\|\cdot\|_{\clubsuit}$ and $\|\cdot\|_{\spadesuit}$ are **equivalent** in X if:

$$\exists m, M > 0 : \quad m \|\cdot\|_{\clubsuit} \leq \|\cdot\|_{\spadesuit} \leq M \|\cdot\|_{\clubsuit} \quad \forall x \in X.$$

Observe that two normed space on the same set with two equivalent norm have the same induced topology: same open sets, same closed sets, same convergent sequences.

Proposition 3.1.29

In a finite dimensional vector space, all norms are equivalent.

Proof. Without loss of generality, we will prove that in \mathbb{R}^N all the norms are equivalent. Consider the norm $\|\cdot\|_1$ and let $\|\cdot\|$ be any other norm; we have to prove that:

$$m \|x\|_1 \leq \|x\| \leq M \|x\|_1 .$$

Second inequality:

Take the canonical basis in \mathbb{R}^N : $\{\mathbf{e}_j\}_{j=1}^N = \{\mathbf{e}_1, \dots, \mathbf{e}_N\}$, so that $\mathbf{x} = \sum_{j=1}^N \alpha_j \mathbf{e}_j$ and $\|\mathbf{x}\|_1 = \sum_{j=1}^N |\alpha_j|$; we have:

$$\|\mathbf{x}\| = \left\| \sum_{j=1}^N x_j \mathbf{e}_j \right\| \leq \sum_{j=1}^N |x_j| \|\mathbf{e}_j\| \leq M \sum_{j=1}^N |x_j| = M \|\mathbf{x}\|_1 ,$$

where $M = \max_{j \in \{1, \dots, N\}} \|\mathbf{e}_j\|$ and so $\|\mathbf{x}\| \leq M \|\mathbf{x}\|_1$.

First inequality:

Set $\phi(\mathbf{x}) = \|\mathbf{x}\|$, by norm continuity ϕ is continuous in \mathbb{R}^N ; in particular, it is continuous on the compact set $K = \{\mathbf{x} \in \mathbb{R}^N : \|\mathbf{x}\|_1 = 1\}$, so we have:

$$|\phi(\mathbf{x}) - \phi(\mathbf{x}_0)| \leq \|\mathbf{x} - \mathbf{x}_0\| \leq M \|\mathbf{x} - \mathbf{x}_0\|_1 .$$

Thus ϕ has a minimum $m \geq 0$ on K due to the Weierstrass theorem, and we have:

$$\phi \left(\frac{\mathbf{x}}{\|\mathbf{x}\|_1} \right) \geq m \quad \forall \mathbf{x} \in \mathbb{R}^N .$$

It must be $m > 0$: indeed, if $m = 0$ for some \mathbf{x}_m , then $\mathbf{x}_m = \mathbf{0} \notin K$.

Finally we have $\|\mathbf{x}\| = \phi(\mathbf{x}) \geq m \|\mathbf{x}\|_1$ for all $\mathbf{x} \in \mathbb{R}^N$.

Alternatively, one could use the BIM corollary [3.2.44 on page 225](#) to prove this second point. ■

It can be proved that any vector space $V \subseteq \mathbb{R}^N$, with $K = \dim V$ is isomorphic to \mathbb{R}^K , that is there exists a linear bijection $F : V \rightarrow \mathbb{R}^K$. In particular, the previous theorem holds for any finite-dimensional vector space on \mathbb{R}^N .

Moreover, consider two vector spaces, namely V_1 and V_2 , whose dimensions are N_1 and N_2 respectively. If they are linearly isomorphic, then $N_1 = N_2$. In particular \mathbb{R} and \mathbb{R}^N , with $N > 1$, cannot be linearly isomorphic, even if they are equivalent as sets.

Examples of equivalent norms are presented together with examples of Banach spaces in next section.

3.1.3 Banach spaces

Definition 3.1.30

A normed vector space $(X, \|\cdot\|)$ is a **Banach space** if it is complete³⁸ with respect to the metric induced by the norm.

Here a first but relevant observation.

Proposition 3.1.31

Every closed subspace of a Banach space is a Banach space itself.

Consider now some examples.

Example 3.1.32. The normed vector spaces $(\mathbb{R}^N, \|\cdot\|_p)$ are Banach spaces $\forall p \in [1, \infty]$. All those norms are equivalent each other.

Example 3.1.33. The normed vector spaces $(l^p, \|\cdot\|_p)$ are Banach spaces $\forall p \in [1, \infty]$.

³⁸See definition [1.2.64 on page 59](#), a metric space is complete if any fundamental sequence converges.

Example 3.1.34. The normed vector space of the continuous functions $(\mathcal{C}([a, b]), \|\cdot\|_\infty)$ is a Banach space. Indeed, if $\|f_n - f\| \rightarrow 0$, then:

$$\forall \varepsilon > 0 \exists n_0 = n_0(\varepsilon) \in \mathbb{N} : |f_n(t) - f(t)| \leq \max_{t \in [a, b]} |f_n(t) - f(t)| < \varepsilon.$$

Moreover, as $\|\cdot\|_\infty$ is not equivalent to the integral norm $\|\cdot\|_1$, $(X, \|\cdot\|_1)$ is not a Banach space.

In $X = \mathcal{C}^1([a, b])$ the norm $\|\cdot\|_{\infty, 1}$ is equivalent to:

$$\|f\| = |f(a)| + \|f'\|_\infty.$$

Example 3.1.35. Let $k \in \mathbb{N}_0$ and define the following space:

$$\mathcal{C}^k([a, b]) := \{f : [a, b] \rightarrow \mathbb{R} : f^{(i)} \in \mathcal{C}([a, b]) \forall i = 1, \dots, k\}.$$

Its norm is

$$\|f\|_{\infty, k} := \sum_{j=0}^k \|f^{(j)}\|_\infty.$$

Then $(\mathcal{C}^k([a, b]), \|\cdot\|_{\infty, k})$ are Banach spaces.

Example 3.1.36. Consider the following space:

$$\mathcal{C}^\infty([a, b]) := \{f : [a, b] \rightarrow \mathbb{R} : \exists f^{(i)} \in \mathcal{C}([a, b]) \forall i \in \mathbb{N}\}.$$

It is a vector space, but there does not exist a norm such that it becomes a Banach space. However, it is a complete metric space with the norm:

$$d_\infty(f, g) = \sum_{j=0}^{\infty} 2^{-j} \frac{\|f^{(j)} - g^{(j)}\|_\infty}{1 + \|f^{(j)} - g^{(j)}\|_\infty}.$$

Also the space of absolute continuous function is Banach.

Example 3.1.37. Let $X = AC[a, b]$. You should prove that

$$\|f\|_\star = |f(a)| + \|f'\|_1$$

is a norm in X , then see the next.

Here we prove that $(AC([a, b]), \|\cdot\|_\star)$ is a Banach space.

Consider a fundamental sequence of absolute continuous function $\{f_n\}_{n \in \mathbb{N}}$.

We have to check that this sequence converges to an absolute continuous function.

Take the sequence $\{f_n(a)\}_{n \in \mathbb{N}}$, it is fundamental and converges in $(\mathbb{R}, |\cdot|)$, so we write $f_n(a) \rightarrow \alpha$.

Consider also $\{f'_n\}_{n \in \mathbb{N}}$, it's fundamental as well and exists $g \in L^1([a, b])$ such that $\|f_n - g\|_1 \rightarrow 0$ for $n \rightarrow \infty$.

By calculus formula we have a relation between those sequences, indeed:

$$f_n(t) = f_n(a) + \int_a^t f'_n(\tau) d\tau \quad t \in [a, b].$$

As f'_n converges to g , it does also the integral, indeed, for all $t \in [a, b]$ we have:

$$\left| \int_a^t (f'_n - g)(\tau) d\tau \right| \leq \int_a^b |(f'_n - g)(\tau)| d\tau = \|f'_n - g\|_1 \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

This implies that f_n converges point-wise (and uniformly) in $[a, b]$ to a function f defined by:

$$f(t) = \alpha + \int_a^t g(\tau) d\tau \quad t \in [a, b],$$

so that $f(a) = \alpha$ and $f' = g$ a.e. in $[a, b]$.

By properties of absolutely continuous functions (see, in particular, [2.4.9 on page 148](#)), $f \in AC([a, b])$ and $\|f_n - f\|_* \rightarrow 0$ as $n \rightarrow \infty$.

Hence, $(AC([a, b]), \|\cdot\|_*)$ is a Banach space.

Moreover, $AC([a, b])$ the following norm is equivalent to the one we present:

$$\|f\| = |f(a)| + \|f'\|_1;$$

also this norm makes $AC([a, b])$ a Banach space.

Example 3.1.38. The space $BV([a, b])$ is a Banach space with respect to the norm

$$\|f\| = \|f\|_1 + V_a^b(f),$$

but $\|\cdot\|_{AC}$ is not a norm in X .

Finally, notice that in an infinite dimensional vector space it's possible to construct non-equivalent norms such that all of them make it a Banach space.

Schauder bases Similarly to the algebraic basis for vector space, can we write any element of a Banach space as a sum of a series of a given sequence of elements? Yes, what we obtain is a sort of weak version of the algebraic basis, as we will see this new basis is for a set which is dense in the original space.

Definition 3.1.39

Let $(X, \|\cdot\|)$ be a normed vector space on \mathbb{R} , $\{x_n\}_{n \in \mathbb{N}} \subset X$, and set:

$$E = \left\{ \sum_{n=0}^k \alpha_n x_n : \alpha_0, \dots, \alpha_k \in \mathbb{R}, k \in \mathbb{N} \right\}.$$

Then the sequence $\{x_n\}_{n \in \mathbb{N}}$ is a **Schauder basis** if E is dense in X .

In other words, a sequence is a Schauder basis in X if the set of its finite linear combinations with rational coefficients is dense in X .

One can easily figure out the following result:

Proposition 3.1.40

If a normed vector space has a Schauder basis, then it is separable.

In general, the reverse is not true.³⁹

Observe that $\{x^n\}_{n \in \mathbb{N}}$ is a Schauder basis for $(\mathcal{C}([a, b]), \|\cdot\|_\infty)$ (see the Stone–Weierstrass theorem 3.1.27 on page 180).

Characterization of Banach spaces Here we provide one characterization through the following theorem. It states that a normed vector space $(X, \|\cdot\|)$ is a Banach space if and only if any absolutely convergent series converges in X .⁴⁰

Theorem 3.1.41

Let $(X, \|\cdot\|)$ be a normed vector space.

Then:

$$(X, \|\cdot\|) \text{ is a B-space} \iff [\forall \{x_n\}_{n \in \mathbb{N}} \subset X, \sum_{n \in \mathbb{N}} \|x_n\| < +\infty \implies \sum_{n \in \mathbb{N}} x_n < +\infty]$$

³⁹Proved by Per Henrik Enflo (Stockholm, 1944) in 1975. Thanks to this proof, he won a goose as a prize.

⁴⁰Remember that a series $\sum_{n \in \mathbb{N}} x_n$ converges absolutely if $\sum_{n \in \mathbb{N}} \|x_n\|$ converges.

Proof. Necessary condition \implies :

Having that $(X, \|\cdot\|)$ is Banach, it's enough to prove that $\sum_{n \in \mathbb{N}} x_n$ is fundamental. Again by hypothesis, $\sum_{n \in \mathbb{N}} \|x_n\| < +\infty$.

Define $S_n = \sum_{i=0}^n x_i$; taking $m > n$, we have:

$$\|S_m - S_n\| = \left\| \sum_{i=n+1}^m x_i \right\| \leq \sum_{i=n+1}^m \|x_i\| \rightarrow 0 \quad \text{as } n, m \rightarrow +\infty;$$

hence $\sum_{n \in \mathbb{N}} x_n$ is fundamental and, being in a Banach space, it converges.

Sufficient condition \longleftarrow :

Consider a Cauchy sequence $\{x_n\}_{n \in \mathbb{N}}$. We can construct a sub-sequence $\{x_{n_h}\}_{h \in \mathbb{N}_0}$ such that:

$$\|x_{n_{h+1}} - x_{n_h}\| \leq \frac{1}{h^2} \quad \forall h \in \mathbb{N}_0.$$

Then, by hypothesis, the series $\sum_{h \in \mathbb{N}_0} (x_{n_{h+1}} - x_{n_h})$, which is telescopic, converges to some $x \in X$, and so the sequence of partial sums does:

$$S_{n_{h+1}} - S_{n_1} \rightarrow x \in X \quad \text{as } h \rightarrow \infty,$$

which means $x_{n_{h+1}} \rightarrow x + x_{n_1} \in X$ as $h \rightarrow \infty$.

From a known result, since we had a Cauchy sequence and we've been able to extract a converging subsequence, then also the Cauchy sequence converges, namely $\{x_n\}_{n \in \mathbb{N}}$ converges. \blacksquare

Finite dimensional Banach spaces Here we have another topological notion that we will try to analyze in the context of Banach spaces. We have seen this property yet, recall definitions of compact space [1.2.69 on page 61](#) and sequentially compact space [1.2.77 on page 62](#), as well Bolzano–Weierstrass theorem [1.2.20 on page 43](#), the theorem of characterization of compact spaces [1.2.80 on page 63](#) and Heine–Borel's [1.2.82 on page 64](#).

Bolzano–Weierstrass states that in a metric space on \mathbb{R}^N , with any metric, from any bounded sequence can be extracted a converging subsequence. Now we know that some metrics coupled with \mathbb{R}^N constitutes a Banach space. The following theorem uses the “Bolzano–Weierstrass” property to characterize finite-dimensional Banach spaces.

Theorem 3.1.42 (characterization of finite-dimensional Banach spaces)

A Banach space $(X, \|\cdot\|)$ is finite-dimensional if and only if from any bounded sequence it is possible to extract a converging subsequence.

Frederic Riesz's theorem ⁴¹ From the previous results we can deduce that any compact set in a Banach space is closed and bounded. The following result try to understand how the converse can be true, and an answer is provided by its corollaries.

Theorem 3.1.43 (Frederic Riesz)

Let $(X, \|\cdot\|)$ be a a normed vector space.

If the closure of the unit ball $B \subset X$ is compact then X is finite-dimensional.

To prove this we will use the following.

Lemma 3.1.44

If $(X, \|\cdot\|)$ is a normed vector space and $Y \subsetneq X$ is a closed (linear) subspace then, for all $\varepsilon > 0$, there exists $x \in X$ such that:

$$\|x\| = 1 \quad \text{and} \quad d(x, Y) \geq 1 - \varepsilon \quad \forall y \in Y.$$

Proof. Let $y \in X \setminus Y$, and fix $y \in Y$. We have that $d(y, Y) = \|x - y\| > 0$ as Y is closed.

Suppose that $\varepsilon \in (0, 1)$, then choose $z \in Y$ for which:

$$d \leq \|y - z\| \leq \frac{d}{1 - \varepsilon}.$$

Setting:

$$x = \frac{y - z}{\|y - z\|} \notin Y,$$

we have for any $u \in Y$ that:

$$\|x - u\| = \left\| \frac{y - z}{\|y - z\|} - u \right\| \geq \frac{d}{\|y - z\|} \geq 1 - \varepsilon,$$

since $z + \|y - z\| u \in Y$.

If $\varepsilon > 1$ then the proof is trivial. ■

⁴¹Frigeys Riesz was an hungarian mathematician whose name is often spelled "Frederic". Other Riesz theorem are named after his younger brother Marcel.

Now we can proceed.

Proof of the Frederic Riesz's theorem. By contradiction suppose that X is infinite-dimensional.

Then we can find a sequence of finite-dimensional subspaces $\{E_n\}_{n \in \mathbb{N}}$ such that $E_{n-1} \subsetneq E_n$.

Thanks to the previous lemma, we can construct a sequence $\{x_n\}_{n \in \mathbb{N}}$ picking up an element from each set as follows:

$$x_n \in E_n, \quad \|x_n\| = 1, \quad d(x_n, E_{n-1}) \geq \frac{1}{2}.$$

In particular, we have $\|x_n - x_m\| \geq \frac{1}{2}$ for $n \neq m$. This means that $\{x_n\}_{n \in \mathbb{N}}$ is bounded but has no convergent sequence. This is a contradiction with the characterization of finite-dimensional Banach spaces (see theorem [3.1.42 on the preceding page](#)), so the proof is complete. ■

Frederic Riesz's theorem has many consequences.

Corollary 3.1.45

A Banach space is finite-dimensional if and only if any bounded and closed subset is compact.

Corollary 3.1.46

Let $(X, \|\cdot\|)$ be an infinite-dimensional normed vector space.

Then any compact set has an empty interior; that is it doesn't contain any ball.

A compactness criterion for the space of continuous functions Our goal now is to deduce a compactness criterion for the *infinite-dimensional* Banach spaces of continuous functions. The following theorem is the first step.

Theorem 3.1.47 (Ascoli–Arzelà)

Consider the metric space $(\mathcal{C}([a, b]), \|\cdot\|_\infty)$, and let $\{f_n\}_{n \in \mathbb{N}} \subset \mathcal{C}([a, b])$.

If $\{f_n\}_{n \in \mathbb{N}}$ is:

- **bounded**, namely:

$$\exists M > 0 : \sup_{n \in \mathbb{N}} \|f_n\|_\infty \leq M;$$

- **equicontinuous**, namely:

$$\forall \varepsilon > 0 \exists \delta = \delta(\varepsilon) > 0 :$$

$$\forall t_1, t_2 \in [a, b] : |t_1 - t_2| < \delta \implies |f_n(t_1) - f_n(t_2)| < \varepsilon \quad \forall n \in \mathbb{N};$$

then there exists a subsequence $\{f_{n_h}\}_{h \in \mathbb{N}}$ such that $f_{n_h} \rightarrow f$ in $(\mathcal{C}([a, b]), \|\cdot\|_\infty)$.

Notice that as $f_{n_h} \rightarrow f$ in norm $\|\cdot\|_\infty$, the convergence is uniform.

Proof. The goal of this proof is to extract a subsequence which converges uniformly, that is a convergence with respect to the infinity norm.

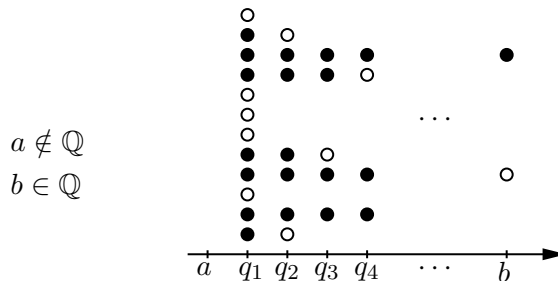
Step 1: Let $\{q_j\}_{j \in \mathbb{N}}$ be an enumeration of $\mathbb{Q} \cap [a, b]$.

Consider first $\{f_n(q_0)\}_{n \in \mathbb{N}}$: because of boundedness, it is bounded in \mathbb{R} , so we can use Bolzano–Weierstrass (see theorem 1.2.20 on page 43) and assert that there exists a subsequence $\{f_{n_{k_0}}(q_0)\}_{k_0 \in \mathbb{N}}$ which converges.

Consider now $\{f_{n_{k_0}}(q_1)\}_{k_0 \in \mathbb{N}}$: it is bounded as well, and there exists $\{f_{n_{k_1}}(q_1)\}_{k_1 \in \mathbb{N}}$ which converges.

Notice that $\{f_{n_{k_1}}(q_0)\}_{k_1 \in \mathbb{N}}$ also converges.

Repeating this argument for every element in $\{q_j\}_{j \in \mathbb{N}}$ using a Cantor diagonalization argument (the one used to prove that \mathbb{Q} is countable), we get a sequence of functions $\{f_{n_h}(t)\}_{h \in \mathbb{N}}$ which converges point-wise for $t = q_j$.



Step 2: Then fix $\varepsilon > 0$. Via equicontinuity, we find $\delta = \delta(\varepsilon) > 0$ such that:

$$|f_{n_h}(t_1) - f_{n_h}(t_2)| < \frac{\varepsilon}{3} \quad \forall t_1, t_2 \in [a, b] : |t_1 - t_2| < \delta \quad \forall h \in \mathbb{N}.$$

Consider now a family of intervals $I_i = \{[a_j, b_j]\}_{j=1}^N$ such that $b_j - a_j < \delta$ and $\cup_{j=1}^N [a_j, b_j] = [a, b]$. Then, for any i , we can choose $q_i \in I_i$, because \mathbb{Q} is dense in \mathbb{R} .

Observe that there exists $h_0(\varepsilon) \in \mathbb{N}$, independent of j , such that⁴²:

$$|f_{n_h}(q_j) - f_{n_{h'}}(q_j)| \leq \frac{\varepsilon}{3} \quad \forall h, h' > h_0 \quad \forall j = 1, \dots, N.$$

Step 3: Let $t \in [a, b]$. Then $t \in I_i$ for some $i \in 1 : N$, thus $|t - q_i| < \delta$. Therefore we have, for any $h, h' > h_0$:

$$\begin{aligned} |f_{n_h}(t) - f_{n_{h'}}(t)| &\leq |f_{n_h}(t) - f_{n_h}(q_j)| + |f_{n_h}(q_j) - f_{n_{h'}}(q_j)| + |f_{n_{h'}}(q_j) - f_{n_{h'}}(t)| \\ &\leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \\ &= \varepsilon \end{aligned}$$

Thus $\{f_{n_h}\}$ is a uniform Cauchy sequence, that is a Cauchy sequence with respect to $\|\cdot\|_\infty$, and it converges in the Banach space $(\mathcal{C}([a, b]), \|\cdot\|_\infty)$. ■

Also it can be shown that the boundedness and equicontinuity are also necessary for compactness in $(\mathcal{C}([a, b]), \|\cdot\|_\infty)$.

Actually, equicontinuity and equiboundedness are also necessary conditions: namely, if $K \subset \mathcal{C}([a, b])$ is compact, then any $\{f_n\} \subset K$ satisfies them.

How this theorem is related to compactness? Apparently in the theorem formulation there is no mention to compactness but we see in the following example that exists a strong relation.

Example 3.1.48. Consider the following subset of $(\mathcal{C}([a, b]), \|\cdot\|_\infty)$:

$$E_M = \{f \in \mathcal{C}^1([a, b]) : \|f\|_\infty + \|f'\|_\infty \leq M\}$$

for some given $M > 0$.

Thanks to the Ascoli–Arzelà theorem, we can prove that E_M is sequentially compact, and so it is also compact. Indeed, take a sequence $\{f_n\}_{n \in \mathbb{N}} \subset E_M$: it is bounded and, owing to the Lagrange theorem, we have:

$$|f_n(x) - f_n(y)| \leq M|x - y| \quad \forall x, y \in [a, b], \quad \forall n \in \mathbb{N}.$$

Thus the sequence is also equicontinuous so it contains a subsequence which converges uniformly to some $f \in E_M$.

⁴²Since this can be done for any j , it's enough to take the biggest, as N is finite.

3.1.4 L^p spaces

As we have seen in Banach spaces examples, we are interested in investigating whether functional spaces are Banach. Our goal in this section is to prove that the spaces related to Lebesgue-integrable functions are Banach. First recall the definition of the space of Lebesgue-integral function \mathcal{L}^1 (see definition 2.3.1 on page 118) and the definition of the space L^1 (see definition 2.3.30 on page 132). Here we define many other spaces related to those; a relevant fact is that all these spaces have a vector space structure.

Definition 3.1.49

Let $(\Omega, \mathfrak{M}, \mu)$ be complete measure space and fix $p \in [1, \infty)$.

$$\mathcal{L}^p(\Omega, \mathfrak{M}, \mu) := \{f : \Omega \rightarrow \mathbb{R} \text{ measurable} : |f|^p \in \mathcal{L}^1(\Omega, \mathfrak{M}, \mu)\};$$

or, equivalently:

$$\mathcal{L}^p(\Omega, \mathfrak{M}, \mu) := \{f : \Omega \rightarrow \mathbb{R} \text{ measurable} : \int_{\Omega} |f|^p d\mu < \infty\}.$$

In addition, we define:

$$L^p(\Omega, \mathfrak{M}, \mu) := \frac{\mathcal{L}^p(\Omega, \mathfrak{M}, \mu)}{\sim}$$

where the equivalence relation is $f \sim g \iff f = g$ almost everywhere with respect to μ in Ω .

Since the inequality

$$\left(\frac{x+y}{2}\right)^p \leq \frac{1}{2}|x|^p + \frac{1}{2}|y|^p \quad \forall x, y \geq 0$$

holds, then $L^p(\Omega, \mathfrak{M}, \mu)$ is a vector space on \mathbb{R} with respect to the standard operations: $\{f\} + \{g\}$, $\alpha\{f\}$.

For each of those spaces we can define its associated norm.

Definition 3.1.50

Let $p \in [1, +\infty)$.

Then

$$\|f\|_p := \left(\int_{\Omega} |f|^p d\mu \right)^{\frac{1}{p}}$$

is a norm in $L^p(\Omega, \mathfrak{M}, \mu)$.

The reader should check this definition satisfy all the property required to a norm, except for the triangular inequality property which coincides with the Minkowsky's theorem (see [3.1.56 on page 194](#)), which will be presented here.

Definition 3.1.51

We define:

$$\mathcal{L}^{\infty}(\Omega, \mathfrak{M}, \mu) := \{f : \Omega \rightarrow \mathbb{R} \text{ measurable} : \text{ess sup } f < +\infty\};$$

and with the same equivalence relation of the previous definition we set:

$$L^{\infty}(\Omega, \mathfrak{M}, \mu) := \frac{\mathcal{L}^{\infty}(\Omega, \mathfrak{M}, \mu)}{\sim}.$$

Notice that also L^{∞} is a vector space on \mathbb{R} .

Definition 3.1.52

The following

$$\|f\|_{\infty} := \text{ess sup}_{\Omega} (|f|)$$

is a norm in $L^{\infty}(\Omega, \mathfrak{M}, \mu)$.

Then we have for $p \in [1, \infty]$ that $(L^p(\Omega, \mathfrak{M}, \mu), \|\cdot\|_p)$ are normed vector spaces. Notice that L^p with those norms are not normed vector spaces as $\|f\| = 0$ implies that $f = 0$ except for a null measure set, while the definition requires that this holds anywhere in Ω .

From those norm it's easy to obtain a notion of metric for each p , called d_p . We have already introduced those in the chapter about topology ([1.2 on page 34](#)). With L^p spaces, those distances defines metric spaces. However, \mathcal{L}^p aren't a metric spaces with respect to d_p as $d_p(f, g) = 0$ implies only that $f = g$ a.e. in Ω .

Relevant inequalities Here are presented three well-known inequalities which are extremely useful for the development of the theory. In order to

make the steps more clear, here we do not use the norm notation, but we added this convention at the end of each proposition.

Definition 3.1.53

Let $p \in (1, +\infty)$.

The number $q \in (1, +\infty)$ is called the **conjugate** of p if

$$\frac{1}{p} + \frac{1}{q} = 1.$$

By extension, the conjugate of $p = 1$ is $q = \infty$ and of $p = \infty$ is $q = 1$.

Sometimes the conjugate of p is written as p^* .

Proposition 3.1.54 (Young's inequality)

For all $a, b > 0$ we have:

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q},$$

where $p \in (1, \infty)$ and q is its conjugate index.

Proof. Fix $b > 0$ and take $\phi(a) := \frac{a^p}{p} + \frac{b^q}{q} - ab$ on $(0, \infty)$.

Proof follows from the fact that ϕ is convex with a positive absolute minimum. ■

Proposition 3.1.55 (Hölder's inequality)

Consider two functions $f \in L^p(\Omega, \mathfrak{M}, \mu)$, $g \in L^q(\Omega, \mathfrak{M}, \mu)$, with $p \in (1, \infty)$ and q its conjugate index.

Then $f, g \in L^1(\Omega, \mathfrak{M}, \mu)$, and:

$$\int_{\Omega} |fg| \, d\mu \leq \left(\int_{\Omega} |f|^p \, d\mu \right)^{1/p} \left(\int_{\Omega} |g|^q \, d\mu \right)^{1/q}.$$

In terms of norms, this equality can be written as:

$$\|fg\|_1 \leq \|f\|_p \cdot \|g\|_q.$$

Proof. Suppose $p \in (1, +\infty)$ and $f, g > 0$ a.e. Consider:

$$F(t) = \frac{f(t)}{\left(\int_{\Omega} |f(t)|^p d\mu\right)^{1/p}} \quad \text{and} \quad G(t) = \frac{g(t)}{\left(\int_{\Omega} |g(t)|^q d\mu\right)^{1/q}}$$

Using Young's inequality (3.1.54 on the preceding page), we get:

$$\begin{aligned} |FG| &\leq \frac{1}{p}|F|^p + \frac{1}{q}|G|^q \\ \int_{\Omega} |FG| d\mu &\leq \frac{1}{p} \int_{\Omega} |F|^p d\mu + \frac{1}{q} \int_{\Omega} |G|^q d\mu \\ &\leq \frac{1}{p} \int_{\Omega} \left| \frac{f(t)}{\left(\int_{\Omega} |f(t)|^p d\mu\right)^{1/p}} \right|^p d\mu + \frac{1}{q} \int_{\Omega} \left| \frac{g(t)}{\left(\int_{\Omega} |g(t)|^q d\mu\right)^{1/q}} \right|^q d\mu \\ &= \frac{1}{p} + \frac{1}{q} = 1 \\ \int_{\Omega} |fg| d\mu &\leq \left(\int_{\Omega} |f|^p d\mu \right)^{1/p} \left(\int_{\Omega} |g|^q d\mu \right)^{1/q} \end{aligned}$$

When $p = 1, q = \infty$, the proof is trivial:

$$\int_{\Omega} |fg| \leq \int_{\Omega} |f| \operatorname{ess\,sup}_{\Omega} |g| d\mu = \operatorname{ess\,sup}_{\Omega} |g| \int_{\Omega} |f| d\mu$$

When $p = +\infty$:

$$\int_{\Omega} |fg| d\mu \leq \operatorname{ess\,sup}_{\Omega} |f| \int_{\Omega} |g| d\mu$$

■

Another proof consist of considering:

$$\|fg\| = \int_{\Omega} \frac{\|g\|_q^{\frac{1}{p}}}{\|f\|_q^{\frac{1}{q}}} f \cdot \frac{\|f\|_p^{\frac{1}{q}}}{\|g\|_q^{\frac{1}{p}}} g,$$

apply the young inequality and operate the proper calculations.

Proposition 3.1.56 (Minkowsky's inequality)

Let $f, g \in L^p(\Omega, \mathfrak{M}, \mu)$ and $p \in (1, \infty)$.

Then:

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

In terms of norms, this equality can be written as:

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p$$

Observe that this result shows the triangular inequality for the p norms.

Proof. The prove in case of $p = 1$ is trivial, here the general case:

$$\begin{aligned} \int_{\Omega} |f + g|^p d\mu &= \int_{\Omega} |f + g| |f + g|^{p-1} d\mu \\ &\leq \int_{\Omega} |f| |f + g|^{p-1} d\mu + \int_{\Omega} |g| |f + g|^{p-1} d\mu \\ &\leq \left(\int_{\Omega} |f|^p d\mu \right)^{\frac{1}{p}} \left(\int_{\Omega} |f + g|^p d\mu \right)^{\frac{p-1}{p}} \\ &\quad + \left(\int_{\Omega} |g|^p d\mu \right)^{\frac{1}{p}} \left(\int_{\Omega} |f + g|^p d\mu \right)^{\frac{p-1}{p}} \\ &= \left[\left(\int_{\Omega} |f|^p d\mu \right)^{\frac{1}{p}} + \left(\int_{\Omega} |g|^p d\mu \right)^{\frac{1}{p}} \right] \left(\int_{\Omega} |f + g|^p d\mu \right)^{\frac{p-1}{p}}. \end{aligned}$$

As $f + g \in L^p(\Omega, \mathfrak{M}, \mu)$ the thesis follows. ■

Relationship between L^p spaces Now we try to understand how those spaces are related and some of their properties. The L^p spaces, if μ is finite, make up an ascending chain (see definition 1.1.49 on page 32) with respect to the inclusion. The greater is p , the less functions belongs. Indeed we have the following result:

Proposition 3.1.57

If $\mu(\Omega) < +\infty$ and $1 \leq s \leq r \leq \infty$, we have:

$$L^r(\Omega, \mathfrak{M}, \mu) \subseteq L^s(\Omega, \mathfrak{M}, \mu)$$

and the norm can be controlled:

$$\|f\|_s \leq (\mu(\Omega))^{\frac{1}{s} - \frac{1}{r}} \|f\|_r.$$

Proof. Suppose $r \leq \infty$ and take $f \in L^r(\Omega, \mathfrak{M}, \mu)$ and apply the Hölder's

inequality with $q = \frac{r}{s}$ and $p = \frac{r}{r-s}$:

$$\|f\|_s^s = \int_{\Omega} |f|^s d\mu = \int_{\Omega} 1 \cdot |f|^s \leq \left(\int_{\Omega} d\mu \right)^{\left(\frac{r-s}{r}\right)} \left(\int_{\Omega} |f|^{s \cdot \frac{r}{s}} d\mu \right)^{\frac{s}{r}} = \mu(\Omega)^{\left(\frac{r-s}{r}\right)} \|f\|_r^s$$

where and the thesis follows. In case of $r = \infty$, we could repeat the same argument, but there is a simpler way:

$$\int_{\Omega} |f|^s d\mu \leq \int_{\Omega} \operatorname{ess\,sup}_{\Omega} (|f|^s) d\mu \leq \int_{\Omega} (\operatorname{ess\,sup}_{\Omega} |f|)^s d\mu = \mu(\Omega) \|f\|_{\infty}^s.$$

■

In such case, L^1 contains all functions that belong to L^p for any p .

However, if $\mu(\Omega) = \infty$, then the former inclusion does not hold in general, consider the following example.

Example 3.1.58. Consider $(\mathbb{R}, \mathcal{L}(\mathbb{R}), \lambda)$ and let:

$$f_{\alpha}(x) := \frac{1}{x^{\alpha}} \mathbb{1}_{[1, \infty]} \quad \text{with } \alpha \in (0, 1).$$

So $f_{\alpha} \in L^r$ for any $r > \alpha^{-1}$, but $f_{\alpha} \notin L^s$ for any $s \in [1, \alpha^{-1}]$.

The l^p spaces The chain of inclusions can be descending in a very special case. Indeed, we can define

$$l^p := L^p(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu_c)$$

and the following proposition holds:

Proposition 3.1.59

For $1 \leq s \leq r \leq \infty$ we have:

$$l^s \subseteq l^r, \quad \|f\|_r \leq \|f\|_s.$$

The inclusion is strict if $s \neq r$.

Proof. We recall that l^p spaces are correspondent to $L^p(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu_c)$, with the counting measure.

Step 1: Let us prove that $l^p \subsetneq l^{\infty}$ for any $p \in [1, +\infty)$.

Consider $x \in l^p$, then⁴³

$$\sum_{n \in \mathbb{N}} |x_n|^p < +\infty \implies |x_n| \rightarrow 0 \text{ as } n \rightarrow +\infty$$

In particular, $\{x_n\}_{n \in \mathbb{N}}$ is bounded, so that $\sup_{n \in \mathbb{N}} |x_n| < +\infty$, hence $x \in l^\infty$. Moreover $l^p \neq l^\infty$, indeed consider the sequence $x = \{x_n\}_{n \in \mathbb{N}}$ defined as $x_n = 1$ for any $n \in \mathbb{N}$, which belongs to l^∞ (its norm is 1) but not to l^p , since $\sum_{n \in \mathbb{N}} 1^p$ diverges.

Step 2: Let us prove that $l^r \subset l^s$ for any $1 \leq r < s < +\infty$.

Consider $x \in l^r$, we can assume $x \neq 0$ (the sequence with all zeros). Define now

$$y = \{y_n\}_{n \in \mathbb{N}}, \quad y_n = \frac{x_n}{\|x\|_r}.$$

Observe that $y \in l^r$, indeed l^r is a vector space, and

$$\|y\|_r = \left\| \frac{x}{\|x\|_r} \right\|_r = \frac{1}{\|x\|_r} \|x\|_r = 1$$

This implies that

$$|y_n| \leq 1, \quad \forall n \in \mathbb{N} \implies |y_n|^s \leq |y_n|^r \implies \sum_{n \in \mathbb{N}} |y_n|^s \leq \sum_{n \in \mathbb{N}} |y_n|^r < +\infty,$$

hence $y \in l^s$ and $x = y \cdot \|x\|_r \in l^s$ (also l^s is a vector space).

Moreover, $l^r \neq l^s$, indeed consider

$$x = \{x_n\}_{n \in \mathbb{N}}, \quad x_n = \frac{1}{n^{1/r}}$$

then

$$\begin{aligned} \sum_{n \in \mathbb{N}} |x_n|^r &= \sum_{n \in \mathbb{N}} \frac{1}{n} = +\infty \implies x \notin l^r \\ \sum_{n \in \mathbb{N}} |x_n|^s &= \sum_{n \in \mathbb{N}} \frac{1}{n^{s/r}} < +\infty \implies x \in l^s \end{aligned}$$

■

Remember that a function defined on \mathbb{N} can be seen as a sequence.

⁴³This very first point is what can't be done in L^p spaces, namely if $f \in L^p$ then not necessarily $f(x) \rightarrow 0$ as $x \rightarrow \infty$. Indeed take $p = 1$ and consider $f(x) = \mathbf{1}_{\{n, n+1/n^2\}}(x)$ where $n \in \mathbb{N}$. You can easily check that its L^1 norm converges ($\sum_{n \in \mathbb{N}} 1/n^2 < +\infty$), however its limit as $x \rightarrow \infty$ is not 0.

L^p spaces are Banach spaces The following is a very relevant result.

Theorem 3.1.60

$(L^p(\Omega, \mathfrak{M}, \mu), \|\cdot\|_p)$ is a Banach space for any $p \in [1, +\infty]$.

Proof. Case $p \in [1, +\infty)$:

We are going to use the characterization of Banach spaces through the convergence of series (see theorem 3.1.41 on page 185): a series converges only if converges the series of absolute values.

Step 0: Let $\{f_n\}_{n \in \mathbb{N}} \subset L^p$ be such that $\sum_{n \in \mathbb{N}} \|f_n\|_p < +\infty$.

We need to prove that $\sum_{n \in \mathbb{N}} f_n$ converges in $L^p(\Omega, \mathfrak{M}, \mu)$, from the characterization we have the thesis.

Step 1, convergence in Ω : Set $g_h(t) = \sum_{n=0}^h |f_n(t)|$; from Minkowsky's inequality we have that $g_h \in L^p$ (see proposition 3.1.56 on page 194):

$$\|g_h\|_{L^p} \leq \sum_{n=0}^h \|f_n\|_p \leq \sum_{n \in \mathbb{N}} \|f_n\|_p < +\infty \quad \forall h \in \mathbb{N}.$$

Since g_h is an increasing sequence, by monotone convergence (see theorem 2.2.5 on page 109) we have:

$$\lim_{h \rightarrow +\infty} \int_{\Omega} |g_h(t)|^p d\mu = \int_{\Omega} \lim_{h \rightarrow +\infty} |g_h(t)|^p d\mu = \int_{\Omega} g(t)^p d\mu$$

where $g(t) = \lim_{h \rightarrow +\infty} g_h(t)$ and, as its norm is bounded, $g \in L^p$ (the absolute value can be removed as g is positive by definition).

Thus $|g(t)| < +\infty$ for almost any t in Ω .

Therefore we have

$$\sum_{n \in \mathbb{N}} f_n(t) < +\infty \quad \text{for almost any } t,$$

the series converges almost everywhere in Ω .

Step 2, convergence in L^p : Set now:

$$S(t) = \sum_{n \in \mathbb{N}} f_n(t) \quad \text{and} \quad S_h(t) = \sum_{n=0}^h f_n(t).$$

For almost any $t \in \Omega$, using the triangular inequality, we have:

$$|S(t) - S_h(t)|^p = \left| \sum_{n=h+1}^{+\infty} f_n(t) \right|^p \leq \left(\sum_{n \in \mathbb{N}} |f_n(t)| \right)^p = (g(t))^p \in L^1.$$

Then, using dominated convergence (see theorem 2.3.2 on page 119), we have $\|S - S_h\|_p \rightarrow 0$ as $h \rightarrow +\infty$, that is $\sum_{n \in \mathbb{N}} f_n$ converges in L^p .

By the characterization theorem, the proof of the considered case is complete.

Case $p = \infty$, step 1:

Let $\{f_n\}_{n \in \mathbb{N}} \subset L^\infty(\Omega, \mathfrak{M}, \mu)$ be a Cauchy sequence, and take those two sets:

$$\begin{aligned} A_n &= \{t \in \Omega : |f_n(t)| > \|f_n\|_\infty\} \\ B_{m,n} &= \{t \in \Omega : |f_n(t) - f_m(t)| > \|f_n - f_m\|_\infty\} \end{aligned}$$

where the inequalities are given by definition of the essential supremum. Notice that we defined those two sequences of sets which catch discontinuity points in the domain, for each f_n .

We have

$$\mu(A_n) = \mu(B_{m,n}) = 0 \quad \forall m, n \in \mathbb{N}, \quad \text{and} \quad E = \left(\bigcup_{n \in \mathbb{N}} A_n \right) \cup \left(\bigcup_{m, n \in \mathbb{N}} B_{m,n} \right)$$

has zero measure.

Then, for all $t \in E^c$, we have that the sequence $\{f_n(t)\}_{n \in \mathbb{N}}$ converges to some measurable f . Moreover $f \in L^\infty(\Omega, \mathfrak{M}, \mu)$ since $\{f_n(t)\}_{n \in \mathbb{N}}$ is bounded in $L^\infty(\Omega, \mathfrak{M}, \mu)$, as it is a fundamental sequence.

Step 2: On the other hand, for any $\varepsilon > 0$ there is $n_0 = n_0(\varepsilon) \in \mathbb{N}$ such that, for all $t \in E^c$, we have:

$$|f_n(t) - f_m(t)| \leq \|f_n - f_m\|_\infty < \varepsilon \quad \forall m, n > n_0.$$

As $m \rightarrow \infty$ we have:

$$|f_n(t) - f| \leq \varepsilon \quad \forall n > n_0, \quad \forall t \in E^c$$

from which we deduce that $\|f - f_n\|_\infty \rightarrow 0$ as $n \rightarrow \infty$.

And the proof is complete. ■

The case $p \in (0, 1)$ When L^p spaces have p less than 1 things are not going straight. Those spaces does not have a vector space structure and is not possible to define a norm on them; but it is feasible to prove that $d_p(f, g) = \int_{\Omega} |f - g|^p d\mu$ is a metric and the corresponding $L^p(\Omega, \mathfrak{M}, \mu)$ is complete with respect to d_p .

However, the reverse Minkowsky's inequality holds:

$$\left(\int_{\Omega} (|f| + |g|)^p d\mu \right)^{\frac{1}{p}} \geq \left(\int_{\Omega} |f|^p d\mu \right)^{\frac{1}{p}} + \left(\int_{\Omega} |g|^p d\mu \right)^{\frac{1}{p}}.$$

Thus $(\int_{\Omega} |f|^p d\mu)^{\frac{1}{p}}$ is not a norm, because it does not respect the triangular inequality. Anyway, inclusion chain follows the same rules as for $p > 1$; if $\mu(\Omega) < \infty$ then we have:

$$L^r(\Omega, \mathfrak{M}, \mu) \subseteq L^s(\Omega, \mathfrak{M}, \mu)$$

for $0 < s \leq r \leq \infty$.

3.1.5 Convergence and separability in L^p spaces

Convergence Here some results on sequence of function convergence for those spaces.

Proposition 3.1.61

Let $p \in [1, +\infty)$, $\{f_n\} \subset L^p(\Omega, \mathfrak{M}, \mu)$ be a L^p -converging sequence. Then $\{f_n\}$ contains a subsequence which converges a.e. in Ω .

Proof. The case $p = \infty$ is obvious while the case $p = 1$ is already proved (see 2.3.34 on page 139). Let's argue for the other cases. As $\{f_n\}_{n \in \mathbb{N}} \rightarrow f$, we can apply the Chebyshev's inequality. Consider $\delta > 0$, we have:

$$\begin{aligned} \int_{\Omega} |f_n - f|^p d\mu &\geq \int_{\{t \in \Omega : |f_n - f| > \delta\}} |f_n(t) - f(t)|^p d\mu \\ &\geq \delta^p \mu(\{t \in \Omega : |f_n - f| > \delta\}). \end{aligned}$$

This shows that $\{f_n\}_{n \in \mathbb{N}}$ converges in measures, and the thesis follows (see theorem 2.3.35 on page 139). ■

Moreover we have the following result:

Proposition 3.1.62

Let $\{f_n\} \subset L^p(\Omega, \mathfrak{M}, \mu)$ with $p \in [1, \infty)$.
 If $f_n \rightarrow f$ in L^p and $f_n \rightarrow g$ a.e. in Ω , then $f = g$ a.e. in Ω .

Proof.

$$\begin{aligned} \int_{\Omega} |f - g|^p \, d\mu &= \int_{\Omega} \lim_{n \rightarrow +\infty} |f - f_n|^p \, d\mu \\ &\leq \liminf_{n \rightarrow +\infty} \int_{\Omega} |f - f_n|^p \, d\mu \quad (\text{Fatou's lemma}) \\ &= 0 \end{aligned}$$

Thus $f = g$ a.e. in Ω . ■

A compactness criterion in L^p spaces Here we present a historically relevant result. First we introduce the notion of the shift operator.

Definition 3.1.63

Given a function $f : \mathbb{R} \rightarrow E$ and a shift value $h \in \mathbb{R}$, the **shift operator** is defined as follows:

$$\tau_h f(x) := f(x + h).$$

The shift operator essentially evaluate the function after changing its argument, increasing or decreasing it.

Theorem 3.1.64 (Kolmogorov–Riesz–Fréchet)

Let $p \in [1, +\infty)$, and a bounded set $\mathcal{F} \subset L^p(\mathbb{R}^N, \mathcal{L}(\mathbb{R}^N), \lambda)$.

If^a $\forall \varepsilon > 0 \exists \delta = \delta(\varepsilon) > 0$ such that:

$$\|\tau_h f - f\|_p < \varepsilon \quad \forall f \in \mathcal{F} \quad \forall h \in \mathbb{R}^N : \|h\|_2 < \delta,$$

for all $h \in \mathbb{R}^N$ such that $|h| < \delta$ and any $f \in \mathcal{F}$, then $\mathcal{F}|_{\Omega}$ is precompact^b in $L^p(\mathbb{R}^N, \mathcal{L}(\mathbb{R}^N), \lambda)$ for any $\Omega \subset \mathbb{R}^N$ of finite measure.

Viceversa, if $E|_{\Omega}$ is precompact the above condition holds along with the *tail condition*: for any $\varepsilon > 0$ there exists $\Omega \subset \mathbb{R}^N$ bounded and measurable such that L^p -norm of f on Ω^C is less then ε for all $f \in E$.

^aThis hypothesis can be equivalently formulated as follows: $\|\tau_h f - f\| \rightarrow 0$ as $\|h\|_2 \rightarrow 0$ uniformly in \mathcal{F} .

^bA precompact set, also known as relatively compact subspace, is a subset whose closure is compact.

Separability of $L^p(\mathbb{R})$ Here we discuss of L^p spaces defined on \mathbb{R} . Our goal now is to prove the following theorem (recall definition 1.2.58 on page 58); to do that we need some other result.

Theorem 3.1.65

$L^p(\mathbb{R}^N, \mathcal{L}(\mathbb{R}^N), \lambda)$ is separable for all $p \in [1, +\infty)$.

Observe that $p = \infty$ is excluded, we will face this case later. The proof will handle only the case $N = 1$.

The pathway to prove this result will be the following.

0. $\mathcal{P}_{\mathbb{Q}}([a, b])$ is dense in $\mathcal{C}([a, b])$ with the infinity norm (Stone–Weierstrass theorem).
1. Approximation of finite supported functions with $g \in \mathcal{C}_C(\mathbb{R})$ (Lusin theorem).
2. $\mathcal{S}(\mathbb{R})$ dense in $L^p(\mathbb{R})$ but not countable.
3. $\mathcal{C}_C(\mathbb{R})$ dense in $L^p(\mathbb{R})$ but not countable.
4. $\{P_m \mathbf{1}_{[-n, n]}\}$, the set of algebraic polynomials with restricted support, dense in $L^p(\mathbb{R})$ and countable.

Meanwhile, notice the following, which is a weaker result highlighting the importance of topology:

Proposition 3.1.66

Let $\Omega \in \mathcal{L}(\mathbb{R})$ and take $E := \{v \in L^p(\mathbb{R}) : v = 0 \text{ a.e. in } \Omega^c\}$.

Then E is separable for any $p \in [1, \infty)$ with respect to the inherited topology, that is the topology on E created by intersecting the open sets of $L^p(\mathbb{R})$ with E itself.

Continuous function with compact support Let's build the tools that we will use for the proof.

Definition 3.1.67

Let $g = \Omega \subseteq \mathbb{R} \rightarrow \mathbb{R}$ be continuous.

The following set is called the **support** of g :

$$\text{supp}(g) := \overline{\{x \in \Omega : g(x) \neq 0\}}.$$

In other words, the support is the subset of the domain of a function in which it has a value different from zero.

Example 3.1.68. Let:

$$g(x) := \begin{cases} e^{-\frac{1}{1-x^2}} & \text{if } |x| < 1 \\ 0 & \text{if } |x| \geq 1 \end{cases}$$

In this case $\text{supp}(g) = [-1, 1]$.

Support is hence a set, and we know that there exist many kind of sets. In particular we are interested in the case of such set is compact.

Definition 3.1.69

Let $\Omega \subseteq \mathbb{R}$.

$$\mathcal{C}_C(\Omega) := \{f : \Omega \rightarrow \mathbb{R} \text{ continuous with compact support}\}.$$

We have define the set of continuous functions with compact support; the following result states that any Lebesgue measurable function can be approximated with one of those functions (similar to Stone–Weierstrass [3.1.27](#) on page 180).

Theorem 3.1.70 (Lusin)

Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be Lebesgue measurable and let $\Omega \in \mathcal{L}(\mathbb{R})$ such that $\lambda(\Omega) < +\infty$.

If $f(x) = 0$ for any $\forall x \in \Omega^C$, then for all $\varepsilon > 0$ exists a function $g \in \mathcal{C}_C(\mathbb{R})$ such that:

$$\lambda(\{x \in \mathbb{R} : f(x) \neq g(x)\}) < \varepsilon,$$

and one can always choose g such that:

$$\sup_{\mathbb{R}}(g) \leq \sup_{\mathbb{R}}(f).^{44}$$

Density of simple functions with finite measure support Here we define another set of functions.

Definition 3.1.71

Let $\mathcal{S}(\mathbb{R})$ be the **set of simple functions with finite measure support**, namely:

$$\mathcal{S} := \{s : \mathbb{R} \rightarrow \mathbb{R} \text{ simple function such that } \lambda(\{t \in \mathbb{R}, s(t) \neq 0\}) < +\infty\}.$$

Observe that, if $s \in \mathcal{S}(\mathbb{R})$ then it necessarily takes non-zero values on finite measure sets only. Therefore $s \in L^p(\mathbb{R})$ for all $p \in [1, \infty)$ since its range is finite.

Theorem 3.1.72

The set of simple functions with finite measure support $\mathcal{S}(\mathbb{R})$ is dense in $L^p(\mathbb{R})$ for all $p \in [1, \infty)$.

Proof. Observe first that $\mathcal{S}(\mathbb{R}) \subset L^p(\mathbb{R})$.

Let $f \in L^p$ non-negative. We already know that there exists a sequence of simple functions $\{s_n\}_{n \in \mathbb{N}}$ such that $0 \leq s_n \leq f$ and $s_n \uparrow f$ a.e. in \mathbb{R} as $n \rightarrow +\infty$.

Since $f \in L^p$, $s_n \in L^p$. For s_n to be both piece-wise constant and integrable, it must be zero as $t \rightarrow \pm\infty$, and thus $s_n \in \mathcal{S}$.

Observing that $0 \leq |f - s_n|^p \leq |f|^p$ a.e. in \mathbb{R} , via dominated convergence we have that $\|f - s_n\|_p \rightarrow 0$ as $n \rightarrow +\infty$. ■

Density of $\mathcal{C}_C(\mathbb{R})$ Also the set of continuous function has some interesting density property.

Theorem 3.1.73

The set of continuous function with compact support $\mathcal{C}_C(\mathbb{R})$ is dense in $L^p(\mathbb{R})$ for all $p \in [1, \infty)$.

⁴⁴For further discussion and a proof, see: W. Rudin, Real and Complex Analysis, 1987, page 55, theorem 2.24.

Moreover consider $L^p(\mathbb{R})$ as a metric space with distance $d_p(f, g) = \|f - g\|_p$ is the completion of the metric space $(\mathcal{C}_C(\mathbb{R}), d_p)$ for each $p \in [1, \infty)$. If $p = 1$, this fact essentially says that the Lebesgue integral is the “natural” generalization of the Riemann integral.

Proof. Fix $\varepsilon > 0$ and consider $f \in L^p(\mathbb{R})$. Thanks to the density of $\mathcal{S}(\mathbb{R})$ we can find a simple function $s \in \mathcal{S}(\mathbb{R})$ such that:

$$\|f - s\|_p < \varepsilon.$$

Then, using Lusin’s theorem (3.1.70 on page 203), we can find a function $g \in \mathcal{C}_C(\mathbb{R})$ such that $\|g\|_\infty \leq \|s\|_\infty$ and

$$\lambda\{(x \in \mathbb{R} : g(x) \neq s(x))\} < \frac{\varepsilon^p}{2}.$$

Therefore,

$$\begin{aligned} \|g - s\|_p^p &= \int_{\mathbb{R}} |g - s|^p d\lambda = \int_E |g - s|^p d\lambda \\ &\leq \int_E (|g| + |s|)^p d\lambda \leq \lambda(E) \cdot 2 \|s\|_\infty^p < \frac{\varepsilon^p}{2} 2 \|s\|_\infty^p = (\varepsilon \|s\|_\infty)^p \end{aligned}$$

so we conclude:

$$\|f - g\|_p \leq \|f - s\|_p + \|g - s\|_p < \varepsilon + \varepsilon \|s\|_\infty$$

.

■

Proof of the separability of $L^p(\mathbb{R})$ Now we have all tools to proceed with this proof.

Proof of 3.1.65 on page 202, case $N = 1$ and $p \in [1, \infty)$. Let $f \in L^p(\mathbb{R})$ and fix $\varepsilon > 0$. Because $\mathcal{C}_C(\mathbb{R})$ is dense in $L^p(\mathbb{R})$, there exists $g \in \mathcal{C}_C(\mathbb{R})$ such that

$$\|g - f\|_p < \frac{\varepsilon}{2}.$$

Since $\text{supp}(g)$ is compact, and thus bounded, there exists $n \in \mathbb{N}_0$ such that $\text{supp}(g) \subseteq [-n, n]$. Moreover, there exists a polynomial P_m with rational coefficients such that:

$$\|g - P_m\|_\infty < \frac{\varepsilon}{2(2n)^{\frac{1}{p}}},$$

this because the set of polynomials, which is countable, contains any approximation of a continuous function by Stone–Weierstrass (see theorem 3.1.27 on page 180).

Thus we have (recall proposition on relationship between L^p spaces 3.1.57 on page 195):

$$\begin{aligned} \|f - P_m \mathbf{1}_{[-n,n]}\|_p &\leq \|f - g\|_p + \|g - P_m \mathbf{1}_{[-n,n]}\|_p \\ &\leq \frac{\varepsilon}{2} + \left(\int_{-n}^n |g - P_m \mathbf{1}_{[-n,n]}|^p d\lambda \right)^{\frac{1}{p}} \\ &< \frac{\varepsilon}{2} + \| [g - P_m \mathbf{1}_{[-n,n]}] \|_{\infty} (2n)^{\frac{1}{p}} \\ &< \varepsilon \end{aligned}$$

The set $\{P_m \mathbf{1}_{[-n,n]}\}$ is countable and dense in L^p , and then the thesis is proven. ■

Non-separability of $L^\infty(\mathbb{R})$ We concentrated to specific cases avoiding a too much large generality to avoid excessive complexity. This is the remaining case: the prove here is for the mono-dimensional case but it can easily extended.

Theorem 3.1.74

The space $L^\infty(\mathbb{R}^N, \mathcal{L}(\mathbb{R}^N), \lambda)$ is *not* separable for any N .

Proof of case $N = 1$. Consider the following uncountable set:

$$\{\mathbf{1}_{[-\alpha,\alpha]}\}_{\alpha>0} \subset L^\infty(\mathbb{R}).$$

Observe that $\|\mathbf{1}_{[-\alpha,\alpha]} - \mathbf{1}_{[-\beta,\beta]}\| = 1$ if $\alpha \neq \beta$.

Thus the following family of balls

$$B_\alpha = \left\{ f \in L^\infty(\mathbb{R}) : \|f - \mathbf{1}_{[-\alpha,\alpha]}\|_{\infty} < \frac{1}{2} \right\}$$

are mutually disjoint and uncountable.

Suppose now $L^\infty(\mathbb{R})$ is separable, that is there exists $E \subset L^\infty(\mathbb{R})$ which is countable and dense. Then any ball B_α must contain at least one element of E . The balls are disjoint, and thus every ball must contain a different

element of E . However E is countable and we have uncountably many balls, so we have a contradiction: we cannot find any countable dense set in $L^\infty(\mathbb{R})$. ■

Some final results The following remarks complete our discussion.

In general, $L^\infty(\Omega)$ is not separable when $\mu(\Omega) = \infty$.

The space l^p is separable if $p \in [1, \infty)$: l^∞ is not separable.

The metric space $(\mathcal{C}_C(\mathbb{R}), d_p)$ where $p \in [1, +\infty)$ and $d_p(f, g) = \|f - g\|_p$ is not complete, its completion is $(L^p(\mathbb{R}), d_p)$; the metric space $(\mathcal{C}_C(\mathbb{R}), d_\infty)$ is not complete as well, but its completion is $(\mathcal{C}_0(\mathbb{R}), d_\infty)$ where

$$\mathcal{C}_0(\mathbb{R}) := \{f \in \mathcal{C}(\mathbb{R}) : \lim_{|x| \rightarrow \infty} f(x) = 0\}.$$

The Banach space $(\mathbb{R}^N, \|\cdot\|_p)$ can be identified with an L^p space; indeed, set $E = \{1, \dots, N\}$ and consider $L^p(E, \mathcal{P}(E), \mu_c)$ with $p \in [1, \infty]$.

3.2 Linear operators

The goal here is to study functions between Banach spaces. In linear algebra we have already seen functions between finite spaces, but here the topic is to deal with the infinite dimensionality.

Functions of this kind are called **operators** or **maps**. An operator from a vector space to \mathbb{R} is called **functional**.⁴⁵

3.2.1 Definition of linear operator and boundedness

Throughout this section, let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be two normed vector spaces on \mathbb{R} . We'll use the compact notation " Tx " in place of " $T(x)$ ", if this does not imply contradictions.

Definition 3.2.1

A map $T : X \rightarrow Y$ is called **linear operator** if:

$$T(\alpha x_1 + \beta x_2) = \alpha T x_1 + \beta T x_2 \quad \forall \alpha, \beta \in \mathbb{R}, \quad \forall x_1, x_2 \in X.$$

Linearity immediately allows us to prove that $L(0) = 0$. Indeed, $L(\alpha f) = \alpha L(f)$, choose $\alpha = 0$.

Definition 3.2.2

We say that $T : X \rightarrow Y$ linear operator is **bounded** if there exists $M > 0$ such that:

$$\|Tx\|_Y \leq M \|x\|_X \quad \forall x \in X$$

or, equivalently, if it maps bounded sets into bounded sets.

Observe that, having not defined an "operator norm" yet, the boundedness relies only on the norm of the image set.

Here some examples of linear operators in different spaces.

Example 3.2.3. Consider the finite case $X = \mathbb{R}^N$, $Y = \mathbb{R}^m$.

The well-known representation theorem states that a linear operator $T : X \rightarrow Y$ can be represented by a matrix $A \in \mathbb{R}^{m \times n}$, which depends on the bases we chose for X and Y , such that $Tx = Ax^T$.

⁴⁵Italian translation: *funzionale*.

Apply the euclidean norm:

$$\|Tx\|_2 \leq M \|X\|_2 \quad \text{with} \quad M = \left(\sum_{j=1}^m \sum_{i=1}^n |a_{ij}^2| \right)^{\frac{1}{2}}.$$

Thus any linear T is bounded in any norm, as they are all equivalent.

Example 3.2.4. Consider $X = Y = L^2([0, 1])$. Fix $k \in L^2([0, 1]^2)$, and let:

$$Tu(x) := \int_0^1 k(x, y) u(y) dy.$$

This is known as the Hilbert–Schmidt operator: it is linear and bounded. We want to prove that $Tu \in L^2([0, 1])$ for any $u \in L^2([0, 1])$. Via the Hölder inequality, we get:

$$(Tu(x))^2 = \left(\int_0^1 k(x, y) u(y) dy \right)^2 \leq \left(\int_0^1 k(x, y)^2 dy \right) \left(\int_0^1 u(y)^2 dy \right).$$

Integrating both sides in x we get:

$$\int_0^1 (Tu(x))^2 dx \leq \|k\|_{L^2([0,1]^2)}^2 \|u\|_{L^2([0,1])}^2,$$

where, defining $M = \|k\|_{L^2([0,1]^2)}^2$, we have:

$$\|Tu\|_{L^2([0,1])}^2 \leq M \|u\|_{L^2([0,1])}^2.$$

Then T is linear and bounded from $L^2([0, 1])$ into itself.

Example 3.2.5. Consider $X = L^p(\Omega, \mathfrak{M}, \mu)$, $Y = L^q(\Omega, \mathfrak{M}, \mu)$, with p fixed and conjugate of q ; consider also $g \in Y$.

Then

$$T_g : f \rightarrow T_g(f) = \int_{\Omega} fg d\mu$$

is a linear functional, and it is bounded via the Hölder inequality: $|T_g(f)| \leq \int_{\Omega} |fg| d\mu \leq \|f\|_p \|g\|_q$.

Boundedness is equivalent to continuity See this first result.

Theorem 3.2.6

Let $T : X \rightarrow Y$ be a linear operator. The following statements are equivalent:

- T is bounded;
- T is Lipschitz continuous in X ;
- T is continuous at $x = 0$.

Proof. We will prove the pairwise implications.

A bounded linear operator is continuous in its domain.

Let $x_0 \in X$, $\{x_n\}_{n \in \mathbb{N}} \subset X$ such that $x_n \rightarrow x_0$ in X . Then:

$$\|Tx_n - Tx_0\|_Y = \|T(x_n - x_0)\|_Y \leq M \|x_n - x_0\|_X \rightarrow 0 \quad \text{as } n \rightarrow +\infty$$

Thus $Tx_n \rightarrow Tx_0$ in Y .

A linear operator continuous in its domain is continuous in $x = 0$.

This point is trivial.

A linear operator continuous in $x = 0$ is bounded.

Suppose, by contradiction, that:

$$\forall n \in \mathbb{N}_0 \quad \exists x_n \in X \setminus \{0\} : \|Tx_n\|_Y \geq n \|x_n\|_X .$$

Then take $z_n = \frac{x_n}{n \|x_n\|_X}$. Observe that $z_n \rightarrow 0$ in X . Moreover:

$$Tz_n = T\left(\frac{x_n}{n \|x_n\|_X}\right) = \frac{1}{n \|x_n\|_X} Tx_n$$

and

$$\|Tz_n\|_Y = \frac{1}{n \|x_n\|_X} \|Tx_n\|_Y \geq 1.$$

Notice that, for any linear operator T , we have $T(0) = 0$. However, $Tz_n \not\rightarrow T(0) = 0$, and thus T is not continuous. ■

Then, as when an operator is continuous in 0 then is continuous everywhere, the point 0 can be changed with any other point in this theorem: this, as we have seen in the proof, is possible due to linearity.

Example 3.2.7. Consider the normed vector space $X = (C^1([-1, 1]), \|\cdot\|_\infty)$, which is not a Banach space, and consider a linear functional defined as follows:

$$Tf = f'(0).$$

Let $f_n(t) = \frac{\sin(nt)}{n}$ for $n \in \mathbb{N}_0$: we have $\|f_n\|_\infty \rightarrow 0$ as $n \rightarrow \infty$.

However, $Tf_n = 1$ for any $n \in \mathbb{N}_0$, thus $Tf_n \not\rightarrow T(0) = 0$ and so T is not continuous at 0, and it is *not* bounded.

In this example X is not a Banach space: finding an example of linear unbounded operator between infinite dimensional Banach spaces, even a functional, is technically very difficult: the Axiom of Choice is required.

Sets of linear operators and norms Here we define the sets which contains linear operators.

Definition 3.2.8

Let X and Y two real vector spaces. Then the **set of linear operators** from X to Y is:

$$\mathcal{L}(X, Y) = \{T : X \rightarrow Y : T \text{ linear operator}\}.$$

On the same setting, the **set of linear bounded operators** is:

$$\mathcal{B}(X, Y) = \{T : X \rightarrow Y : T \text{ linear bounded operator}\}.$$

Notice that, if X is infinite-dimensional, then $\mathcal{B}(X, Y) \subsetneq \mathcal{L}(X, Y)$.

Both of them are vector spaces:

Proposition 3.2.9

The spaces $\mathcal{L}(X, Y)$ and $\mathcal{B}(X, Y)$ are vector space with respect to the standard operations.

Now, recalling that boundedness is defined on the norm of the image space, we define a norm for bounded operators:

Definition 3.2.10

The norm for \mathcal{B} is:

$$\|T\|_{\mathcal{B}(X,Y)} := \sup_{\|x\|_X \leq 1} \|Tx\|_Y.$$

If $T \in \mathcal{B}(X, Y)$ then $\|Tx\|_Y \leq M$ for all $x \in X$ such that $\|x\|_X \leq 1$. Then follows that

$$\sup_{\|x\|_X \leq 1} \|Tx\|_Y$$

is finite (the reader should prove here that this is a norm, do it!).

So, with the norm previously defined, $\mathcal{B}(X, Y)$ is also a normed vector space.

This norm can be rewritten in many other forms, providing different interpretations, like specified by the following.

Proposition 3.2.11

These equalities hold:

$$\|T\|_{\mathcal{B}(X,Y)} := \sup_{\|x\|_X \leq 1} \|Tx\|_Y = \sup_{\|x\|_X = 1} \|Tx\|_Y = \sup_{x \in X} \frac{\|Tx\|_Y}{\|x\|_X}.$$

Where the first alternative differs from the definition by the set they pick x .

Proof. First, as $\{x : \|x\|_X \leq 1\} \supseteq \{x : \|x\|_X = 1\}$, we have:

$$\sup_{\|x\|_X \leq 1} \|Tx\|_Y \geq \sup_{\|x\|_X = 1} \|Tx\|_Y.$$

On the other hand, if $\|x\|_X \leq 1$, then we have, if $x \neq 0$,

$$\|Tx\|_Y = \|x\|_X \left\| T \left(\frac{x}{\|x\|_X} \right) \right\|_Y \leq \left\| T \left(\frac{x}{\|x\|_X} \right) \right\|_Y = \frac{1}{\|x\|_X} \|Tx\|_Y.$$

We want to maximize the LHS, which is the same as minimizing the RHS, which can be done by taking $\|x\|_X$ as biggest as possible ($\|x\|_X = 1$). Hence

$$\sup_{\|x\|_X \leq 1} \|Tx\|_Y = \sup_{\|x\|_X = 1} \|Tx\|_Y.$$

The remaining equality is obvious for $x \neq 0$, since

$$\frac{\|Tx\|_Y}{\|x\|_X} = \left\| T \left(\frac{x}{\|x\|_X} \right) \right\|_Y.$$



Having defined those norms an observation is required; the operator norm is defined from the norm of the image space. Then, boundedness of the operator can depend on the norm of the image space.

Dual spaces In next chapters we will deeply develop the concept of duality. We present here this definition for a complete dissertation of the matter.

Definition 3.2.12

Consider X vector space and $Y = \mathbb{R}$.

The space $X' := \mathcal{L}(X, \mathbb{R})$ is said to be the **algebraic dual space** of X .

The space $X^* := \mathcal{B}(X, \mathbb{R})$ is said to be the **topological dual space** of X .

From now on in this book, when we write “dual space” without further specifications, we are referring to *topological* dual space.

Recalling that operators which have \mathbb{R} as image set are called functional, notice that here we defined the sets of linear functionals and bounded functionals respectively.

The set of bounded linear operators can be a Banach space Consider the following theorem.

Theorem 3.2.13

If Y is a Banach space, then $\mathcal{B}(X, Y)$ is a Banach space.

Observe that the condition is on the image space, similarly to the boundedness that depends on the norm of the image space, also completeness relies on the same.

Proof. Let $\{T_n\}_{n \in \mathbb{N}} \subset \mathcal{B}(X, Y)$ be a fundamental sequence.

Then $\{T_n x\}_{n \in \mathbb{N}_0}$ is also a fundamental sequence in Y so it converges to some $y \in Y$ for each $x \in X$.

Set now $Tx = y$; the operator T is clearly linear; we have to prove that T is bounded.

Fixed some $\varepsilon > 0$, there exists $\bar{n} = \bar{n}(\varepsilon) \in \mathbb{N}$ such that:

$$\|T_n x - T_m x\|_Y \leq \|T_n - T_m\|_{\mathcal{B}(X,Y)} \|x\|_X < \varepsilon \|x\|_X \quad \forall n, m \geq \bar{n} \quad \forall x \in X.$$

Now let n go to $+\infty$ and we have:

$$\|Tx - T_m x\| < \varepsilon \|x\|_X \quad \forall m \geq n \quad \forall x \in X.$$

And so

$$\|Tx\|_Y \leq \|Tx - T_m x\|_Y + \|T_m x\|_Y \leq (\varepsilon + M) \|x\|_X \quad \forall x \in X$$

as $\{T_m\}_{m \in \mathbb{N}}$ is bounded by some $M > 0$.

Therefore $T \in \mathcal{B}(X, Y)$ and

$$\|T - T_m\|_{\mathcal{B}(X,Y)} \leq \varepsilon \quad \forall m > n_0.$$

■

Now observe that, since \mathbb{R} with p norm is Banach, then for any X the space $\mathcal{B}(X, \mathbb{R})$ is Banach as well.

3.2.2 Isomorphism

The notion of isomorphism is widely used in mathematics. When we have two analogous structures then a bijection between them which preserves their operation is an isomorphism. Here we discuss the linear isomorphism of normed vector spaces through linear operators and its implications. In this case we have to preserve the operation and the continuity.

Linear isomorphism Here we analyze the intertwining of three properties of an operator, that are linearity, boundedness and invertibility. If an operator satisfy those properties then it preserve the structure of the spaces on which it works. Then we have an isomorphism.

Definition 3.2.14

Let X and Y two normed vector spaces.

If there exists $T \in \mathcal{B}(X, Y)$, such that $T^{-1} \in \mathcal{B}(Y, X)$, then X and Y are **isomorphic** and T is a **linear isomorphism**.

Later we will prove a results which states that any bijective operator in $\mathcal{B}(Y, X)$ has those properties and it is a linear isomorphism (see bounded inverse mapping [3.2.43 on page 224](#)).

Theorem 3.2.15

Every real normed vector space X on \mathbb{R} of dimension N is linearly isomorphic to \mathbb{R}^N with any given norm.

Corollary 3.2.16

Let X be a normed vector space on \mathbb{R} .
Then any finite dimension subspace of X is closed in X .

The proof of these last two results can be a good exercise for the reader. You should do this!

Image and kernel Consider the following definition.

Definition 3.2.17

Let $T \in \mathcal{B}(X, Y)$. We define the following sets:

$\text{Im}(T) := \{y \in Y : \exists x \in X, Tx = y\} \subset Y$, as the **image** of T and
 $\text{Ker}(T) := \{x \in X : Tx = 0\} \subset X$, as the **kernel** of T .

Those are subspaces of Y and X , respectively.

Proposition 3.2.18

If a linear operator is bounded, then its kernel is closed.

Try to prove this!

Proposition 3.2.19

The operator $T \in \mathcal{L}(X, Y)$ is injective if and only if $\text{Ker}(T) = \{0\}$.

Isometries Consider the following definition. Observe that here we have no request on boundedness.

Definition 3.2.20

We say that $T \in \mathcal{L}(X, Y)$ is an **isometry** if:

$$\|Tx\|_Y = \|x\|_X \quad \forall x \in X.$$

Observe that if T is an isometry then is also bounded:

Proposition 3.2.21

If T is an isometry, then $\|T\|_{\mathcal{B}(X,Y)} = 1$ and T is injective. Moreover, if X is a Banach space, $\text{Im}(T)$ is closed.

Example 3.2.22. Consider $Z = L^p(\Omega, \mathfrak{M}, \mu)$ for $p \in (1, \infty)$.

Take $g \in X = L^q(\Omega, \mathfrak{M}, \mu)$ where q is the conjugate of p and define

$$L_g(f) = \int_{\Omega} fg \, d\mu \quad \forall f \in Z.$$

It's easy to check that $L_g \in Y = Z^*$. The linear map $T : X \rightarrow Y$ defined by $T(g) = L_g$ is an isometry, indeed, we have: $|L_g(f)| \leq \|f\|_Z \|g\|_X$ and, via Hölder inequality we get:

$$\|L_g\|_{X^*} = \sup_{\|f\|_Z=1} \left| \int_{\Omega} fg \, d\mu \right| \leq \|g\|_X.$$

Moreover, if we choose

$$\tilde{f} = \frac{|g|^{q-2}g}{\|g\|_X^{q-1}},$$

provided that $g \neq 0$ and $p \neq 1$, we have $\tilde{f} \in Z$ and⁴⁶

$$T(g)(\tilde{f}) = L_g(\tilde{f}) = \|g\|_X \text{ implies } \|T(g)\|_Y = \|g\|_X.$$

Continuous injections Consider the following definition:

Definition 3.2.23

Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed vector spaces, and $X \subset Y$.

Define:

$$J : X \rightarrow Y, \quad Jx = x \quad \forall x \in X.$$

If

$$x_n \xrightarrow{X} x \quad \text{then} \quad x_n \xrightarrow{Y} x,$$

then the map J is called **continuous injection** from X to Y , and it is denoted by

$$X \hookrightarrow Y.$$

This definition holds even when the norms are different.

⁴⁶Try this!

Example 3.2.24. Let $\Omega \in \mathcal{L}(\mathbb{R}^N)$, with $\lambda(\Omega) < +\infty$.

$$L^r(\Omega) \hookrightarrow L^s(\Omega) \quad \forall r \geq s : r, s \in (1, +\infty]$$

Indeed, from $\|f\|_x \leq C_{\lambda(\Omega)} \|f\|_\Omega$, we have that $f_n \xrightarrow{L^r} f \implies f_n \xrightarrow{L^s} f$.

Example 3.2.25. $l^r \hookrightarrow l^s \quad \forall r \leq s (r, s \in [1, \infty])$

Example 3.2.26. $(\mathcal{C}([a, b]), \|\cdot\|_\infty) \hookrightarrow (\mathcal{C}([a, b]), \|\cdot\|_1)$

Example 3.2.27. $BV([a, b]) \hookrightarrow L^1([a, b])$

Example 3.2.28. $AC([a, b]) \hookrightarrow (\mathcal{C}([a, b]), \|\cdot\|_\infty)$

Prove as an exercise the last two examples, using the fundamental theorem of calculus! Remember that $\|f\|_{AC} = \|f\|_1 + \|f'\|_1$.

Proposition 3.2.29

Let $X \hookrightarrow Y$ and $C \subset X$.

Then C closed in Y implies C closed in X .

Proof. Indeed J is continuous, so $J^{-1}(C) = C$ is closed in X . ■

To prove the previous proposition one can argue that, as J is continuous, $J^{-1}(C) = C$ and so it is closed in X .

3.2.3 Baire's categories

This section shows some results which are another insight in the algebraic structures we have worked with. In this section X can be either a normed vector space $(X, \|\cdot\|_X)$ or a metric space (X, d) .

First, recall the definition of nowhere dense set (1.2.56 on page 57); $E \subset X$ is *nowhere dense* if $\overset{\circ}{\bar{E}} = \emptyset$ (the interior of the closure of E is empty), or equivalently if the closure of E does not contain any ball.

Observe that this is not true of dense set, consider for example \mathbb{Q} in \mathbb{R} , the closure of the internal point of \mathbb{Q} is not empty. Indeed, for $X = \mathbb{R}$, \mathbb{N} and \mathbb{Z} are nowhere dense.

For any Banach space X , a finite-dimensional subspace $Y \subsetneq X$ is nowhere dense with respect to the topology of X .

A compact set in an infinite-dimensional Banach space is nowhere dense also.

Every subspace $V \subsetneq \mathbb{R}^N$ of dimension $M < N$ is nowhere dense, with respect to the topology of \mathbb{R}^N .

Both the Cantor set and its generalized counterpart are nowhere dense.

First and second category Consider the following definition.

Definition 3.2.30

We say that $E \subset X$ is **first category** if it is a countable union of nowhere dense sets.

Otherwise we say that E is **second category**.

First category sets are very small in measure: they are also called **meagre sets**.

Example 3.2.31. Consider $X = \mathbb{R}$ and $E = \mathbb{Q}$. Notice that $\bar{E} = \bar{\mathbb{Q}} = \mathbb{R}$, and $\overset{\circ}{\mathbb{Q}} = \mathbb{Q}$, thus \mathbb{Q} is not nowhere dense: \mathbb{Q} is first category, since is a countable union of single points.

Notice that \mathbb{Q} is an F_σ set, but it is not closed.

Moreover, $\mathbb{R} \setminus \mathbb{Q} = \bigcap_{n \in \mathbb{N}} A_n$, where $A_n = \{x_n\}^c$. Every A_n is open and dense, and thus $\mathbb{R} \setminus \mathbb{Q}$ is G_δ , but it is not open (see definition 2.1.6 on page 82).

Proposition 3.2.32

There exists subsets $E \subset \mathbb{R}$ which are either first category and $\lambda(E) > 0$ or second category and $\lambda(E) > 0$.

Also \mathbb{R} can be written as follows:

$$\mathbb{R} = E_1 \cup E_2$$

where E_1 is first category and $\lambda(E_2) = 0$.

Observe that the union of two first category is still first category, and \mathbb{R} is not first category. Thus E_2 must be second category.

Baire's theorem The following is a topological theorem which is a cornerstone in functional analysis.

Theorem 3.2.33 (Baire)

Any complete metric space is second category with respect to its metric.

Proof. Let X be the space in question. By contradiction, suppose:

$$X = \bigcup_{n=1}^{\infty} C_n \quad \text{with } C_n \text{ closed and such that } \overset{\circ}{C}_n = \emptyset \quad \forall n \in \mathbb{N}_0.$$

Then define $A_n = C_n^C$ which are open $\forall n \in \mathbb{N}_0$.

Let $x_1 \in A_1$. Then there exists $\varepsilon_1 < 1$ such that $\overline{B(x_1, \varepsilon_1)} \subset A_1$.

Since C_2 cannot contain all $B(x_1, \varepsilon_1)$, then there exists x_2 contained in the open set $A_2 \cap B(x_1, \varepsilon_1) \neq \emptyset$.

Thus we can find $\varepsilon_2 < \frac{1}{2}$ such that $\overline{B(x_2, \varepsilon_2)} \subset A_2 \cap B(x_1, \varepsilon_1)$.

Iterating this argument, we can construct a decreasing sequence of closed balls:

$$\overline{B(x_1, \varepsilon_1)} \supset \overline{B(x_2, \varepsilon_2)} \supset \overline{B(x_3, \varepsilon_3)} \supset \cdots \supset \overline{B(x_n, \varepsilon_n)} \supset \cdots$$

such that $\varepsilon_n < \frac{1}{2^n}$ and $\overline{B(x_{n+1}, \varepsilon_{n+1})} \subset A_{n+1} \cap B(x_n, \varepsilon_n)$.

Therefore $\{x_n\}_{n \in \mathbb{N}_0}$ is a Cauchy sequence as

$$d(x_n, x_m) \leq \frac{1}{2^{m-1}} \quad n \geq m \geq 1,$$

so $x_n \rightarrow x$ for some $x \in X$, because X is complete. Thus there exists $n_0 \in \mathbb{N}_0$ such that $x \in C_{n_0}$.

On the other hand, we have $x_n \in \overline{B(x_{n_0}, \varepsilon_{n_0})} \subset A_{n_0} = C_{n_0}^C \quad \forall n \geq n_0$, and thus $\{x\} = \bigcap_{n \geq n_0} B(x_n, \varepsilon_n) \subset A_{n_0} = C_{n_0}^C$, which is a contradiction. ■

Corollary 3.2.34

An infinite dimensional Banach space cannot have a countable algebraic basis.

Proof. Suppose that exists a sequence $\{x_n\}_{n \in \mathbb{N}} \subset X$ is a Hamel basis of the Banach space X and consider the space generated by the linear combination of those vectors, namely:

$$V_n = \langle \{x_0, x_1, \dots, x_n, \dots\} \rangle$$

Then $\overset{\circ}{V}_n = \overset{\circ}{V}_n = \emptyset$ for any n , but $X = \bigcup_{n \in \mathbb{N}} V_n$, so it is first category. This contradicts Baire's theorem, and thus the supposition is false. ■

Corollary 3.2.35

Let $(X, \|\cdot\|)$ be a Banach space (or (X, d) a complete metric space). Then, the intersection of any countable family of open dense sets is dense.

Proof. Let $\{A_n\}_{n \in \mathbb{N}}$ such that A_n open and dense ($\bar{A}_n = X$) $\forall n \in \mathbb{N}$. Suppose by contradiction that:

$$C = \overline{\bigcap_{n \in \mathbb{N}} A_n} \subsetneq X$$

Thus C is closed, C° is open, and there exists $B \subset E^C$ closed ball in X such that:

$$\left(\bigcap_{n \in \mathbb{N}} A_n \right) \cap B = \emptyset.$$

Then:

$$\bigcup_{n \in \mathbb{N}} (A_n^C \cap B) = B.$$

We also have:

$$\overline{A_n^C \cap B} \subset \overline{A_n^C} = A_n^{\circ C} = \emptyset$$

and so

$$\overline{A_n^C \cap B} = A_n^{\circ C}.$$

If $A_n^{\circ C}$ wouldn't be the emptyset, there would exist a ball \tilde{B} such that $\tilde{B} \subset A_n^{\circ C}$, but $\bar{A}_n = X$. Therefore, $\overline{A_n^C \cap B} = \emptyset$, $(A_n^C \cap B)$ is nowhere dense, and we deduce that B is first category. However, B is complete metric space since it's closed, and because of Baire's theorem it should be second category, which is a contradiction. ■

As an exercise, consider X Banach space and $C \subset X$ closed. Prove that $C^\circ = \emptyset$ if and only if $A = C^C$ is dense.

3.2.4 Uniform boundedness principle

Here we see how Baire's theorem has a topological relevance.

Definition 3.2.36

A family $\mathcal{F} \subset \mathcal{B}(X, Y)$ is **point-wise bounded** if for each $x \in X$ there exists $M_x > 0$ such that:

$$\sup_{T \in \mathcal{F}} \|Tx\|_Y \leq M_x.$$

Definition 3.2.37

A family $\mathcal{F} \subset \mathcal{B}(X, Y)$ is **uniformly bounded** in $\mathcal{F} \subset \mathcal{B}(X, Y)$ if exists $M > 0$ such that:

$$\sup_{T \in \mathcal{F}} \|T\|_{\mathcal{B}(X, Y)} \leq M.$$

Then we have the following, which is known also as **uniform boundedness principle**.

Theorem 3.2.38 (Banach–Steinhaus)

Let $(X, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y)$ be real Banach spaces, and consider a family $\mathcal{F} \subset \mathcal{B}(X, Y)$.

If \mathcal{F} is point-wise bounded then it is uniformly bounded.

Proof. For each $n \in \mathbb{N}_0$, define:

$$C_n = \{x \in X : \|Tx\|_Y \leq n \ \forall T \in \mathcal{F}\}.$$

Each C_n is closed, indeed consider $x_n \in C_n$, $x_n \rightarrow x$. Since T and the norm are continuous, $\|Tx\|_Y = \lim_{n \rightarrow +\infty} \|Tx_n\|_Y \leq n_0$ then $x \in C_{n_0}$.

Moreover, on account of the point-wise boundedness, we have

$$X = \bigcup_{n \in \mathbb{N}} C_n.$$

Via Baire's theorem, at least one of the C_n is not nowhere dense, namely there is $n_0 \in \mathbb{N}$ such that $\overset{\circ}{C}_{n_0} \neq \emptyset$.

Hence there exists $\varepsilon > 0$ and $x_0 \in X$ for which we find a ball $\overline{\mathcal{B}(x_0, \varepsilon)} \subset C_{n_0}$.

Now, if $\|z\|_X \leq \varepsilon$, then $x_0 + z \in \overline{B(x_0, \varepsilon)}$. Therefore, $\forall T \in \mathcal{F}$:

$$\|Tz\|_Y \leq \|T(x_0 + z)\|_Y - \|Tx_0\|_Y \leq 2n_0 \quad \forall T \in \mathcal{F}.$$

Finally, observe that, for any $T \in \mathcal{F}$ and for $x \neq 0$:

$$\|Tx\|_Y = \frac{\|x\|_X}{\varepsilon} \left\| T \left(\frac{\varepsilon x}{\|x\|_X} \right) \right\|_Y \leq \frac{2n_0}{\varepsilon} \|x\|_X.$$

Thus, setting $M = \frac{2n_0}{\varepsilon}$, we get the thesis. ■

Actually, we do not need $(Y, \|\cdot\|_Y)$ to be a Banach space. Indeed, we applied Baire's theorem using the completeness hypothesis only on X .

Consequences This theorem has many consequences:

Corollary 3.2.39

Let X, Y be Banach spaces and consider $\mathcal{F} \subset \mathcal{B}(X, Y)$.

If \mathcal{F} is not uniformly bounded then there exists a G_δ -set G dense in X such that:

$$\sup_{T \in \mathcal{F}} \|Tx\|_Y = \infty \quad \forall x \in G.$$

Recall that G_δ -sets are countable intersection of open sets.

Proof. Consider the family of closed sets $\{C_n\}_{n \in \mathbb{N}}$ defined in the proof of Banach–Steinhaus theorem [3.2.38 on the previous page](#). Then $\mathring{C}_n = \emptyset \quad \forall n \in \mathbb{N}$: otherwise we could find a uniform bound as before.

Define now:

$$A_n = C_n^c = \{x \in X : \exists T \in \mathcal{F} \text{ s.t. } \|Tx\|_Y > n\};$$

so A_n is open and dense in X .

Then set

$$G = \bigcap_{n \in \mathbb{N}} A_n.$$

Notice that G is G_δ and is dense in X (see corollary [3.2.35 on page 220](#)).

Therefore:

$$\sup_{T \in \mathcal{F}} \|Tx\|_Y > n \quad \forall x \in G \quad \forall n \in \mathbb{N}_0.$$

■

Theorem 3.2.40 (Alternative formulation of Banach–Steinhaus)

Let X, Y be Banach spaces and consider $\mathcal{F} \subset \mathcal{B}(X, Y)$.

Then either \mathcal{F} is uniformly bounded or it's unbounded on a G_δ -set dense in X .

As an exercise, prove that for any fixed $\alpha \in \mathbb{R}$, setting $L_\alpha x = \alpha x$ for all $x \in X$ where X is a Banach space, the family $\{T_\alpha\}_{\alpha \in \mathbb{R}}$ is unbounded at any $x \neq 0$.

Corollary 3.2.41

Let $\{T_n\}_{n \in \mathbb{N}} \subset \mathcal{B}(X, Y)$ such that $\{T_n x\}$ converges in Y .

Then there exists a unique $T \in \mathcal{B}(X, Y)$ such that:

$$T_n x \xrightarrow{Y} T x \quad \forall x \in X \quad \text{as } n \rightarrow \infty.$$

Proof. By hypothesis, $T_n x \rightarrow y$ in Y as $n \rightarrow +\infty$. Therefore we can define a unique $T \in \mathcal{L}(X, Y)$ by setting $Tx = y$. Moreover, every T_n is bounded, and so:

$$\forall x \quad \exists M_x > 0 : \quad \sup_{n \in \mathbb{N}} \|T_n x\|_Y \leq M_x.$$

Via Banach–Steinhaus theorem, $\exists M > 0$ such that $\sup_{n \in \mathbb{N}} \|T_n\|_{\mathcal{B}(X, Y)} \leq M$. Then:

$$\|Tx\|_Y = \lim_{n \rightarrow \infty} \|T_n x\|_Y \leq \left(\limsup_{n \rightarrow \infty} \|T_n\|_{\mathcal{B}(X, Y)} \right) \|x\|_X \leq M \|x\|_X$$

and this implies $T \in \mathcal{B}(X, Y)$. ■

In general $\|T_n - T\|_{\mathcal{B}(X, Y)} \not\rightarrow 0$.

It's easy to check that there exists $T \in \mathcal{L}(X, Y)$ defined as the point-wise limit on X of the sequence $\{T_n\}_{n \in \mathbb{N}}$.

Therefore, thanks to the uniform boundedness principle, we have that $\{T_n\}_{n \in \mathbb{N}}$ is uniformly bounded by some constant $M > 0$.

Hence we have:

$$\|Tx\|_Y = \lim_{n \rightarrow \infty} \|T_n x\|_Y \leq M \|x\|_X$$

so that T is also bounded.

3.2.5 Open mapping, bounded inverse mapping and closed graph theorem

In this chapter we state three theorems which are related to the uniform boundedness principle. As we'll see, those results are equivalent each other.

Open mapping and bounded inverse mapping theorems As we have done with functions, here we discuss the condition on the inverse of a given operator.

Theorem 3.2.42 (open mapping)

Let $(X, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y)$ be Banach spaces.

If $T \in \mathcal{B}(X, Y)$ is surjective, namely $T(X) = Y$, then it maps open sets in open sets.

The proof is based on Baire's theorem and is quite long.⁴⁷

Theorem 3.2.43 (bounded inverse mapping)

If $T \in \mathcal{B}(X, Y)$ bijective, then $T^{-1} \in \mathcal{B}(Y, X)$.

Observe that T is an isomorphism.

Proof. Observe that T^{-1} is well defined from Y to X with $T^{-1} \in \mathcal{L}(Y, X)$. To prove that T^{-1} is bounded, it's enough to prove its continuity. Consider a open set $A \subset X$, then $(T^{-1})^{-1}(A) = T(A)$ is open, thanks to the open mapping theorem. ■

The problem of solving the equation $Tx = y$, where $T \in \mathcal{B}(X, Y)$ and $y \in Y$ is given is well posed if it has a unique solution x for each fixed $y \in Y$ and such x continuously depends on y .

We can say that this problem is well posed if and only if for each $y \in Y$ the equation $Tx = y$ has a solution and $\ker(T) = 0$.

Here we can briefly discuss about norm equivalence in Banach space. Indeed, working with these spaces we can have a simpler criteria than definition 1.2.8 on page 36. The following result shows how.

⁴⁷For further discussion, see: H. Brezis, Functional Analysis, Sobolev Spaces and Partial Differential Equations, 2010, page 35, theorem 2.6.

Corollary 3.2.44

Let $(X, \|\cdot\|_{\spadesuit})$ and $(X, \|\cdot\|_{\clubsuit})$ be two Banach spaces. If there exists $M > 0$ such that

$$\|x\|_{\spadesuit} \leq M \|x\|_{\clubsuit} \quad \forall x \in X,$$

then $\|\cdot\|_{\spadesuit}$ and $\|\cdot\|_{\clubsuit}$ are equivalent.

Proof. Consider the identity operator $I : (X, \|\cdot\|_{\spadesuit}) \rightarrow (X, \|\cdot\|_{\clubsuit})$. As I is bijective and continuous, due to 3.2.43 on the facing page, its inverse $I^{-1} = I$ is continuous and there exists $k > 0$ such that

$$\|x\|_{\clubsuit} \leq k \|x\|_{\spadesuit} \quad \forall x \in X;$$

this thanks to the boundedness, having $\|Ix\| \leq k \|x\|$.

Set now $k = \frac{1}{m}$, we have:

$$m \|x\|_{\clubsuit} \leq \|x\|_{\spadesuit} \quad \forall x \in X,$$

restoring the full condition. ■

Closedness of an operator and closed graph theory Consider this definition.

Definition 3.2.45

If X, Y are vector spaces and $T \in \mathcal{L}(X, Y)$ then:

$$\mathcal{G}(T) = \{(x, y) \in X \times Y : y = Tx\}$$

is called **graph** of T .

Observe that $\mathcal{G}(T)$ is a subspace of $X \times Y$ which is a vector space in a canonical way. It is easy to prove that $T \in \mathcal{B}(X, Y)$ has a closed graph in $X \times Y$. Do it!

Proposition 3.2.46

Let $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y)$ be Banach spaces.

Then the vector space $X \times Y$ is a Banach space with respect to the

norm:

$$\|(x, y)\|_{X \times Y} := \|x\|_X + \|y\|_Y.$$

This norm is equivalent to $\|(x, y)\|_p = (\|x\|_X^p + \|y\|_Y^p)^{\frac{1}{p}}$ with $p \in [1, +\infty)$, or to $\|(x, y)\|_\infty = \max\{\|x\|_X, \|y\|_Y\}$.

Definition 3.2.47

We say that $T \in \mathcal{L}(X, Y)$ is **closed** if

$$x_n \xrightarrow{X} x \quad \text{implies} \quad Tx_n \xrightarrow{Y} Tx.$$

The following result has a very simple proof, but its inverse, which is presented right after, is very important in the development of theory.

Proposition 3.2.48

If $T \in \mathcal{B}(X, Y)$, then T is closed.

Proof. If exists $T^{-1} \in \mathcal{B}(Y, X)$ then, fixing $y_0 \in Y$ and $x_0 = T^{-1}y_0$, we have:

$$\forall \varepsilon > 0 \quad \exists \delta = \delta_0(\varepsilon, y_0) > 0 : Tx \in \mathcal{B}_Y(y_0, \varepsilon) \quad \forall x \in B_X(x_0, \delta).$$

■

Theorem 3.2.49 (closed graph theorem)

If X and Y are Banach spaces, then every closed linear operator is bounded.

To prove the theorem we will use the following norm, known as the **graph norm**:

$$\|x\|_{\mathcal{G}} := \|x\|_X + \|Tx\|_Y.$$

Before the proof check that this is actually a norm

Proof. Step 1: First, we have to prove that $(X, \|\cdot\|_{\mathcal{G}})$ is a Banach space: consider a fundamental sequence $\{x_n\}_{n \in \mathbb{N}} \subset X$: it is a fundamental sequence with respect to $\|x\|_X$ and $\{Tx_n\}_{n \in \mathbb{N}}$ is fundamental in Y .

Then (X, Y) are Banach spaces, and, for some $x \in X$ and some $y \in Y$, we have:

$$x_n \xrightarrow{(X, \|\cdot\|)} x \quad \text{and} \quad Tx_n \xrightarrow{(Y, \|\cdot\|)} y.$$

Thus T is closed, $y = Tx$, so $\{x_m\}_{m \in \mathbb{N}}$ converges with respect to $\{\|\cdot\|_{\mathcal{G}}\}$, namely:

$$x_n \xrightarrow{(X, \|\cdot\|_{\mathcal{G}})} x.$$

Therefore we have $(X, \|\cdot\|_X)$ $(X, \|\cdot\|_{\mathcal{G}})$ Banach spaces.

It is easy to see that:

$$\|x\|_X \leq \|x\|_{\mathcal{G}} \quad \forall x \in X.$$

Step 2: Now we can apply corollary 3.2.44 on page 225, the two norms are equivalent and, in particular, we have:

$$\exists M > 1 : \quad \|x\|_{\mathcal{G}} \leq M \|x\|_X \quad \forall x \in X.$$

Then, by definition of $\|x\|_{\mathcal{G}}$:

$$\exists M > 1 : \quad \|Tx\|_Y \leq (M - 1) \|x\|_X \quad \forall x \in X.$$

and thus T is bounded. ■

We have proven that open mapping implies bounded inverse, which in turn implies closed graph. However, it can also be proven that closed graph implies open mapping: therefore, these theorems are equivalent.

Observe that a non-linear mapping may have a closed graph without being continuous:

Example 3.2.50. Take $f : \mathbb{R} \rightarrow \mathbb{R}$ defined by $f(t) = t^{-1}$ if $t \neq 0$ otherwise $f(0) = 0$.

This isn't linear nor continuous but it has a closed graph.

Sum up Let us retrace our steps: we have used Baire's theorem (1899) to prove the uniform boundedness principle (1927) and the open mapping theorem (Banach–Schauder, 1929), which is equivalent to bounded inverse and closed graph theorems.

Proposition 3.2.51

A linear operator $T \in \mathcal{L}(X, Y)$ is closed if and only if $\mathcal{G}(T)$ is a closed subspace of $X \times Y$.

Hence we conclude $T \in \mathcal{B}(x, y)$ if and only if $\mathcal{G}(T)$ is closed in $X \times Y$.

Example 3.2.52. A non-linear operator can have a closed graph without being continuous. For instance, take:

$$f(t) := \begin{cases} \frac{1}{t} & \text{if } t \neq 0 \\ 0 & \text{if } t = 0 \end{cases}.$$

3.3 Duality

We have already introduced two kind of dual spaces, we will develop those notions and the properties which they imply.

Moreover we will introduce the notion of reflexivity which is a crucial property to develop more tools for applications.

3.3.1 Dual spaces

Recalling the definition [3.2.12 on page 213](#), given the normed vector space X , its **topological dual space** is the space of all linear bounded functional $T : X \rightarrow \mathbb{R}$, namely:

$$X^* := \mathcal{B}(X, \mathbb{R}) = \{T \in \mathcal{L}(X, \mathbb{R}) : T \text{ bounded}\},$$

and its norm is:

$$\|L\|_* := \sup_{\|x\|_X=1} |Lx|.$$

Notice that X^* is always a Banach space (see theorem [3.2.13 on page 213](#)).

Now consider the following two results which we will use them later.

Proposition 3.3.1

If X and Y are isomorphic normed vector spaces, then X^* and Y^* are isomorphic.

Proof. Consider an isomorphism $J : X \rightarrow Y$.

Give $L \in X^*$ and define:

$$\Lambda y = LJ^{-1}y \quad \forall y \in Y.$$

Thus we have a linear map $\tilde{J} : X^* \rightarrow Y^*$ by setting:

$$\tilde{J}L = \Lambda \quad \forall L \in X^*.$$

Now it's easy to prove that \tilde{J} is bounded, injective and surjective.

Therefore, thanks to the bounded inverse mapping theorem (see [3.2.43 on page 224](#)), we have that \tilde{J} is an isomorphism. ■

Moreover, the following proposition holds.

Proposition 3.3.2

If X is infinite dimensional, then $X^* \subsetneq X'$.

Characterization of linear bounded functionals

Theorem 3.3.3

Let $L \in X'$ such that $L \neq 0$.

Then the following statements are equivalent:

1. $L \in X^*$;
2. $\text{Ker}(L)$ is closed;
3. $\overline{\text{Ker}(L)} \subsetneq X$, that is $\text{Ker } L$ is not dense in X .

Proof. We prove the pairwise implications.

Proof of $L \in X^$ implies $\text{Ker}(L)$ is closed:*

Take a sequence x_n belonging to the Kernel, that is, $Lx_n = 0$, converging to some x . We want to show that $Lx = 0$, but since L is bounded (i.e. continuous), this is a sequence of zeros which converges to 0, hence $L(0) = 0$, so $x \in \text{Ker}(L)$.

Proof of $\text{Ker}(L)$ is closed implies $\text{Ker } L$ is not dense in X :

By contradiction, suppose $\overline{\text{Ker}(L)} = X$. As $\text{Ker}(L)$ is closed by hypothesis we have $\text{Ker}(L) = X$, but then it would be $L \equiv 0$, which goes against the hypothesis.

Proof of $\text{Ker } L$ is not dense in X implies $L \in X^$:*

Again by contradiction, let $L \notin X^*$, and $x \in X$ such that $Lx = \alpha \neq 0$, namely $x \notin \text{Ker}(L)$.

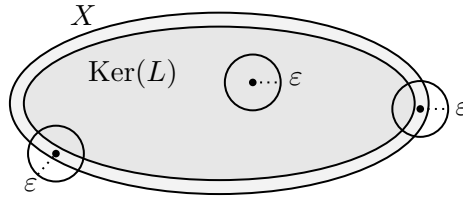
Let $\varepsilon > 0$. Since $L \notin X^*$, namely L is unbounded, then $\exists y \in B(0, \varepsilon)$ such that $Ly = \beta$ with $|\beta| > |\alpha|$.

Set $\gamma = -\frac{\alpha}{\beta}$: we have $|\gamma| < 1$.

Observe now that

$$L(x + \gamma y) = Lx + \gamma Ly = 0,$$

and thus $z = x + \gamma y \in \text{Ker}(L)$, with $z \in B(x, \varepsilon)$.



Then⁴⁸ $\overline{\text{Ker}(L)} = X$, which is a contradiction. ■

Duality in L^p spaces The following two theorems, the first for the case $p \in (1, \infty)$ and the second for the case $p = 1$, discuss duality for functional spaces L^p .

Notice that if we are working on L^p , with $p \in (1, \infty)$, then we can find $g \in L^q$ for which

$$Lg(f) = \int_{\Omega} fg \, d\mu \quad \text{where } Lg \in (L^p)^*.$$

The following theorem goes further providing a complete characterization.

Theorem 3.3.4

Let $(\Omega, \mathfrak{M}, \mu)$ a complete measure space.
 Consider $L^p(\Omega, \mathfrak{M}, \mu)$ for any fixed $p \in (1, \infty)$.
 If $\Lambda \in (L^p(\Omega, \mathfrak{M}, \mu))^*$ then there exists a unique $g \in L^q(\Omega, \mathfrak{M}, \mu)$
 where q is the conjugate index of p , such that:

$$\Lambda(f) = \int_{\Omega} fg \, d\mu \quad \forall f \in L^p(\Omega, \mathfrak{M}, \mu).$$

Moreover, we have:

$$\|\Lambda\|_{(L^p(\Omega, \mathfrak{M}, \mu))^*} = \|g\|_{L^q(\Omega, \mathfrak{M}, \mu)}.$$

Hence $(L^p(\Omega, \mathfrak{M}, \mu))^*$ and $L^q(\Omega, \mathfrak{M}, \mu)$ are isometrically isomorphic through $T : \Lambda \mapsto g$.

This one instead develops the case in which $p = 1$ and $q = \infty$.

⁴⁸Remember that saying that $\forall x \in X, \forall \varepsilon > 0, \exists z \in \text{Ker}(L), z \in B(x, \varepsilon)$ is a way to say that $\text{Ker}(L)$ is dense in X .

Theorem 3.3.5

Let $\Omega \in \mathcal{L}(\mathbb{R}^N)$ and consider $L^1(\Omega, \mathcal{L}(\Omega), \lambda)$.

If $\Lambda \in (L^1(\Omega, \mathfrak{M}, \mu))^*$ then there exists a unique $g \in L^\infty(\Omega, \mathfrak{M}, \mu)$ such that:

$$\Lambda(f) = \int_{\Omega} fg d\mu \quad \forall f \in L^1(\Omega, \mathcal{L}(\Omega), \lambda).$$

Moreover, we have:

$$\|\Lambda\|_{(L^1(\Omega, \mathcal{L}(\Omega), \lambda))^*} = \|g\|_{L^\infty(\Omega, \mathcal{L}(\Omega), \lambda)}.$$

Hence $(L^1(\Omega, \mathfrak{M}, \mu))^*$ and $L^\infty(\Omega, \mathfrak{M}, \mu)$ are isometrically isomorphic through $T : \Lambda \mapsto g$.

Pay attention to the main difference: the second result only holds in \mathbb{R}^N with the Lebesgue measure, while the first one holds in a general metric space.

We need more tools to discuss the case $p = \infty$, we will come back to this at page 235.

3.3.2 Hahn–Banach theorem and consequences

In this section we see the first very relevant theorem for dual spaces: we will see some of its implications, which are many. It was proved by these two mathematician independently in late Twenties. Here is reported the analytic version, but it was also proved a geometric version.⁴⁹

Theorem 3.3.6 (Hahn–Banach)

Let $(X, \|\cdot\|)$ be a real normed vector space and $Y \subset X$ a non-empty subspace.

If $L_0 : Y^* \rightarrow \mathbb{R}$ then there exists $L \in X^*$, called the extension of L_0 such that:

$$\|L\|_{X^*} = \|L\|_{Y^*}, \quad L|_Y = L_0.$$

Proof. First step: construct an extension by adding a dimension to the original subspace.

⁴⁹For further discussion, see: H. Brezis, Functional Analysis, Sobolev Spaces and Partial Differential Equations, 2010, page 4, section 1.2.

Let $x \notin Y$ and set the following:

$$Z_1 = \{y + \lambda x : y \in Y, \lambda \in \mathbb{R}\},$$

$$L_1(y + \lambda x) = L_0y + \lambda\beta,$$

where Z is a subspace of X and $\beta \in \mathbb{R}$ is such that we have a control for the norm:

$$|L_1(y + \lambda x)| = |L_0y + \lambda\beta| \leq \|L_0\|_{Y^*} \|y + \lambda x\|,$$

for all $y \in Y$ and all $\lambda \in \mathbb{R}$. This gives us $\|L_1\|_{Z_1^*} \leq \|L_0\|_{Y^*}$. On the other hand:

$$Z_1^* \supset Y^* \implies \|L_1\|_{Z_1^*} \geq \|L_1\|_{Y^*} = \|L_0\|_{Y^*}$$

Hence we have the equality:

$$\|L_1\|_{Z_1^*} = \|L_0\|_{Y^*}, \quad L_1|_Y = L_0.$$

Second step: extension to the complete space.

Consider now the non-empty family of all possible extensions:

$$\mathcal{S} = \{(L, Z) : Y \subseteq Z \subseteq X, L \in Z^*, \|L\|_{Z^*} = \|L_0\|_{Y^*}, L|_Y = L_0\}.$$

Notice that $(L_1, Z_1) \in \mathcal{S}$. Introduce the partial⁵⁰ order relation:

$$(L', Z') \leq (L'', Z'') \iff Z' \subset Z'' \text{ with } L''|_{Z'} = L'.$$

Consider a chain (totally ordered subset) $\mathcal{C} \neq \emptyset$ of \mathcal{S} and set:

$$\tilde{Z} = \bigcup_{Z:(L,Z) \in \mathcal{C}} Z.$$

Observe that \tilde{Z} is a subspace of X , since at every step we're adding to the union another subspace.

If $x \in \tilde{Z}$ then x belongs to some Z such that $(L, Z) \in \mathcal{C}$ and also to all Z' such that $(L', Z') \in \mathcal{C}$ and $Z \subset Z'$.

⁵⁰The inclusion induces a partial order: having a set included in the other or viceversa are not the only two possibilities.

Recalling that $L'|_Z = L$ we can define a linear bounded operator $\tilde{L} : \tilde{Z} \rightarrow \mathbb{R}$ by setting:

$$\tilde{L}x = Lx,$$

where L is associated to the Z where we first encounter x .

It's clear that (\tilde{L}, \tilde{Z}) is an upper bound for \mathcal{C} .

Hence \mathcal{S} is an inductive set (see definition 1.1.49 on page 32) and, by Zorn's Lemma (see proposition 1.1.51 on page 32), it has a maximal element, namely (L, Z) .

Suppose that $Z \subsetneq X$, then we can construct (see the beginning of this proof) an extension of (L, Z) but this contradicts its maximality.

Thus $Z = X$ and L is the required extension. ■

Notice that we did not even require completeness for X .

Relevant consequences Using the Hahn–Banach theorem we can prove the following three corollaries.

Corollary 3.3.7

If $x \in X$, with $x \neq 0$, then there exists $L_x \in X^*$ such that $\|L_x\|_{X^*} = 1$ and $L_x x = \|x\|$.

Proof. Let $Y = \{\lambda x : \lambda \in \mathbb{R}\}$ and $L_0(\lambda x) = \lambda \|x\|$.

So L_0 is linear and $|L_0(\lambda x)| \leq \|\lambda x\|$ ($M = 1$).

Notice that $\|L_0\|_{(Y)^*} = 1$ so we can simply apply Hahn-Banach theorem. ■

Corollary 3.3.8

Let $x, z \in X$ are such that $Lx = Lz$ for all $L \in X^*$.

Then $x = z$, that is the elements of X^* separate the points in X .

Proof. By contradiction: let $\tilde{x} = x - z \neq 0$. Then, using the previous corollary, we find $L \in X^*$ such that $L\tilde{x} = \|\tilde{x}\| \neq 0$, and $Lx \neq Lz$, which is absurd. ■

Corollary 3.3.9

Let $Y \subsetneq X$ be a closed subspace and $x \notin Y$.

Then $L \in X^*$ such that $Lx \neq 0$ and $L|_Y = 0$.

Proof. Let

$$Z = \{y + \lambda x : y \in Y, \lambda \in \mathbb{R}\} \subset X \text{ and } L_0(\lambda x + y) = \lambda.$$

Then L_0 is linear, $L_0x = 1 \neq 0$, and $\text{Ker}(L_0) = Y$ is closed.

Thus L_0 is bounded on Z (due to 3.3.3 on page 230), and we can apply Hahn–Banach theorem getting the thesis. ■

Duality in L^p spaces In the previous chapter we discussed duality when $p \in (1, \infty)$ (3.3.4 on page 231) and when $p = 1$ (3.3.5 on page 232), now with these results we can discuss the remaining case: $p = \infty$.

Consider a function $g \in L^1(\Omega, \mathcal{L}(\Omega), \lambda)$, we set:

$$\Lambda_g(f) = \int_{\Omega} fgd\lambda \quad \forall f \in L^{\infty}(\Omega, \mathcal{L}(\Omega), \lambda).$$

We have that $\Lambda_g(f) \in (L^{\infty}(\Omega, \mathcal{L}(\Omega), \lambda))^*$ with the norm:

$$\|\Lambda_g\|_{(L^{\infty}(\Omega, \mathcal{L}(\Omega), \lambda))^*} = \|g\|_{L^1(\Omega, \mathcal{L}(\Omega), \lambda)}.$$

Observe that the isometry $T : g \mapsto \Lambda_g$ is not surjective. Indeed, consider $L : (\mathcal{C}_C(\mathbb{R}^N), \|\cdot\|_{\infty}) \rightarrow \mathbb{R}$ defined by

$$L(f) = f(0) \quad \forall f \in \mathcal{C}_C(\mathbb{R}^N).$$

Thanks to the Hahn–Banach theorem, L can be extended to an element $\Lambda \in (L^{\infty}(\mathbb{R}^N))^*$. However, there is no $g \in L^1(\mathbb{R}^N)$ such that:

$$\Lambda(f) = \int_{\mathbb{R}^N} fgd\lambda \quad \forall f \in L^{\infty}(\mathbb{R}^N).$$

Suppose that such $g \in L^1(\mathbb{R}^N)$ exists. Then:

$$\int_{\mathbb{R}^N} fgd\lambda = 0 \quad \forall f \in \mathcal{C}_C^{\infty}(\mathbb{R}^N) \text{ such that } f(0) = 0.$$

We have the following result which is analogous to the vanishing lemma.

Proposition 3.3.10

If $\Omega \subseteq \mathbb{R}^N$ is an open set and $u \in L^1(\Omega)$ is such that:

$$\int_{\Omega} f u d\lambda = 0 \quad \forall f \in \mathcal{C}_C^{\infty}(\Omega)$$

then $u = 0$ almost everywhere in Ω .

Reloading the previous discussion we can take $\Omega = \mathbb{R}^N \setminus \{0\}$ and conclude $g = 0$ almost everywhere in \mathbb{R}^N .

Therefore $L(f) = f(0) = 0$ for all $f \in \mathcal{C}_C^{\infty}(\mathbb{R}^N)$, and we have a contradiction.

3.3.3 Reflexivity

Since from a space we have constructed its topological dual, we can construct the topological dual of this last one. This is called bidual as we will state soon. But how many times we can redo this process? At a certain point must we stop? Sometimes yes, because the bidual is very similar to the original space, or even it can be the original space itself.

Definition 3.3.11

Consider $(X, \|\cdot\|)$ be a real normed vector space. The dual of its dual, namely $(X^{**}, \|\cdot\|_{**})$, is called **bidual**.

Now we will try to figure out how X^{**} is related to X ?

Fix $x \in X$ and let

$$\Lambda_x : X^* \rightarrow \mathbb{R}, \quad \Lambda_x L \mapsto Lx \quad \forall L \in X^*.$$

Observe that Λ_x is linear and also bounded: indeed, we have:

$$|Lx| \leq \|L\|_* \|x\| \quad \forall L \in X^*.$$

Thus we have $\Lambda_x \in X^{**}$ and $\|\Lambda_x\|_{**} \leq \|x\|$.

Moreover, we can find $L \in X^*$ such that $\|L\|_* = 1$ and $Lx = \|x\|$ (see corollary 3.3.7 on page 234), and thus the maximum is achieved:

$$\|\Lambda_x\|_{**} = \|x\|.$$

Definition 3.3.12

Let $\tau : X \rightarrow X^{**}$ a map defined as $\tau x = \Lambda_x$ where $\Lambda_x L \mapsto Lx$ $\forall L \in X^*$. Then τ is a linear isometry called **canonical map**, and we have:

$$\|\Lambda_x\|_{**} = \|x\|.$$

Observe that $\tau(X)$ is a closed subspace of X^{**} but it does not coincide with it in general. If they coincide we have:

Definition 3.3.13

A normed vector space X linearly isomorphic and isometric to X^{**} is said to be **reflexive**. In that case we have:

$$\tau(X) = X^{**}.$$

There are some Banach spaces X which are not reflexive, but are linearly isometric and isomorphic to X^{**} through a different isomorphism.⁵¹

Moreover observe that any finite dimensional Banach space $(X, \|\cdot\|)$ is reflexive as they are linearly isomorphic to $(\mathbb{R}^N, \|\cdot\|_2)$.

Proposition 3.3.14

If a normed vector space X is reflexive, then it is also a Banach space.

Indeed, X is isomorphic to its bidual, which is always a Banach space.

Theorem 3.3.15

Let $(X, \|\cdot\|)$ be reflexive.

If X is linearly isomorphic to Y , then Y is reflexive too.

Proof. Let $J : X \rightarrow Y$ be a linear isomorphism, and $L \in X^*$. Then (see 3.3.1 on page 229) we know that there exists $\tilde{J} : X^* \rightarrow Y^*$.

Consider $\tilde{\Lambda} \in Y^{**}$, $\tilde{L} \in Y^*$. Observe that

$$\tilde{\Lambda}\tilde{L} = \tilde{\Lambda}\tilde{J}L$$

We define $\Lambda : X^* \rightarrow \mathbb{R}$, $\Lambda \in X^{**}$, $L \mapsto \tilde{\Lambda}\tilde{L}$.

Since X is reflexive, there exists a unique $x \in X$ such that $\Lambda L = Lx$. So

⁵¹This was proved by R. C. James in 1951

we have

$$\tilde{\Lambda}\tilde{L} = \Lambda L = Lx = LJ^{-1}y = \tilde{L}y$$

■

Theorem 3.3.16

A space X is reflexive if and only if X^* is reflexive.

Proof. Necessary condition \implies :

By contradiction: suppose $\tau(X) \subsetneq X^{**}$, where τ is the canonical map. Since τ is an isometry and X^* is a Banach space (see 3.2.13 on page 213) and thus $\tau(X)$ is a closed subspace (see proposition 3.2.21 on page 216).

Therefore, there exists $G : X^{**} \rightarrow \mathbb{R}$, such that $G \not\equiv 0$ and $G|_{\tau(X)} = 0$ (see 3.3.9 on page 234).

Observe that G belongs to the tridual of X , namely: $G \in (X^{**})^*$.

We have:

$$G\Lambda_x = 0 \quad \forall x \in X,$$

where $\Lambda_x = \tau x \in \tau(X) \subsetneq X^{**}$.

But X^* is reflexive, thus $\tau(X^*) = X^{***}$ and, using the canonical map between X^* and X^{***} , we have that there exists only one $L \in X^*$ such that:

$$G\Lambda_x = \Lambda_x L \quad \forall \Lambda \in X^{**}$$

with $L = \tau^{-1}(G)$.

Finally, using the canonical map between X^* and X^{**} , one can observe that

$$G\Lambda_x = \Lambda_x L = Lx = 0 \quad \forall x \in X.$$

Therefore $L \equiv 0$ and $G \equiv 0$ as $\|L\|_* = \|G\|_{***}$ which is a contradiction.

Sufficient condition \impliedby :

If X is reflexive then it is linear isometric and isomorphic to X^{**} . Thus we can apply the previous problem: X^{**} is reflexive and then X^* is reflexive. ■

Theorem 3.3.17

If X is reflexive, then any closed subspace of X is reflexive.

Proof. Let $Y \subset X$ be a non-empty closed subspace of X . Y is a Banach space.

Let also $\Lambda_0 \in Y^{**}$, and the mapping $\Lambda_{\sharp} : X^* \rightarrow \mathbb{R}$, $L \mapsto \Lambda_0(L|_Y) \in \mathbb{R}$. $\Lambda_{\sharp} \in X^{**}$. X is reflexive, and thus there exists a unique $x_0 \in X$ such that $\Lambda_{\sharp}L = Lx_0$.

Suppose $x_0 \notin Y$ by contradiction. Then we can find $L_0 \in X^*$ such that $L_0x_0 \neq 0$, $L_0|_Y = 0$ (see proposition 3.3.9 on page 234). However

$$L_0x_0 = \Lambda_{\sharp}L_0 = \Lambda_0(L_0|_Y) = 0$$

Contradiction, therefore $x_0 \in Y$.

Now, for any $\tilde{L} \in Y^*$, there exists its Hahn-Banach extension $L \in X^*$, and we have:

$$\Lambda_0\tilde{L} = \Lambda_0(L|_Y) = \Lambda_{\sharp}L = Lx_0 = \tilde{L}x_0.$$

■

Relation between dual spaces and separability Now we have the tools to set up a proper discussion on how separability is kept or inherited by the dual.

Theorem 3.3.18

If X^* is separable then X is separable as well.

Proof. Let $\{L_n\}_{n \in \mathbb{N}} \subset X^*$ be dense in X^* .

Then for any $n \in \mathbb{N}$, there exists $x_n \in X$ such that $\|x_n\| = 1$ and

$$|L_nx_n| \geq \frac{1}{2} \|L_n\|_* = \frac{1}{2} \sup_{\|x\|=1} |L_nx|.$$

Set now:

$$E = \left\{ \sum_{j=0}^N \alpha_j x_j : N \in \mathbb{N}, \alpha_j \in \mathbb{Q} \right\}$$

We have E is countable and $Y = \bar{E} \subset X$ is a closed subspace.

By contradiction, suppose $Y \subsetneq X$. We find $L \in X^*$ such that $L \neq 0$ and $L|_Y = 0$ (see corollary 3.3.9 on page 234). Since $\{L_n\}_{n \in \mathbb{N}}$ is dense, we have:

$$\forall \varepsilon > 0 \quad \exists n_0 \in \mathbb{N} : \|L - L_{n_0}\| < \varepsilon.$$

But the corresponding x_{n_0} belongs to $E \subseteq Y$, and thus $Lx_{n_0} = 0$. Therefore:

$$\begin{aligned} \frac{1}{2} \|L_{n_0}\|_* &\leq |L_{n_0}x_{n_0}| = |L_{n_0}x_{n_0} - Lx_{n_0}| \\ &\leq \|L_{n_0} - L\|_* \|x_{n_0}\| < \varepsilon. \end{aligned}$$

Summing up, we have

$$\|L\|_* \leq \|L - L_{n_0}\|_* + \|L_{n_0}\|_* \leq \varepsilon + 2\varepsilon \quad \forall \varepsilon > 0$$

and thus $L \equiv 0$, which is a contradiction. ■

Theorem 3.3.19

If X is separable and reflexive then X^* is separable.

Proof. If X is reflexive and separable, then immediately we have that X^{**} is separable, you can check this.

Then, from the previous point, X^* is separable. ■

Reflexivity on L^p spaces Also for functional spaces we can have reflexivity. Indeed, we can show that $L^p(\Omega, \mathfrak{M}, \mu)$ is reflexive for any $p \in (1, \infty)$. In general $L^1(\Omega, \mathfrak{M}, \mu)$ and $L^\infty(\Omega, \mathfrak{M}, \mu)$ are not reflexive. Consider the case $\Omega = \mathbb{R}$ with the Lebesgue measure, L^1 is separable but L^∞ is not. If L^1 were reflexive, it's separable, then (3.3.19) $(L^1)^*$ would be separable. This would make $L^\infty \approx (L^1)^*$ separable, which is false. On the other hand, if L^∞ were reflexive, then $(L^1)^*$ would be reflexive, then (3.3.16 on page 238) also L^1 would be reflexive: this is false as we've just seen.

Sufficient condition for reflexivity We know that if X is reflexive, then any bounded sequence contains a weakly-converging subsequence.

We can apply the first consequence of the Hahn–Banach theorem (see proposition 3.3.7 on page 234), to X^{**} , $\forall L \in X^* \exists \Lambda_L \in X^{**}$ with norm 1 and such that $\Lambda_L L = \|L\|_*$.

If X is reflexive, then $\forall L \in X^* \exists x = \tau^{-1}(\Lambda_L)$ with norm 1 and s.t. $Lx = \Lambda_L L = \|L\|_*$, i.e. one for which the norm is attained.

Definition 3.3.20

Consider a real Banach space $(X, \|\cdot\|)$.

We say that X is **strictly convex**, s.c., if for all distinct $x, y \in X$

such that $\|x\| \leq 1$ and $\|y\| \leq 1$ it follows:

$$\left\| \frac{x+y}{2} \right\| < 1.$$

We say that X is **uniformly convex**, u.c., if for all $\varepsilon > 0$ there exists $\delta > 0$ such that, if $x, y \in X$ are such that $\|x\| \leq 1$, $\|y\| \leq 1$, $\|x - y\| > \varepsilon$ then

$$\left\| \frac{x+y}{2} \right\| < 1 - \delta.$$

This last property entails that if two points x and y are in the closed unit ball, even on the boundary, then their mean point must lie deep inside that same unit ball.

Observe that a uniform convex space is also strictly convex.

The space $(\mathbb{R}^N, \|\cdot\|_p)$ with $N \geq 2$ is uniformly convex if and only if $p \in (1, \infty)$, but for $p = 1$ and $p = \infty$ is not even strictly convex.

It can be shown that, given a real normed vector space X , the Hahn–Banach extension is unique if X^* is strictly convex.⁵²

Theorem 3.3.21 (Milman–Pettis)

Any real uniform convex Banach space is reflexive.

The converse is not true. More precisely, there are some infinite-dimensional reflexive Banach spaces which are not linearly isomorphic to a uniformly convex space.⁵³

Remember that any finite-dimensional Banach space is reflexive: uniform continuity isn't, indeed, a necessary condition.

Now consider the \mathbb{R}^N cases; the spaces $(\mathbb{R}^2, \|\cdot\|_\infty)$ and $(\mathbb{R}^2, \|\cdot\|_1)$ both have “square” unit balls: we can check that they are not uniformly convex spaces. However, we know that both of them are reflexive and linearly isomorphic to $(\mathbb{R}^2, \|\cdot\|_2)$ which is uniformly convex. Indeed, we can prove that $(\mathbb{R}^n, \|\cdot\|_2)$ is reflexive for any n .⁵⁴

⁵²A.E. Taylor proved in 1939 that the condition is sufficient while S. Foguel in 1958 proved that is also necessary.

⁵³Proven by M.M. Day in 1941.

⁵⁴This via uniform convexity, via equivalence of weak and strong convergence and Bolzano-Weierstrass' theorem or via James' theorem.

Spaces $L^p(\Omega, \mathfrak{M}, \mu)$ with $p \in (1, +\infty)$ are reflexive.

Clarkson's inequalities We could also prove it via Milman–Pettis and the following inequalities.

Proposition 3.3.22

Let $f, g \in L^p(\Omega, \mathfrak{M}, \mu)$ with $p \in (1, \infty)$.

The **Clarkson's inequalities** holds:

$$\left\| \frac{f+g}{2} \right\|_p^p + \left\| \frac{f-g}{2} \right\|_p^p \leq \frac{1}{2} \left(\|f\|_p^p + \|g\|_p^p \right) \quad p \in [2, \infty)$$

$$\left\| \frac{f+g}{2} \right\|_p^q + \left\| \frac{f-g}{2} \right\|_p^q \leq \left(\frac{1}{2} \|f\|_p^p + \frac{1}{2} \|g\|_p^p \right)^{q/p} \quad p \in (1, 2)$$

Where q is p 's conjugate.

Characterization of reflexivity As we have seen uniform convexity is only a sufficient conditions but there exist some necessary and sufficient conditions for reflexivity.

Theorem 3.3.23 (James)

Let X be a Banach space. Then X reflexive if and only if any $L \in X^*$ has a maximum on the unit ball.

The implication is trivial: if $L \in X^*$ then by applying HB first corollary (3.3.7 on page 234) on X^* there exists $\Lambda \in X^{**}$ such that $\|\Lambda\|_{**} = 1$ and $\Lambda L = \|L\|_*$; therefore, for $x = \tau^{-1}(\Lambda)$ we have $Lx = \|L\|_*$ with $\|x\| = 1$.

3.3.4 Weak convergence

From now on, let $(X, \|\cdot\|)$ be a Banach space.

Definition 3.3.24

We say that the sequence $\{x_n\} \subset X$ **weakly converges** to $x \in X$ if, for all $L \in X^*$ we have:

$$Lx_n \rightarrow Lx \text{ as } n \rightarrow \infty;$$

in such case we will write:

$$x_n \rightharpoonup X.$$

From now on, we will refer to $x_n \rightarrow x$ as **strong convergence**.

Observe that the strong convergence implies the weak one. Indeed, if $x_n \rightarrow x$ in X , then $|Lx_n - Lx| \leq \|L\|_* \|x_n - x\| \rightarrow 0$ for any $L \in X^*$, and thus $x_n \rightharpoonup x$.

In general, the converse is not true; consider the following example:

Example 3.3.25. Take $X = L^2([-1, 1])$ and let $f_n(t) = \sin(nt)$, with $n \in \mathbb{N}$. Clearly $f_n \not\rightarrow 0$, but it is easy to prove that $f_n \rightharpoonup 0$: since $(L^1)^* \approx L^\infty$, we have:

$$\int_{-1}^1 f_n(t)g(t) dt \rightarrow 0 \quad \forall g \in L^\infty([-1, 1]).$$

Check this as an exercise.

You can show also that if X is finite dimensional then the converse holds.

Indeed: consider $(\mathbb{R}^N, \|\cdot\|_2)$. If $x_n \rightharpoonup x$, then $Lx_n \rightarrow Lx \quad \forall L \in (\mathbb{R}^N)^*$.

But, $\forall L \in (\mathbb{R}^N)^*$, $\exists! a \in \mathbb{R}^N : Ly = \langle a, y \rangle$, thus:

$$\langle a, x_n \rangle \rightarrow \langle a, x \rangle \quad \forall a \in \mathbb{R}^N$$

Taking $a = e_j$, we have $x_n^j \rightarrow x^j$, i.e. strong convergence in each coordinate.

This implies that $x_n \rightarrow x$ in $\|\cdot\|_\infty$ so, by equivalence: $x_n \rightarrow x$ in $\|\cdot\|_2$ and in every norm which makes \mathbb{R}^N a Banach space.

By isometry, we deduce that in any finite dimensional Banach space the weak convergence implies the strong one.

Finally notice that weak and strong convergence are equivalent in l^1 .

Basic properties We can easily deduce some properties of this new kind of convergence.

Proposition 3.3.26

The weak limit is unique.

Proof. Suppose $(x_n \rightharpoonup x) \wedge (x_n \rightharpoonup \tilde{x})$ with $x \neq \tilde{x}$.

Then for any $L \in X^*$, $(Lx_n \rightarrow Lx)$ and $(Lx_n \rightarrow L\tilde{x})$.

The strong limit is unique, so $Lx = L\tilde{x}$ for any $L \in X^*$, and finally using HB second corollary 3.3.8 on page 234 we get $x = \tilde{x}$. ■

Proposition 3.3.27

If $x_n \rightharpoonup x$, then $\{x_n\}$ is bounded in X .

Proof. A sequence $\{Lx_n\}$ is bounded for each $L \in X^*$ since it's a convergent sequence in \mathbb{R} .

Set $T_n L = Lx_n$ for any $L \in X^*$. Applying the uniform boundedness principle (see theorem 3.2.38 on page 221) to $\{T_n\} \subset X^*$ we have that there exists $M > 0$ such that:

$$\|T_n\|_{**} \leq M.$$

Moreover, for each $n \in \mathbb{N}$, we can find (3.3.7 on page 234) $L_n \in X^*$ such that $\|L_n\|_* = 1$ and $L_n x_n = \|x_n\|$. So we have:

$$\|x_n\| = |L_n x_n| = |T_n L_n| \leq \|T_n\|_{**} \underbrace{\|L_n\|_*}_{=1} \leq M.$$

■

Proposition 3.3.28

The norm function is lower semi-continuous with respect to the weak convergence, namely if $x_n \rightharpoonup x$ then:

$$\|x\| \leq \liminf_{n \rightarrow \infty} \|x_n\|.$$

Proof. Using HB first corollary (3.3.7 on page 234), let $L \in X^*$ such that norm $\|L\| = 1$ and $Lx = \|x\|$. We have that:

$$0 < \|x\| = Lx = \lim_{n \rightarrow +\infty} Lx_n = \lim_{n \rightarrow +\infty} |Lx_n| = \liminf_{n \rightarrow +\infty} |Lx_n| \leq \liminf_{n \rightarrow \infty} \|x_n\|.$$

We can put the absolute value in $|Lx_n|$ because the limit is the norm of x , which is non negative. ■

Proposition 3.3.29

If $x_n \rightharpoonup x$ and $L_n \xrightarrow{X^*} L$, then $L_n x_n \rightarrow Lx$.

Proof. As $n \rightarrow +\infty$ we have:

$$L_n x_n - L x_n = |(L_n - L)|x_n \leq \|L_n - L\|_* M \rightarrow 0$$

$$\begin{aligned} L_n x_n - L x &= L_n x_n - L x_n + L x_n - L x \\ &= (L_n x_n - L x_n) + L(x_n - x) \rightarrow 0. \end{aligned}$$

■

What happens to the weak convergence when $Y \neq \mathbb{R}$? We have the following result.

Proposition 3.3.30

Let X, Y be Banach spaces and $T \in \mathcal{B}(X, Y)$. Then $x_n \rightharpoonup x$ implies $T x_n \rightharpoonup T x$. In this case we say that T is **weak-weak continuous**.

Proof. Let $L \in Y^*$.

The mapping $x \mapsto L T x$ is an element of X^* .

Now set $\Lambda x = L T x$ with $\Lambda \in X^*$.

We have $\Lambda x_n \rightarrow \Lambda x$ as $x_n \rightharpoonup x$; so we have:

$$L T x_n \rightarrow L T x \quad \forall L \in Y^*$$

that is the definition of $T x_n \rightharpoonup T x$.

■

Proposition 3.3.31

Let X be reflexive. If $\{L x_n\}$ converges for any $L \in X^*$, then there exists a unique $x \in X$ such that $x_n \rightharpoonup x$.

Proof. Fix $n \in \mathbb{N}$, then let $\{T_n\}_{n \in \mathbb{N}} \subset X^{**}$ such that $T_n L = L x_n$ for each $L \in X^*$.

We have (see corollary 3.2.41 on page 223) that $\{T_n\}$ converges point-wise to some $T \in X^{**}$.

Setting $x = \tau^{-1}(T)$ we have $L x_n = T_n L \rightarrow T L = L x$ for any $L \in X^*$. The thesis is proven.

■

Weak* convergence Analogously, we can define a weak convergence notion also for functionals.

Definition 3.3.32

We say that $\{L_n\} \subset X^*$ **weakly* converges** to $L \in X^*$ if, for all $x \in X$ we have:

$$L_n x \rightarrow Lx \text{ as } n \rightarrow \infty;$$

in such case we will write:

$$L_n \xrightarrow{*} L.$$

In X^* , both weak convergence and weak* convergence are defined, but the latter is weaker. Indeed:

$$\begin{aligned} L_n \xrightarrow{*} L &\iff L_n x \rightarrow Lx \quad \forall L \in X \\ L_n \rightharpoonup L &\iff \Lambda L_n \rightarrow \Lambda L \quad \forall \Lambda \in X^{**} \\ L_n \xrightarrow{*} L &\implies \Lambda L_n \rightarrow \Lambda L \quad \forall \Lambda \in \tau(X) \end{aligned}$$

where $\Lambda L_n = L_n x$, $\Lambda L = Lx$. Therefore, such convergences are equivalent if and only if $\tau(X) = X^{**}$, that is if and only if X is reflexive.

If X is not reflexive we still have $L_n \rightharpoonup L \implies L_n \xrightarrow{*} L$.

Arguing as before we have the following properties

Proposition 3.3.33

- The weak* limit is unique;
- If $L_n \xrightarrow{*} L$ then $\{L_n\}$ is bounded in X^* ;
- The norm is lower semi-continuous with respect to the weak* convergence, that is, if $L_n \xrightarrow{*} L$ then $\|L\| \leq \liminf_{n \rightarrow \infty} \|L_n\|_*$;
- If $(x_n \rightarrow x)$ and $(L_n \xrightarrow{*} L)$ then $L_n x_n \rightarrow Lx$.

A weak* compactness criterion**Theorem 3.3.34** (Banach–Alaoglu)

Let X be a separable Banach space.

Then any bounded sequence $\{L_n\}_{n \in \mathbb{N}} \subset X^*$ contains a subsequence which weakly* converges to some $L \in X^*$.

Separability is necessary, consider the following example.

Example 3.3.35. Take $X = l^\infty$ and $\{L_n\}_{n \in \mathbb{N}} \subset X^*$ such that:

$$L_n x = x_n \quad \forall x = \{x_n\}_n \in l^\infty.$$

Notice that L_n is obviously bounded, namely $\|L_n\|_* \leq 1$, because $|L_n x| \leq \|x\|_\infty$.

Anyway L_n does not contain any weakly* convergent sub-sequence: if there exists any $\{L_{n_k}\}_k$, taking $x = \{(-1)^{n_k}\}_{n_k \in \mathbb{N}}$ we reach a contradiction $L_{n_k} x = (-1)^{n_k}$ does not converge as $n \rightarrow +\infty$.

Proof. First, let $M = \sup_{n \in \mathbb{N}} \|L_n\|_* \in [0, +\infty)$.

Consider a sequence $\{x_k\}_{k \in \mathbb{N}} \subset X$ dense in X .

Take the sequence $\{L_n x_0\}_{n \in \mathbb{N}} \subset \mathbb{R}$ which is bounded. By Bolzano–Weierstrass theorem (1.2.20 on page 43) there exists a converging subsequence $\{L_{n_{j_0}} x_0\}_{j_0 \in \mathbb{N}}$.

Take also the bounded sequence $\{L_{n_{j_0}} x_1\}$: there exists a converging subsequence $\{L_{n_{j_1}} x_1\}_{j_1 \in \mathbb{N}}$.

By diagonalization (see for analogy the proof of Ascoli–Arzelà theorem 3.1.47 on page 188) we can extract $\{L_{n_h}\}_{h \in \mathbb{N}}$ such that $\{L_{n_h} x_k\}_{j \in \mathbb{N}}$ converges for any $k \in \mathbb{N}$ with respect to j .

Consider $x \in X$ and fix $\varepsilon > 0$. Via separability, we can find x_k such that $\|x - x_k\| \leq \frac{\varepsilon}{2M}$.

Observe now that:

$$|L_{n_i} x - L_{n_j} x| \leq |L_{n_i} x - L_{n_i} x_k| + |L_{n_i} x_k - L_{n_j} x_k| + |L_{n_j} x_k - L_{n_j} x|.$$

We have:

$$\begin{aligned} |L_{n_i} x - L_{n_i} x_k| &\leq M \|x - x_k\| \leq \frac{\varepsilon}{2}, \\ |L_{n_j} x_k - L_{n_j} x| &\leq M \|x - x_k\| \leq \frac{\varepsilon}{2}. \end{aligned}$$

Moreover, $\{L_{n_j} x_k\}_{j \in \mathbb{N}}$ converges as it is a fundamental sequence, and we can find $j' \in \mathbb{N}$ such that:

$$|L_{n_i} x_k - L_{n_j} x_k| \leq \varepsilon \quad \forall i, j \geq j'.$$

Summing up:

$$|L_{n_i} x - L_{n_j} x| \leq 2\varepsilon \quad \forall i, j \geq j'.$$

Therefore $\{L_{n_j}x\}$ is a Cauchy sequence and converges. Via an implication of the uniform boundedness principle (see corollary 3.2.41 on page 223), there exists $L \in X^*$ such that $L_{n_j}x \rightarrow Lx \ \forall x \in X$. An this concludes the proof. ■

Example 3.3.36. Consider the space $(\Omega, \mathcal{L}(\Omega), \lambda)$ with $\Omega \subseteq \mathbb{R}^N$ and $\lambda(\Omega) > 0$, and a sequence $\{f_n\}_{n \in \mathbb{N}} \subset L^\infty(\Omega)$.

Set:

$$L_n(g) = \int_{\Omega} f_n g \, d\lambda \quad \forall g \in L^1(\Omega).$$

If $\{f_n\}_{n \in \mathbb{N}}$ is bounded, then $\{L_n\}_{n \in \mathbb{N}}$ is bounded in $(L^1(\Omega))^*$.

By Banach–Alaoglu theorem, as $L^1(\Omega)$ is separable, we can extract a subsequence $\{L_{n_h}\}_{h \in \mathbb{N}}$ which weakly* converges to an $L \in (L^1(\Omega))^*$, namely:

$$L_{n_h}(g) \rightarrow L(g) \quad \forall g \in L^1(\Omega) \quad \text{as } h \rightarrow \infty.$$

Moreover, we know that exists a unique $f \in L^\infty(\Omega)$ such that:

$$L(g) = \int_{\Omega} f g \, d\lambda \quad \forall g \in L^1(\Omega).$$

Therefore the weak* convergence can be written in terms of integrals as follows:

$$\int_{\Omega} f_{n_h} g \, d\lambda \rightarrow \int_{\Omega} f g \, d\lambda \quad \forall g \in L^1(\Omega, \mathfrak{M}, \mu) \quad \text{as } h \rightarrow \infty.$$

This is an example of the statement “for any bounded sequence in $L^\infty(\Omega)$ we can extract a weakly* convergent subsequence”.

Remember also that $L^1(\Omega, \mathfrak{M}, \mu)$ is separable if, for instance, $\Omega \in \mathcal{L}(\mathbb{R}^N) = \mathfrak{M}$. It may not be separable in some extremely pathological cases.

Considering also the reflexivity Reflexivity is a strong assumption. Can we say more if X is also reflexive?

Proposition 3.3.37

If X is separable and reflexive, then any of its bounded sequences contains a weakly-converging sub-sequence.

Proof. Let $\{x_n\}_{n \in \mathbb{N}} \subset X$ a bounded sequence.

As X^* is separable we can apply the Banach–Alaoglu theorem to $\{\tau(x_n)\}_{n \in \mathbb{N}} \subset X^{**}$, which is bounded since x_n is a bounded sequence and the canonical map is isometric, and find $\{\tau(x_{n_h})\}_{h \in \mathbb{N}}$ such that

$$\tau(x_{n_h}) \xrightarrow{*} \Lambda$$

as $h \rightarrow \infty$ for some $\Lambda \in X^{**}$.

Thanks to the reflexivity of X we have that $x = \tau^{-1}(\Lambda)$ and

$$x_{n_h} \rightharpoonup x$$

as $h \rightarrow \infty$. ■

Another characterization for reflexivity Actually we can say much more: separability is not necessary in statement [3.3.37 on the preceding page](#). Moreover the converse holds:

Theorem 3.3.38 (Eberlin–Šmulian)

If X is a Banach space in which any bounded sequence contains a weakly-converging subsequence, then X is reflexive.

Let $\{f_n\}_n \subseteq L^p(\Omega, \mathcal{L}(\Omega), \lambda)$ with $p \in (1, +\infty)$ and $\lambda(\Omega) > 0$.

We know that L^p is separable, and hence $(L^p)^* \approx L^q$, with p, q conjugates: operators on L^p can be represented as integrals with an appropriate L^q function.

If $\{f_n\}$ is bounded, then there exists a subsequence $\{f_{n_k}\}$ and a function $f \in L^p$ such that $\int_{\Omega} f_{n_k} g \, d\lambda \rightarrow \int_{\Omega} f g \, d\lambda \, \forall g \in L^q$, and, in particular, $L_g f = \int_{\Omega} f g \, d\lambda$ and $L_g \in (L^p)^* \approx L^q$.

Summing up, if $p \in (1, \infty)$, from any bounded sequence in $L^p(\Omega)$ we can extract a weakly convergent subsequence.

3.3.5 Linear compact operators

Now we see a further property for linear operator. Here we require an improvement of the image of the operator that paves the way to new results.

Throughout this section, let $(X, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y)$ be Banach spaces.

Definition 3.3.39

A linear operator $K \in \mathcal{L}(X, Y)$, $K : X \rightarrow Y$ is **compact** if for any bounded set $B \subseteq X$ the image KB is precompact in Y , i.e. \overline{KB} is compact. Their space is defined as

$$\mathcal{K}(X, Y) := \{K \in \mathcal{L}(X, Y) \text{ compact}\}.$$

The term precompact means that its closure is compact.

In other words, any bounded sequence $\{y_n\}_n \subseteq KB$ contains a converging subsequence which limit needs not to be in KB .

If $X = Y$, then often it is written as $\mathcal{K}(X)$.

Proposition 3.3.40

Any compact operator is bounded.

Proof. Take the unit ball $B = B_1(0) \subseteq X$.

As KB is precompact, \overline{KB} is compact, so it is also bounded.

Therefore there exists $M > 0$ such that:

$$\|K\|_* = \sup_{x \in \overline{B}} \|Kx\| \leq M$$

that is, $K \in \mathcal{B}(X, Y)$. ■

Definition 3.3.41

A linear operator is **finite-rank** if its image is finite dimensional.

Proposition 3.3.42

A finite-rank bounded operator is also compact.

Indeed, notice that if $T \in \mathcal{B}(X, Y)$ and Y is finite-dimensional then T is compact.

However, there are compact operators which are not finite-rank: consider $(C([a, b]), \|\cdot\|_\infty)$ and, for each $u \in C([a, b])$, set:

$$Ku(t) = \int_a^t u(s)ds \quad \forall t \in [a, b].$$

The linear operator $K : C([a, b]) \rightarrow C([a, b])$ is compact (see Ascoli–Arzelà theorem 3.1.47 on page 188), indeed we have $\text{Im}(K) = \{f \in C^1([a, b]) : f(a) = 0\}$.

Note also that if X and Y are infinite-dimensional, then $\mathcal{K}(X, Y) \subsetneq \mathcal{B}(X, Y)$. An example of bounded, non-compact operator is the identity map, if $X = Y$ infinite-dimensional.

Definition 3.3.43

If Y is a subspace of X and the identity $I : Y \rightarrow X$ is compact we say that Y is **compactly embedded** in X and we write:

$$Y \hookrightarrow\hookrightarrow X.$$

Proposition 3.3.44

Let X be an infinite dimensional Banach space.

Then a compact operator $K : X \rightarrow X$ cannot be bijective.

Proof. By contradiction, take $\{y_n\}_{n \in \mathbb{N}} \subset B$, where B is the unit ball of X . Consider now $x_n = K^{-1}y_n$: the sequence $\{x_n\}_{n \in \mathbb{N}}$ is bounded, as K^{-1} is bounded.

Since $\{y_n\}_{n \in \mathbb{N}}$ contains a convergent sub-sequence, B is compact, which is a contradiction to 3.1.43 on page 187. ■

Characterization for compact operators To prove directly the compactness is not an easy task, anyway we can find a criterion to do the job.

Theorem 3.3.45 (Compactness characterization)

If $K \in \mathcal{K}(X, Y)$ and $x_n \rightharpoonup x$, then:

$$Kx_n \rightarrow Kx.$$

This means that K is **weak-strong continuous**.

Moreover, if X is reflexive and $K \in \mathcal{B}(X, Y)$ is weak-strong continuous, then K is also compact.

Proof. The two parts will be proved separately.

Proof of the first part:

Suppose $K \in \mathcal{K}(X, Y)$ and $x_n \rightharpoonup x$. Being also bounded (see 3.3.40 on

page 250), K is weak-weak continuous (see 3.3.30 on page 245), namely $Kx_n \rightharpoonup Kx$.

If $\{x_n\}$ is bounded, we have that $\{Kx_n\}$ is also bounded and contains a strongly convergent subsequence, and $\{Kx_n\}$ has a non-empty and bounded class limit.

Let y be a limit point: there exists $\{Kx_{n_h}\}_{h \in \mathbb{N}}$ such that $Kx_{n_h} \rightarrow y$ and in particular $Kx_{n_h} \rightharpoonup y$.

Therefore $y = Kx$ is the only limit point: we have that K is weak-strong continuous, namely $Kx_n \rightarrow Kx$.

Proof of the second part:

Suppose now that $K \in \mathcal{B}(X, Y)$ is weak-strong continuous, consider a bounded set $E \subset X$ and the sequence $\{y_n\}_{n \in \mathbb{N}} \subset KE$.

Then there exists $x_n \in E$ such that $Kx_n = y_n$ for every $n \in \mathbb{N}$.

Using the reflexivity of X we can apply the corollary 3.3.37 on page 248 to the BA theorem⁵⁵ and find $\{x_{n_h}\}_{h \in \mathbb{N}}$ such that $x_{n_h} \rightharpoonup x$. Hence $Kx_{n_h} \rightarrow Kx$ and KE is precompact. This entails $K \in \mathcal{K}(X, Y)$. ■

We have seen that $\mathcal{K}(X, Y)$ is a subspace of $\mathcal{B}(X, Y)$. We are now going to prove that $\mathcal{K}(X, Y)$ is closed.

Theorem 3.3.46

The space of compact operators $\mathcal{K}(X, Y)$ is a closed subspace of $\mathcal{B}(X, Y)$.

This means that $\mathcal{K}(X, Y)$ is a Banach space with respect to the induced norm.

Proof. For simplicity, we will prove the theorem when X is also reflexive. Consider a converging sequence $\{K_n\}_{n \in \mathbb{N}} \subset \mathcal{K}(X, Y)$, such that there exists:

$$K \in \mathcal{B}(X, Y) : \|K_n - K\|_{\mathcal{B}(X, Y)} \rightarrow 0 \quad \text{as } n \rightarrow +\infty.$$

We are left to prove that K is compact, namely $K \in \mathcal{K}(X, Y)$, by showing that it is weak-strong continuous (see the theorem stating the characterization of the compactness 3.3.45 on the previous page).

⁵⁵Remember that separability was not necessary, even though to prove it without this assumption, we can't take advantage of the Banach-Alaoglu theorem.

Let $\{x_n\}$ such that $x_n \rightharpoonup x$; as $\{x_n\}_{n \in \mathbb{N}}$ is bounded there exists $M > 0$ such that $\|x_n\|_X \leq M$, and we have:

$$\begin{aligned} \|Kx_n - Kx\|_Y &\leq \|Kx_n - K_j x_n\|_Y + \|K_j x_n - K_j x\|_Y + \|K_j x - Kx\|_Y \\ &\leq 2M \|K - K_j\|_{\mathcal{B}(X,Y)} + \|K_j x_n - K_j x\|_Y \end{aligned}$$

Let $\varepsilon > 0$ be fixed. Then exists $j_0 \in \mathbb{N}$ such that

$$\|K_j - K\|_{\mathcal{B}(X,Y)} < \frac{\varepsilon}{3M} \quad \forall j \geq j_0.$$

Therefore:

$$\|Kx_n - K_j x_n\|_Y \leq M \cdot \|K - K_j\|_{\mathcal{B}(X,Y)} < \frac{\varepsilon}{3} \quad \forall j \geq j_0.$$

Moreover, since $x_n \rightharpoonup x$, via lower semi-continuity we have:

$$\|K_j x - Kx\|_Y \leq \liminf_n \|K_j x_n - Kx_n\|_Y \leq \frac{\varepsilon}{3} \quad \forall j \geq j_0.$$

Fix $j = j_0$. Since $K_{j_0} \in \mathcal{K}(X, Y)$, there exists $n_0 \in \mathbb{N}$ such that:

$$\|K_{j_0} x_n - K_{j_0} x\|_Y < \frac{\varepsilon}{3} \quad \forall n \geq n_0.$$

Summing up:

$$\forall \varepsilon > 0 \quad \exists n_0 = n_0(\varepsilon) \in \mathbb{N}: \quad \|Kx_n - Kx\|_Y < \varepsilon \quad \forall n \geq n_0$$

Thus $Kx_n \xrightarrow{Y} Kx$, and K is compact. ■

The approximation problem The previous theorem has an immediate consequence:

Proposition 3.3.47

If we have a converging sequence of finite-rank linear operators, then its limit is a compact operator.

The converse of this proposition is known as **approximation property**: is any $K \in \mathcal{K}(X, Y)$ the limit of a sequence of finite rank operators, with respect to the operator norm?

In general it is not: Per Enflo proved in 1973 that there exists a Banach space which is separable but it does not have any Schauder basis and so the approximation property does not hold.

We know that the approximation property holds if Y has a Schauder basis or if Y is a Hilbert spaces (see definition 3.4.5 on page 260).

First example of compact operator We are now going to introduce a class of linear compact operators, from L^p to L^q .

Consider $(\Omega, \mathcal{L}(\Omega), \lambda)$ with $\Omega \in \mathcal{L}(\mathbb{R}^N)$ and $G \in L^q(\Omega \times \Omega)$ with respect to the Lebesgue measure in \mathbb{R}^{2N} for some $q \in (1, \infty)$.

For any $u \in L^p(\Omega)$, where p is the conjugate of q , set

$$K_G u(x) = \int_{\Omega} G(x, y) u(y) dy$$

for all $u \in L^p(\Omega)$ and for almost any $x \in \Omega$. The function G is called **kernel** of the operator K_G .

Using Hölder's inequality we have:

$$|K_G u(x)|^q \leq \left(\int_{\Omega} |G(x, y)| |u(y)| dy \right)^q \leq \|G(x, \cdot)\|_q^q \|u\|_p^q$$

for almost any $x \in \Omega$. This implies that:

$$\|K_G\|_q \leq \|G\|_q \|u\|_p \quad \forall u \in L^p(\Omega).$$

Then we deduce that K_G is linear and bounded from $X = L^p(\Omega)$ and $Y = L^q(\Omega)$.

Let $\{u_n\}_{n \in \mathbb{N}} \subset L^p(\Omega)$ be such that $u_n \rightharpoonup u$.

Define the converging sequence:

$$\Phi_n(x) = K_G(u_n - u)(x) = \int_{\Omega} G(x, y)(u_n - u)(y) dy \rightarrow 0$$

for almost any $x \in \Omega$. Notice that we have:

$$|\Phi_n(x)|^q \leq \|G(x, \cdot)\|_q^q \|u_n - u\|_p^q \leq M \|G(x, \cdot)\|_q^q = F(x)$$

for almost any $x \in \Omega$ since $\{u_n\}_{n \in \mathbb{N}}$ is bounded in $L^p(\Omega)$.

On the other hand, $F \in L^1(\Omega)$. Therefore (see dominated convergence theorem [2.3.2 on page 119](#)):

$$\|K_G(u_n - u)\|_q \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Hence K is weak-strong continuous so it's compact since $L^p(\Omega)$ is reflexive.

In case of $p = q = 2$ we have the following definition.

Definition 3.3.48

Let $G \in L^2(\Omega \times \Omega)$, for all $u \in L^p(\Omega)$ the operator

$$K_G u(x) = \int_{\Omega} G(x, y)u(y) \, dy$$

is called **Hilbert–Schmidt** operator with kernel G .

The set of all Hilbert–Schmidt operators is a subspace of $\mathcal{K}(L^2(\Omega))$.

Second example of compact operator Let's set:

$$X_p = \{f \in AC([a, b]) : f' \in L^p((a, b))\} \subseteq AC([a, b])$$

where $p \in [1, \infty]$. We know that X_p can be identified with some *Sobolev space*

$$W^{1,p}((a, b)) := \{f \in L^p([a, b]) : Df \in L^p((a, b))\}$$

where Df is the distributional derivative, namely:

$$\int_a^b Df \phi \, dt = - \int_a^b f \phi' \, dt \quad \phi \in \mathcal{C}_C^\infty((a, b)).$$

In particular we have that $AC([a, b])$ can be identified with $W_1^1((a, b))$ since each equivalence class contains one and only one continuous representative.

It's easy to prove that $W^{1,p}(a, b)$ is a Banach space with respect to the norm:

$$\|f\|_{\spadesuit, p} = |f(a)| + \|Df\|_p$$

which is equivalent to

$$\|f\|_{1, p} = \|f(a)\|_p + \|Df\|_p.$$

Consider now the identity application:

$$I : (W^{1,p}((a, b)), \|\cdot\|_{\spadesuit, p}) \rightarrow (\mathcal{C}([a, b]), \|\cdot\|_\infty), \quad If = f.$$

Using Ascoli–Arzelà theorem (3.1.47 on page 188) we can prove that I is compact if $p \in (1, \infty]$. Thus we have

$$W^{1,p}((a, b)) \hookrightarrow \mathcal{C}([a, b]) \quad \forall p > 1.$$

Let $\{f_n\}_{n \in \mathbb{N}} \subset W^{1,p}((a, b))$ be bounded by a constant $M > 0$. Observe that

$$f_n(t) = f_n(a) + \int_a^t Df_n(r) \, dr \quad t \in [a, b].$$

Then, using Hölder's inequality, we get

$$\begin{aligned} |f_n(t)| &\geq |f_n(a)| + \int_a^b |Df_n(r)| \, dr \\ &\geq |f_n(a)| + (b-a)^{\frac{1}{q}} \|Df_n\|_p \\ &\quad \max\{1, (b-a)^{\frac{1}{q}}\} \|f_n\|_{\clubsuit, p} \end{aligned}$$

where $q \in [1, \infty)$ is the conjugate of p . Thus we find:

$$\|f_n\|_\infty \leq \max\{1, (b-a)^{\frac{1}{q}}\} M \quad \forall n \in \mathbb{N}.$$

Observe now that

$$f_n(t) - f_n(s) = \int_s^t Df_n(r) \, dr \quad t, s \in [a, b].$$

Using again Hölder's inequality we find:

$$\begin{aligned} |f_n(t) - f_n(s)| &\leq \left| \int_s^t |Df_n(r)| \, dr \right| \\ &\leq |t-s|^{\frac{1}{q}} \|Df_n\|_p \\ &\leq M |t-s|^{\frac{1}{q}}. \end{aligned}$$

Then $\{f_n\}_{n \in \mathbb{N}}$ is bounded and equicontinuous. Using Ascoli–Arzelà theorem we deduce that there exists a subsequence $I f_{n_h} = f_{n_h}$ for $h \in \mathbb{N}$ and $f \in C([a, b])$ such that $\|f_{n_h} - f\|_\infty \rightarrow 0$ as $h \rightarrow \infty$. Thus I is compact.

The Fredholm alternative in Banach spaces Surjectivity and injectivity of an operator are strongly related to the solvability of a certain equation somehow associated to the operator itself. The following results will handle a scenario where the operator $I - K$ for some K compact.

Theorem 3.3.49

Let X be a Banach space and $K \in \mathcal{K}(X)$. Then

- $\text{Ker}(I - K)$ is finite dimensional;
- $\text{Im}(I - K)$ is closed;
- $\text{Ker}(I - K) = \{0\}$ if and only if $\text{Im}(I - K) = X$.

This result will be proven for Hilbert spaces (see [3.4.54 on page 297](#)).

Corollary 3.3.50 (Fredholm's alternative)

Let $K \in \mathcal{K}(X)$. Given $y \in X$, consider the functional equation:

$$x - Kx = y$$

Then either the equation has a unique solution for any $y \in Y$ or $x - Kx = 0$ has at $n > 1$ linearly independent solutions.

Corollary 3.3.51

Let X finite dimensional and $K \in \mathcal{K}(X)$.

Then the operator K is injective if and only if it is also surjective.

Otherwise if X is infinite-dimensional then there are linear operator which are injective but non surjective or viceversa.

Example 3.3.52. Consider $X = l^2$ and $x = \{x_1, x_2, \dots, x_n, \dots\} \in X$, take:

$$Rx = \{0, x_1, x_2, \dots, x_n, \dots\}, \quad Lx = \{x_2, x_3, \dots, x_n, \dots\}.$$

We have that R , the right shift, is injective but not surjective while L , the left shift, is surjective but not injective.

3.4 Hilbert spaces

Up until now we dealt with spaces, specifically Banach spaces, which are normed and have certain properties of convergence with respect to the norm. Now we will bring this to a step further, and we will introduce another operation, called scalar (or inner) product. This will lead us to Hilbert Spaces, which have very strong properties and are often a very good framework when solving partial differential equations. We will see that \mathbb{R}^2 is a Hilbert space, and the scalar product which we will talk about, in that space is the well known euclidean inner product related to the notion of angle.

3.4.1 Hilbert spaces

Let's start from the very beginning. In this section consider X as a vector space on \mathbb{R} .

Definition 3.4.1

An application $p : X \times X \rightarrow \mathbb{R}$ is called **scalar product** in X if, for all $x, y, z \in X$ and for all $\alpha, \beta \in \mathbb{R}$ the following properties hold:

- **positivity:**

$$p(x, x) \geq 0;$$

- **annihilation:**

$$p(x, x) = 0 \iff x = 0;$$

- **symmetry:**

$$p(x, y) = p(y, x);$$

- **linearity on the first component:**

$$p(\alpha x + \beta y, z) = \alpha p(x, z) + \beta p(y, z).$$

In this case we set

$$\langle x, y \rangle = p(x, y).$$

Scalar product is known also as **inner product**. Observe that the first two properties are in case of both components are with the same argument. Moreover, thanks to symmetry property, linearity holds for both components.

Definition 3.4.2

A vector space X endowed with a scalar product, namely

$$(H, \langle \cdot, \cdot \rangle),$$

is called **pre-Hilbert space**, and their elements are called **vectors**.

On \mathbb{R}^2 the scalar inner product is defined: $x \cdot y = \langle x, y \rangle := x_1y_1 + x_2y_2$.

This product allows us to define angles between vectors: letting $a = (a_1, a_2)$, $b = (b_1, b_2)$, we geometrically have $a \cdot b = \|a\| \|b\| \cos(\theta)$. This motivates the following definition of angle through the scalar product:

$$\cos(\theta) = \frac{a \cdot b}{\sqrt{a \cdot a} \sqrt{b \cdot b}}.$$

The following is a famous inequality we will use often from now on.

Proposition 3.4.3 (Cauchy–Schwarz or Bunakowsky inequality)

Let $(H, \langle \cdot, \cdot \rangle)$ be a pre-Hilbert space.

Then:

$$|\langle x, y \rangle| \leq \sqrt{\langle x, x \rangle} \sqrt{\langle y, y \rangle} \quad \forall x, y \in H.$$

Proof. Fix two vectors $x, y \in X$ and consider the non-negative function

$$t \mapsto \langle x + ty, x + ty \rangle.$$

Then the inequality is given by the following, valid for all $t \in \mathbb{R}$:

$$\langle x + ty, x + ty \rangle = \langle x, x \rangle + 2t \langle x, y \rangle + t^2 \langle y, y \rangle \geq 0,$$

the discriminant should be negative so the inequality follows. ■

Defining the Hilbert spaces It's easy to understand that the scalar product has same properties in common with the norm, indeed, the scalar product of the same vector returns always a real non-negative number. Furthermore, if such vector is the null vector, the scalar product is zero. Using those properties is it possible to define a new norm from the scalar product?

Definition 3.4.4

Let $x \in H$. We define the **norm induced by the scalar product** as follows:

$$\|x\| := \sqrt{\langle x, x \rangle}.$$

To check that $\|\cdot\|$ is actually a norm in H (see definition 3.1.10 on page 175), we have only to check the homogeneity, which is given by the linearity of the scalar product, and that the triangular inequality holds, namely $\|x + y\| \leq \|x\| + \|y\|$, which is obtained from the Cauchy–Schwarz inequality as follows:

$$\|x + y\|^2 = \langle x + y, x + y \rangle = \langle x, x \rangle + \langle y, y \rangle + 2\langle x, y \rangle \leq \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| = (\|x\| + \|y\|)^2.$$

Having this definition we can rewrite the Cauchy–Schwarz inequality as follows:

$$|\langle x, y \rangle| \leq \|x\| \cdot \|y\|,$$

where the norm is induced by the scalar product.

Now we can define the Hilbert spaces:

Definition 3.4.5

We say that a pre-Hilbert space is a **Hilbert space** if it is complete with respect to the norm induced by its scalar product.

Previously we argued the concept of angle. From Cauchy–Schwarz inequality, we have:

$$-1 \leq \frac{\langle x, y \rangle}{\|x\| \|y\|} \leq 1 \quad \forall x, y \in H : x, y \neq 0.$$

Thus there exists a unique angle $\theta \in [0, \pi]$ such that $\cos(\theta) = \frac{\langle x, y \rangle}{\|x\| \|y\|}$. This is an abstract generalization of the concept of angle between two or three dimensional vectors.

Notable examples of Hilbert spaces Here we present for some spaces a definition of scalar product which make the space an Hilbert one.

- Consider the set \mathbb{R}^N : it becomes an Hilbert space when endowed with the *Euclidean scalar product*:

$$\langle x, y \rangle = \sum_{i=1}^N x_i y_i \quad \forall x, y \in \mathbb{R}^N.$$

- The space $L^2(\Omega, \mathfrak{M}, \mu)$ is a Hilbert space with:

$$\langle f, g \rangle = \int_{\Omega} fg \, d\mu.$$

- The space l^2 is a Hilbert space with:

$$\langle \{x_n\}_{n \in \mathbb{N}}, \{y_n\}_{n \in \mathbb{N}} \rangle = \sum_{n \in \mathbb{N}} x_n y_n.$$

Moreover, space of continuous functions $\mathcal{C}([0, 1])$ is a *pre*-Hilbert space with:

$$\langle f, g \rangle = \int_0^1 fg \, dx.$$

It is not complete with respect to the norm induced by this scalar product.

Parallelogram identity and minimal distance The following identity is a characterization of Hilbert spaces, as specified by the subsequent theorem.

Proposition 3.4.6 (parallelogram identity)

Let H be a pre-Hilbert space.

Then:

$$\|x - y\|^2 + \|x + y\|^2 = 2\|x\|^2 + 2\|y\|^2 \quad \forall x, y \in H.$$

Theorem 3.4.7 (Von Neumann)

Let $(X, \|\cdot\|)$ be a Banach-space.

If $\|\cdot\|$ satisfies the parallelogram identity, then $\|\cdot\|$ is induced by the following inner product:

$$\langle x, y \rangle = \frac{1}{2}(\|x + y\|^2 + \|x\|^2 + \|y\|^2)$$

and $(X, \langle \cdot, \cdot \rangle)$ is a Hilbert space.⁵⁶

⁵⁶For further discussion and many more references, see: H. Brezis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, 2010, page 144, “Characterization of Hilbert spaces”.

The main difficulty of proving this results lays in proving that the one presented is actually an inner product, in particular that is linear.

Notice that, via parallelogram identity, every Hilbert space is uniformly convex and thus reflexive (see Milman-Pettis theorem 3.3.21 on page 241), indeed since

$$\|x + y\|^2 = 2 \underbrace{\|x\|^2}_{\leq 1} + 2 \underbrace{\|y\|^2}_{\leq 1} + \underbrace{(-\|x - y\|^2)}_{< -\varepsilon^2} \leq 4 - \varepsilon^2$$

then

$$\sqrt{\left\| \frac{x + y}{2} \right\|^2} < \sqrt{1 - \frac{\varepsilon^2}{4}} = 1 - \delta(\varepsilon)$$

Example 3.4.8. Consider $(C([0, 1], \|\cdot\|_\infty))$, this is a Banach space, but not an Hilbert space.

Indeed, take f and g as follows:

$$f(x) = \begin{cases} 4x & 0 \leq x \leq \frac{1}{4} \\ -4(x - \frac{1}{2}) & \frac{1}{4} < x \leq \frac{1}{2} \\ 0 & \frac{1}{2} < x \leq 1 \end{cases} \quad g(x) = \begin{cases} 0 & 0 \leq x \leq \frac{1}{2} \\ 4(x - \frac{1}{2}) & \frac{1}{2} < x \leq \frac{3}{4} \\ -4(x - 1) & \frac{3}{4} < x \leq 1 \end{cases}$$

Then $\|f - g\|_\infty = 1 = \|f + g\|_\infty$, and $\|f\|_\infty = \|g\|_\infty = 1$: the parallelogram identity does not hold.

The following theorem is a fundamental brick for the next development of the theory. First let's define a simple concept, the theorem will characterize it.

Definition 3.4.9

Let X be a normed vector space and $V \subset X$ one of his closed non-empty subsets.

For any $x \in X$ the **minimal distance of V from x** is:

$$d(x, V) = \inf_{v \in V} \|x - v\|.$$

This definition is very simple, the distance between a set and an element is the distance of the element from the closest element belonging to the set. Observe that the distance of a set from one of his point is zero.

Theorem 3.4.10 (minimal distance)

Let $(H, \langle \cdot, \cdot \rangle)$ be a Hilbert space, and $V \subset H$ be a closed non-empty subspace.

Then for any $x \in H$ there exists a unique $\bar{v} \in V$ such that:

$$d(x, V) = \|x - \bar{v}\|.$$

Proof. We have to prove that \bar{v} exists unique such that:

$$\inf_{v \in V} \|x - v\| = \|x - \bar{v}\|.$$

Take a point $x \in H$. We are looking for an element which realize the inferior.

Existence:

Consider a sequence $\{v_n\}_{n \in \mathbb{N}} \subset V$ such that:

$$\lim_{n \rightarrow \infty} \|x - v_n\| = d(x, V).$$

Suppose $m > n$ and observe that $\frac{v_n + v_m}{2} \in V$ due to vector space structure. Therefore:

$$\left\| x - \frac{v_n + v_m}{2} \right\| \geq d(x, V)$$

which implies

$$\|2x - (v_m + v_n)\| \geq d(x, V)$$

and then

$$-\|2x - (v_m + v_n)\|^2 \leq -4d(x, V)^2.$$

Now we use this result as follows:

$$\begin{aligned} \|v_m - v_n\|^2 &= \|v_m - x + x - v_n\|^2 \\ &= 2(\|x - v_m\|^2 + \|x - v_n\|^2) - \|2x - (v_m + v_n)\|^2 \\ &\leq 2(\underbrace{\|x - v_n\|^2}_{\rightarrow d} + \underbrace{\|x + v_m\|^2}_{\rightarrow d}) - 4d^2 \\ &\rightarrow 0 \end{aligned}$$

as $m, n \rightarrow +\infty$. Observe that in the second equality the parallelogram identity has been used in the form $\|a + b\|^2 = 2\|a\|^2 + 2\|b\|^2 - \|a - b\|^2$, where $a = x - v_m$ and $b = x - v_n$.

Therefore $\{v_n\}_{n \in \mathbb{N}} \subset V$ is a Cauchy sequence and it converges to a $\bar{v} \in V$, as V is closed by hypothesis.

By the continuity of the norm we have that the inferior is reached, namely:

$$\|x - v^*\| = d(x, V).$$

Uniqueness:

Finally we have to prove that \bar{v} is unique; suppose we have $\|x - \bar{v}_1\| = d(x, V)$ and $\|x - \bar{v}_2\| = d(x, V)$.

Using the parallelogram identity again we have:

$$\|\bar{v}_1 - \bar{v}_2\|^2 = \|\bar{v}_1 - x + x - \bar{v}_2\|^2 = 4d^2 - \|2x - \bar{v}_1 + \bar{v}_2\|^2 \leq 0$$

and thus $\bar{v}_1 = \bar{v}_2$. ■

This theorem holds even if V is a convex, closed, non-empty subset of H .

A space V is convex if for all $u, v \in V$, for any $\lambda \in [0, 1]$ we have:

$$(1 - \lambda)u + \lambda v \in V.$$

The proof argues exactly in the same way, with the only difference when we state that $\frac{v_n + v_m}{2} \in V$. In fact, in the original proof, this holds because we assume V is a *subspace*, but in reality *convexity* is enough since if $v_n, v_m \in V$ then choosing $\lambda = \frac{1}{2}$ we still obtain $\frac{1}{2}v_n + (1 - \frac{1}{2})v_m = \frac{v_n + v_m}{2} \in V$.

3.4.2 Orthogonality and projection

In this chapter we will explain the orthogonality in the context of Hilbert spaces and begin the discussion on projection on subspaces.

General definition As we know from geometry, also in this context we can define orthogonality and orthonormality. In addition, here we see how to characterize the Hilbert spaces.

Definition 3.4.11

We say that x and y are **orthogonal** if

$$\langle x, y \rangle = 0;$$

in such case we write $x \perp y$.

Moreover, if $x \perp y$ and $\|x\| = \|y\| = 1$, then we say that x and y

are **orthonormal**.

Proposition 3.4.12 (Pythagoras' theorem)

Let $x, y \in H$, $x, y \neq 0$.

If $x \perp y$, then:

$$\|x \pm y\|^2 = \|x\|^2 + \|y\|^2.$$

Proof. Consider: $\|x + y\|^2 = \|x\|^2 + 2\langle x, y \rangle + \|y\|^2 = \|x\|^2 + \|y\|^2$.
Same in the case of minus. ■

Extending to set This notion can be extended from points to sets:

Definition 3.4.13

Let $(H, \langle \cdot, \cdot \rangle)$ be a Hilbert space, and $V \subset H$.

Then we define the **orthogonal set** of V as follows:

$$V^\perp := \{x \in H : \langle x, v \rangle = 0, \forall v \in V\}.$$

Proposition 3.4.14

The set V^\perp is always a closed subspace of V .

Moreover, we have:

$$\bar{V}^\perp = V^\perp \text{ and } (\overline{\text{span}V})^\perp = V^\perp.$$

The proof is easy and left as an exercise to the reader.

Definition 3.4.15

Let $(H, \langle \cdot, \cdot \rangle)$ be a Hilbert space, and $V, W \subset H$ be subspaces.

We say that V and W are **orthogonal subspaces**, and we write

$V \perp W$, if:

$$V = W^\perp \text{ and } W = V^\perp$$

that means:

$$v \perp w \quad \forall v \in V, w \in W.$$

In this case we set the subspace **direct sum** of V and W as follows:

$$V \oplus W := \{v + w : v \in V, w \in W\}.$$

Proposition 3.4.16

If $V \perp W$, then $V \cap W = \{0\}$, and any element $x \in V \oplus W$ has a unique representation as $v + w$.

Projections These definitions concludes our introduction to orthogonality. Now we see this notion in action.

Theorem 3.4.17 (projection theorem)

Let $(H, \langle \cdot, \cdot \rangle)$ be a Hilbert space, and $V \subset H$ be a closed subspace. Then we can write H as:

$$H = V \oplus V^\perp.$$

This is also called **orthogonal decomposition** of H .

Observe that, on account of this theorem, if V is a subspace of H , then:

$$(V^\perp)^\perp = \bar{V}.$$

Indeed, we have:

$$\begin{aligned} V^\perp \text{ closed} &\implies H = V^\perp \oplus (V^\perp)^\perp = (V^\perp)^\perp \oplus V^\perp, \\ \bar{V} \text{ closed} &\implies H = \bar{V} \oplus (\bar{V})^\perp = \bar{V} \oplus V^\perp. \end{aligned}$$

Note also that $V^\perp = \{0\}$ if and only if $\bar{V} = H$.

This is because $V^\perp = \{0\}$ so that $(V^\perp)^\perp = H$ and $\bar{V} = H$.

Viceversa, from $\bar{V} = H$ we find $V^\perp = (V^\perp)^\perp = \{0\}$. To sum up, we say that V is dense in H if and only if $V^\perp = \{0\}$.

Proof. Let $x \in H$. Due to minimal distance theorem [3.4.10 on page 263](#), supposing V non-empty, there exists a unique $v \in V$ such that $d(x, V) = \|x - v\|$.

Set now $w := x - v$ and choose $\lambda \neq 0$ such that for any $u \in V$ we have $\lambda \langle w, u \rangle \geq 0$.

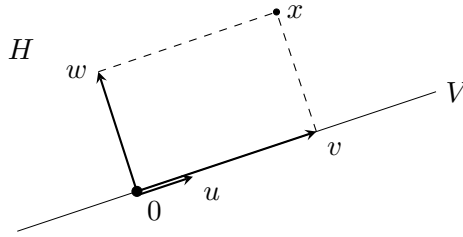


Figure 3.1: Geometric intuition behind the proof of the projection theorem.

Then we have:

$$\begin{aligned}
 \inf_{y \in V} \|x - y\|^2 &= \|w\|^2 \\
 &\leq \left\| x - \underbrace{(v + \lambda u)}_{\in V} \right\|^2 \\
 &= \|w - \lambda u\|^2 \\
 &= \|w\|^2 - 2\lambda \langle w, u \rangle + \lambda^2 \|u\|^2.
 \end{aligned}$$

Then we have $-2\lambda \langle w, u \rangle + \lambda^2 \|u\|^2 \geq 0$, thus $\lambda^2 \|u\|^2 \geq 2\lambda \langle w, u \rangle$ which implies

$$|\langle w, u \rangle| < \frac{|\lambda|}{2} \|u\|^2.$$

Letting λ go to zero, we get:

$$\langle w, u \rangle = 0 \quad \forall u \in V.$$

Thus we have proven that $x = v + w$ where $w \in W = V^\perp$.

We are left to prove that $V = W^\perp$ so that $H = V \oplus V^\perp$.

Observe that $V \subset W^\perp$.

On the other hand if $x \in W^\perp$ then:

$$x = u^* + w^*, \quad u^* \in V, w^* \in W.$$

Thus $x - u^* \in W \cap W^\perp = \{0\}$ and this gives $x = u^*$ so that $x \in V$. ■

On account of this theorem we can give the following definition.

Definition 3.4.18 (orthogonal projectors)

Let $(H, \langle \cdot, \cdot \rangle)$ be a Hilbert space, and $V \subset H$ be a closed subspace. Let also v the unique element of V such that $\|x - v\| = d(x, V)$. Define the following **projector** of H onto V :

$$P_V : H \rightarrow V \quad P_V x = v.$$

In the same way we can define the projector of H onto V^\perp :

$$P_{V^\perp} : H \rightarrow V^\perp \quad P_{V^\perp} x = x - v = x - P_V x.$$

Notice that $x = P_V x + P_{V^\perp} x$.

P_V is linear and bounded, namely:

$$P_V \in \mathcal{B}(H, H).$$

- boundedness is easy:

$$\|x\|^2 = \|P_V x\|^2 + \|P_{V^\perp} x\|^2 \implies \|P_V x\| \leq \|x\|$$

- linearity is a bit harder:

$$\left. \begin{array}{l} \alpha x - P_V(\alpha x) \in V^\perp \\ \alpha x - \alpha P_V(x) \in V^\perp \end{array} \right\} \implies \alpha x - P_V(\alpha x) - [\alpha x - \alpha P_V(x)] \in V^\perp$$

but also

$$\alpha \underbrace{P_V(x)}_{\in V} - \underbrace{P_V(\alpha x)}_{\in V} \in V$$

thus

$$\alpha P_V(x) - P_V(\alpha x) \in V^\perp \cap V = \{0\} \implies \alpha P_V(x) = P_V(\alpha x)$$

A similar argument can be done to prove that $P_V(x) + P_V(y) = P_V(x + y)$.

3.4.3 Dual of a Hilbert Space and Riesz's representation theorem

In this chapter we will introduce a theorem which play an important role in application.

accoppiare al precedente?

Start with an observation. Consider a Hilbert Space, namely $(H, \langle \cdot, \cdot \rangle)$ and fix one of his elements $y \in H$. Define the operator L_y as follows:

$$|L_y x| = |\langle y, x \rangle| \quad \forall x \in H.$$

Then this operator belongs to the dual of H and we will write $L_y \in H$. As $|L_y x| = |\langle y, x \rangle| \leq \|y\| \|x\|$, we have that $\|L_y\|_{H^*} \leq \|y\|$: with the fact that H is reflexive this implies that $\|L_y\|_* = \|y\|$.

Now we have to acknowledge that there exists an isometry $J : H \rightarrow H^*$, $y \mapsto L_y = \langle \cdot, y \rangle$.

This reasoning can be revert: the converse is also true and we have

Theorem 3.4.19 (Riesz's representation theorem)

Let $(H, \langle \cdot, \cdot \rangle)$ be a Hilbert space.

An operator $L \in H^*$ if and only if there exists a unique $y \in H$ such that $Lx = \langle y, x \rangle$ for all $x \in H$.

In other words, $J : H \rightarrow H^*$ is surjective.

Observe that the reflexivity of H can be proven using this theorem. Note also that as $n \rightarrow \infty$ we have

$$\langle x_n, y \rangle \rightarrow \langle x, y \rangle \quad \forall y \in H \text{ if and only if } x_n \rightarrow x.$$

We will prove the “only if” part as the “if” part can be deduced from the introduction.

Proof. Proof of the necessary condition \Leftarrow .

If $L \equiv 0$, then we choose $y = 0$ and the thesis is obtained.

General case.

Consider $L \neq 0$. Then $\text{Ker}(L)$ is a closed proper subspace of H , that is $\{0\} \subsetneq \text{Ker}(L)^\perp$.

Then, applying the projection theorem and then normalizing, we can find $z \in \text{Ker}(L)^\perp$ such that $\|z\| = 1$; then for any $x \in H$ since:

$$L \left(x - \frac{Lx}{Lz} z \right) = Lx - \frac{Lx}{Lz} Lz = 0,$$

we have

$$x - \frac{Lx}{Lz} z \in \text{Ker}(L).$$

a better organization is more suitable for a book

Thus:

$$\left\langle x - \frac{Lx}{Lz}z, z \right\rangle = 0,$$

which implies

$$\langle x, z \rangle = \left\langle \frac{Lx}{Lz}z, z \right\rangle = \frac{Lx}{Lz} \text{ and } Lx = \langle x, (Lz)z \rangle \quad \forall x \in H.$$

We choose $y = (Lz)z \in Y$.

The element y is unique.

If $Lx = \langle y_1, x \rangle = \langle y_2, x \rangle$ for all $x \in H$, then $\langle y_1 - y_2, x \rangle = 0 \quad \forall x \in H$.

Choosing $x = y_1 - y_2$, we have $\|y_1 - y_2\|^2 = 0$, and finally $y_1 = y_2$. ■

We can say that H and H^* is linearly and isometrically isomorphic, through the linear isometry and isomorphism we introduced before J . This isomorphism is called **Riesz map**. It can be defined also as $J : H^* \rightarrow H$ by setting

$$JL = y \text{ where } Lx = \langle x, JL \rangle \quad \forall x \in H.$$

As it is a linear bijection and an isometric isomorphism we have $\|JL\| = \|L\|_*$.

Proposition 3.4.20

Consider the following inner product on H^* :

$$\langle L_1, L_2 \rangle_* := \langle J^{-1}L_1, J^{-1}L_2 \rangle \quad \forall L_1, L_2 \in H^*.$$

Then $(H^*, \langle \cdot, \cdot \rangle_*)$ is a Hilbert space.

The induced norm is the existing norm in H^* :

$$\langle L_1, L_2 \rangle_* = \|JL_1\|^2 = \|L_1\|_*^2$$

because $\|L_1\|_* = \sqrt{\langle L_1, L_1 \rangle_*} = \sqrt{\langle J^{-1}L_1, J^{-1}L_1 \rangle} = \|J^{-1}L_1\| = \|L_1\|_{H^*}$.

Taking advantage of the Riesz's representation theorem you can prove that any Hilbert space is reflexive: this is a really tough exercise.

3.4.4 Proof of the Radon–Nikodym Theorem

The theorem is named after Johann Radon, who proved the theorem for the special case where the underlying space is \mathbb{R}^N in 1913, and after Otto Nikodym who proved the general case in 1930. It is a representation theorem: it provides, under suitable assumptions, a link between two measures.

The theorem is explained in [2.2.3 on page 114](#), for a better use of this book we copy-pasted it here.

Theorem 3.4.21 (Radon–Nikodym)

Let $(\Omega, \mathfrak{M}, \mu)$ be a complete measure space, and ν, μ two measures on (Ω, \mathfrak{M}) .

If μ is σ -finite and $\nu \ll \mu$, then the Radon–Nikodym derivative $\frac{d\nu}{d\mu}$ exists.

This proof belongs to John von Neumann (1903–1957).

Proof. Set up:

For simplicity we consider the w Consider μ, ν finite, namely $\mu(\Omega), \nu(\Omega) < \infty$, such that $\nu \ll \mu$.

Let $\lambda = \mu + \nu$. Then, for every non-negative and measurable function f we have the **additivity with respect to measure**:

$$\int_{\Omega} f d\lambda = \int_{\Omega} f d\mu + \int_{\Omega} f d\nu$$

which is a direct consequence of the definition of Lebesgue interval.

First step, use of the Hölder inequality and the Riesz's representation theorem:

Suppose $f \in L^2_{\lambda}(\Omega)$. Applying the Hölder inequality we have:

$$\begin{aligned} \left| \int_{\Omega} f d\nu \right| &\leq \int_{\Omega} |f| d\nu \\ &\leq \int_{\Omega} |f| d\lambda \\ &\leq \left(\int_{\Omega} |f|^2 d\lambda \right)^{\frac{1}{2}} (\lambda(\Omega))^{\frac{1}{2}} \\ &= \|f\|_{L^2_{\lambda}(\Omega)} (\lambda(\Omega))^{\frac{1}{2}} \end{aligned}$$

$< +\infty$.

Thus $F : f \mapsto \int_{\Omega} f d\nu$ is a linear bounded functional on $L_{\lambda}^2(\Omega)$, from which we deduce:

$$\|F\|_{(L_{\lambda}^2(\Omega))^*} \leq (\lambda(\Omega))^{\frac{1}{2}}.$$

Applying Riesz's representation theorem (3.4.19 on page 269) there exists $g \in L_{\lambda}^2(\Omega)$ such that:

$$F(f) = \int_{\Omega} f d\nu = \int_{\Omega} fg d\lambda \quad \forall f \in L_{\lambda}^2(\Omega).$$

For any $E \in \mathfrak{M}$ such that $\lambda(E) > 0$, consider $f = \mathbf{1}_E$; we have:

$$0 \leq \nu(E) = \int_{\Omega} f d\nu = \int_{\Omega} fg d\lambda = \int_E g d\lambda = F(\mathbf{1}_E) \leq \lambda(E).$$

So $\nu(E) \leq \lambda(E)$ and then:

$$0 \leq \frac{1}{\lambda(E)} \int_E g dx \leq 1.$$

Second step:

As E is arbitrary we have $0 \leq g(x) \leq 1$ almost everywhere with respect to the measure λ , and thus with respect to ν and μ .

We can suppose that there exists $\tilde{\Omega} \in \mathfrak{M}$ such that $0 \leq g(x) \leq 1$ for any $x \in \tilde{\Omega}$.

So $\lambda(\tilde{\Omega}^c) = 0$, and we define $g(x) = 1$ for all $x \in \tilde{\Omega}^c$.

Then, for any $f \in L_{\lambda}^2(\Omega)$:

$$\int_{\Omega} f d\nu = \int_{\Omega} fg d\lambda = \int_{\Omega} fg d\mu + \int_{\Omega} fg d\nu$$

and we can rewrite in this way:

$$\int_{\Omega} (1 - g)f d\nu = \int_{\Omega} fg d\mu.$$

Third step, convergences:

Let $A = \{x \in \Omega : 0 \leq g(x) < 1\}$, and $B = A^c$.

Taking $f = \mathbb{1}_B$ in the previous identity we find $\mu(B) = 0$.
 By hypothesis $\nu \ll \mu$, and thus $\nu(B) = 0$.

Define

$$\Phi_n = \sum_{j=0}^{n+1} g^j.$$

Take now $f = (1 + g + g^2 + \cdots + g^n)\mathbb{1}_E$ in the same identity, with $E \in \mathfrak{M}$, $n \in \mathbb{N}$. We get:

$$\int_E (1 - g^{n+1}) \, d\nu = \int_E \Phi_n \, d\mu.$$

Observe that as $n \rightarrow +\infty$ and $0 \leq 1 - g^{n+1} \leq 1$ in Ω we have $1 - g^{n+1} \rightarrow 1$ almost everywhere with respect to ν and, due to $\nu(B) = 0$:

$$\int_E (1 - g^{n+1}) \, d\nu = \int_{E \cap A} (1 - g^{n+1}) \, d\nu + \underbrace{\int_{E \cap B} (1 - g^{n+1}) \, d\nu}_{=0 \, \forall n} \rightarrow \nu(E \cap A) = \nu(E).$$

First, applying dominated convergence theorem which entails:

$$\lim_{n \rightarrow \infty} \int_E (1 + g^{n+1}) \, d\nu = \nu(E).$$

Second, as $\Phi_n \uparrow \Phi$ monotonically and $\Phi = \lim_n \Phi_n$ is measurable, we can use monotone convergence theorem and obtain:

$$\int_E \Phi_n \, d\mu \rightarrow \int_E \Phi \, d\mu.$$

Thus we can conclude deducing:

$$\nu(E) = \int_E \Phi \, d\mu \quad \forall E \in \mathfrak{M}.$$

Note also that Φ is measurable and belongs to $L^1_\mu(\Omega)$ as ν and μ are finite. ■

3.4.5 Bilinear Forms and Lax Milgram lemma

Here we will generalize the concept of inner product. In fact, the following definition will make it clear that the inner product is an operation which belongs to a much general class.

Definition 3.4.22

Let $(H, \langle \cdot, \cdot \rangle)$ be a Hilbert space.

An application $a : H \times H \rightarrow \mathbb{R}$ is called **bilinear form** if:

$$u \mapsto a(u, v) \text{ is linear for each fixed } v \in H,$$

$$v \mapsto a(u, v) \text{ is linear for each fixed } u \in H.$$

Definition 3.4.23

A bilinear form $a : H \times H \rightarrow \mathbb{R}$ is **coercive** if there exists $\alpha > 0$ such that:

$$a(u, u) \geq \alpha \|u\|^2 \quad \forall u \in H.$$

Proposition 3.4.24

A bilinear form $a : H \times H \rightarrow \mathbb{R}$ is continuous if and only if there exists $M > 0$ such that:

$$|a(u, v)| \leq M \|u\| \|v\| \quad \forall u, v \in H.$$

Observe that the inner product of an Hilbert space is a continuous and coercive bilinear form.

Consider $H = \mathbb{R}^N$ with the euclidean scalar product $\langle \cdot, \cdot \rangle_2$ and a matrix $A \in \mathbb{R}^{N \times N}$. Then $a(x, y) = \langle Ax, y \rangle_2$ for all $x, y \in \mathbb{R}^N$ is a continuous bilinear form.

The following is a lemma which generalize the last representation theorem.

Theorem 3.4.25 (Lax–Milgram)

Let $a : H \times H \rightarrow \mathbb{R}$ be a continuous and coercive bilinear form.

Then for any $l \in H^*$ there exists a unique $u \in H$ such that:

$$a(u, v) = Lv \quad \forall v \in H.$$

Moreover we have

$$\|u\| \leq \frac{1}{\alpha} \|L\|_*$$

where α is the coercivity constant.

Proof. Fix $u \in H$ and observe that:

$$v \in H : v \mapsto a(u, v) \in \mathbb{R}$$

is an element of H^* . Therefore there is a unique $h \in H$ such that:

$$a(u, v) = \langle h, v \rangle \quad \forall v \in H.$$

Thus we can define a linear operator $A : H \rightarrow H$ by setting $Au = h$. This operator is also bounded. Indeed:

$$\|Au\|^2 = \langle Au, Au \rangle = a(u, Au) \leq M \|u\| \|Au\|$$

so that $\|Au\| \leq M \|u\|$.

Moreover, we have:

$$\alpha \|u\|^2 \leq a(u, u) = \langle Au, u \rangle \leq \|Au\| \|u\|$$

so A is also injective.

On the other hand, there is a unique $w \in H$ such that (see Riesz's representation theorem [3.4.19 on page 269](#)):

$$Lv = \langle w, v \rangle \quad \forall v \in V.$$

Therefore the thesis can be reformulated as follows: for any $w \in H$ find a unique $u \in H$ such that:

$$\langle Au, v \rangle = \langle w, v \rangle \quad \forall v \in V$$

so we're left to show that A is surjective.

Let's prove first that $\text{Im}(A)$ is a closed subspace of H .

Indeed let $\{y_n\}_{n \in \mathbb{N}} \subset \text{Im}(A)$ be such that $y_n \rightarrow y$ as $n \rightarrow \infty$.

We know that there exists a unique $u_n \in H$ such that $Au_n = y_n$.

Moreover, we have

$$\alpha \|u_n - u_m\| \leq \|Au_n - Au_m\| = \|y_n - y_m\|$$

therefore $\{u_n\}_{n \in \mathbb{N}}$ converges in H to some u and, because A is continuous, it follows

$$Au_n \rightarrow Au \quad n \rightarrow \infty.$$

Hence $y = Au \in \text{Im}(A)$ and $\text{Im}(A)$ is closed.

We can now prove that A is surjective.

By contradiction suppose that $\text{Im}(A) \subsetneq H$. Then there exists $y_0 \in H$ such that $y_0 \neq 0$ and $y_0 \in \text{Im}(A)^\perp$. However, we have

$$0 = \langle Ay_0, y_0 \rangle = a(y_0, y_0) \geq \alpha \|y_0\|^2 \implies y_0 = 0.$$

This is a contradiction.

Finally, we recall that

$$\alpha \|u\| \leq \|Au\| = \|w\| = \|L\|_*.$$

■

Inner product Even if a scalar product is a continuous and coercive bilinear form, a continuous and coercive bilinear form is not necessarily a scalar product: this because it might be not symmetric.

Due to this point, the Lax–Milgram lemma can be seen as a generalization of the Riesz representation theorem (3.4.17 on page 266).

Proposition 3.4.26

Let $a : H \times H \rightarrow \mathbb{R}$ be a continuous and coercive bilinear form.

If a is symmetric, namely:

$$a(u, v) = a(v, u) \quad \forall u, v \in H.$$

then a is a scalar product in H and its induced norm is equivalent to the norm in H .

Proof. By coercivity and the other hypothesis we have:

$$\alpha \|u\|^2 \leq a(u, u) \leq M \|u\|^2 \quad \forall u \in H$$

where $\|\cdot\|$ is the norm in H .

Hence $a(u, u) \geq 0$ for all $u \in H$ and $a(u, u) = 0$ if and only if $u = 0$.

Thus $a(\cdot, \cdot)$ is a scalar product in H and its induced norm

$$\|u\|_a = \sqrt{a(u, u)}$$

is equivalent to $\|\cdot\|$. ■

An application of the Lax–Milgram lemma All this theory has a practical purpose. Here we will see in action the tools we built solving a differential problem.

Consider the following boundary value problem: find u such that

$$\begin{cases} -(\alpha(x)u' + \beta(x)u)' + \gamma(x)u = f(x) & x \in (a, b) \\ u(a) = u(b) = 0 \end{cases}$$

where $\alpha, \beta \in C^1([a, b])$ and $\gamma, f \in C([a, b])$.

Suppose $w \in C^2([a, b])$ is a classical solution to the problem.

Consider $\phi \in C^\infty((a, b))$ multiply the equation by ϕ and integrate over (a, b) . Integrating by parts we get the identity:

$$\int_a^b [(\alpha(x)w'(x) + \beta(x)w(x))\phi'(x) + \gamma(x)w(x)\phi(x)] dx = \int_a^b f(x)\phi(x) dx.$$

Setting

$$V_0 = \{v \in H^1((a, b)) : v(a) = v(b) = 0\}$$

where $H^1((a, b)) = W^{1,2}((a, b))$, we observe that V_0 is a closed subspace of $H^1((a, b))$. Moreover the scalar product

$$(v_1, v_2) = \int_a^b v_1'(x)v_2'(x) dx$$

induces a norm

$$\|v\|_0 = \|v'\|_2$$

which is equivalent to the standard norm in $H^1((a, b))$.

(This because the Poincaré's inequality $\|v\|_2 \leq (b-a)\|v'\|_2$.)

We can prove that $C_c^\infty((a, b))$ is dense in V_0 (this is a density theorem: density theorems allow to establish approximations) which allows us to deduce that our solution w satisfies the identity

$$\int_a^b [(\alpha(x)w'(x) + \beta(x)w(x))v'(x) + \gamma(x)w(x)v(x)] dx = \int_a^b f(x)v(x) dx \quad \forall v \in V_0.$$

We can now look at our problem from a more general standpoint.

Let us set, for all $u, v \in V_0$

$$a(u, v) = \int_a^b [(\alpha(x)u'(x) + \beta(x)u(x))v'(x) + \gamma(x)u(x)v(x)] dx$$

and observe that $a : V_0 \times V_0 \rightarrow \mathbb{R}$ is a bilinear form.

Thus we can give to the original problem a weaker formulation, namely find $w \in V_0$ such that

$$a(w, v) = Lv \quad \forall v \in V_0$$

where $L \in V_0^*$ is defined by

$$Lv = \int_a^b f(x)v(x) dx \quad \forall v \in V_0.$$

Now we want to apply Lax–Milgram lemma. Let's check the hypothesis.

Observe that $a(\cdot, \cdot)$ is continuous as

$$\begin{aligned} |a(u, v)| &\leq \int_a^b |(\alpha(x)u'(x) + \beta(x)u(x))v'(x) + \gamma(x)u(x)v(x)| dx \\ &\leq \|\alpha\|_\infty \int_a^b |u'v'| dx + \|\beta\|_\infty \int_a^b |uv'| dx + \|\gamma\|_\infty \int_a^b |uv| dx \\ &\leq \|\alpha\|_\infty \|u'\|_2 \|v'\|_2 + \|\beta\|_\infty \|u\|_2 \|v'\|_2 + \|\gamma\|_\infty \|u\|_2 \|v\|_2 \end{aligned}$$

for all $u, v \in V_0$.

Then, using again the Poincaré's inequality we have:

$$|a(u, v)| \leq \|\alpha\|_\infty \|u'\|_2 \|v'\|_2 + (b-a) \|\beta\|_\infty \|u'\|_2 \|v'\|_2 + (b-a)^2 \|\gamma\|_\infty \|u'\|_2 \|v'\|_2.$$

Hence, setting

$$M = \|\alpha\|_\infty + (b-a) \|\beta\|_\infty + (b-a)^2 \|\gamma\|_\infty,$$

we have

$$|a(u, v)| \leq M \|u\|_0 \|v\|_0 \quad \forall u, v \in V_0.$$

We can add other assumption to the coefficients α , β and γ that hold for any $x \in [a, b]$, like the following:

$$\alpha_0 = \min_{x \in [a, b]} \alpha(x) > (b - a) \|\beta\|_\infty \quad \text{and} \quad \gamma(x) \geq 0.$$

Collecting the results we obtained we have:

$$\begin{aligned} a(u, u) &= \int_a^b \alpha(x)(u'(x))^2 dx + \int_a^b \beta(x)u(x)u'(x) dx + \int_a^b \gamma(x)(u(x))^2 dx \\ &\geq \alpha_0 \|u'\|_2 + \int_a^b \beta(x)u(x)u'(x) dx. \end{aligned}$$

We can find an upper bound for the middle term as well:

$$\left| \int_a^b \beta(x)u(x)u'(x) dx \right| \leq \|\beta\|_\infty \|u\|_2 \|u'\|_2 \leq (b - a) \|\beta\|_\infty \|u'\|_2^2,$$

so we get:

$$a(u, u) \geq [\alpha_0 - (b - a) \|\beta\|_\infty] \|u'\|_2$$

we can set:

$$\alpha = [\alpha_0 - (b - a) \|\beta\|_\infty] > 0.$$

Then, finally we have:

$$a(u, u) \geq \alpha \|u\|_0.$$

and we can conclude that $a(\cdot, \cdot)$ is coercive.

Now we can apply the Lax–Milgram lemma and we deduce the following proposition: If α , β , γ , $f \in C([a, b])$ and

$$\min_{x \in [a, b]} \alpha(x) > (b - a) \|\beta\|_\infty \quad \text{and} \quad \gamma(x) \geq 0 \quad \forall x \in [a, b]$$

then there exists a unique $u \in V_0$ such that:

$$a(u, v) = Lv \quad \forall v \in V_0.$$

The solution u given by the Lax–Milgram lemma is a weak solution to the original boundary value problem.

We can further weaken the assumptions on the coefficients by taking $\alpha, \beta, \gamma \in L^\infty((a, b))$.

In addition f can be taken in V_0^* : assume $f \in L^2((a, b))$ or even a linear bounded functional link, for instance,

$$Lv = v(x_0) \quad \forall v \in V_0$$

where $x_0 \in (a, b)$. It's easy to check that $L \in V_0^*$.

Consider now the linear application $K : L^2((a, b)) \rightarrow L^2((a, b))$ defined by $K : f \mapsto u$ where u satisfies

$$a(u, v) = \int_a^b f v \quad \forall v \in V_0.$$

We have that $\text{Im}(K) \subseteq V_0$ and we know that

$$V_0 \hookrightarrow C([a, b]) \hookrightarrow L^2((a, b)).$$

Therefore

$$V_0 \hookrightarrow L^2((a, b)).$$

Hence K is a compact operator.

From Lax–Milgram we have also a stability estimate:

$$\|u\|_0 \leq \frac{1}{\alpha} \|L\|_{V_0^*}.$$

If $f \in L^2((a, b))$ then, using Poincaré's inequality:

$$|Lv| = \left| \int_a^b f v \, dx \right| \leq \|f\|_2 \|v\|_2 \leq (b-a) \|f\|_2 \|v\|_0$$

so that

$$\|L\|_* \leq (b-a) \|f\|_2$$

and

$$\|u\|_0 \leq \frac{(b-a)}{\alpha} \|f\|_2.$$

If

$$Lv = v(x_0) \quad \forall v \in V_0$$

where $x_0 \in (a, b)$, then, using Hölder's inequality:

$$|Lv| = |v(x_0)| = \left| \int_a^{x_0} v'(y) \, dy \right| \leq (b-a)^{\frac{1}{2}} \|v\|_0$$

so that

$$\|L\|_* \leq (b-a)^{\frac{1}{2}}$$

and

$$\|u\|_0 \leq \frac{(b-a)^{\frac{1}{2}}}{\alpha}.$$

The bilinear form $a(\cdot, \cdot)$ is not symmetric because of β .

If $\beta \equiv 0$ then we deal with following boundary value problem

$$\begin{cases} -(\alpha(x)u')' + \gamma(x)u = f(x) & x \in (a, b) \\ u(a) = u(b) = 0 \end{cases}$$

and its weak formulation has a unique solution provided that, for instance

$$\alpha = \min_{x \in [a, b]} \alpha(x) > 0, \quad \gamma(x) \geq 0 \quad x \in [a, b].$$

However, in this case we can just apply the Riesz representation theorem since $a(\cdot, \cdot)$ is an inner product in V_0 whose induced norm is equivalent to $\|\cdot\|_0$.

3.4.6 Bases in Hilbert spaces

In this section we will see that elements of Hilbert spaces can be represented with respect to a series of vectors, known as basis. The idea is the same of Linear Algebra, but now we are dealing with infinite dimensional spaces, thus things require a higher level of technicality and not every operation can be done without some specific assumptions.

Definition 3.4.27

Let $(H, \langle \cdot, \cdot \rangle)$ be a Hilbert spaces.

A set of orthogonal vectors $E \subset H$ is **complete** if, given a vector $x \in H$ such that $\langle x, u \rangle = 0$ for all $u \in E$, then $x = 0$.

A subset is an **orthogonal basis** if E consists of orthogonal vectors and is complete.

An orthonormal basis whose vectors have unitary norm (an then

consists of orthonormal vectors) is an **orthonormal basis**.

If E is an orthonormal basis of H then $\overline{\text{span}E}^\perp = E^\perp = \{0\}$. Thus $\overline{\text{span}E} = H$.

One can prove that, if H is separable, a countable orthogonal basis is a Schauder basis.

Theorem 3.4.28

If a Hilbert space contains at least two distinct vectors, then it has an orthonormal basis.

We are going to prove the theorem in a constructive way, using Gram-Schmidt orthogonalization, supposing H separable.

If H is non-separable,⁵⁷ then the proof requires Zorn's Lemma.

Proof. Let $\{v_n\}_{n \in \mathbb{N}} \subset H$ be a dense set.

Set

$$F_h = \langle v_1, \dots, v_h \rangle \quad \forall h \in \mathbb{N}_0.$$

Observe that for all $h \in \mathbb{N}_0$, F_h is a closed subspace of H and $F_h \subset F_{h+1}$.

Take $v_1 \in V_1$ and set

$$e_1 = \frac{v_1}{\|v_1\|}.$$

In $V_2 \supsetneq V_1$, thanks to the projection theorem, we can find $e_2 \in V_1^\perp$ such that $\|e_2\| = 1$.

Therefore $\{e_1, e_2\}$ is an orthonormal basis in V_2 .

Then we proceed by induction, getting an orthonormal basis of H . ■

Theorem 3.4.29

All the orthonormal basis in a Hilbert space have the same cardinality.

Definition 3.4.30

The cardinality of an orthonormal basis of an Hilbert space $(H, \langle \cdot, \cdot \rangle)$ is called **orthogonal dimension** of H , and is denoted by $\dim_\perp H$.

If H is also separable, then $\dim_\perp H \leq \aleph_0$. Indeed, if $u, v \in D$, $u \neq v$, then $\|u - v\|^2 = \|u\|^2 + \|v\|^2 = 2 > 0$. So if $\dim_\perp H \leq \aleph_0$ we cannot find a countable dense set.

⁵⁷“Non-separable Hilbert spaces are quite exotic.” - M. G.

Example 3.4.31. Let $H = l^2$, with $\langle x, y \rangle = \sum_{i \in \mathbb{N}} x_i y_i$. Define the set of sequences $\{e^j\}_{j \in \mathbb{N}}$ as follows:

$$e_i^j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

Then $\{e_j\}_{j \in \mathbb{N}}$ is an orthonormal basis of l^2 and we have:

$$x = \sum_{j \in \mathbb{N}} x_j e^j \quad \forall x = \{x_n\}_{n \in \mathbb{N}} \in l^2.$$

Example 3.4.32. If $H = L^2((-\pi, \pi))$ then we can prove that

$$\left\{ \frac{1}{\sqrt{2\pi}}, \frac{\sin(nt)}{\sqrt{\pi}}, \frac{\cos(nt)}{\sqrt{\pi}} \right\}_{n \in \mathbb{N}_0}$$

is an orthonormal basis in H .

Fourier basis Now we will use the tools we developed to introduce the Fourier series, which allows to represent continuous function in a different way. In particular, Fourier series is a very powerful tool in engineering since it helps breaking down almost any kind of function in a sum of smooth pieces, which are much easier to treat. This is also very useful when solving partial differential equations for the same reason.

Theorem 3.4.33 (Bessel's inequality)

Let $\{u_n\}_{n \in \mathbb{N}}$ be a set of orthonormal vectors in H .

Then, for all $x \in H$, we have

$$\sum_{n \in \mathbb{N}} |\langle x, u_n \rangle|^2 \leq \|x\|^2$$

Proof. The absolute value is only really needed when the field is \mathbb{C} , while in this treaty we only used \mathbb{R} . We will however provide a complete proof for the sake of understanding where each term comes from. The reader may revise some concepts, like conjugate in complex numbers, before reading the proof.

$$0 \leq \left\| x - \sum_{n=0}^k \langle x, u_n \rangle u_n \right\|^2$$

$$\begin{aligned}
&= \left\langle x - \sum_{n=0}^k \langle x, u_n \rangle u_n, x - \sum_{n=0}^k \langle x, u_n \rangle u_n \right\rangle \\
&= \langle x, x \rangle - 2 \underbrace{\left\langle x, \sum_{n=0}^k \langle x, u_n \rangle u_n \right\rangle}_{(*)} + \underbrace{\left\langle \sum_{n=0}^k \langle x, u_n \rangle u_n, \sum_{n=0}^k \langle x, u_n \rangle u_n \right\rangle}_{(**)} = (***)
\end{aligned}$$

The second term becomes

$$\begin{aligned}
(*) &= \sum_{n=0}^k \langle x, \langle x, u_n \rangle u_n \rangle && \text{(sesquilinearity)} \\
&= \sum_{n=0}^k \overline{\langle \langle x, u_n \rangle x, u_n \rangle} = \sum_{n=0}^k \overline{\langle x, u_n \rangle \langle x, u_n \rangle} && \text{(conjugate symmetry)} \\
&= \sum_{n=0}^k \langle u_n, x \rangle \overline{\langle x, u_n \rangle} && \text{(conjugate)} \\
&= \sum_{n=0}^k |\langle x, u_n \rangle|^2 && \text{(conjugate product)}
\end{aligned}$$

While the third

$$(**) = \left\| \sum_{n=0}^k \langle x, u_n \rangle u_n \right\|^2 \leq \sum_{n=0}^k |\langle x, u_n \rangle|^2 \underbrace{\|u_n\|^2}_{=1}$$

Thus

$$(***) \leq \|x\|^2 - \sum_{n=0}^k |\langle x, u_n \rangle|^2$$

Letting k to infinity we get the thesis. ■

Definition 3.4.34

Let $(H, \langle \cdot, \cdot \rangle)$ be a separable Hilbert space and $\{u_n\}_{n \in \mathbb{N}}$ one of its orthonormal basis.

We define the **Fourier coefficients** of $x \in H$ with respect to $\{u_n\}_{n \in \mathbb{N}}$ as follows:

$$x_n := \langle x, u_n \rangle.$$

Theorem 3.4.35

Let $(H, \langle \cdot, \cdot \rangle)$ be a separable Hilbert space and $\{u_n\}_{n \in \mathbb{N}}$ one of its orthonormal basis.

Then we can express any $x \in H$ with the following series:

$$x = \sum_{n \in \mathbb{N}} \langle x, v_n \rangle v_n.$$

The series is known as **Fourier series**.

Observe that we can also write:

$$x = \sum_{n \in \mathbb{N}} x_n u_n \quad \forall x \in H \quad \text{and} \quad \langle x, y \rangle = \sum_{n \in \mathbb{N}} x_n y_n \quad \forall x, y \in H.$$

Proof. Observe that thanks to Bessel's inequality, $\sum_{n \in \mathbb{N}} |\langle x, v_n \rangle|^2 = \sum_{n \in \mathbb{N}} |x_n|^2$ converges.

Then the sequence of partial sums $S_n = \sum_{j=0}^n x_j v_j$ is fundamental:

$$S_n - S_m = \left\| \sum_{j=m+1}^n \langle x, v_j \rangle v_j \right\|^2 \leq \sum_{j=m+1}^n |\langle x, v_j \rangle|^2 \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Hence $\sum_{n \in \mathbb{N}} x_n$ converges in H to some \tilde{x} .

Fix $m \in \mathbb{N}$ for any $n > m$ we have:

$$\left\langle x - \sum_{j=0}^n x_j v_j, v_m \right\rangle = x_m - \left\langle \sum_{j=0}^n x_j v_j, v_m \right\rangle = x_m - x_m = 0.$$

Taking $n \rightarrow +\infty$, we have $\langle x - \tilde{x}, v_m \rangle = 0$ for all $m \in \mathbb{N}$, and finally $x = \tilde{x}$ because $\{u_n\}_{n \in \mathbb{N}}$ is complete. ■

Theorem 3.4.36 (Parseval identity)

Let $(H, \langle \cdot, \cdot \rangle)$ be a separable Hilbert space and $\{u_n\}_{n \in \mathbb{N}}$ one of its orthonormal basis.

Then we can write the inner product as the series of the products of the Fourier coefficients, namely:

$$\langle x, y \rangle = \sum_{n \in \mathbb{N}} \langle x, v_n \rangle \langle y, v_n \rangle \quad \forall x, y \in H.$$

This is known as **Parseval identity**.

Proof. Thanks to orthonormality, for any $n \in \mathbb{N}$ fixed, we have:

$$\left\langle \sum_{j=1}^n x_j v_j, \sum_{j=1}^n x_j v_j \right\rangle = \sum_{j=0}^n \langle x, v_j \rangle \langle y, v_j \rangle$$

Letting $n \rightarrow +\infty$, the first term converges to $\langle x, y \rangle$ since the scalar product is continuous in $H \times H$.

The second term is also an inner product in l^2 , being continuous converges to $\sum_{n \in \mathbb{N}} x_n y_n$.

We got the thesis. ■

From the previous two results we have the following:

Corollary 3.4.37

Let $(H, \langle \cdot, \cdot \rangle)$ be a separable Hilbert space.

If $\{u_n\}_{n \in \mathbb{N}}$ is an orthonormal basis on H then:

$$u_n \rightharpoonup 0 \text{ as } n \rightarrow \infty.$$

However $\{u_n\}_{n \in \mathbb{N}}$ does not converges strongly to 0.

Proof. For any $x \in H$, by Parseval identity we obtain:

$$\|x\|^2 = \sum_{n \in \mathbb{N}} x_n^2$$

then $x_n \rightarrow 0$ which is $\langle x, u_n \rangle \rightarrow 0$ for each $x \in H$ as $n \rightarrow \infty$. Thus $u_n \rightharpoonup 0$. As $\|u_n\| = 1$, the sequence cannot converge strongly to 0. ■

Example 3.4.38. We see that $H = L^2((-\pi, \pi))$ has the following orthonormal basis:

$$\left\{ \frac{1}{\sqrt{2\pi}}, \frac{\sin(nt)}{\sqrt{\pi}}, \frac{\cos(nt)}{\sqrt{\pi}} \right\}_{n \in \mathbb{N}_0}.$$

Therefore any function $f \in L^2((-\pi, \pi))$ can be written as a sum of its Fourier series, namely

$$f(t) = \frac{a_0}{\sqrt{2\pi}} + \sum_{n=1}^{\infty} \frac{a_n}{\sqrt{\pi}} \cos(nt) + \sum_{n=1}^{\infty} \frac{b_n}{\sqrt{\pi}} \sin(nt)$$

and the convergence is in $L^2((-\pi, \pi))$.

The Fourier coefficient are given by:

$$a_0 = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} f(t) dt,$$

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos(nt) dt, \quad b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin(nt) dt.$$

One can prove that if $f \in \mathcal{C}^2([-\pi, \pi])$ then its Fourier series converges to f uniformly.

Just integrate by part the integrals defining a_n and b_n twice.

Finite dimensional projectors Here we will see how we can also project elements onto finite dimensional spaces. This will be useful to make many useful counterexamples and understand some concepts.

Let $(H, \langle \cdot, \cdot \rangle)$ be a Hilbert space, and $\{v_n\}_{n \in \mathbb{N}}$ be an orthonormal basis.

Set

$$H_N = \text{span}\{v_1, \dots, v_N\}.$$

This is a closed subspace: an explicit representation of the projection on it is:

$$P_{H_N} x = \sum_{n=1}^N \langle x, v_n \rangle v_n = \sum_{n=1}^N x_n v_n.$$

Notice that $P_{H_N} \in \mathcal{K}(H)$ since it is finite-rank, and

$$\|x - P_{H_N} x\| \rightarrow 0$$

as $N \rightarrow +\infty$, but $\{P_N\}_{N \in \mathbb{N}_0}$ does not converge in $\mathcal{B}(H)$ if $\dim_{\perp} H = \aleph_0$. Indeed it would converge to the identity operator which is not compact while the projectors P_{H_N} are.

The space l^2 is a model of a separable Hilbert space

Remember that any separable Hilbert space has a Schauder basis.

bla bla
bla

Proposition 3.4.39

Let $(H, \langle \cdot, \cdot \rangle)$ be a separable Hilbert space.

Then it is linearly isomorphic and isometric to l^2 .

Proof. Let $\{v_n\}_{n \in \mathbb{N}} \subset H$ be an orthonormal basis of H . The mapping $\mathcal{F} : H \rightarrow l^2$, defined by

$$x \mapsto \mathcal{F}x = \{x_n\}_{n \in \mathbb{N}}$$

where x_n is the n -th Fourier coefficient of x , is well defined, due to Bessel's inequality.

Notice that it is also injective and surjective.

Let $\{a_n\}_{n \in \mathbb{N}} \in l^2$ and consider

$$\sum_{n \in \mathbb{N}} a_n u_n$$

where $\{u_n\}_{n \in \mathbb{N}}$ is an orthonormal basis in H .

The series converges to some $x \in H$, indeed:

$$\left\| \sum_{n \in \mathbb{N}} a_n u_n \right\|^2 \leq \sum_{n \in \mathbb{N}} |a_n|^2 \|u_n\|^2 = \sum_{n \in \mathbb{N}} |a_n|^2 = \|\{a_n\}\|_{l^2}^2 < +\infty$$

Moreover, for any fixed $m \in \mathbb{N}$ we have:

$$\left\langle \sum_{k=0}^n a_k u_k, u_m \right\rangle = a_m$$

for $n > m$.

Letting n go to ∞ we find $x_m = \langle x, u_m \rangle = a_m$.

Thus

$$\mathcal{F}x = \{a_n\}_{n \in \mathbb{N}}.$$

Finally we also use the bounded inverse mapping theorem (see [3.2.43 on page 224](#)) to prove that the inverse is also bounded. ■

3.4.7 Linear Bounded Operators in Hilbert Spaces

Now we will generalize one of the most powerful theorem of Linear Algebra to infinite dimensional case. The crucial element we have to build is about how to calculate the norm of an operator. In case of bounded operators we can use a bilinear form, the scalar product.

Proposition 3.4.40

Let $(H, \langle \cdot, \cdot \rangle)$ be a separable Hilbert space, and $T \in \mathcal{B}(H)$. Then we have:

$$\|T\|_{\mathcal{B}(H)} = \sup_{\|x\|=1, \|y\|=1} |\langle Tx, y \rangle|.$$

Proof. Observe that (see the CS inequality 3.4.3 on page 259)

$$|\langle Tx, y \rangle| \leq \|Tx\| \|y\| \leq \|T\| \|x\| \|y\| \implies \sup_{\|y\|=\|x\|=1} |\langle Tx, y \rangle| \leq \|T\|.$$

On the other hand, taking $y = \frac{Tx}{\|Tx\|}$, with $Tx \neq 0$, we have the equality:

$$\sup_{\|y\|=\|x\|=1} |\langle Tx, y \rangle| \geq \sup_{\|x\|=1} \left| \left\langle Tx, \frac{Tx}{\|Tx\|} \right\rangle \right| = \sup_{\|x\|=1} \frac{\|Tx\|^2}{\|Tx\|} = \|T\|$$

■

Let $\{v_n\}_{n \in \mathbb{N}}$ be an orthonormal basis in $(H, \langle \cdot, \cdot \rangle)$, separable Hilbert space. Then let

$$x = \sum_{n \in \mathbb{N}} x_n v_n.$$

The sequence of partial sums converges in H to x , and T is continuous, thus

$$Tx = \sum_{n \in \mathbb{N}} x_n T v_n.$$

But in turn $T v_n = \sum_{m \in \mathbb{N}} \langle T v_n, v_m \rangle v_m$.

Set $T_{nm} = \langle T v_n, v_m \rangle$ to have $Tx = \sum_{n \in \mathbb{N}} \sum_{m \in \mathbb{N}} x_n T_{nm} v_m$.

Therefore, being $y = \sum_{n \in \mathbb{N}} y_n v_n$, we have $\langle Tx, y \rangle = \sum_{n, m \in \mathbb{N}} T_{nm} x_n y_m$, and:

$$\|T\|_{\mathcal{B}(H)} = \sup_{\substack{\|\{x_n\}\|_l^2=1 \\ \|\{y_n\}\|_l^2=1}} \left| \sum T_{nm} x_n y_m \right|.$$

Where T_{nm} are the Fourier coefficients of T .

Definition 3.4.41

Let $(H, \langle \cdot, \cdot \rangle)$ be a separable Hilbert space.

We say that $T \in \mathcal{B}(H)$ is **symmetric** if

$$\langle Tx, y \rangle = \langle x, Ty \rangle \quad \forall x, y \in H.$$

Observe that T is symmetric if and only if $T_{nm} = T_{mn}$.

Proposition 3.4.42

If $T \in \mathcal{B}(H)$ is symmetric, then $\|T\|_{\mathcal{B}(H)} = \sup_{\|x\|=1} |\langle Tx, x \rangle|$.

Proof. Set

$$c = \sup_{\|x\|=1} |\langle Tx, x \rangle|;$$

it is easy to realize that (see the CS inequality 3.4.3 on page 259) $c \leq \|T\|_{\mathcal{B}(H)}$.

Vice-versa we have, for any $x, y \in H$:

$$\begin{aligned} 4 \langle Tx, y \rangle &= \langle T(x + y), x + y \rangle - \langle T(x - y), x - y \rangle \\ &\leq c(\|x + y\|^2 + \|x - y\|^2) \\ &= 2c(\|x\|^2 + \|y\|^2) \end{aligned}$$

where at the last equality we have applied the parallelogram identity.

Suppose now $Tx \neq 0$ (otherwise the proof is trivial) and take

$$y = \frac{\|x\|}{\|Tx\|} Tx.$$

Then we have:

$$\|x\| \|Tx\| \leq c \|x\|^2$$

which implies

$$\|Tx\| \leq c \|x\|$$

and then

$$\|T\|_{\mathcal{B}(H)} \leq c.$$

Therefore $\|T\|_{\mathcal{B}(H)} = c$ and this concludes the proof. ■

Example 3.4.43. Let $V \subset H$ be a closed subspace. The projector $P_V \in \mathcal{B}(H)$ is symmetric, indeed:

$$\begin{aligned}\langle P_V x, y \rangle &= \langle P_V x, P_V y + P_{V^\perp} y \rangle \\ &= \langle P_V x, P_V y \rangle \\ &= \langle P_V x + P_{V^\perp} x, P_V y \rangle \\ &= \langle x, P_V y \rangle.\end{aligned}$$

Moreover, if V is finite dimensional, notice that P_V is compact and symmetric.

Example 3.4.44. Consider the space $H = L^2(\Omega)$ where $\Omega \in \mathcal{L}(\mathbb{R}^N)$. Let $G \in L^2(\Omega \times \Omega)$, and K_G be the associated Hilbert–Schmidt operator:

$$K_G u(t) = \int_0^1 G(t, s) u(s) \, d\lambda.$$

We already know that $K_G \in \mathcal{K}(H)$. You can prove that: K_G is symmetric if and only if $G(t, s) = G(s, t)$ a.e. in $\Omega \times \Omega$.

Eigenvalues and eigenvectors

Definition 3.4.45

Let $T \in \mathcal{B}(H)$.

A real number $\lambda \in \mathbb{R}$ is an **eigenvalue** for T if there exists $u \in H$, such that $u \neq 0$ and:

$$Tu = \lambda u.$$

In this case u is the **eigenvector** of T associated with λ .

The linear subspace generated by all the eigenvectors associated with λ is called **eigenspace** of λ and is denoted with E_λ .

Proposition 3.4.46

Any eigenspace E_λ is closed.

You can prove this result. Do it!⁵⁸

Moreover for any eigenvalue and its eigenvector we have

$$\frac{\|Tu\|}{\|u\|} = |\lambda|$$

⁵⁸Hint: use the continuity of T .

and

$$|\lambda| \leq \|T\|_{\mathcal{B}(H)}.$$

Definition 3.4.47

The dimension of the eigenspace associated to λ , namely $\dim E_\lambda$, is called **geometric multiplicity** of λ .

Notice that in Hilbert spaces there is no such thing as algebraic multiplicity of an eigenvalue.

Proposition 3.4.48

Let $T \in \mathcal{B}(H)$ symmetric.

Given two distinct eigenvalues of T , their associated eigenvectors are mutually orthogonal, and therefore are linearly independent.

Moreover we have also that the eigenvalues of T are at most countable.

Proof. Let u_1 and u_2 be the eigenvector associated with λ_1 and λ_2 , where $\lambda_1 \neq 0$ (otherwise just swap the indexes).

Then we have:

$$\langle u_1, u_2 \rangle = \frac{1}{\lambda_1} \langle Tu_1, u_2 \rangle = \frac{1}{\lambda_1} \langle u_1, Tu_2 \rangle = \frac{\lambda_2}{\lambda_1} \langle u_1, u_2 \rangle.$$

As $\frac{\lambda_2}{\lambda_1} \neq 0$, we have that necessary $\langle u_1, u_2 \rangle = 0$.

Finally, as H is separable, the eigenvalues are at most countable. ■

Compactness

Proposition 3.4.49

Let $K \in \mathcal{K}(H)$ be symmetric.

Then $\|K\|_{\mathcal{B}(H)}$ or $-\|K\|_{\mathcal{B}(H)}$ is an eigenvalue of K .

Proof. First, observe that if $\|K\|_{\mathcal{B}(H)} = 0$ then the thesis is trivial.

Consider $c = \|K\|_{\mathcal{B}(H)}$ and suppose $c > 0$. We know that $c = \sup_{\|x\|=1} |\langle Kx, x \rangle|$ and thus, by definition, there exists a maximizing sequence $\{x_n\}_{n \in \mathbb{N}} \subset H$ such that $\|x_n\| = 1$ and

$$|\langle Kx_n, x_n \rangle| \rightarrow c.$$

Since $\{x_n\}$ is bounded and H is reflexive there exist $\{x_{n_h}\}_{h \in \mathbb{N}}$ such that $x_{n_h} \rightharpoonup x$ and $\|x\| < 1$ (see Eberlein–Šmulian theorem 3.3.38 on page 249). Moreover, via weak-strong continuity of K we have $Kx_{n_h} \rightarrow Kx$ (see theorem 3.3.45 on page 251).

Hence

$$|\langle Kx_{n_h}, x_{n_h} \rangle| \rightarrow |\langle Kx, x \rangle| = c \neq 0,$$

and $x \neq 0$.

Suppose, without loss of generality, $c = \langle Kx, x \rangle$. Observe that:

$$\begin{aligned} \|Kx - cx\|^2 &= \|Kx\|^2 + c^2 \|x\|^2 - 2c \langle Kx, x \rangle \\ &= \|Kx\|^2 + c^2 \|x\|^2 - 2c^2 \\ &\leq 2c^2 \|x\|^2 - 2c^2 \\ &\leq 0 \end{aligned}$$

So $Kx = cx$, that is x is the eigenvector associated with c . ■

Therefore, if $K \in \mathcal{K}(H)$ symmetric, its norm can be equivalently defined as follows:

$$\|K\|_{\mathcal{B}(H)} = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } K\}.$$

Proposition 3.4.50

Let $K \in \mathcal{K}(H)$ and $\lambda \neq 0$ be an eigenvalue of K .

Then the geometric multiplicity of λ is finite, namely:

$$\dim E_\lambda \leq \infty.$$

Proof. Suppose by contradiction that the dimension is infinite: let $\{v_n\}_{n \in \mathbb{N}}$ be an orthonormal basis of E_λ .

Then $Kv_n = \lambda v_n \forall n \in \mathbb{N}$, $v_n \rightharpoonup 0$ and $\|v_n\| = 1$ for all $n \in \mathbb{N}$.

Thus $Kv_n \rightarrow 0$ but $v_n \not\rightarrow 0$, which is absurd. ■

The spectrum

Definition 3.4.51

Let $T \in \mathcal{B}(H)$ be symmetric.

Define the **spectrum** of T as:

$$\sigma(T) := \{\lambda \in \mathbb{R} : \lambda \text{ is an eigenvalue of } T\}.$$

Moreover define the **resolvent** of T :

$$\mathcal{R}(T) := \sigma(T)^c.$$

It is easy to see that if $\mu \in \mathcal{R}(K)$ then $\text{Ker}(\mu I - K) = \{0\}$ and so is injective. Note also that if $T \in \mathcal{K}(H)$ is symmetric then $\sigma(T) \neq 0$.

Theorem 3.4.52 (spectral theorem)

Let $K \in \mathcal{K}(H)$ be symmetric. Then:

- either $\sigma(K)$ is finite,
- or $\sigma(K)$ is a sequence $\{\lambda_n\}_{n \in \mathbb{N}}$ such that $\lambda_n \rightarrow 0$.

In both cases 0 can be either an eigenvalue or not.

Moreover, we can always choose the eigenvectors in such a way that they form an orthonormal basis on H .

If H has infinite dimension and $\sigma(T)$ is finite the 0 is necessarily an eigenvalue and E_0 has infinite dimension. Otherwise if $\sigma(T)$ is a sequence and 0 is an eigenvalue then E_0 can have either finite or infinite dimension.

Proof. Let λ_1 be an eigenvalue such that (see 3.4.49 on page 292) $|\lambda_1| = \|T\|_{\mathcal{B}(H)}$, and E_{λ_1} be its eigenspace.

If $\lambda_1 = 0$, then $\text{Ker}(T) = H$ and $T \equiv 0$. Therefore $\sigma(T) = 0$, any orthonormal basis of H is a basis of the eigenspace, and the theorem is proven.

If $\lambda_1 \neq 0$, we know that E_{λ_1} has finite dimension and setting $H_1 = E_{\lambda_1}^\perp$ we have:

$$H = E_{\lambda_1} \oplus E_{\lambda_1}^\perp.$$

We have

$$\langle Ty, x \rangle = \langle y, Tx \rangle = \langle y, \lambda_1 x \rangle = 0 \quad \forall y \in H, x \in E_{\lambda_1}$$

and so $T(H_1) \subset H_1$ (prove it).

Consider now $T_1 = T|_{H_1} \in \mathcal{K}(H_1)$: if $\|T_1\|_{\mathcal{B}(H_1)} = 0$ we have finished, otherwise we consider the non-zero eigenvalue λ_2 such that:

$$|\lambda_2| = \|T_1\|_{\mathcal{B}(H_1)} \leq |\lambda_1| = \|T\|_{\mathcal{B}(H)}$$

where λ_2 has finite geometric multiplicity.

Now set $H_2 = E_{\lambda_2}^\perp$: we have

$$H_1 = E_{\lambda_2} \oplus H_2.$$

Consider $T_2 = T_1|_{H_2}$: either we find that the norm of the restricted operator is zero after a finite number of steps, and the proof is finished, or we can construct a sequence of non-zero eigenvalues $\{\lambda_n\}_{n \in \mathbb{N}_0}$ such that

$$|\lambda_{n+1}| \leq |\lambda_n| \quad \forall n \in \mathbb{N}_0.$$

To conclude the proof of this case, first we have to show that $\lambda_n \rightarrow 0$.

Suppose by contradiction that

$$\Lambda = \inf_{n \in \mathbb{N}_0} |\lambda_n| > 0.$$

We want it to be 0. For each E_{λ_n} consider an orthonormal basis of eigenvectors and denote such whole sequence with $\{w_k\}_{k \in \mathbb{N}_0}$, obviously it is bounded and we have $\|w_k\| = 1$.

As H is reflexive there exists (see [3.3.37 on page 248](#)) a subsequence $\{w_{k_h}\}$ such that $w_{k_h} \rightharpoonup w$.

Then

$$T w_{k_h} = \lambda_{k_h} w_{k_h}$$

converges strongly to $T w$ (each eigenvalue is counted with its geometric multiplicity).

Therefore:

$$\|\lambda_{k_h} w_{k_h} - \lambda_{k_{h'}} w_{k_{h'}}\|^2 = |\lambda_{k_h}|^2 + |\lambda_{k_{h'}}|^2 \geq 2\Lambda^2,$$

but $\|T w_{k_h} - T w_{k_{h'}}\|^2 \rightarrow 0$ as $h, h' \rightarrow \infty$: this is a contradiction because it implies $\Lambda = 0$.

So $\lambda_n \rightarrow 0$ as n goes to ∞ .

Observe now that if $x \in H_n$ for all $n \in \mathbb{N}_0$ then

$$\langle x, w_k \rangle = 0 \quad \forall k \in \mathbb{N}_0$$

and

$$\|T x\| = \|T_n x\| \leq \|T_n\|_{\mathcal{B}(H_n)} \|x\| = |\lambda_n| \|x\| \quad \forall n \in \mathbb{N}_0,$$

letting n go to ∞ we find $T x = 0$, namely $x \in \text{Ker}(T)$.

Then, either $\text{Ker}(K) = \{0\}$, and zero is not an eigenvalue, and the sequence $\{w_k\}_{k \in \mathbb{N}_0}$ is an orthonormal basis for H , or $\text{Ker}(K) \neq \{0\}$, and zero is an eigenvalue, with finite or infinite multiplicity, and we have to add an orthonormal basis of $\text{Ker}(T)$ in order to have an orthonormal basis for H . ■

Consequences of the spectral theorem This result has two immediate and important consequences: the **diagonalization** and the **approximation property**.

Proposition 3.4.53 (diagonalization)

Let $(H, \langle \cdot, \cdot \rangle)$ be a separable Hilbert space, and $T \in \mathcal{K}(H)$ be symmetric.

If $\dim_{\perp} H = N$, then the spectral theorem 3.4.52 on page 294 coincides with the homonymous theorem of Linear Algebra: T can be identified as a Matrix once a basis is fixed.

If $\dim_{\perp} H = \aleph_0$, let $\{w_n\}_{n \in \mathbb{N}}$ be an orthonormal basis of eigenvectors of T , then $t_{n,m} := \langle Kw_n, w_n \rangle = \lambda_n \delta_{n,m}$, where each λ_n is counted with its multiplicity, we have

$$Tx = \sum_{n \in \mathbb{N}} \lambda_n x_n w_n.$$

The **approximation property** is the following: consider $(H, \langle \cdot, \cdot \rangle)$ separable Hilbert space, and $T \in \mathcal{K}(H)$ be symmetric.

Suppose also $\dim_{\perp} H = \aleph_0$ and consider $\{w_n\}_{n \in \mathbb{N}}$ as an orthonormal basis of eigenvectors of T . then we know that:

$$Tx = \sum_{n \in \mathbb{N}} \lambda_n x_n w_n.$$

Therefore by setting $T_N x = \sum_{n=1}^N \lambda_n x_n w_n$ we have:

$$\begin{aligned} \|Tx - T_N x\|^2 &= \sum_{n=N}^{+\infty} |\langle (k - k_n)x, w_n \rangle|^2 \\ &= \sum_{n=N+1}^{+\infty} (\lambda_n x_n \langle w_n, w_n \rangle)^2 \end{aligned}$$

maybe this is also a proposition

$$\begin{aligned} &\leq \left(\sup_{n \geq N+1} \lambda_n^2 \right) \sum_{n=N+1}^{+\infty} x_n^2 \quad (\{\lambda_n\} \text{ is decreasing}) \\ &\leq \lambda_{N+1}^2 \sum_{n=0}^{+\infty} x_n^2 = \lambda_{N+1}^2 \|x\|^2, \end{aligned}$$

so that

$$\|T - T_N\|_{\mathcal{B}(H)} \leq \lambda_{N+1}^2 \rightarrow 0 \text{ as } N \rightarrow +\infty.$$

Fredholm's alternative We recall that in Banach spaces we stated without proving corollary 3.3.50 on page 257. In this scenario, namely Hilbert Spaces, we can give a similar result and a proof.

Theorem 3.4.54

Let $(H, \langle \cdot, \cdot \rangle)$ be a separable Hilbert space, and $T \in \mathcal{K}(H)$ be symmetric.

Let $\mu \in \mathbb{R}, \mu \neq 0$.

Then, either for any $y \in H$ there exists a unique $x \in H$ such that

$$\mu x - Tx = y.$$

or μ is an eigenvalue of T .

Moreover

$$\text{Ker}(\mu I - T)^\perp = \text{Im}(\mu I - T).$$

This implies that if $\mu \neq 0$ then

$$\text{Ker}(\mu I - T) \text{ is finite dimensional}$$

and

$$\text{Ker}(\mu I - T) = \{0\} \iff \text{Im}(\mu I - T) = H.$$

The first one is because if μ is not an eigenvalue, we are in the first alternative, then we can choose $y = 0$ and there will be one and only one x such that $Tx - \mu x = 0$, namely only one element in the Kernel: $\text{Ker}(\mu I - T) = \{0\}$. If instead μ is an eigenvalue, if we consider an element of the Kernel, $x \in \text{Ker}(\mu I - T)$, this element is the eigenvector associated with μ , and then $\text{Ker}(\mu I - T) = E_\mu$, with $\mu \neq 0$. But then (see 3.4.50 on page 293) this space must be finite dimensional.

The second one is entailed by the 'Moreover' part of the theorem.

Proof. If $\dim_{\perp} H \neq \aleph_0$ then the thesis follows from Fredholm's alternative (this would be a theorem of Linear Algebra).

Let $\{w_n\}_{n \in \mathbb{N}}$ be an orthonormal basis of eigenvectors of T and $\{\lambda_n\}_{n \in \mathbb{N}}$ be the sequence of corresponding eigenvalues (each counted with its multiplicity). We have:

$$x = \sum_{n \in \mathbb{N}} x_n w_n \quad Tx = \sum_{n \in \mathbb{N}} \lambda_n x_n w_n \quad y = \sum_{n \in \mathbb{N}} y_n w_n;$$

then

$$\mu x - Tx = \sum_{n \in \mathbb{N}} (\mu - \lambda_n) x_n w_n.$$

If $y \in H$, to find a solution to $\mu x - Tx = y$ is equivalent to solve the system of infinite equations

$$(\mu - \lambda_n) x_n = y_n \quad n \in \mathbb{N}_0.$$

If $\mu \neq \lambda_n$ for any $n \in \mathbb{N}_0$ then the previous system is uniquely solvable by taking $x_n = (\mu - \lambda_n)^{-1} y_n$.

Alternatively, if $\mu = \lambda$ for some $n \in \mathbb{N}_0$, remember first that λ_n is counted with its multiplicity m , that is:

$$\lambda_n = \lambda_{n+1} = \dots = \lambda_{n+m}.$$

The key idea here is that:

$$\begin{aligned} \text{Ker}(\mu I - T) &= \{x : \mu x - Tx = 0\} \\ &= \{x : \mu x = Tx\} \\ &= \{\text{eigenvectors associated with } \mu\} \end{aligned}$$

Then we may have no solutions if $y \notin \text{Ker}(\mu I - T)^{\perp}$, i.e. if y is not orthogonal to all the eigenvectors associated with μ .

Otherwise, there are m independent solutions.

We also conclude that the equation is solvable if and only if $y \in \text{Ker}(\mu I - T)^{\perp}$, namely if $y_j, j = n, \dots, n+m$ are all equal to zero, that is, y orthogonal to all the eigenvectors associated with μ . In this latter case, we obviously have infinite solutions. ■

From the previous theorem we have the following result:

Proposition 3.4.55

Let $(H, \langle \cdot, \cdot \rangle)$ be a separable Hilbert space, and $T \in \mathcal{K}(H)$ be symmetric.

If $|\mu| \leq \|T\|_{\mathcal{B}(H)}$ then μ can be an eigenvalue of T .

If $|\mu| > \|T\|_{\mathcal{B}(H)}$ then μ is not an eigenvalue of T and the equation $\mu x - Tx = y$ is uniquely solvable for all $y \in H$.

Example 3.4.56. Consider $H = L^2(\Omega)$, with $\Omega \in \mathcal{L}(\mathbb{R}^N)$ and consider:

$$(K_G u)(x) = \int_{\Omega} G(t, s)u(s) ds$$

with $G \in L^2(\Omega \times \Omega)$ such that $G(t, s) = G(s, t)$ almost everywhere in $\Omega \times \Omega$: we deduce that $K_G \in \mathcal{K}(H)$ and is symmetric.

Therefore, if $v \in H$ and we want to check the solvability of the equation

$$\mu u - K_G u = v$$

for some $\mu \neq 0$ which is an eigenvalue for K_G .

Then we need to check the following conditions

$$\int_{\Omega} v w_j dt = 0 \quad j = 1, \dots, m$$

where $\{w_1, \dots, w_m\}$ is an orthonormal basis of $\text{Ker}(\mu I - K_G)$.

Example 3.4.57. This example complete the example of application of Lax–Milgram lemma 3.4.5 on page 277. Consider a bilinear form $a : V_0 \times V_0 \rightarrow \mathbb{R}$ defined by:

$$a(u, v) = \int_a^b [(\alpha(x)u'(x) + \beta(x)u(x))v'(x) + \gamma(x)u(x)v(x)] dx$$

and assume $\alpha, \gamma \in L^\infty((a, b))$, like the following:

$$\alpha_0 = \inf_{x \in [a, b]} \alpha(x) > 0 \text{ and } \gamma(x) \geq 0 \text{ for almost any } x \in [a, b].$$

Then $a(\cdot, \cdot)$ is a symmetric continuous and coercive bilinear form.

Moreover, its induced norm is equivalent to $\|\cdot\|_0$.

By Lax–Milgram (see theorem 3.4.25 on page 274) we have that for any $f \in L^2((a, b))$ there is a unique $u \in V_0$ such that:

$$a(u, v) = \langle f, v \rangle := \int_a^b f(x)v(x) \, dx \quad \forall v \in V_0.$$

Thus linear operator $K : L^2((a, b)) \rightarrow L^2((a, b))$ defined by

$$Kf = u \in V_0$$

is compact.

Notice that K is also symmetric, indeed, we have:

$$a(Kf, v) = \langle f, v \rangle, \quad a(Kg, v) = \langle g, v \rangle \quad \forall v \in V_0.$$

If we choose $v = Jg$ in the first identity and $v = Jf$ in the second one, we get:

$$a(Kf, Kg) = \langle f, Kg \rangle, \quad a(Kg, Kf) = \langle g, Kf \rangle$$

but

$$a(Kf, Kg) = a(Kg, Kf)$$

so that

$$\langle f, Kg \rangle = \langle g, Kf \rangle.$$

Thanks to the spectral theorem and recalling that K is injective since $\text{Ker } K = \{0\}$, we can say that K has a sequence of non-zero eigenvalues $\{\lambda_n\}_{n \in \mathbb{N}}$ such that $\lambda_n \rightarrow 0$.

Moreover we can construct an orthonormal basis of $L^2((a, b))$ which consists of eigenfunctions $\{u_n\}_{n \in \mathbb{N}}$, namely

$$Ku_n = \lambda_n u_n \text{ implies } a(u_n, v) = \mu_n \langle u_n, v \rangle \quad \forall v \in V_0$$

where $\mu_n = \lambda_n^{-1}$. Notice that these eigenfunctions form an orthogonal basis of V_0 as well.

As an exercise you can calculate the eigenvalues and eigenfunctions of K in case of $(a, b) = (0, 2\pi)$ and $\alpha = 1$ and $\gamma = 0$.

Examples on the spectra of non-compact linear bounded operators If $T \in \mathcal{K}(H)$ is symmetric, then $\sigma(T)$ is always non-empty and countable. This is not always true if compactness fails.

Example 3.4.58. The identity operator $I : H \rightarrow H$ is symmetric and non-compact if H has finite dimension. Notice that its spectrum is 1 and its eigenspace is the whole H .

Example 3.4.59. Consider $H = L^2((-1, 1))$ and consider

$$(Tf)(t) = tf(t) \text{ a.e. in } (-1, 1).$$

Then $T \in \mathcal{B}(H)$ and is symmetric, but not compact.

Suppose $\exists f \neq 0$ and $\lambda \in \mathbb{R}$ such that $Tf = \lambda f$ for almost any $t \in [-1, 1]$. Then $tf(t) = \lambda f(t)$: it must be $f \equiv 0$, and $\sigma(T) = \emptyset$.

Example 3.4.60. Consider $H = l^2$ and let:

$$\begin{aligned} T_R(x) &:= \{0, x_1, x_2, \dots, x_n, \dots\} & \forall x = \{x_1, x_2, \dots, x_n, \dots\} \in H, \\ T_L(x) &:= \{x_2, x_3, x_4, \dots, x_n, \dots\} & \forall x = \{x_1, x_2, \dots, x_n, \dots\} \in H. \end{aligned}$$

It's easy to prove that $T_R, T_L \in \mathcal{B}(H)$ but they are not neither compact nor symmetric.

It's easy to check that $\sigma(T_R) = \emptyset$.

Concerning T_L , from $T_L x = \lambda x$ we deduce:

$$\lambda x_1 = x_2, \lambda x_2 = x_3, \dots, \lambda x_n = x_{n+1}, \dots$$

for all $n \in \mathbb{N}_0$.

Then $x = \{1, \lambda, \lambda^2, \dots, \lambda^{n+1}, \dots\}$ is an eigenvalue for T_L if and only if $x \in l^2$, that is if and only if $|\lambda| < 1$. Thus

$$\sigma(T_L) = (-1, 1).$$

Example 3.4.61. Consider $H = l^2$. Let T_r . Suppose $\exists x \neq 0$ such that $T_r x = \lambda x$ for some $\lambda \in \mathbb{R}$. Then $0 = \lambda x_1$, $x_1 = \lambda x_2$, and so on. If $\lambda = 0$, then $x = 0$, and λ is not an eigenvalue. If $\lambda \neq 0$, then it must be $x_1 = 0$, $x_2 = 0$, and so on; then again $x = 0$. Therefore T_r has no eigenvalues: $\sigma(T_r) = \emptyset$.

Example 3.4.62. Consider again $H = l^2$, and define $T_l(\{x_n\}_{n \in \mathbb{N}}) := \{x_2, x_3, \dots, x_n, \dots\}$. Suppose $\exists x \neq 0$ such that $T_r x = \lambda x$ for some $\lambda \in \mathbb{R}$. Then $\lambda x_2 = x_1$, $\lambda x_3 = x_2$, and in general $\lambda x_{n+1} = x_n$.

If $\lambda = 0$, then $(x_1, 0, \dots, 0, \dots, 0)$ is an eigenvector for any $x_1 \neq 0$. Therefore $\lambda = 0$ is an eigenvalue.

If $\lambda \neq 0$, we have $x_{n+1} = \lambda^n x_1$. Moreover:

$$x = (1, \lambda, \lambda^2, \dots, \lambda^n, \dots) \in l^2 \iff \sum_{n \in \mathbb{N}} \lambda^{2n} < +\infty \iff |\lambda| < 1$$

Therefore $\sigma(T_l) = (-1, 1)$: the spectrum is continuous.

Appendices

A Limits

Given a sequence $\{x_n\}_{n \in \mathbb{N}}$ we define:

$$\limsup_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \left(\sup_{k \geq n} x_k \right) = \inf_{n \geq 0} \left(\sup_{k \geq n} x_k \right),$$

and

$$\liminf_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \left(\inf_{k \geq n} x_k \right) = \sup_{n \geq 0} \left(\inf_{k \geq n} x_k \right).$$

The limit exists if the \liminf and \limsup coincide, that is:

$$\lim_{n \rightarrow \infty} x_n = \limsup_{n \rightarrow \infty} x_n = \liminf_{n \rightarrow \infty} x_n.$$

Talking about functions, consider $f : (X, d) \rightarrow \mathbb{R}^*$, we define:

$$\limsup_{n \rightarrow \infty} f_n(x) := \inf_{n \geq 0} \left(\sup_{k \geq n} f_k(x) \right), \quad \liminf_{n \rightarrow \infty} f_n(x) := \sup_{n \geq 0} \left(\inf_{k \geq n} f_k(x) \right).$$

We can build the notion of \liminf and limit in the same manner as we have done with sequences.

For a more general context consider a sequence of sets $\{A_n\}_n$. Then we can define:

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k > n} A_k \quad \text{and} \quad \liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k > n} A_k.$$

Further results One can prove that:

$$\liminf(-\star) = -\limsup(\star).$$

Another result is the following: given two sequences a_n, b_n then:

$$\liminf_{n \rightarrow \infty} (a_n + b_n) \geq \liminf_{n \rightarrow \infty} (a_n) + \liminf_{n \rightarrow \infty} (b_n);$$

if a_n converges to a we have also that the equality holds:

$$\liminf_{n \rightarrow \infty} (a_n + b_n) = a + \liminf_{n \rightarrow \infty} (b_n).$$

check the
limsup
case

B Additional proofs

Proposition B.0.1

The set of polynomials with integer coefficients is countable.

Proof. First consider the set P_n of polynomials of degree n with nonnegative integer coefficients. First, since there are infinitely many primes, there exists an injective map $p : \mathbb{N} \rightarrow Q$ which enumerates the set of prime numbers $Q \subset \mathbb{N}$. Then there is a map

$$f : P_n \rightarrow \mathbb{N}$$

given by

$$f(a_n x^n + \cdots + a_1 x + a_0) = 2^{a_0} 3^{a_1} 5^{a_2} \cdots p(n+1)^{a_n}$$

By unique factorization, this map is one-to-one (the same argument as the proof that $A \times B$ is countable if A and B are countable – one could also show that P_n is equinumerous to $\mathbb{N} \times \mathbb{N} \times \mathbb{N} \times \cdots \times \mathbb{N}$ where there are n factors). Thus P_n is countable. Finally, the set of polynomials P can be expressed as

$$P = \bigcup_{n=0}^{\infty} P_n$$

which is a union of countable sets, and hence countable. ■

Proposition B.0.2

The set of algebraic numbers is countable.

Proof. By B.0.1, we know that the set P of polynomials is countable. Each polynomial of degree n has at most n roots, thus for any polynomial p , the set R_p of roots of p is countable. Thus the set A of algebraic numbers can be expressed as

$$A = \bigcup_{p \in P} R_p$$

is a countable union of countable sets, and hence countable. ■

