Real Analysis

Arnau Mas

2019

These are notes gathered during the subject *Anàlisi Real i Funcional* as taught by Joan Orobitg between September 2019 and January 2020.

Part I

Measure Theory

Chapter 1

Measure spaces

Definition 1.1 (σ-algebra)**.** We say a family of subsets $\mathcal{A} \subseteq \mathcal{P}(X)$ of a set X is a σ-algebra over it if

(i) $\emptyset, X \in \mathcal{A}$,

(ii) $\mathcal{A}$ is closed under countable unions, i.e. if there is a countable set $\{A_i\}_{i\in\mathbb{N}} \subseteq \mathcal{A}$ then $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{A}$,

(iii) $\mathcal{A}$ is closed under countable intersections, i.e. if there is a countable set $\{A_i\}_{i\in\mathbb{N}} \subseteq \mathcal{A}$ then $\bigcap_{i\in\mathbb{N}} A_i \in \mathcal{A}$,

(iv) $\mathcal{A}$ is closed under complements, i.e. if $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$.

$\triangle$

Notice that if a collection of subsets is closed under countable unions and under complements then it is also closed under intersections since by De Morgan's laws a countable intersection is the complement of the union of complements. Similarly if it is closed under intersections and complements it is closed under unions. So when showing that a certain collection is a σ-algebra it is enough to show that is closed under one of unions or intersections.

Example 1.1. The following are all examples of σ-algebras.

(i) For any set X, $\mathcal{P}(X)$ is a σ-algebra called the *discrete σ-algebra.* It is the *finest* σ-algebra since any other possible σ-algebra over X is contained in it.

(ii) On the other hand, the *coarsest* σ-algebra over any set X is simply $\{\emptyset, X\}$, meaning any other possible σ-algebra contains it. It is called the *trivial σ-algebra.*

(iii) If $\mathcal{A}_1$ and $\mathcal{A}_2$ are σ-algebras over a set X then so is $\mathcal{A}_1 \cap \mathcal{A}_2$.

(*iv*) Given a family of subsets $S \subseteq \mathcal{P}(X)$ then the σ-algebra generated by it is the intersection of all σ-algebras that contain it and it is the smallest σ-algebra that contains S. We write it $\sigma(S)$.

(*v*) The *Borel σ-algebra* over a topological space X is the σ-algebra generated by the open sets of X, written $\mathcal{B}(X)$. Since a closed set is the complement of an open set the family of closed sets also generates the Borel σ-algebra.

$\bigtriangledown$

The pair formed by a set and its σ-algebra is called a *measurable space*.

Definition 1.2 (Measure)**.** Let $(X, \mathcal{A})$ be a measurable space. A measure is a map $\mu\colon A \to [0, \infty]$ such that the following are true

(*i*) $\mu(\emptyset) = 0$.

(*ii*) If $\{A_i\}_{i\in\mathbb{N}} \subseteq \mathcal{A}$ is a family of pairwise disjoint sets of finite measure then

$$\mu\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \sum_{i\in\mathbb{N}} \mu(A_i).$$

$\triangle$

A measurable space equipped with a measure is called a *measure space*.

Example 1.2. The following are all examples of measures.

(*i*) Consider a measurable space X with the discrete σ-algebra. Then, for any subset $A \subseteq X$ we define $\mu(A) = |A|$ if A is finite and $\mu(A) = \infty$ if A is infinite. This is the *counting measure*, for obvious reasons.

(*ii*) On a finite measurable space X with the discrete σ-algebra we define for any subset $A \subseteq X$

$$\mu(A) = \frac{|A|}{|X|}.$$

This is a special case of a probability measure since $\mu(X) = 1$. In fact a probability is exactly a measure satisfying $\mu(X) = 1$.

(*iii*) On a measurable space X with any σ-algebra fix a point $x \in X$ and define $\mu(A) = 1$ if $x \in A$ and $\mu(A) = 0$ otherwise. This is called the *Dirac measure*.

$\bigtriangledown$

Chapter 2

The Lebesgue measure

A question worth asking is whether any measurable space can be made into a measure space. Carathéodory's extension theorem gives an affirmative answer to the question provided we give a starting point for the measure. Roughly speaking the starting point consists of specifying the measure of a collection of subsets, subject to some requirements, which can then be extended to a measure on the whole of the σ-algebra. In this section we explore a particular case of this construction on $\mathbb{R}^n$ which is known as the *Lebesgue measure*.

The main motivation behind the Lebesgue measure is to rigorously generalise the idea of length —in the case of $\mathbb{R}$—, area —in the case of $\mathbb{R}^2$— and volume —in the case of $\mathbb{R}^n$— to arbitrary dimension and for as many subsets as possible. The starting point will be the rectangles (segments in $\mathbb{R}$, rectangles in $\mathbb{R}^2$, prisms in $\mathbb{R}^3$— for which their volume is clear: it is simply the product of the length of the sides.

2.1 The Lebesgue exterior measure

As we said, the starting point for the construction will be rectangles. Let's lay down the precise definitions.

Definition 2.1 (Interval)**.** An *interval* I is a subset of $\mathbb{R}$ with the property that if $a, b \in I$ then $c \in I$ whenever $a < c < b$. It can be shown that if an interval is bounded then it must be one of $[a, b]$, (a, b), $[a, b)$ or $(a, b]$. a and b are called the *endpoints* of the interval and we will write $\langle a, b\rangle$ for any interval with endpoints a and b.

The length of an interval with endpoints a and b is defined to be $|b - a|$. △

Definition 2.2 (Rectangle)**.** An (n-dimensional) *rectangle* R is the product of n

intervals, that is

$$R = \langle a_1, b_1\rangle \times \cdots \times \langle a_n, b_n\rangle.$$

The *volume* of a rectangle is defined to be

$$v(R) = |b_1 - a_1| \cdot \cdots \cdot |b_n - a_n| \,.$$

$\triangle$

These two definitions of (hopefully) clear concepts are perhaps needlessly fussy but it pays to be precise in the beginning.

Definition 2.3 (Exterior measure)**.** We define the *Lebesgue exterior measure* or simply *exterior measure* as

$$\inf \left\{ \sum_{j=1}^{\infty} v(R_j) \mid A \subseteq \bigcup_{j=1}^{\infty} R_j,\ R_j \text{ rectangles} \right\}.$$

We will denote it by $m^*(A)$. $\triangle$

The intuition behind the exterior measure is as follows: given any set, cover it with rectangles and add up their volumes. Then try to refine the covering by acheiving less total area. The infimimum of the volumes of all possible covers is the exterior measure. In two dimensions this describes trying to literally cover the set by a patchwork of rectangles finer and finer that approximates the area of the set in question.

Also, given that the volume of a rectangle is always positive, the set we are taking the infimum of is bounded below by zero and so its infimimum always exists and is non-negative. Thus the exterior measure exists for any set.

Example 2.1. It is impractical to use the definition to directly compute the exterior measure of a given set. However here we calculate the exterior measure of various classes of sets which constitute relatively easy examples.

(i) The exterior measure of a point is 0. Indeed, let $a \in \mathbb{R}^n$ and consider the square of center a and side ϵ, $Q_\epsilon(a)$[1]. Then $Q_\epsilon(a)$ is certainly a cover of $\{a\}$ and has volume ϵ^n. That is, $m^*(\{a\}) \leq \epsilon^n$. Since ϵ can be as small as we wish we conclude $m^*(\{a\}) = 0$.

(ii) A segment in $\mathbb{R}^n$ (with $n > 1$) has exterior measure 0. If the segment has length L then we can cover it with a rectangle of length L and whose all other

[1]More precisely, if $a = (a_1, \cdots, a_n)$ then $Q_\epsilon(a) = (a_1 - \frac{\epsilon}{2}, a_1 + \frac{\epsilon}{2}) \times \cdots \times (a_1 - \frac{\epsilon}{2}, a_1 + \frac{\epsilon}{2})$.

sides have length δ. Then its total volume is $L\delta^{n-1}$ and the exterior measure of the segment is bounded by it. And since δ can be made as small as we want, we conclude the exterior measure must be 0. The details of the proof are a little cumbersome but the idea is hopefully clear.

(iii) In general any (sufficiently well-behaved) bounded subset of a hyperspace of dimension k inside $\mathbb{R}^n$ with $k < n$ has zero exterior measure. The idea is a generalisation of the previous two examples: the set can be covered by an n-dimensional hypercube in such a way that $n-k$ of its sides can be shrunk as much as one whishes and so the total volume of the cube goes to zero. Again, this is a little handwavy but the argument can be made precise.

(iv) Any countable set set of $\mathbb{R}^n$ has zero exterior measure. Since a countable set is a countable union of points, cover one of the points with a square of volume $\frac{\epsilon}{2}$, the next one with a square of volume $\frac{\epsilon}{4}$, the following with a square of volume $\frac{\epsilon}{8}$ and so on. The total volume of the cover is

$$\sum_{n=1}^{\infty} \frac{\epsilon}{2^n} = \epsilon$$

which can be made as small as one wishes.

$\bigtriangledown$

Lemma 2.1. *The exterior measure of any set is the same even if only open covers are considered.*

Proof. Let's for the moment write $M^*(A)$ for the outer measure of a set considering only open covers. It is clear that $m^*(A) \leq M^*(A)$ since a cover with open rectangles of A is still a cover by rectangles of A and so $M^*(A)$ should be at least as big as $m^*(A)$.

Now we prove the reverse inequality, $M^*(A) \leq m^*(A)$. If $\{R_i\}_{i=1}^{\infty}$ is a cover of A by rectangles then

$$\sum_{i=1}^{\infty} v(R_i) = \sum_{i=1}^{\infty} v(\mathring{R}_i)$$

since a rectangle and its interior have the same endpoints. In general, however, it is not the case that

$$\bigcup_{i=1}^{\infty} R_i \subseteq \bigcup_{i=1}^{\infty} \mathring{R}_i$$

since the interior of a set is contained in the set itself and not the other way (as should be the case) and in fact we might not even cover A anymore.

To mend this we can simply dilate the interiors. In detail, given a rectangle $R = \langle a_1, b_1 \rangle \times \cdots \times \langle a_n, b_n \rangle$ to be

$$\lambda R = \lambda\langle a_1, b_1 \rangle \times \cdots \times \lambda\langle a_n, b_n \rangle$$

where by definition

$$\lambda\langle a, b \rangle = \left\langle \frac{a+b}{2} - \lambda\frac{b-a}{2}, \frac{a+b}{2} + \lambda\frac{b-a}{2} \right\rangle.$$

This is all a very complicated way of saying we slightly inflate the rectangles while keeping their centers the same. It should be clear that if $\lambda > 1$ then $R \subseteq \lambda\mathring{R}$ and $v(\lambda R) = \lambda^n v(R)$. And so

$$A \subseteq \bigcup_{i=1}^{\infty} R_i \subseteq \bigcup_{i=1}^{\infty} \lambda\mathring{R}_i.$$

Then, since $\{\lambda\mathring{R}_j\}_{j=1}^{\infty}$ is a cover of A by open rectangles we have

$$M^*(A) \leq \sum_{j=1}^{\infty} v(\lambda\mathring{R}_j) = \lambda^n \sum_{j=1}^{\infty} v(R_j).$$

Letting $\lambda \to 1$ we obtain that for any cover of A by rectangles $\{R_j\}_{j=1}^{\infty}$ then

$$M^*(A) \leq \sum_{j=1}^{\infty} v(R_j)$$

and so $M^*(A) \leq m^*(A)$ as we wanted. □

Lemma 2.2. *The exterior measure of a rectangle is exactly its volume.*

Proof. Let R be a rectangle. It is clear that $m^*(R) \leq v(R)$ since R covers itself. We need to show then that $v(R) \leq m^*(R)$. If $\{R_j\}_{j=1}^{\infty}$ is a cover of R then we wish to conlcude that

$$v(R) \leq \sum_{j=1}^{\infty} v(R_j).$$

We can without loss of generality assume that we are dealing with a closed rectangle since $v(R) = v(\bar{R})$. And since rectangles are bounded they are compact. Using the previous lemma we need only consider covers by open rectangles and by compactness we can further limit our scope to finite open covers. Now all that is left is to show that if a rectangle is covered by a finite amount of other rectangles then their combined volume is greater than that of the the original rectangle.

Let, then, $R = \langle a_1, b_1 \rangle \times \cdots \times \langle a_n, b_n \rangle$ be a rectangle and $\{R_j\}_{j=1}^{N}$ be a cover of R with $R_j = \langle a_1^j, b_1^j \rangle \times \cdots \times \langle a_n^j, b_n^j \rangle$. Then we take the projection onto the i-th dimension and we have that

$$\langle a_i, b_i \rangle \subseteq \bigcup_{j=1}^{N} \langle a_i^j, b_i^j \rangle.$$

Let $A_i = \min\{a_i^1, \cdots, a_i^N\}$ and $B_i = \max\{b_i^1, \cdots, b_i^N\}$. Then we have $\left|b_i^j - a_i^j\right| \leq |B_i - A_i|$ and

Finish this proof

This last part of the proof consists mainly of technical details. The main insight is that since we need only look at closed rectangles and open covers then we can, by compactness, reduce potentially infinite covers to finite ones and then we can take maximums and minimums without concern. $\square$

Proposition 2.3 (Properties of the exterior measure)**.** *The following are some properties of the exterior measure*

(*i*) $m^*(\emptyset) = 0$.

(*ii*) *The exterior measure is increasing, that is if* $A \subseteq B$ *then* $m^*(A) \leq m^*(B)$.

(*iii*) *The exterior measure is countably subadditive, i.e.*

$$m^*\left(\bigcup_{j=1}^{\infty} A_j\right) \leq \sum_{j=1}^{\infty} m^*(A_j)\,.$$

(*iv*) *The exterior measure is invariant under translations.*

(*v*) *If* $A \subseteq \mathbb{R}^n$ *and* $\lambda \in \mathbb{R}$ *then* $m^*(\lambda A) = \lambda^n m^*(A)$[2].

(*vi*) *If a set* A *satisfies* $\mathring{R} \subseteq A \subseteq \bar{R}$ *for a rectangle* R *then* $m^*(A) = v(R)$.

Proof. (*i*) follows from the fact that any cover is a cover of the empty set. (*ii*) is because any cover of B is a cover of A so $m^*(A)$ must be less than $m^*(B)$.

Proving (*iii*) requires a bit more work. We may assume that every one of the A_j has finite exterior measure since otherwise we are dealing with a vacuous statement. Let $\{R_i^j\}_{i=1}^{\infty}$ be a cover of A_j such that

$$\sum_{i=1}^{\infty} v(R_i^j) \leq m^*(A_j) + \frac{\epsilon}{2^j}.$$

Then

$$\bigcup_{j=1}^{\infty} A_j \subseteq \bigcup_{j=1}^{\infty}\bigcup_{i=1}^{\infty} R_i^j$$

so

$$m^*\left(\bigcup_{j=1}^{\infty} A_j\right) \leq \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} v(R_i^j) \leq \sum_{j=1}^{\infty} m^*(A_j) + \frac{\epsilon}{2^j} = \sum_{j=1}^{\infty} m^*(A_j) + \epsilon.$$

[2]The notation λA does not refer to a dilation as used in the proof of lemma 2.1 but rather to the image of A under scalar multiplication by λ which is more standard.

And then by letting $\epsilon \to 0$ we obtain the countable subadditivity.

It should be clear that the volume of a rectangle is invariant under translations. This means that if we have a cover of a set A then we can transform it, by a translation, into a cover of the translated set $A + x$ of the same total volume, and viceversa. And so it follows that $m^*(A) = m^*(A + x)$. This proves *(iv)*.

The proof of *(v)* is very similar. Again, it should be clear that if we scale a rectangle of dimension n by a factor of λ then its volume picks up a factor of λ^n. So, given a cover of A with total volume V we can scale it by λ and we obtain a cover of λA with volume $\lambda^n V$, and viceversa. Thus we see that $m^*(\lambda A) = \lambda^n m^*(A)$.

It follows immediately from *(ii)* that $m^*\left(\mathring{R}\right) \leq m^*(A) \leq m^*\left(\bar{R}\right)$. And then, using lemma 2.2 we find $m^*\left(\mathring{R}\right) = m^*\left(\bar{R}\right) = v(R)$ and so $m^*(A) = v(R)$. □

2.2 Measurable sets

If the measure we are constructing is to be a useful generalization of the notion of volume we should expect the measure of the union of disjoint sets to be the sum of their measures. With the exterior measure this is the case for most well-behaved sets, but there exist counterexamples. The solution to this problem is to restrict ourselves to a smaller class of sets which we will call the measurable sets.

Definition 2.4 (Measurable set)**.** We say a set $E \subseteq \mathbb{R}^n$ is *measurable* if for any other set $A \subseteq \mathbb{R}^n$ it is true that

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

The set A is sometimes called a *test set*. △

Notice that because the exterior measure is subadditive we get the other inequality for free so we could have required equality in the definition of a measurable set without being more restrictive.

Example 2.2. The following are various examples of measurable sets

(i) $\mathbb{R}^n$ is measurable since $\mathbb{R}^n \cap A = A$ and $(\mathbb{R}^n)^c \cap A = \emptyset$.

(ii) Similarly $\emptyset$ is also measurable.

(iii) Any set of zero exterior measure, also called a *null set*, is measurable. Indeed, if $m^*(E) = 0$ since $E \cap A \subseteq E$ then $m^*(E \cap A) = 0$ and

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

$\triangledown$

The collection of measurable subsets of $\mathbb{R}^n$ forms a σ-algebra. To prove this we will first show a preliminary result.

Proposition 2.4. *The collection of measurable subsets is stable under finite unions. Furthermore the exterior measure is finitely additive, that is if $E_1, \cdots, E_n$ are measurable sets and are pairwise disjoint then*

$$m^*\left(\bigcup_{k=1}^n E_k\right) = \sum_{k=1}^n m^*(E_k)\,.$$

Proof. Let $\mathcal{M}$ denote the set of measurable subsets of $\mathbb{R}^n$. It is sufficient to show that if $E, F \in \mathcal{M}$ then $E \cup F \in \mathcal{M}$ since we can then prove by induction that any finite union of measurable sets is measurable. For any $A \subseteq \mathbb{R}^n$ we have

$$m^*(A\cap(E\cup F))+m^*(A\cap(E\cup F)^c) = m^*((A\cap E)\cup(A\cap F))+m^*(A\cap E^c\cap F^c)\,.$$

We may use the identity of sets $(A\cap E)\cup(A\cap F) = (A\cap E)\cup(A\cap E^c\cap F)$ and subadditivity to get

$$\begin{aligned} m^*(A\cap(E\cup F)) + m^*(A\cap(E\cup F)^c) = \\ &= m^*((A\cap E)\cup(A\cap E^c\cap F)) + m^*(A\cap E^c\cap F^c) \\ &\leq m^*(A\cap E) + m^*((A\cap E^c)\cap F) + m^*((A\cap E^c)\cap F^c) \\ &= m^*(A\cap E) + m^*(A\cap E^c) = m^*(A)\,, \end{aligned}$$

where we have used the measurability of E and F in the last two steps.

We now prove the finite additivity. Let $E_1, \cdots, E_n$ be pairwise disjoint measurable sets. Then using $A\cap(\bigcup_{k=1}^n E_k)$ as a test set and E_n as the measurable set we have, by definition of measurability

$$\begin{aligned} m^*\left(A\cap\left(\bigcup_{k=1}^n E_k\right)\right) &= m^*\left(A\cap\left(\bigcup_{k=1}^n E_k\right)\cap E_n\right) + m^*\left(A\cap\left(\bigcup_{k=1}^n E_k\right)\cap E_n^c\right) \\ &= m^*(A\cap E_n) + m^*\left(A\cap\left(\bigcup_{k=1}^{n-1} E_k\right)\right)\,. \end{aligned}$$

By induction we find

$$m^*\left(A\cap\left(\bigcup_{k=1}^n E_k\right)\right) = \sum_{k=1}^n m^*(A\cap E_k)$$

and taking $A = \bigcup_{k=1}^n E_k$ we obtain

$$m^*\left(\bigcup_{k=1}^n E_k\right) = \sum_{k=1}^n m^*(E_k)$$

as we wanted. $\square$

Proposition 2.5. *The collection $\mathcal{M}$ of measurable subsets of $\mathbb{R}^n$ is a σ-algebra.*

Proof. We have already seen in example 2.2 that $\mathbb{R}^n$ and $\emptyset$ are both measurable. It is also immediate that the complement of a measurable set is also measurable. All that remains to be shown is that $\mathcal{M}$ is closed under countable unions.

Let $\{E_k\}_{k=1}^{\infty}$ be a countable family of measurable subsets. A first observation is that we may, without loss of generality, assume that the E_k are pairwise disjoint. Indeed, define $F_1 = E_1$ and $F_k = E_k - \bigcup_{j=1}^{k-1} E_j$. The F_k are disjoint by construction. They are also all measurable by virtue of being finite intersections and unions of measurable sets and more importantly

$$\bigcup_{k=1}^{\infty} E_k = \bigcup_{k=1}^{\infty} F_k.$$

What this means is that any union of measurable sets is equal to the union of some other pairwise disjoint measurable sets. So if we manage to show that countable unions of pairwise disjoint measurable sets are measurable we are done.

Let's get to it then. We want to show that

$$m^*(A) \geq m^*\left(A \cap \left(\bigcup_{k=1}^{\infty} E_k\right)\right) + m^*\left(A \cap \left(\bigcup_{k=1}^{\infty} E_k\right)^c\right).$$

We start from the fact that finite unions of measurable sets are measurable and so for any $n \in \mathbb{N}$

$$m^*(A) \geq m^*\left(A \cap \left(\bigcup_{k=1}^{n} E_k\right)\right) + m^*\left(A \cap \left(\bigcup_{k=1}^{n} E_k\right)^c\right).$$

Now notice that

$$\bigcap_{k=1}^{N} E_k \subseteq \bigcap_{k=1}^{\infty} E_k$$

so when we take complements the inclusion reverses and we get

$$\left(\bigcap_{k=1}^{\infty} E_k\right)^c \subseteq \left(\bigcap_{k=1}^{N} E_k\right)^c.$$

Therefore

$$m^*\left(A \cap \left(\bigcup_{k=1}^{n} E_k\right)^c\right) \geq m^*\left(A \cap \left(\bigcup_{k=1}^{\infty} E_k\right)^c\right)$$

which takes care of the second term.

For the first term, we use the distributivity of intersection over unions and finite additivity —remember we are assuming the E_k to be pairwise disjoint— to get

$$m^*\left(A \cap \left(\bigcup_{k=1}^{n} E_k\right)\right) = m^*\left(\bigcup_{k=1}^{n} A \cap E_k\right) = \sum_{k=1}^{n} m^*(A \cap E_k).$$

All together reads

$$m^*(A) \geq \sum_{k=1}^{n} m^*(A \cap E_k) + m^*\left(A \cap \left(\bigcup_{k=1}^{\infty} E_k\right)^c\right).$$

By taking the limit $n \to \infty$ we get

$$m^*(A) \geq \sum_{k=1}^{\infty} m^*(A \cap E_k) + m^*\left(A \cap \left(\bigcup_{k=1}^{\infty} E_k\right)^c\right).$$

Finally, we use subadditivity to arrive at the desired result,

$$\begin{aligned} m^*(A) &\geq \sum_{k=1}^{\infty} m^*(A \cap E_k) + m^*\left(A \cap \left(\bigcup_{k=1}^{\infty} E_k\right)^c\right) \\ &\geq m^*\left(\bigcup_{k=1}^{\infty} A \cap E_k\right) + m^*\left(A \cap \left(\bigcup_{k=1}^{\infty} E_k\right)^c\right) \\ &\geq m^*\left(A \cap \left(\bigcup_{k=1}^{\infty} E_k\right)\right) + m^*\left(A \cap \left(\bigcup_{k=1}^{\infty} E_k\right)^c\right). \end{aligned}$$

□

We now have a σ-algebra defined on $\mathbb{R}^n$, which makes into a measurable space. To make it into a measure space we restrict the exterior measure to the measurable sets. We have to check that this is an honest to goodness measure. We have already seen that the measure of the empty set is 0, but we have to show that the measure is additive for disjoint unions.

Proposition 2.6. *The exterior measure of the union of disjoint sets is the sum of their measures. That is, if E_k are pairwise disjoint measurable sets then*

$$m^*\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} m^*(E_k).$$

Proof. We showed in the proof of the previous proposition that

$$m^*(A) \geq \sum_{k=1}^{\infty} m^*(A \cap E_k) + m^*\left(A \cap \left(\bigcup_{k=1}^{\infty} E_k\right)^c\right).$$

Take as a test $A = \bigcup_{k=1}^{\infty} E_k$, then, using the fact that the E_k are pairwise disjoint,

$$m^*\left(\bigcup_{k=1}^{\infty} E_k\right) \geq \sum_{k=1}^{\infty} m^*(E_k) + m^*(\emptyset) = \sum_{k=1}^{\infty} m^*(E_k).$$

The reverse inequality is a statement of subadditivity. So we get

$$m^*\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} m^*(E_k).$$

□

2.3 The structure of measurable sets

In this section we will give a number of results that shed light into the nature of measurable sets. The σ-algebra $\mathcal{M}$ is not exactly the Borel σ-algebra $\mathcal{B}(\mathbb{R})$. However measaurable sets are really close to Borel sets, as we will see.

Theorem 2.7. *For any subset $E \subseteq \mathbb{R}^n$ the following are equivalent:*

(*i*) *E is measurable.*

(*ii*) *For all $\epsilon > 0$ there is an open set $G_\epsilon \supseteq E$ such that $m^*(G_\epsilon - E) < \epsilon$.*

(*iii*) *For all $\epsilon > 0$ there is a closed set $F_\epsilon \subseteq E$ such that $m^*(E - F_\epsilon) < \epsilon$.*

(*iv*) *For all $\epsilon > 0$ there are an open set G_ϵ and a closed set F_ϵ such that $F_\epsilon \subseteq E \subseteq G_\epsilon$ and $m^*(G_\epsilon - F_\epsilon) < \epsilon$.*

Proof. Let's first show (*i*) $\implies$ (*ii*). We tackle first the case $m^*(E) < \infty$. Then, by definition of the exterior measure there is a cover of E by open rectangles, $\{R_j\}_{j=1}^{\infty}$ such that

$$\sum_{j=1}^{\infty} v(R_j) < m^*(E) + \epsilon.$$

Let $G_\epsilon = \bigcup_{j=1}^{\infty} R_j$. This is an open set that contains E. Thus $G_\epsilon \cup E = E$ and since E is measurable by hypothesis we have

$$m^*(G_\epsilon - E) = m^*(G_\epsilon) - m^*(E) \leq \sum_{j=1}^{\infty} v(R_j) - m^*(E) < \epsilon.$$

There is a usual trick to deal with the case $m^*(E) = \infty$, which is to write it as a countable union of sets of finite measure. Let $E_N := E \cap B(0, N)$. Then every E_N has finite measure since it is, by construction, contained inside of a ball of finite radius. Furthermore we have

$$E = \bigcup_{N=1}^{\infty} E_N.$$

The inclusion $\bigcup_{N=1}^{\infty} E_N \subseteq E$ is clear, since $E_N \subseteq E$. For the other inclusion we use that if $x \in E$ then $x \in E_N$ for $N \geq \|x\|$.

Since every E_N is measurable and has finite measure there exists for each $N \geq 1$ an open set $G_N \supseteq E_N$ such that

$$m^*(G_N - E_N) < \frac{\epsilon}{2^N}.$$

It should be clear that the G_N are an open cover of E. Let G_ϵ be their union, so that $G_\epsilon \supseteq E$. Now we have

$$G_\epsilon - E = \bigcup_{N=0}^{\infty} (G_N - E) \subseteq \bigcup_{N=0} (G_N - E_N)$$

thus

$$m^*(G_\epsilon - E) \leq m^*\left(\bigcup_{N=1}^{\infty} (G_N - E_N)\right) \leq \sum_{N=1}^{\infty} m^*(G_N - E_N) < \sum_{N=1}^{\infty} \frac{\epsilon}{2^N} = \epsilon$$

as we wanted.

We now prove *(ii)* $\implies$ *(i)*. For each $n \in \mathbb{N}$ let $G_n \supseteq E$ be an open set such that $m^*(G_n - E) < \frac{1}{n}$ using our hypothesis *(ii)*. Then let $N := \bigcap_{N=1}^{\infty} G_N - E$. It follows N is a null set, since for each $n \in \mathbb{N}$ we have

$$m^*(N) = m^*\left(\bigcap_{n=1}^{\infty} G_n - E\right) \leq m^*(G_n - E) < \frac{1}{n}$$

so it must be $m^*(N) = 0$. In particular N is measurable.

Now notice that $N \cup E = \bigcap_{n=1}^{\infty} G_n$, thus, since N and E are disjoint by construction

$$E = \bigcap_{n=1}^{\infty} G_n - N. \tag{2.3.1}$$

This means E is measurable. Indeed, every G_n is measurable since it is open, so their intersection is also measurable. We showed before that N is measurable since it is null, so E must be measurable since it is a difference of measurable sets.

— * —

Let's show *(i)* $\iff$ *(iii)*. For the implication *(i)* $\implies$ *(iii)* apply *(ii)* to E^c. That is, there is an open set $G_\epsilon \supseteq E^c$ such that $m^*(G_\epsilon - E^c) < \epsilon$. Now $F_\epsilon := G_\epsilon^c$ is closed and

$$F_\epsilon = G_\epsilon^c \subseteq (E^c)^c = E.$$

Furthermore

$$E - F_\epsilon = E \cap F_\epsilon^c = E \cap G_\epsilon = (E^c)^c \cap G_\epsilon = G_\epsilon - E^c$$

which means $m^*(E - F_\epsilon) < \epsilon$ as we wanted to show.

To prove the converse, *(iii)* $\implies$ *(i)* we use an argument very similar to the proof of *(ii)* $\implies$ *(i)*. For every $n \in \mathbb{N}$ let $F_n \subseteq E$ be a closed set such that $m^*(E - F_n) < \frac{1}{n}$. Then let

$$N := E - \bigcup_{n=1}^{\infty} F_n.$$

It is easy to see that N is a null set, and therefore measurable. Finally from

$$E = N \cup \bigcup_{n=1}^{\infty} F_n \tag{2.3.2}$$

one concludes E is measurable.

— * —

Finally we will prove that *(iv)* is equivalent to *(ii)* and *(iii)*, which in turn means it is equivalent to *(iv)*.

It is easy to see that *(iv)* implies *(ii)* and *(iii)*. Indeed, if G_ϵ and F_ϵ satisfy the hypotheses of *(iv)* then

$$m^*(E - F_\epsilon) \leq m^*(G_\epsilon - F_\epsilon) < \epsilon$$

and

$$m^*(G_\epsilon - E) \leq m^*(G_\epsilon - F_\epsilon) < \epsilon.$$

We now prove the converse. If there exist $G_{\epsilon/2}$ and $F_{\epsilon/2}$ satisfying the conditions in *(ii)* and *(iii)* then it is easy to show that $G_{\epsilon/2} - F_{\epsilon/2} = (G_{\epsilon/2} - E) \cup (E - F_{\epsilon/2})$ which implies

$$m^*\Big(G_{\epsilon/2} - F_{\epsilon/2}\Big) \leq m^*\Big(G_{\epsilon/2} - E\Big) + m^*\Big(E - F_{\epsilon/2}\Big) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

□

Corollary 2.8. *A set $E \subseteq \mathbb{R}^n$ is measurable if and only if it is a countable union of open sets minus a null set. Equivalently, E is measurable if and only if it is the union of a null set and a countable union of closed sets.*

Proof. It should be clear that a countable union of open sets minus a null set is measurable, as is a countable union of closed sets and a null set. The proof of the converse is essentially in the proof of theorem 2.7, namely eqs. (2.3.1) and (2.3.2). □

Chapter 3

The Lebesgue integral and integrable functions

Theorem 3.1 (Monotone Convergence). *Let $f_n \colon E \subseteq \mathbb{R} \to [0, \infty]$ an increasing sequence of positive measurable functions, i.e. $f_n \leq f_{n+1}$ a.e., which converges pointwise almost everywhere to a measurable function f. Then*

$$\int_E f = \lim_{n\to\infty} \int_E f_n.$$

Proof. Observe that since the sequence of the f_n is increasing then so is the sequence of their integrals and therefore it has a limit —although it is potentially not finite—. In particular we have

$$\lim_{n\to\infty} \int_E f_n \leq \int_E f$$

since each of the f_n is bounded by f. Therefore we need only show the reverse inequality,

$$\int_E f \leq \lim_{n\to\infty} \int_E f_n \tag{3.0.1}$$

The theorem is easy to prove when the domain of integration has finite measure and the convergence of the f_n to f is almost everywhere uniform. In this case we have that for any $\epsilon > 0$ there is a term beyond which

$$f(x) - f_n(x) \leq \epsilon$$

holds for almost all $x \in E$. When we integrate over E, ignoring the null set where chapter 3 does not hold, we get

$$\int_E f_n \geq \int_E f - \int_E \epsilon = \int_E f - \epsilon m(E).$$

Letting $\epsilon \to 0$ and $n \to \infty$ we get the result we wanted.

Let's now tackle the general case. We will show it by way of the definition of the Lebesgue integral, i.e. by working with simple functions that are bounded by f. Let, then, s be a positive measurable simple function on E such that $s \leq f$. We also introduce a parameter $c \in (0,1)$ to give us some wiggle room. Then define

$$E_n = \{x \in E \mid cs(x) \leq f_n(x)\}.$$

Since $E_n = (cs - f_n)^{-1}([-\infty, 0])$ it is measurable. Notice that beyond a certain n the E_n are not empty. Indeed, for all $x \in E$ we have $s(x) \leq f(x)$, thus $cs(x) < f(x)$ and so, since $f_n(x)$ converges to $f(x)$, beyond some n we must have $cs(x) < f_n(x) < f(x)$ which means $x \in E_n$. Also, it should be clear that $E_n \subseteq E_{n+1}$ because of the monotonicity the f_n. Finally, we have

$$E - N = \bigcup_{n=1}^{\infty} E_n$$

where N is a set of measure zero, which reflects the fact that the f_n converge to f almost everywhere on E.

Consider a disjoint representation of s, $s = \sum_{k=1}^{N} \lambda_k \chi_{A_k}$ where the $A_k \subseteq E$ are pairwise disjoint measurable sets. Then we have

$$\int_{E_n} s = \sum_{k=1}^{N} \lambda_k m(A_k \cap E_n).$$

Then when $n \to \infty$ since the E_n increase to $E - N$, thus $A_k \cap E_n$ increases to $A_k - N$. Then, by continuity from below of the Lebesgue measure

$$m(A_k \cap E_n) \xrightarrow{n\to\infty} m(A_k - N) = m(A_k)$$

Therefore, by the definition of the integral of a simple function,

$$\int_{E_n} s = \sum_{k=1}^{N} \lambda_k m(A_k \cap E_n) \xrightarrow{n\to\infty} \sum_{k=1}^{N} \lambda_k m(A_k) = \int_E s. \tag{3.0.2}$$

On E_n we have $cs \leq f_n$ which means

$$c \int_{E_n} s \leq \int_{E_n} f_n \leq \int_E f_n.$$

Now we take the limit $n \to \infty$ and using eq. (3.0.2) we find

$$c \int_E s \leq \lim_{n\to\infty} \int_E f_n.$$

And by letting $c \to 1$ we obtain

$$\int_E s \leq \lim_{n\to\infty} \int_E f_n.$$

We have just shown that the integral of any simple function bounded by f is bounded by the limit of the integrals $\int_E f_n$. Thus, the supremum over all simple functions bounded by f, which is, by definition the integral of f, is also bounded by it,

$$\int_E f \leq \lim_{n\to\infty} \int_E f_n,$$

as we wanted. $\square$

The Monotone Convergence Theorem implies a couple of other important results.

Theorem 3.2 (Beppo Levi)**.** *Let $f_n \to E \subseteq \mathbb{R} \to [0,\infty]$ be positive measurable functions. Then*

$$\int_E \sum_{n=1}^{\infty} f_n = \sum_{n=1}^{\infty} \int_E f_n.$$

Proof. Let F_N denote the N-th partial sum of the series. Since the f_n are positive, the F_N are increasing. We can then apply Monotone Convergence, theorem 3.1. $\square$

This result shows one of the advantages of the Lebesgue integral in front of the Riemann integral. In the Riemann theory of integration, one required uniform convergence to be able to exchange a series with an integral, whereas in this case pointwise convergence suffices. Notice, however, that the terms must be postive.

Theorem 3.3 (Fatou's Lemma)**.** *Given a sequence of positive measurable functions, $f_n \colon E \subseteq \mathbb{R}^n \to [0,\infty]$ then*

$$\int_E \liminf_{n\to\infty} f_n \leq \liminf_{n\to\infty} \int_E f_n.$$

Proof. We have, by definition

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

Observe as well that

$$\inf_{m\geq n} f_n \leq \inf_{m\geq n+1} f_m$$

which means we can apply Monotone Convergence to get

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

When we write $\inf_{m\geq n} f_m$ we mean that we are taking the infimimum pointwise, which means $\inf_{m\geq n} f_m$ need not coincide with any of the f_n. At each point, however, we have for all $m \geq n$, essentially by definition of the infimum,

$$\inf_{m\geq n} f_m(x) \leq f_m(x)$$

which means

$$\int_E \inf_{m\geq n} f_m \leq \int_E f_m$$

thus

$$\int_E \inf_{m\geq n} f_m \inf_{m\geq n} \int_E f_m.$$

Finally we put it all together to get the desired bound

$$\int_E \liminf_{n\to\infty} f_n = \lim_{n\to\infty} \int_E \inf_{m\geq n} f_m \leq \lim_{n\to\infty} \inf_{m\geq n} \int_E f_m = \liminf_{n\to\infty} \int_E f_n.$$

□

Theorem 3.4 (Dominated Convergence)**.** *Let $f_n \colon E \subseteq \mathbb{R}^n \to \bar{\mathbb{R}}$ be a sequence of measurable functions which converge pointwise to a function $f \colon E \to \bar{\mathbb{R}}$. If the f_n are almost everywhere dominated by an integrable function $g \colon E \to \bar{\mathbb{R}}$, that is*

$$|f_n(x)| \leq g(x) \ \textit{a.e. on } E$$

then f is integrable and

$$\int_E |f - f_n| \xrightarrow{n\to\infty} 0.$$

In particular

$$\int_E f_n \xrightarrow{n\to\infty} \int_E f.$$

Proof. Observe that the f_n are integrable: they are measurable by assumption and

$$\int_E |f_n| \leq \int_E g < \infty$$

since g is integrable. Similarly, f is also integrable. It is measurable by being the limit of measurable functions and since $|f_n(x)| \leq g(x)$ a.e. it follows $|f(x)| \leq g(x)$ a.e., which means f is integrable.

Let $h_n = |f_n - f|$. We have

$$h_n(x) = |f_n(x) - f(x)| \leq 2g(x) \text{ a.e. on } E.$$

Now, $h_n(x)$ goes to 0 as $n \to \infty$ everywhere on E so

$$\int_E 2g = \int_E \lim_{n\to\infty} (2g - h_n).$$

Since $2g(x) - h_n(x) \geq 0$ a.e. we can apply Fatou's Lemma —restricting ourselves to E minus the null set where $2g(x) - h_n(x) \geq 0$ may not hold—, to find

$$\int_E 2g = \int_E \lim_{n\to\infty} (2g - h_n) \leq \liminf_{n\to\infty} \int_E 2g - h_n$$

since $\lim_{n\to\infty}(2g - h_n) = \liminf_{n\to\infty}(2g - h_n)$. Now, using standard properties of the limit inferior,

$$\begin{aligned}\int_E 2g \leq \liminf_{n\to\infty} \int_E 2g - h_n &= \liminf_{n\to\infty} \left(\int_E 2g - \int_E h_n\right) \\ &= \int_E 2g + \liminf_{n\to\infty} \left(-\int_E h_n\right) \\ &= \int_E 2g - \limsup_{n\to\infty} \int_E h_n.\end{aligned}$$

From this it follows that $\limsup_{n\to\infty} \int_E h_n \leq 0$, which means it is actually 0 since the h_n are positive. In fact, since they are positive we have that $\liminf_{n\to\infty} \int_E h_n \geq 0$. Since the limit superior is always greater than the limit inferior, we conclude both are 0 for $\int_E h_n$ which means

$$\int_E h_n = \int_E |f_n - f| \xrightarrow{n\to\infty} 0,$$

as we wanted.

Finally we have

$$\left|\int_E f_n - \int_E f\right| = \left|\int_E f_n - f\right| \leq \int_E |f_n - f| \to 0$$

which implies

$$\int_E f_n \to \int_E f$$

as we wished. □

3.1 Differentiation under the integral sign

Theorem 3.5. *Let $f\colon E \times I \to \mathbb{R}$ be a measurable function where $E \subseteq \mathbb{R}^n$ is measurable and $I \subseteq \mathbb{R}$ is an interval. Write f_x for the function*

3.2 Applications of the theory of integration

Once we have the various results of the theory of Lebesgue integration at out disposal we can restate a number of concepts in terms of the Lebesgue integral. These include the convolution of functions, differentiation under the integral sign and the change of variable theorem.

3.2.1 Convolution

The operation of convolution is defined on integrable functions.

Definition 3.1 (Convolution)**.** Let $f, g \in \mathcal{L}^1(\mathbb{R}^n)$ be two integrable functions. We define their *convolution* $f * g$ as

$$(f * g)(x) := \int_{\mathbb{R}^n} f(x-y)g(y) \,\mathrm{d}y.$$

$\triangle$

If the functions f and g are continous and have compact support it is clear that their convolution is finite everywhere. But in fact we can show that the convolution of any two integrable functions is always integrable.

Proposition 3.6. *The convolution of any two integrable funcions is always integrable. In other words, if $f, g \in \mathcal{L}^1(\mathbb{R}^n)$ then $f * g \in \mathcal{L}^1(\mathbb{R}^n)$.*

Proof. To show that $f * g$ is integrable we need to show that it is measurable and absolutely integrable. We can directly compute the integral of $|f * g|$:

$$\begin{aligned}
\int_{\mathbb{R}^n} |f * g| &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |f(x-y)|\,|g(y)| \,\mathrm{d}y \,\mathrm{d}x \\
&= \int_{\mathbb{R}^n} |g(y)| \int_{\mathbb{R}^n} |f(x-y)| \,\mathrm{d}x \,\mathrm{d}y \\
&= \int_{\mathbb{R}^n} |g(y)| \,\mathrm{d}y \int_{\mathbb{R}^n} |f(x-y)| \,\mathrm{d}x && \text{by Tonelli's Theorem} \\
&= \int_{\mathbb{R}^n} |g| \int_{\mathbb{R}^n} |f| < \infty && \text{since } f, g \in \mathcal{L}^1(\mathbb{R}^n).
\end{aligned}$$

In the last step we made use of the change of variable theorem which we haven't shown yet. However it should not be difficult to convince yourself that it is true in this case since $f(x-y)$ is merely f translated by y, so its integral over $\mathbb{R}^n$ should not change.

It is clear that $h(x, y) = f(x-y)g(y)$ is measurable since it is the product of measuarable functions . We have just shown that it is absolutely integrable, so, by Fubini's Theorem we find that the integral of h with respect to y, that is, $f * g$ is measurable. From this we conclude that $f * g \in \mathcal{L}^1(\mathbb{R}^n)$. $\square$

Not as clear

The operation of convolution has a number of nice properties, namely it is associative and commutative.

Proposition 3.7. *Convolution is an associative and commutative operation. That is, for any $f, g, h \in \mathcal{L}^1(\mathbb{R}^n)$ we have*

$$(f * g) * h$$

and

$$f * g = g * f.$$

Proof. We show commutativity first.

$$\begin{aligned}
(f * g)(x) &= \int_{\mathbb{R}^n} f(x-y)g(y)\,\mathrm{d}y \\
&= \int_{\mathbb{R}^n} f(z)g(x-z)\,\mathrm{d}z && \text{using the change } z = x - y \\
&= \int_{\mathbb{R}^n} g(x-z)f(z)\,\mathrm{d}z = (g * f)(x).
\end{aligned}$$

$$\begin{aligned}
((f * g) * h)(x) &= \int_{\mathbb{R}^n} (f * g)(x-y)h(y)\,\mathrm{d}y \\
&= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(x-z)g(z)\,\mathrm{d}z \right) h(y)\,\mathrm{d}y && =
\end{aligned}$$

□

Part II

Banach Spaces

Chapter 4

Definition and Examples

The theory of Banach and Hilbert spaces generalises the basic analytical and geometric ideas and tools of Euclidean space to a more general class of spaces. By Euclidean space we mean $\mathbb{R}^n$ with the topology that is induced by the standard inner product, i.e. the topology generated by open balls. Our starting point will be normed vector spaces. From now on E will denote a vector space over either the real or the complex numbers. If the distinction is meaningful at any point we will make note of it. We will refer to elements of the base field simply as scalars, unless whether they are real or complex is relevant.

4.1 Normed spaces

Definition 4.1 (Norm)**.** A norm on a vector space E is a function $\|\cdot\| : E \to [0, \infty)$ such that

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

(ii) $\|x + y\| \leq \|x\| + \|y\|$, which is known as the *triangle inequality*,

(iii) for any scalar λ, $\|\lambda x\| = |\lambda| \, \|x\|$.

△

A vector space equipped with a norm is called a *normed space*. A norm can be used to define a *metric* or *distance*, which is the function

$$\begin{aligned} d \colon E \times E &\longrightarrow [0, \infty) \\ (x, y) &\longmapsto \|x - y\| \,. \end{aligned}$$

It is easy to check that d indeed satisfies the definition of a metric:

(i) By definition $d(x, y) = 0$ if and only if $\|x - y\| = 0$, which is equivalent to $x - y = 0$ and thus $x = y$.

(ii) $d(x, y) = \|x - y\| = |-1| \, \|y - x\| = \|y - x\| = d(y, x)$.

(iii) $d(x, y) = \|x - z\| = \|x - y + y - z\| \leq \|x - y\| + \|y - z\| = d(x, y) + d(y, z)$.

We have that a normed space is also a metric space and therefore a topological space, with the topology generated by the open balls

$$B(x, r) := \{y \in E \mid \|x - y\| < r\}.$$

Thus, a subset $U \subseteq E$ is open if and only if for every $x \in E$ there exists $r > 0$ such that $B(x, r) \subseteq U$.

Note that while a norm always induces a metric, it is not true that every metric comes from a norm.

There is further structure we could give to our vector space, namely a *scalar product*. This is a positive definite symmetric bilinear form $\langle \cdot \, , \cdot \rangle : E \times E \to \mathbb{R}$. This requires that E be a real vector space. The analog for complex spaces is a *Hermitian product*, which is a positive definite, conjugate symmetric bilinear form $\langle \cdot \, , \cdot \rangle : E \times E \to \mathbb{C}$. We will deal with these later on when discussing Hilbert spaces. For now note that a scalar product infuces a norm by

$$\|x\| := \sqrt{\langle x \, , x \rangle}.$$

4.2 Convergence and completeness

Since normed spaces are metric spaces we can speak of convergence.

Definition 4.2 (Convergence)**.** We say a sequence (x_n) in a normed space E *converges* to $x \in E$ if for every $\epsilon > 0$ ther exists $N \in \mathbb{N}$ such that when $n \geq N$ one has

$$\|x_n - x\| < \epsilon.$$

$\triangle$

The intuition behind this definition is exactly the same as in the case of Euclidean space: the terms of the sequence get arbitrarily close to a certain point x which is naturally called the *limit* of x_n. We will write $\lim_{n\to\infty} x_n = x$ or $x_n \xrightarrow{n\to\infty} x$ when (x_n) converges to x.

There is a long list of facts about limits and convergence that are shown for Euclidean spaces in first and second year Analysis which are also true for any normed space, namely those that only make use of the normed vector space structure of Euclidean space. Because of this, their proofs can be repeated verbatim for any normed space. For example, the limit of a sum is the sum of limits, scalars can move in and out of limits...

An important class of sequences are Cauchy sequences:

Definition 4.3 (Cauchy sequence)**.** A sequence (x_n) is a *Cauchy sequence* if for any $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $n, m > N$ one has

$$\|x_n - x_m\| < \epsilon.$$

△

It is easy to show that a convergent sequence is a Cauchy sequence. Indeed, if $x_n \to x$ then for any $\epsilon > 0$ there is $N \in \mathbb{N}$ such that if $n, m > N$

$$\|x_n - x_m\| \leq \|x_n - x\| + \|x_m - x\| < 2\epsilon.$$

The converse if, of course, not true in general. The most famous offender is the set of real numbers. If there exists a Cauchy sequence that fails to be convergent it means that our space has a sort of hole. Indeed, the terms of a Cauchy sequence all get arbitrarily close to each other so if they fail to converge it means they are circling a point which should be there but isn't.

A useful fact to keep in mind is that the norms of the terms of a Cauchy sequence are convergent.

Proposition 4.1. *If x_n is a Cauchy sequence in a normed space E, then the sequence $\|x_n\|$ is convergent in $\mathbb{R}$.*

Proof. This follows from the Cauchy condition and the reverse triangle inequality. Indeed, for sufficiently large n and m we have

$$\|x_n - x_m\| < \epsilon.$$

Then

$$|\|x_n\| - \|x_m\|| \leq \|x_n - x_m\| < \epsilon$$

thus the real sequence $\|x_n\|$ is Cauchy and thus convergent. □

A metric space in which every Cauchy sequence is convergent is called *complete*, and a complete normed space is known as *Banach space*.

4.3 Examples

The examples of Banach spaces come in various different flavours. In this section we will discuss three of these: spaces of functions on a compact domain, sequence spaces and Lebesgue spaces.

4.3.1 Function spaces

Consider a compact set, $K \subseteq \mathbb{R}^m$. Then we can consider various sets of functions defined on K. We will write, as is standard, $B(K)$ for the set of all bounded functions on K, $C^0(K)$ for the set of all continuous functions on K and $C^n(K)$ for the set of n times differentiable functions with continuous n-th partial derivatives on K. All of these sets are actually vector spaces with pointwise addition and scalar multiplication. And in fact we have the chain of inclusions

$$B(K) \supseteq C^0(K) \supseteq C^1(K) \supseteq \cdots \supseteq C^n(K) \supseteq \cdots .$$

All of these are standard results from elementary real analysis. On $B(K)$ we can define the norm

$$\|f\|_\infty := \sup_{x \in K} |f(x)| .$$

Which we will refer to as the *uniform norm*. This is indeed a norm —check it!—, which makes $B(K)$ into a normed space. Restricting it to each of the $C^n(K)$ also makes them into normed spaces. Convergence with respect to this norm is known as *uniform convergence*.

With the uniform norm, however, only $B(K)$ and $C^0(K)$ are complete, therefore Banach spaces. Let's show it. Let (f_n) be a Cauchy sequence of bounded functions on K, i.e. for all $\epsilon > 0$ there is $N \in \mathbb{N}$ such that if $n, m > N$ then

$$\|f_n - f_m\| = \sup_{x \in K} |f_n(x) - f_m(x)| < \epsilon. \tag{4.3.1}$$

One often says that the sequence is *uniformly Cauchy*, as opposed to pointwise Cauchy, which would mean that the sequence $f_n(x)$ is Cauchy at every point $x \in K$. Equation (4.3.1) implies that the sequence is pointwise Cauchy. Indeed, at any $x \in K$ we have

$$|f_n(x) - f_m(x)| < \|f_n - f_m\|_\infty < \epsilon$$

for sufficiently large n and m. In particular, this means the sequence of real numbers $(f_n(x))$ is convergent since it is Cauchy. This gives us a candidate for the limit of the f_n by defining

$$\begin{aligned} f\colon K &\longrightarrow \mathbb{R} \\ x &\longmapsto \lim_{n\to\infty} f_n(x).\end{aligned}$$ [1]

If f is indeed the uniform limit of the f_n then we must show that f is bounded and that $\|f - f_n\|_\infty \xrightarrow{n\to\infty} \infty$. Let's first show that f is bounded, so $f \in B(K)$. Using Proposition 4.1 we find that the sequence $\|f_n\|_\infty$ is convergent. It then follows that for all $x \in K$

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

which means f is bounded, as we wanted.

Let's now show that the convergence is uniform. At every $x \in K$ we have, for sufficiently large n and m

$$|f_n(x) - f_m(x)| < \epsilon$$

which means that when we take the limit $m \to \infty$ we obtain

$$|f_n(x) - f(x)| \leq \epsilon$$

thus $\|f_n - f\| \leq \epsilon$ for a sufficiently large n, which shows the convergence is indeed uniform.

This shows $B(K)$ is complete. To prove that $C^0(K)$ is complete we must show that if the f_n are contiuous then their uniform limit is also continuous, not just bounded. This is fairly straightforward once we know the convergence of the f_n is uniform. There is $N \in \mathbb{N}$ beyond which $\|f - f_n\|_\infty < \epsilon$, and since the f_n are continuous, if x and y are sufficiently close then $|f_n(x) - f_n(y)| < \epsilon$. Putting this all together we find

$$\begin{aligned} |f(x) - f(y)| &\leq |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)| \\ &\leq \|f - f_n\|_\epsilon + |f_n(x) - f_n(y)| + \|f_n - f\|_\infty < 3\epsilon \end{aligned}$$

so f is also continuous.

As we mentioned before, $C^n(K)$ is not complete with respect to the uniform norm for $n \geq 1$. For now let's take K to be a closed interval, say $[a, b]$. There exist sequences of differentiable functions whose limit is not differentiable, and furthermore, there are sequences of differentiable functions whose limit is also differentiable but its derivative need not be the limit of the derivatives of the functions of the sequence. This then shows that $C^1([a, b])$. To ensure that derivatives interact well with limits one needs to require not only uniform convergence of the functions but also uniform convergence of the derivatives, which then ensures that the derivative of the limit is the limit of the derivatives. This sort of convergence is captured by the so-called C^1 norm

$$\|f\|_{C^1} := \|f\|_\infty + \|f'\|_\infty$$

which then makes $C^1([a,b])$ into a Banach space. More generally, the C^n norm

$$\|f\|_{C^n} := \sum_{k=0}^{n} \left\|f^{(k)}\right\|_\infty$$

makes $C^n([a,b])$ into a normed space. The appropriate generalisation to an m-dimensional fomain must take into account all possible partial derivatives.

4.3.2 Sequence Spaces

Another common family of Banach spaces are the so-called sequence spaces. They are all subsets of the space of real[2] sequences, $\mathbb{R}^{\mathbb{N}}$. This is a vector space with pointwise addition and scalar multiplication. There are various subspaces of $\mathbb{R}^{\mathbb{N}}$ which can be made into Banach spaces with the appropriate norm. First, we have the space ℓ^∞ of bounded sequences with the uniform norm defined as

$$\|x\|_\infty := \sup_{n\in\mathbb{N}} |x(n)|\,.$$[3]

The proof that this is indeed a Banach space is essentially identical to the proof that $B(K)$ is a Banach space, so I will spare you the details. The main steps are to show that, given a Cauchy sequence x_n in ℓ^∞, the sequence of images $x_n(m)$ is Cauchy, which gives a pointwise limit, which is then shown to be bounded and finally one shows that the sequence x_n converges to its pointwise limit uniformly.

For $p > 1$ one talks about the ℓ^p spaces, which are the spaces of sequences x such that

$$\sum_{n=1}^{\infty} |x(n)|^p < \infty$$

each with the so-called p-norm

$$\|x\|_p := \left(\sum_{n=1}^{\infty} |x(n)|\right)^{\frac{1}{p}}\,.$$

[2]All of these examples also work if one replaces real with complex

[3]Since we will be dealing with sequences of sequences we adopt the convention to write $x(n)$ for the n-th term of the sequence x, as opposed to the more standard x_n, and reserve the subscript to refer to the terms of a sequence in $\mathbb{R}^{\mathbb{N}}$.

Showing that the p-norms satisfy the triangle inequality is not trivial —at least not for $p > 1$—. We will see that in fact they do when we prove Minkowski's inequality.

For the sake of illustration, I will show the completeness of ℓ^1. This will be standard fare. To begin, we have a Cauchy sequence in ℓ^1, x_n. The Cauchy condition with respect to the 1-norm means that for sufficiently large n and m one has

$$\|x_n - x_m\|_1 = \sum_{k=1}^{\infty} |x_n(k) - x_m(k)| < \epsilon.$$

Since all of the terms of the series are postive this means, in particular, that for any $k \in \mathbb{N}$

$$|x_n(k) - x_m(k)| < \|x_n - x_m\| < \epsilon$$

meaning the sequence $x_n(k)$ is Cauchy for any k and therefore convergent. We then define the pointwise limit of the x_n as

$$x(k) := \lim_{n\to\infty} x_n(k).$$

We need to show that $x \in \ell^1$ and that $x_n \to x$ uniformly.

To show that $x \in \ell^1$ we will make use of Proposition 4.1, which gives us that the sequence $\|x_n\|_1$ is convergent and in particular bounded. That is, there is $M \in \mathbb{R}$ such that $\|x_n\|_1 \leq M$ for all $n \in \mathbb{N}$. Let's write this out: for all $n \in \mathbb{N}$ we have

$$\sum_{k=1}^{\infty} |x_n(k)| < M. \tag{$*$}$$

Because every term in the series is positive we can cut it off wherever we want, meaning for any $K \in \mathbb{N}$ we have the bound

$$\sum_{k=1}^{K} |x_n(k)| < M.$$

Now, by definition, the norm of x is

$$\|x\|_1 = \sum_{k=1}^{\infty} |x(k)| = \sum_{k=1}^{\infty} \left|\lim_{n\to\infty} x_n(k)\right| = \sum_{k=1}^{\infty} \lim_{n\to\infty} |x_n(k)|.$$

Because all of the terms of this series are positive, to show it converges it suffices to show that its partial sums are bounded, but this follows immediately from the bound in ($*$):

$$\sum_{k=1}^{K} \lim_{n\to\infty} |x_n(k)| = \lim_{n\to\infty} \sum_{k=1}^{K} |x_n(k)| < M.$$

Thus $\|x\|_1 < \infty$ and $x \in \ell^1$.

4.4 Inequalities

Definition 4.4 (Conjugate exponents)**.** We say two numbers $p, q \in (1, \infty)$ are *conjugate exponents* if

$$\frac{1}{p} + \frac{1}{q} = 1.$$

By convention, the conjugate exponent of 1 is taken to be ∞, and vice versa. △

Note that 2 is its own conjugate exponent.

Theorem 4.2 (Young's Inequality)**.** *If p and q are conjugate exponents then the inequality*

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

holds for any two $a, b \in \mathbb{R}$.

Proof. This follows from the fact that the logarithm is a concave function, for

$$\log\left(\frac{a^p}{p} + \frac{b^q}{q}\right) \geq \frac{1}{p}\log\left(a^p\right) + \frac{1}{q}\log\left(b^q\right) = \log(a) + \log(b) = \log(ab).$$

And since the logarithm is increasing,

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

as we wanted. □

Theorem 4.3 (Hölder's Inequality)**.** *For any two measurable functions, f and g, on a measurable domain $E \subseteq \mathbb{R}$, and conjugate exponents $p, q \in (1, \infty)$ one has the inequality*

$$\int_E |fg| \leq \left(\int_E |f|^p\right)^{\frac{1}{p}} \left(\int_E |g|^q\right)^{\frac{1}{q}}$$

Proof. If $\int_E |f|^p$ is infinite then the inequality is trivially true, as it is if $\int_E |g|^q$ is infinite. On the other hand, if either of the integrals on the right is zero then either f or g is zero almost everywhere, which means the integral on the left is zero and the inequality is also true. We can then asssume that both integrals are finite and strictly positive.

Let

$$\tilde{f} := \frac{f}{\left(\int_E |f|^p\right)^{\frac{1}{p}}}$$

and

$$\tilde{g} := \frac{g}{\left(\int_E |g|^q\right)^{\frac{1}{q}}}.$$

It is clear that both $\left|\tilde{f}\right|^p$ and $|\tilde{g}|^q$ integrate to 1 over E. Using Young's Inequality we have

$$\left|\tilde{f}\right| |\tilde{g}| \leq \frac{1}{p} \left|\tilde{f}\right|^p + \frac{1}{q} |\tilde{g}|^q .$$

From this follows, integrating over E,

$$\int_E \left|\tilde{f}\tilde{g}\right| \leq \frac{1}{p} \int_E \left|\tilde{f}\right|^p + \frac{1}{q} \int_E |\tilde{g}|^q \leq \frac{1}{p} + \frac{1}{q} = 1.$$

And this implies

$$\int_E |fg| \leq \left(\int_E |f|^p\right)^{\frac{1}{p}} \left(\int_E |g|^q\right)^{\frac{1}{q}}$$

□

Corollary 4.4 (Hölder's Inequality for L^p spaces)**.** *Given $f \in L^p(E)$ and $g \in L^q(E)$ where p and q are conjugate exponents we have that $fg \in L^1(E)$ and*

$$\|fg\|_1 \leq \|f\|_p \|g\|_q .$$

Proof. The case where $p, q \in (1, \infty)$ is Hölder's Inequality. And for the case where $p = 1$ and $q = \infty$ we have

$$\|fg\|_1 = \int_E |fg| \leq \int_E |f| \|g\|_\infty = \|g\|_\infty \int_E |f| = \|f\|_1 \|g\|_\infty .$$

□

Chapter 5

Bounded Operators

5.1 Bounded Operators

We now ask what the appropriate structure preserving maps are between Banach spaces and more generally normed spaces. They should certainly be linear, since normed spaces are vector spaces. They also are, of course, normed, so the maps we will be interested in should somehow preserve that. We could simply ask that they be continuous or we could be stricter and ask that they preserve the norm. Or we could settle somewhere in-between and ask that they be nondecreasing in norm, meaning $\|Tx\| \leq \|x\|$. However, the agreed upon notion in the field is that of boundedness.

Definition 5.1 (Bounded operator)**.** Let E and F be normed spaces. We say a map $T\colon E \to F$ is *bounded* if there exists $M > 0$, sometimes called a bounding constant, such that for all $x \in E$

$$\|Tx\| \leq M \|x\|\,[1].$$

We call a bounded linear map a *bounded operator*. △

When dealing with linear maps in the context of normed spaces it is customary to ommit parentheses when writing the image of an element ny a map, that is, we will generally write Tx instead of $T(x)$.

— ∗ —

[1] Strictly speaking, since E and F both come with their own norms we should somehow distinguish between them. In most cases, however, it is clear which norm we are talking about from the context —for instance, $\|Tx\|$ must refer to the norm of F since $Tx \in F$— so we will mostly dispense with any distinction unless required.

As it turns out, linear maps are bounded if and only if they are continuous, so really we are settling for the weaker class of structure preserving maps: continuous and linear.

Proposition 5.1. *Let E and F be normed spaces and $T\colon E \to F$ a linear map between the two. Then the following are all equivalent:*

(i) T is bounded.

(ii) T is continuous.

(iii) T is continuous at 0.

(iv) The image of the unit ball of E by T, $T(B(0,1))$ is bounded.

(v) T sends bounded sets to bounded sets.

Proof. Let's first show (ii) $\iff$ (iii). Obviously, if T is continuous then it is continuous at 0. Suppose, then, T is continuous at 0. Let x_n be any sequence in E that converges to x. Then $x_n - x$ converges to 0. Because T is continuous at 0, $T(x_n - x) \to T(0)$. Because T is linear, $T(0) = 0$, and again because of linearity, $T(x_n - x) = T(x_n) - T(x)$. Thus $T(x_n)$ converges to $T(x)$, which implies T is continuous at $x \in E$, therefore continuous.

Now we prove (iii) $\implies$ (i). T being continuous at 0 means that for any $\epsilon > 0$ there is $\delta > 0$ such that if $\|x\| < \delta$ then $\|Tx\| < \epsilon$. Choose $\epsilon = 1$. Then, for any $x \in E$ we have

$$\left\| \frac{\delta}{2} \frac{x}{\|x\|} \right\| = \frac{\delta}{2} < \delta$$

thus

$$\left\| \frac{\delta}{2} \frac{Tx}{\|x\|} \right\| < 1$$

which means, for any $x \in E$ we have

$$\|Tx\| < \frac{2}{\delta} \|x\|$$

which shows T is bounded.

Conversely, if T is bounded it is easy to show it is continuous at 0. If M is a bounding constant of T then, for any $\epsilon > 0$, provided $\|x\| < \frac{\epsilon}{M}$ we have

$$\|Tx\| < M \|x\| < \epsilon$$

meaning T is continuous at 0 —since $T(0) = 0$—.

We now show (*i*) $\iff$ (*iv*). If T is bounded with bounding constant M, then, for any $x \in B(0,1)$ we have

$$\|Tx\| \leq M \|x\| < M$$

which means $T(B(0,1)) \subseteq B(0,M)$, which shows $T(B(0,1))$ is bounded. Conversely, if $T(B(0,1)) \subseteq B(0,M)$ then for any $x \in E$ we have

$$\frac{\|Tx\|}{\|x\|} = \left\| T\frac{x}{\|x\|} \right\| < M$$

therefore $\|Tx\| < M \|x\|$, showing T is bounded.

Finally we prove (*iv*) $\iff$ (*v*). (*v*) $\implies$ (*iv*) is obvious, since the unit ball is by definition bounded. On the other hand, if $A \subseteq E$ is bounded then there is $R > 0$ such that $A \subseteq B(0,R) = R(B(0,1))$, thus

$$T(A) \subseteq T(B(0,R)) = RT(B(0,1))$$

which is bounded, provided $T(B(0,1))$ is bounded. □

5.2 Compactness

Theorem 5.2. *A metric space X is compact if and only if every sequence in X has a convergent subsequence.*[2]

Proof. ($\implies$) We will show the contrapositive: if X has a sequence with no convergent subsequence then it cannot be compact. Let (x_n) be such a sequence. This means, in particular, that the sequence must have infinitely many different terms —i.e., its image cannot be finite— since otherwise we would be able to find a convergent subsequence. Since no subsequence of (x_n) is convergent, for any point $y \in X$ there must be an open ball $B(y, r_y)$ which only contains finitely many different terms of the sequence —otherwise y would be an accumulation point of the sequence and therefore the limit of a subsequence—. Thus we have an open cover of X,

$$X = \bigcup_{y \in X} B(y, r_y).$$

If we were able to extract a finite subcover from this cover, say

$$X = \bigcup_{n=1}^{N} B(y_n, r_{y_n})$$

[2]This property is usually called *sequential compactness*.

then, by the pigeonhole principle, at least one of these balls would have to contain infinitely many different terms of the sequence, which is not the case. Thus X is not compact.

($\Longleftarrow$) Let's now assume that X is sequentially compact. This implies that it is totally bounded. Indeed, if it weren't, we could find a sequence in X with no convergent subsequence. Let $\epsilon > 0$ be such that X can not be covered by a finite number of balls with radius ϵ or less. Take $x_1 \in X$ and consider the ball $B(x_1, \epsilon)$. There must then exist a point $x_2 \in X$ outside of this ball. Since $B(x_1, \epsilon) \cup B(x_2, \epsilon)$ cannot cover X, there must exist $x_3 \in X$ outside of these balls. In general, there must exist, for each $n \in \mathbb{N}$ a point x_n such that

$$x_n \in X - \bigcup_{k=1}^{n-1} B(x_k, \epsilon).$$

This gives a sequence with the property that

$$\|x_n - x_m\| \geq \epsilon$$

when $m \geq n$. It is easy to see that none of its subsequences is Cauchy and therefore cannot converge.

Let's show now that X is compact. Consider an open cover of X,

$$X = \bigcup_{\alpha \in I} U_\alpha.$$

We have to show that we can extract from it a finite subcover. For any $x \in X$ let

$$R(x) := \sup\{r > 0 \mid \exists \beta \in I\colon B(x, r) \subseteq U_\beta\},$$

that is, $R(x)$ is the largest possible radius such that a ball with that radius fits inside one of the open sets in the cover. We will show that $\epsilon := \inf_{x \in X} R(x) > 0$. By definition of the infimum, there must be a sequence x_n in X such that $R(x_n) \xrightarrow{n\to\infty} \epsilon$. Using the hypothesis of the sequential compactness of X, we extract a convergent subsequence, (x_{n_k}), with limit x. By definition of convergence, there is $K \in \mathbb{N}$ such that

$$x_{n_k} \in B(x, R(x))$$

provided $k \geq K$, therefore $R(x)$ must be strictly greater than 0, since otherwise $B(x, R(x))$ could not contain points other than x.

Since $\epsilon > 0$, because X is totally bounded, there must be a finite set of points $y_1, \ldots, y_N \in X$ such that

$$X \subseteq \bigcup_{n=1}^{N} B(y_n, \epsilon) \subseteq B(y_n, R(y_n)) \subseteq \bigcup_{n=1}^{N} U_n.$$

□

Part III

Hilbert Spaces

Chapter 6

Inner and Hermitian Products

As we mentioned earlier, Hilbert spaces are Banach spaces with the special property that the norm they are equipped with comes from an inner product, in the real case, and a Hermitian product in the complex case.

6.1 Definition

Definition 6.1 (Inner product). An *inner product* on a real vector space E is a map $\phi\colon E \times E \to \mathbb{R}$ which is

(i) bilinear, that is, for all $u, v, w \in E$ and $\lambda, \mu \in \mathbb{R}$:

$$\phi(\lambda u + \mu v, w) = \lambda\phi(u, w) + \mu\phi(v, w)$$
$$\phi(u, \lambda v + \mu w) = \lambda\phi(u, v) + \mu\phi(u, w),$$

(ii) symmetric, meaning for all $u, v \in E$

$$\phi(u, v) = \phi(v, u),$$

(iii) and positive definite, so that for all $u \in E$

$$\phi(u, u) \geq 0$$

and $\phi(u, u) = 0$ if and only if $u = 0$.

$\triangle$

The properties of such an inner product mimic those of the standard dot product on $\mathbb{R}^n$. This gives our space a geometry that is richer than what a norm gives, as we will see. In the complex case, the correct notion is that of a Hermitian product:

Definition 6.2. A *Hermitian product* on a complex vector space E is a map $\phi\colon E\times E\to\mathbb{C}$ which is

(i) sesquilinear, meaning that for any $u, v, w \in E$ and $\lambda, \mu \in \mathbb{C}$

$$\begin{aligned}\phi(\lambda u+\mu v, w) &= \lambda\phi(u,w)+\mu\phi(v,w)\\ \phi(u, \lambda v+\mu w) &= \bar{\lambda}\phi(u,v)+\bar{\mu}\phi(u,w),\end{aligned}$$

where $\bar{\lambda}$ is the complex conjugate of λ,

(ii) conjugate symmetric, so that for any $u, v \in E$

$$\phi(u,v)=\overline{\phi(v,u)}, \tag{6.1.1}$$

(iii) and positive definite.

△

This definition requires some remarks. Firstly, note that linearity in the first argument and conjugate symmetry imply sesquilinearity. One often says that a Hermitian product is *linear* in its first argument and *semilinear*[1] in its second[2]. This means a Hermitian product is altogether sesquilinear, sesqui being Latin for one and a half, as opposed to bilinear. Finally, because of conjugate symmetry, the product of a vector with itself is always real, for a complex number is equal to its conjugate if and only if it is real. Therefore the requirement of positive-definiteness is sound.

It is customary to write the inner product with angle brackets, so

$$\langle u\,, v\rangle := \phi(u,v).$$

In what follows you should almost always have a Hermitian product in mind. In the case any result is specialised to an inner product, so it makes use of the fact it is over the reals, it will be duly noted. Spaces equipped with an inner or Hermitian product are called, unsurprisingly, inner product spaces, and less often, pre-Hilbert spaces (which refers to the fact that they are not quite product spaces, not that they were worked on before Hilbert).

[1] This just means that scalars are conjugated when they come out

[2] On the other hand, physicists tend to use a different convention and assume semilinearity in the first argument instead

6.2 The norm induced by an inner product

An inner product induces a norm by

$$\|u\| := \sqrt{\langle u\,,u\rangle}.$$

We can check it satisfies two of the properties of a norm immediately. For one, because of positive-definiteness, $\langle u\,,u\rangle$ is always positive, so $\|u\|$ is positive as well. And if $\|u\| = 0$ then it must be $\langle u\,,u\rangle = 0$, which implies $u = 0$, again by positive-definiteness.

Additionally, for any scalar $\lambda \in \mathbb{C}$ one has

$$\|\lambda u\| = \sqrt{\langle \lambda u\,,\lambda u\rangle} = \sqrt{\lambda\bar{\lambda}\,\langle u\,,u\rangle} = \sqrt{|\lambda|^2\,\langle u\,,u\rangle} = |\lambda|\sqrt{\langle u\,,u\rangle} = |\lambda|\,\|u\|\,.$$

The proof of subadditivity does not follow as easily. It requires a very important inequality of inner products, namely the Cauchy-Schwarz inequality.

Theorem 6.1 (Cauchy-Schwarz Inequality)**.** *Let E be an inner product space. Then for any $u, v \in E$*

$$|\langle u\,,v\rangle|^2 \leq \langle u\,,u\rangle\,\langle v\,,v\rangle\,.$$

Proof. For any $u, v \in E$ and $t \in \mathbb{C}$ we have, by positive-definiteness

$$\begin{aligned} 0 \leq \langle u - tv\,,u - tv\rangle &= \langle u\,,u\rangle + t\bar{t}\,\langle v\,,v\rangle - t\,\langle v\,,u\rangle - \bar{t}\,\langle u\,,v\rangle \\ &= \langle u\,,u\rangle + |t|^2\,\langle v\,,v\rangle - t\,\langle v\,,u\rangle - \bar{t}\,\langle u\,,v\rangle\,. \end{aligned}$$

Thus

$$t\,\langle v\,,u\rangle + \bar{t}\,\langle u\,,v\rangle \leq \langle u\,,u\rangle + |t|^2\,\langle v\,,v\rangle\,.$$

It is now a question of making an astute choice of t. Letting $t = \frac{\langle u\,,v\rangle}{\langle v\,,v\rangle}$ we find, for the LHS,

$$\begin{aligned} \frac{\langle u\,,v\rangle}{\langle v\,,v\rangle}\,\langle v\,,u\rangle + \overline{\frac{\langle u\,,v\rangle}{\langle v\,,v\rangle}}\,\langle u\,,v\rangle &= \frac{\langle u\,,v\rangle}{\langle v\,,v\rangle}\,\langle v\,,u\rangle + \frac{\langle v\,,u\rangle}{\langle v\,,v\rangle}\,\langle u\,,v\rangle \\ &= \frac{2\,|\langle u\,,v\rangle|^2}{\langle v\,,v\rangle}. \end{aligned}$$

Meanwhile, the RHS becomes

$$\langle u\,,u\rangle + \frac{|\langle u\,,v\rangle|^2}{\langle v\,,v\rangle^2}\,\langle v\,,v\rangle = \langle u\,,u\rangle + \frac{|\langle u\,,v\rangle|^2}{\langle v\,,v\rangle}.$$

Thus

$$\frac{2\,|\langle u\,,v\rangle|^2}{\langle v\,,v\rangle} \leq \langle u\,,u\rangle + \frac{|\langle u\,,v\rangle|^2}{\langle v\,,v\rangle}$$

or, equivalently,

$$|\langle u\,,v\rangle|^2 \leq \langle u\,,u\rangle\,\langle v\,,v\rangle\,.$$

□

The Triangle Inequality then follows. Indeed, taking square roots in Cauchy-Schwarz (which we can do since everything is positive, we find

$$|\langle u\,,v\rangle| \leq \sqrt{\langle u\,,u\rangle\,\langle v\,,v\rangle} = \|u\|\,\|v\|\,.$$

And from this follows

$$\begin{aligned}
\|u+v\|^2 &= \langle u+v\,,u+v\rangle \\
&= \langle u\,,u\rangle + \langle v\,,v\rangle + \langle u\,,v\rangle + \langle v\,,u\rangle \\
&= \|u\|^2 + \|v\|^2 + 2\,\mathrm{Re}\,\langle u\,,v\rangle \\
&\leq \|u\|^2 + \|v\|^2 + 2\,|\langle u\,,v\rangle| \\
&= \|u\|^2 + \|v\|^2 + 2\,\|u\|\,\|v\| \\
&= (\|x\| + \|y\|)^2
\end{aligned}$$

which becomes the Triangle Inequality after taking square roots. This then shows that the Hermitian product does indeed induce a norm.

An inner product space which is complete with respect to the norm induced by its inner product[3] is called a *Hilbert space*. A Hilbert space is in particular a Banach space, so all of the theory developed in the previous cahpters still holds true.

— ∗ —

It is possible to write the product of two vectors simply in terms of their norms, using the so-called polarisation identities.

Theorem 6.2 (Polarisation Identity)**.** *In an inner product space E the* polarisation identity

$$\langle u\,,v\rangle = \frac{1}{4}\left(\|u+v\|^2 - \|u-v\|^2 + i\,\|u+iv\|^2 - i\,\|u-iv\|^2\right)$$

holds for any $u, v \in E$.

If E is a real vector space, rather than a complex vector space, then the polarisation identity becomes

$$\langle u\,,v\rangle = \frac{1}{4}\left(\|u+v\|^2 - \|u-v\|^2\right).$$

[3] More often than not we will simply say it is complete with respect to the inner product

Proof. This is a direct computation using the definition of the norm induced by the inner product. Let's work out the first two terms

$$\|u+v\|^2 = \langle u+v\,, u+v\rangle = \langle u\,,u\rangle + \langle v\,,v\rangle + \langle u\,,v\rangle + \langle v\,,u\rangle$$
$$\|u-v\|^2 = \langle u-v\,, u-v\rangle = \langle u\,,u\rangle + \langle v\,,v\rangle - \langle u\,,v\rangle - \langle v\,,u\rangle\,.$$

Subtracting these two yields

$$\|x+y\|^2 - \|x-y\|^2 = 2\langle u\,,v\rangle + 2\langle v\,,u\rangle\,.$$

If we have an inner product, rather than a Hermitian product, then we are done because of symmetry. If we don't we need to get rid of the second term, which is what the other two terms in the polarisation expression do:

$$i\|u+iv\|^2 = i\langle u+iv\,, u+iv\rangle = i\langle u\,,u\rangle + i\langle v\,,v\rangle + \langle u\,,v\rangle - \langle v\,,u\rangle$$
$$i\|u-iv\|^2 = i\langle u-iv\,, u-iv\rangle = i\langle u\,,u\rangle + i\langle v\,,v\rangle - \langle u\,,v\rangle + \langle v\,,u\rangle$$

for when we subtract we get

$$i\|u+iv\|^2 - i\|u-iv\|^2 = 2\langle u\,,v\rangle - 2\langle v\,,u\rangle\,.$$

And then polarisation follows. □

Another useful identity satisfied by norms induced by an inner product is the parallelogram law.

Theorem 6.3 (Parallelogram Law)**.** *The following identity holds for a norm induced by an inner product:*

$$\|u+v\|^2 + \|u-v\|^2 = 2\|u\|^2 + 2\|v\|^2\,.$$

Proof. This is again a simple computation:

$$\begin{aligned}\|u+v\|^2 + \|u-v\|^2 &= \langle u+v\,, u+v\rangle + \langle u-v\,, u-v\rangle \\ &= 2\langle u\,,u\rangle + 2\langle v\,,v\rangle + \langle v\,,u\rangle + \langle u\,,v\rangle - \langle v\,,u\rangle - \langle u\,,v\rangle \\ &= 2\|u\|^2 + 2\|v\|^2\,.\end{aligned}$$

□

This identity can be used to show that many of the normed spaces we have dealt with before are not inner product spaces. Take $C^0(K)$ with the uniform norm, for instance: one can find a function $f \in C^0(K)$ which is zero outside of a ball centered aroound a point $x \in K$ and such that $\|f\|_\infty = |f(x)| = 1$. We could then pick another $g \in C^0(K)$ which is zero outside of a ball centered at $y \in C^0(K)$ which does not intersect the ball in which f is nonzero, and such that $\|g\|_\infty = |f(y)| = 1$. Then

$$\|f+g\|_\infty + \|f-g\|_\infty = 2 \neq 4 = 2\|f\|_\infty + 2\|g\|_\infty$$

which shows $\|\cdot\|_\infty$ cannot come from an inner product on $C^0(K)$. On the other hand, $L^2(E)$ is an inner product space, for its norm is induced by the product

$$\langle f\,, g\rangle = \int_E fg$$

in the real case and

$$\langle f\,, g\rangle = \int_E f\bar{g}$$

in the real case. It's easy to see that these are indeed an inner product and a Hermitian product, respectively.

A remarkable fact is that the converse is true, meaning a norm which satisfies the Parallelogram Law must be induced by an inner product, and we can use the Polarisation Identity to recover said inner product. The proof of this is non-trivial. As a consequence, it can be shown that the spaces $L^p(X)$ and ℓ^p are inner product spaces exactly when $p = 2$.

6.3 Orthogonality and orthogonal complements

Once we have an inner product at our disposal we can define an important geometrical notion on our space, that of orthogonality.

Definition 6.3 (Orhtogonality)**.** We say two non-zero vectors u and v of an inner product space E are orthgonal if $\langle u\,, v\rangle = 0$. △

A generalised version of Pythagoras' Theorem holds in inner product spaces.

Theorem 6.4 (Pythagoras)**.** *If $u_1, \ldots, u_n \in E$ are pairwise orthogonal vectors in an inner product space E then*

$$\|x_1 + \cdots + x_n\|^2 = \|x_1\|^2 \ldots \|x_n\|^2 .$$

Proof. Let's first show the result for $n = 2$. We have

$$\begin{aligned}\|x+y\|^2 &= \langle x+y\,, x+y\rangle \\ &= \langle x\,, x\rangle + \langle y\,, y\rangle + \langle x\,, y\rangle + \langle y\,, x\rangle \\ &= \langle x\,, x\rangle + \langle y\,, y\rangle \\ &= \|x\|^2 + \|y\|^2 .\end{aligned}$$

The general case then follows by induction and by observing that if x_n is orthogonal to $x_1, \ldots, x_n$ then it is orthogonal to their sum. $\square$

— ∗ —

Definition 6.4 (Orthogonal complement)**.** We define the *orthogonal complement* of a subset $A \subseteq E$ of an inner product space E as the set of all vectors in E orthogonal to every vector in A:

$$A^{\perp} := \{x \in E \mid \forall a \in A\colon\ \langle x\,, a\rangle = 0\}.$$

$\triangle$

Note that the set A need not be a linear subspace of E. Nonetheless, the orthogonal complement of any subset is a closed linear subspace. To prove this we make use of the following result.

Proposition 6.5. *The inner product is continuous and for any $x \in E$, the map*

$$\begin{aligned} \langle \cdot\,, x\rangle : H &\longrightarrow \mathbb{C} \\ y &\longmapsto \langle y\,, x\rangle \end{aligned}$$

is a bounded operator with norm $\|x\|$.

Proof. The continuity of the inner product follows from the following estimate:

$$\begin{aligned} |\langle x\,, y\rangle - \langle a\,, b\rangle| &= |\langle x\,, y\rangle - \langle a\,, y\rangle + \langle a\,, y\rangle - \langle a\,, b\rangle| \\ &= |\langle x - a\,, y\rangle - \langle a\,, y - b\rangle| \\ &\leq |\langle x - a\,, y\rangle| - |\langle a\,, y - b\rangle| \\ &\leq \|x - a\|\,\|x - a\| + \|a\|\,\|b - y\| \end{aligned}$$

where we used Cauchy-Schwarz in the last step.

The linearity of $\langle \cdot\,, x\rangle$ is a consequence of the sesquilinearity of the inner product. The boundedness follows from Cauchy-Schwarz, since for any $a \in E$

$$|\langle a\,, x\rangle| \leq \|x\|\,\|a\|$$

therefore $\|\langle \cdot\,, x\rangle\| \leq \|x\|$. And to show equality observe that, by definition,

$$|\langle x\,, x\rangle| = \|x\|^2$$

thus $\|\langle \cdot\,, x\rangle\| = \|x\|$. $\square$

We can then show that the orthogonal complement satisfies the following useful properties.

Proposition 6.6. *Let E be an inner product space. Then the follwoing are true*

(i) *If $A \subseteq B \subseteq E$ are any two subsets then*

$$B^{\perp} \subseteq A^{\perp}.$$

(ii) *The orthogonal complement of any subset of E is a closed linear subspace of E.*

Proof. (i) If $x \in B^{\perp}$ then x is orthogonal to every element of B, which means, in particular, it is orthogonal to any element of A. Thus $x \in A^{\perp}$. (ii) If $x, y \in A^{\perp}$ then, for any $\lambda, \mu \in \mathbb{R}$ and $a \in A$

$$\langle \lambda x + \mu y \, , a \rangle = \lambda \langle x \, , a \rangle + \mu \langle y \, , a \rangle = 0$$

so $\lambda x + \mu y \in A^{\perp}$. Since $0 \in A^{\perp}$ we conclude $A^{\perp}$ is a linear subspace. To show it is closed, consider a sequence (x_n) in $A^{\perp}$ which converges to $x \in E$. If we can show $x \in A^{\perp}$ then we are done. But this follows from the continuity of the inner product we just proved: indeed, for any $a \in A$

$$\langle x \, , a \rangle = \left\langle \lim_{n\to\infty} \, , a \right\rangle = \lim_{n\to\infty} \langle x_n \, , a \rangle = \lim_{n\to\infty} 0 = 0$$

so $x \in A^{\perp}$.

Alternatively, observe that

$$A^{\perp} = \bigcap_{a \in A} \langle \cdot \, , a \rangle^{-1} (\{0\}).$$

As we proved before, the linear form $\langle \cdot \, , a \rangle$ is bounded, therefore continuous, thus $\langle \cdot \, , a \rangle^{-1} (\{0\})$ is closed since $\{0\}$. Thus $A^{\perp}$ is also closed since it is the intersection of closed sets. $\square$

Chapter 7

Projection onto subspaces

7.1 Existence and uniqueness of optimal approximations

Recall that given a closed subset A of a normed space[1], so in particular of an inner product space H, we can talk of the distance between a point $x \in H$ and A, which is

$$d(x, A) = \inf_{a \in A} \|x - a\| .$$

The closedness of A is necessary if we want the property that $d(x, A) = 0$ if and only if $x \in A$. We have, then, that $d(x, A) \geq 0$ provided $x \notin A$. This does not mean, however, that there exists $a \in A$ such that $\|x - a\| = d(x, A)$. If this is the case, though, we say a is an *optimal approximation* of x in A. This means a is the closest we will ever get to x if we restrict ourselves to A.

In this section we will investigate the existence and uniqueness of optimal approximations depending on the properties of A.

Proposition 7.1. *If $A \subseteq H$ is a closed and convex[2] and x has an optimal approximation in A then it is unique.*

Proof. Let $y, z \in A$ be optimal approximations of x in A. To prove uniqueness we need to show that $y = z$. By definition of optimal approximation, we have

$$\|y - x\| + \|z - x\| = d(x, A) > 0,$$

the last inequality because A is closed.

[1]We could consider this for metric spaces if we wanted to be extra general but normed spaces suffice

[2]Recall that a subset A of a vector space is convex if, for any two points a and b in A then the segment between them is also in A, and the segment between a and b is the set of all linear combinations of a and b such that their coefficients add up to 1.

The idea of the proof is to use the fact that if two points y and z are equidistant from a third point x, then any other point in the segment between y and z is closer to x than y and z. Then, if y and z are optimal approximations of x then, because of the convexity of A, there must be other points in A which are closer to x. This cannot be the case unless the segment between z and y contains no points other than z and y themselves, which implies y and z are equal.

Let's write this out. By virtue of A being convex, the point $\frac{y+z}{2}$ lies in A. This means

$$\left\| \frac{y+z}{2} - x \right\| \geq d(x, A).$$

On the other hand, using the Parallelogram Law we have

$$2 \|y-x\|^2 + 2\|z-x\|^2 = \|y-z\|^2 + \left\|y+z-2x^2\right\| = \|y-z\|^2 + 4\left\| \frac{y+z}{2} - x \right\|^2.$$

Since $\|y-x\|^2 = \|z-x\|^2 = d(x,A)^2$, we have

$$4\left\| \frac{y+z}{2} - x \right\|^2 = 4d(x,A)^2 - \|y-z\|^2.$$

But using $\left\| \frac{y+z}{2} - x \right\| \geq d(x,A)$ one finds $\|y-z\| = 0$, which means, of course, that $y = z$, as we wanted. □

The requirement that A be convex is important. Indeed, consider the unit circle S^1 in $\mathbb{R}^2$. Then every point of S^1 is distance 1 from the origin, thus they all are optimal approximations of it. S^1 is not convex, however.

When x has a unique optimal approximation in a set A we write it $P_A x$ and call it its *projection onto* A.

Theorem 7.2. *Let H be a Hilbert space and $A \subseteq H$ a subset. The map*

$$\begin{aligned} P_A \colon H &\longrightarrow A \\ x &\longmapsto P_A x \end{aligned}$$

is well-defined provided A is closed and convex.

Proof. All that remains to be shown is that an optimal approximation in A exists, since the previous proposition shows that it is unique if it does exist. For this we need to use the fact that A is closed.

Since $d(x,A)$ is defined as an infimum, there exists a sequence y_n in A such that $\|y_n - a\| \xrightarrow{n\to\infty} d(x,A)$. Using the Parallelogram Law we find

$$\begin{aligned} \|y_n - y_m\|^2 + 4\left\| \frac{y_n + y_m}{2} - x \right\|^2 &= \|y_n - y_m\|^2 + \|y_n + y_m - 2x\|^2 \\ &= 2\|y_n - x\|^2 + 2\|y_m - x\|^2. \end{aligned}$$

Because A is convex, $\frac{y_n+y_m}{2} \in A$ and thus

$$\left\| \frac{y_n + y_m}{2} - x \right\| \geq d(x, A).$$

This means

$$\begin{aligned} \|y_n - y_m\|^2 &\leq 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - 4d(x, A)^2 \\ &= 2(\|y_n - x\|^2 - d(x, A)^2) + 2(\|y_m - x\|^2 - d(x, A)^2). \end{aligned}$$

And now just use the fact that $\|y_n - x\|$ converges to $d(x, A)$ to conclude y_n is Cauchy. Because H is a Hilbert space it is complete and therefore y_n converges to $y \in H$. But because A is closed, in fact $y \in A$. And, thus

$$d(x, A) = \lim_{n\to\infty} \|y_n - x\| = \left\| \lim_{n\to\infty} y_n - a \right\| = \|y - a\|.$$

It follows, then, that $y = P_A x$. □

Corollary 7.3. *The projection onto a closed linear subspace of a Hilbert space is well defined.*

Proof. This follows from the fact that a linear subspace is convex. □

For the case of linear subspaces, we can give a useful characterisation of the projection in terms of the orthogonal complement.

Proposition 7.4. *Let $F \subseteq H$ be a closed linear subspace of a Hilbert space. Then $y \in F$ is the projection of $x \in H$ onto F if and only if $x - y \in F^\perp$.*

Proof. ($\Longrightarrow$) Say $y \in F$ is such that $y - x \in F^\perp$. Then, for any other $z \in F$ we have $\langle x - y\,, y - z\rangle = 0$. Then, since $x - z = x - y + y - z$ use Pythagoras to conclude

$$\|x - z\|^2 = \|x - y\|^2 + \|y - z\|^2 \geq \|x - y\|^2.$$

This means $\|x - y\| \leq \inf_{z\in F} \|x - z\|$. But since $y \in F$ we conclude

$$\|x - y\| = \inf_{z\in F} \|x - z\| = d(x, F)$$

which means $y = P_F x$.

($\Longleftarrow$) The proof of the converse is a bit more involved. Suppose $y = P_F x$. This means $\|y - x\| = d(x, F)$. We need to show $\langle y - x\,, z\rangle = 0$ for all $z \in F$. Since $y = P_F x$ we have, for all $\lambda \in \mathbb{C}$ and $z \in F$

$$\|x - (y + \lambda z)\|^2 \geq \|x - y\|^2.$$

Then

$$\begin{aligned}\|x-y\|^2 \leq \|x-(y+\lambda z)\|^2 &= \langle x-(y+\lambda z)\,, x-(y+\lambda z)\rangle \\ &= \|x-y\|^2 + \|\lambda z\|^2 - \langle x-y\,, \lambda z\rangle - \langle \lambda z\,, x-y\rangle \\ &= \|x-y\|^2 + \|\lambda z\|^2 - 2\,\mathrm{Re}\,(\lambda\,\langle z\,, x-y\rangle)\,.\end{aligned}$$

Cancelling $\|x-y\|^2$ and rearranging, this is equivalent to

$$|\lambda z|^2\,\|z\|^2 \geq 2\,\mathrm{Re}\,(\lambda\,\langle z\,, x-y\rangle)\,.$$

Let's pick $\lambda = s\,\langle x-y\,, z\rangle$ with $s \in \mathbb{R}$. Then this means

$$s^2\,\|z\|^2\,|\langle x-y\,, z\rangle|^2 \geq 2s\,|\langle x-y\,, z\rangle|^2\,.$$

Now, since $\frac{cs^2}{2s} \xrightarrow{s\to 0}$ for any $c > 0$ we can find $s \in \mathbb{R}$ such that $2s \leq \|z\|^2\,s^2$ (why?). This must mean $\langle x-y\,, z\rangle = 0$ since otherwise the inequality could not possibly hold for any $s \in \mathbb{R}$. Therefore $x - y \in F^\perp$, as we wished. □

All of these facts lead to the following important result in the theory of Hilbert spaces.

Theorem 7.5 (Projection Theorem)**.** *Given any closed linear subspace F of a Hilbert space H then for any $x \in H$*

$$P_F x + P_{F^\perp} x = x.$$

Proof. First, since both F and $F^\perp$ are closed linear subspaces, the projections onto both are well defined. If we show $P_{F^\perp} x = x - P_F x$ then we are done. For any $y \in F^\perp$ we have

$$\|x-y\|^2 = \|x - P_F x + P_F x - y\|^2$$

We showed $x - P_F x \in F^\perp$, which means $x - P_F x - y$ is also in $F^\perp$, thus

$$\langle P_F x\,, x - P_F x - y\rangle = 0,$$

which in turns implies, by Pythagoras

$$\|x-y\|^2 = \|P_F x\|^2 + \|x - P_F x - y\|^2 \geq \|P_F x\|^2 = \|x - (x - P_F x)\|^2\,.$$

This lets conclude $P_F^\perp = x - P_F x$. □

Corollary 7.6. *Any Hilbert space H can be decomposed as*

$$H = F \oplus F^\perp$$

where F is a closed linear subspace of H.

Proof. By the Projection Theorem, we have $H = F + F^{\perp}$, so all we need to show is $F \cap F^{\perp} = \langle 0 \rangle$. And this is true. Indeed, if $x \in F \cap F^{\perp}$ then, in particular, $\langle x\,, x \rangle = 0$. And this means $x = 0$, by positive-definiteness. □