Real Analysis

Corran Webster

April 26, 2006

Contents

Introduction

This is a massive set of notes, so approaching this may be somewhat daunting. It's also somewhat incomplete, and I probably won't be covering all of the material in these notes in this class.

So here's a general guide to what we're going to be doing.

We're going to start off covering Chapter 1 pretty much completely. There isn't much material that is missing from this chapter, and it's the core of Math 707.

The next objective after that is to generalize measure and integration to arbitrary σ-algebras. This requires a little topology, since many important examples of σ-algebras are closely related to topological spaces. However, I won't be covering everything in the chapter on topology, and you'll notice that there are few or no proofs in that chapter, so it will of necessity be a fairly shallow look at topology. I strongly encourage you to take a course in topology at some point in your studies.

With topology out of the way, we can then define general σ-algebras, and general measures on them, and define integrals using these measures in a way which is analogous to Chapter 1. Most of the key results of Chapter 1 have corresponding results, and we also get results about product spaces (giving integration on $\mathbb{R}^n$ almost for free). Towards the end of this section, some of the proofs are omitted, but hopefully I should have time to fill in the details before we get to them in class.

The last part of the notes cover the theory of L^p spaces, which are prototypical Banach and Hilbert spaces. This is likely to be material which we cover in 708, and a lot of this is concerned with verifying basic facts about certain norms that were discussed in 458/658 last year, such as the fact that the triangle inequality holds for them.

The appendices contain notation and basic definitions that you hopefully have seen in undergraduate classes, but you may not have seen some of this discussion before. I'll take time to discuss some of these issues in class, particularly if people are confused about things. It's worth noting that the definitions are not restricted to just analysis, but also include some algebraic concepts which are useful, particularly when we get to talking about Banach and Hilbert spaces.

Chapter 1

Lebesgue Measure on $\mathbb{R}$

1.1 Introduction

In your study of integration, you should be aware that there are functions, like the characteristic function of the rational numbers, $\chi_{\mathbb{Q}}$, which are not Riemann integrable on any interval. This may seem to be a contrived example, but functions like $\chi_{\mathbb{Q}}$ which are discontinuous at every point are common in many real-world applications. For example, a function giving the forces experienced by a particle subject to Brownian motion would be discontinuous everywhere and would almost certainly not be Riemann integrable. The same would be true of functions which model "white noise" and functions which model static in a communication channel. In fact, a strong argument can be made that such nowhere-continuous functions are far more common than the elementary functions you learned to integrate in calculus.

You may also know that the behaviour of the Riemann integral is not very good when considering sequences of functions. You should know that uniform limits of Riemann integrable functions preserve the integral, but pointwise limits do not. While the previous problem may seem more important from a standpoint of applying integration to models of the real world, the "bad" behaviour of the Riemann integral when looking at sequences of functions is a more significant stumbling block from the theoretical viewpoint.

The following example illustrates that at least some of the time, being able to take pointwise limits would allow us to find integrals of more functions.

Example 1.1.1

You should know that the characteristic function of the rational numbers $\chi_{\mathbb{Q}}$ is not Riemann integrable on any interval.

However, observe that $\chi_{\mathbb{Q}}$ is the pointwise limit of an increasing sequence of functions χ_{Q_n}, where

$$Q_n = \{q_k : 1 \leq k \leq n\}$$

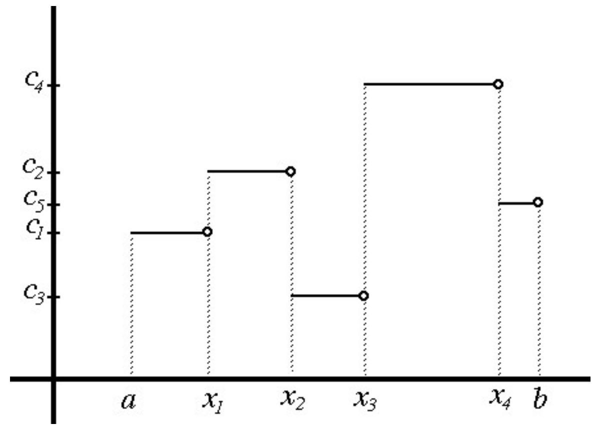

Figure 1.1: A step function of the form described in Equation (1.1)

and $\{q_n\}_{n=1}^{\infty} = \mathbb{Q}$ (for example, $n \mapsto q_n$ could be an enumeration of the rational numbers). We can check that

$$\int_a^b \chi_{Q_n}(x)\ dx = 0$$

for all n. It seems reasonable that the integral of $\chi_{\mathbb{Q}}$ *ought* to be 0. $\diamond$

To overcome the problem, there are two basic approaches. One is to investigate what is happening in the limiting process to find a reasonable way to extend the integral to more functions[1]. This involves learning at least enough functional analysis to be able to talk about linear maps on vector spaces which also have topology. While you will hopefully eventually study this theory, there is an alternative way to approach the problem.

One way to think of the definition of the Riemann integral, is that we are approximating the integrand f by step functions of the form

$$\varphi(x) = \sum_{k=1}^{n} c_n \chi_{[x_{n-1}, x_n)}(x), \tag{1.1}$$

where c_n are constants (usually chosen so that $c_n = f(x_n*)$, for some $x_n* \in [x_{n-1}, x_n]$), and $\mathcal{P} = \{x_0 < x_1 < x_2 < \ldots < x_n\}$ is a partition of $[a, b]$. We can easily integrate these simple functions φ

$$\int_a^b \varphi(x)\ dx = \sum_{k=1}^{n} c_n (x_k - x_{k-1}),$$

[1]This method is sometimes called the Daniell integral, although it is completely equivalent to the Lebesgue integral that we will discuss.

to get a Riemann sum. We hope that as the function φ approximates f better and better, the Riemann sums converge to some number which is what we define the Riemann integral of f to be.

This is deliberately vague, but it indicates a way forward. The integral of φ is given by the sum of areas of the rectangles of the form $[x_{k-1}, x_k) \times [0, c_k]$ (see Figure 1.1). If we could replace the intervals $[x_{k-1}, x_k)$ by more general sets E_k, then we could replace the area of the rectangles by the areas of the sets $E_k \times [0, c_k]$.

More generally, if we were integrating a function $f : \mathbb{R}^n \to \mathbb{R}$ in some box, the Riemann sum uses the volumes of boxes of the form form $[x^1_{k_1-1}, x^1_{k_1}) \times [x^2_{k_2-1}, x^2_{k_2}) \times \cdots \times [x^n_{k_n-1}, x^n_{k_n}) \times [0, c)$, and would hope to replace these by sets of the form $E \times [0, c)$, where $E \subseteq \mathbb{R}^n$ and then the volume of this set would be the volume of E times c.

In a perfect world, this would work for any set $E \subset \mathbb{R}^n$. The world is not perfect, but it does give us our first objective: we want to measure the "size", or "volume", of an arbitrary set.

In other words, we would like to abstract the idea of "area," "volume," or "length" in much the same way that a metric abstracts the idea of distance, or a norm abstracts the idea of the length of a vector. In other words we would like to find some simple axioms which we can work with to prove basic facts which match our intuition about volumes.

To simplify discussion, we'll talk in terms of volume, although it should be readily apparent that these ideas apply to more than just volume. A little thought should tell you that there are some basic properties which should hold for a good definition of volume:

1. the volume of a set is at least 0, and a set can have infinite volume.

2. The volume of the empty set is 0.

3. Given a countable collection of *disjoint* sets, the volume of the union is the sum of the volume of each set.

4. Given two sets which are not disjoint, the volume of their union is the sum of the volumes, less the volume of the intersection (providing that the volume of the intersection is finite). This is the "inclusion-exclusion" principle which should be familiar from your undergraduate work.

5. The volume of a subset is less than or equal to the volume of the set which contains it.

6. In $\mathbb{R}^n$ the volume of a box is the product of the side lengths.

7. The volume of a set is invariant under translations, rotations and reflections.

Letting $m(E)$ denote the "volume" of the set E, we can translate these into the following mathematical conditions:

1. m is a function from sets to $[0, \infty]$.

2. $m(\emptyset) = 0$.

3. Given disjoint sets $E_1, E_2, \dots, E_k, \dots$,

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

4. If $m(E \cap F) < \infty$, then $m(E \cup F) = m(E) + m(F) - M(E \cap F)$.

5. If $E \subseteq F$, then $m(E) \leq m(F)$.

6. If I_k is an interval from a_k to b_k, $m(I_1 \times I_2 \times \cdots \times I_n) = \prod_{k=1}^{n} |b_k - a_k|$.

7. If $u : \mathbb{R}^n \to \mathbb{R}^n$ is an isometric transformation (ie. a translation, rotation, reflection or combination thereof), then $m(u(E)) = m(E)$.

Note:: *since m can take th value $+\infty$ on some sets, an want to at least be able to values, we have to use the r for extended real numbers discussed in Section A.2.*

The third statement is probably the most controversial: why *countable* collections? Hopefully you would agree with the statement for finite collections of sets. It is easy to see that it is quite unreasonable for uncountable collections: consider the singleton sets $\{x\}$ where $x \in \mathbb{R}$; they are disjoint, have no volume, yet their union has infinite volume. Countable disjoint unions of sets are reasonable at least some of the time, however. We know, for example, that the sets $U_n = [2^{-n-1}, 2^{-n})$ for $n = 1, 2, \dots, \infty$ have volume 2^{-n-1}, are disjoint, and their union is the open interval $(0, 1)$ and we know that

$$\sum_{n=1}^{\infty} 2^{-n-1} = 1 = m((0, 1)).$$

Hopefully this quick example convinces you that countability is at least a reasonable thing to want.

In any case, these are good starting points. Unfortunately, there is one severe problem: this "volume" m cannot possibly be defined for *all* sets. In fact, working with just $\mathbb{R}$ and only four of the assumptions:

1. $m : \mathcal{P}(\mathbb{R}) \to [0, \infty]$

2. if $E_1, E_2, \dots$ are disjoint subsets of $\mathbb{R}$, then

$$m(E_1 \cup E_2 \cup \dots) = m(E_1) + m(E_2) + \dots.$$

3. $m([0, 1)) = 1$.

4. If E and F are subsets of $\mathbb{R}$ which are congruent under translation and/or reflection (ie. "length-preserving" transformations), then $m(E) = m(F)$.

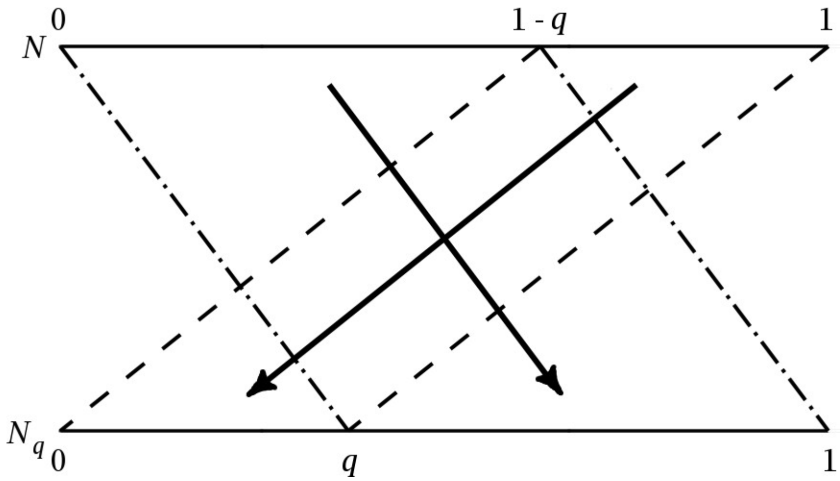

Figure 1.2: The transformation of N to N_q from Example 1.1.2

we can get a contradiction.

Example 1.1.2

Consider an equivalence relation on $[0, 1)$ where $x \sim y$ iff $x - y \in \mathbb{Q}$. So, for example, all rational numbers are equivalent to each other. But $1/\pi$ and $1/\sqrt{2}$ are not.

Let N be a subset of $[0, 1)$ consisting of one representative from each equivalence class under $\sim$.

Now for every rational number $q \in Q = [0, 1) \cap \mathbb{Q}$, let

$$N_q = \{x + q : x \in N, 0 \le x < 1 - q\} \cup \{x - (1 - q) : x \in N, 1 - q \le x < 1\},$$

ie. to get N_q, shift $N \cap [0, 1 - q)$ q units to the right and shift $N \cap [1 - q, 1)$ left by $1 - q$ units, and take the union (see Figure 1.1). One can show that given q, $r \in Q$, with $q \ne r$, that N_q and N_r are disjoint. (See Section 1.11 for the proof of this claim.)

Let $y \in [0, 1)$. Then by construction, there is a unique $x \in N$ such that $x \in [y]$, so by the definition of equivalence, $y - x$ is a rational number. If $y \ge x$, then let $q = y - x \in Q$, so that $1 - q = 1 - (y - x) = x + (1 - y) > x$ and $y = x + q$, from which we conclude that $y \in N_q$. On the other hand, if $y < x$, then let $q = 1 - (x - y)$, so that $1 - q = x - y \ge x$ and $y = x - (1 - q)$, from which we conclude that $y \in N_q$. So every $y \in [0, 1)$ lies in N_q for some $q \in Q$.

Therefore $[0, 1) = \bigcup_{q \in Q} N_q$, and the N_q are disjoint. In other words, $[0, 1)$ is a countable disjoint union of the sets N_q.

Now by the disjoint union property of m,

$$m(N_q) = m(N_q \cap [0, q)) + m(N_q \cap [q, 1)).$$

Since $N_q \cap [1-q, 1)$ is the set $N \cap [0, 1-q)$ translated right by q units, and $N_q \cap [0, q)$ is the set $N \cap [1-q, 1)$ translated left by $1-q$ units, we have by the translation property

$$m(N_q \cap [0, q)) = m(N \cap [1-q, 1)) \qquad \text{and} \qquad m(N_q \cap [q, 1)) = m(N \cap [0, 1-q)).$$

Therefore, using the disjoint union property once more,

$$m(N_q) = m(N \cap [1-q, 1)) + m(N \cap [0, 1-q)) = m(N).$$

Now, since $[0, 1)$ is the disjoint union of the sets $m(N_q)$ for $q \in Q$, we conclude that.

$$m([0, 1)) = \sum_{q \in Q} m(N_q) = \sum_{q \in Q} m(N).$$

Now either, $m(N) = 0$, in which case $m([0, 1)) = 0$; or $m(N) > 0$, in which case $m([0, 1)) = \infty$. Either way, this contradicts our third assumption that $m([0, 1)) = 1$.

So there is no such volume function m. $\diamond$

One thing to note is that this construction depends upon the axiom of choice: to get the set N we are choosing an element from each of an uncountable collection of equivalence classes. The fact that you probably didn't notice that the axiom of choice was being used might indicate that it's not an unreasonable axiom. However, if you deny the axiom of choice and equivalent axioms, you don't have these problems. In this course, we will use the axiom of choice from time to time, because it is a well-accepted part of standard analysis.

Our solution to the problem, then, is to avoid sets like N from the above example. We want to find a class of sets which includes all "reasonable" sets (like $\mathbb{Q}$, and "fractal" sets), but avoids bad sets like N. The way that we will do this is by building up from sets we know are safe: intervals.

Exercises

1.1.1. Show that $\chi_{\mathbb{Q}}$ is not Riemann integrable. Show that the functions χ_{Q_n} of Example 1.1.1 converge pointwise to $\chi_{\mathbb{Q}}$. Show that

$$\int_a^b \chi_{Q_n}(x)\, dx = 0.$$

1.1.2. Find Riemann integrable functions f_n which converge pointwise to some Riemann integrable function f, but for which

$$\lim_{n \to \infty} \int_a^b f_n(x)\, dx \neq \int_a^b f(x)\, dx.$$

1.1.3. Let f_n be an increasing sequence of Riemann integrable functions, and let f_n converge to f pointwise, where f is also Riemann integrable. Show that

$$\lim_{n\to\infty} \int_a^b f_n(x)\ dx = \int_a^b f(x)\ dx.$$

1.1.4. Find another function which is not Riemann integrable.

1.1.5. Let f be a Riemann integrable function on $[a,b]$. Let $\mathcal{F}$ be the set of all step functions of the form given by (1.1). Show that

$$\int_a^b f(x)\ dx = \sup \left\{ \int_a^b \varphi(x)\ dx : \varphi \in \mathcal{F}, \varphi \le f \right\}.$$

and

$$\int_a^b f(x)\ dx = \inf \left\{ \int_a^b \varphi(x)\ dx : \varphi \in \mathcal{F}, \varphi \ge f \right\}.$$

1.1.6. Show that if f is Riemann integrable on $[a,b]$, then one can find an increasing sequence φ_n of step functions of the form given by (1.1) with $\varphi_n \le f$ such that $\varphi_n \to f$ pointwise, and

$$\int_a^b f(x)\ dx = \lim_{n\to\infty} \int_a^b \varphi_n(x)\ dx.$$

1.1.7. Find a sequence φ_n of step functions of the form given by (1.1) which converge pointwise to a Riemann integrable function f, but for which

$$\lim_{n\to\infty} \int_a^b \varphi_n(x)\ dx \ne \int_a^b f(x)\ dx.$$

1.1.8. Let $f(x) = 1/\sqrt{x}$ on $(0,1]$ and $f(0) = 0$. Show that $f(x)$ is Riemann integrable. Show that f cannot be a uniform limit of step functions of the form given by (1.1).

1.1.9. Show that the relation $x \sim y$ iff $x - y \in \mathbb{Q}$ from Example 1.1.2 is an equivalence relation as claimed.

1.1.10. Based on the proposed properties we want for the volume of a set, what should the volume of $\mathbb{Q}$ be?

Notes: Exercise 1.1.8 shows that uniform limits of step functions are too weak to give us all Riemann integrable functions; Exercise 1.1.7 shows that pointwise limits of step functions are strong enough to give us all Riemann integrable functions as limits of step functions, but integrals are not preserved by them; but Exercises 1.1.3 and 1.1.6 shows that increasing pointwise limits of step functions are nicely behaved.

Some of these exercises are quite hard.

1.2 Measures and σ-Algebras

For the rest of this section we will just work with $\mathbb{R}$, as opposed to $\mathbb{R}^n$ for simplicity. However, we will still say that we are trying to measure the "volume" of a set in $\mathbb{R}$, even though it is more properly some sort of "total length," because volume gives a better mental image of the properties we want.

We start be defining a class of sets for which we know we can safely define the volume: finite unions of intervals.

An ***interval*** in $\mathbb{R}$, is a connected subset of $\mathbb{R}$. In other words it is a set of the form (a, b), $(a, b]$, $[a, b)$, $[a, b]$, $(-\infty, b)$, $(-\infty, b]$, (a, ∞), $[a, \infty)$ or $(-\infty, \infty)$, where a, $b \in \mathbb{R}$ and $a \leq b$. We let $\mathcal{I}$ be the collection of all intervals.

Note:: *the definition of an interval includes the empty set as a trivial interval of the form (a, a). This makes many arguments easier, as we don't have to consider the empty set as a special case.*

Definition 1.2.1
*A subset of $\mathbb{R}$ is **elementary** if it is a finite union of intervals. Let $\mathcal{E}$ be the collection of all elementary subsets of $\mathbb{R}$.*

Note that any set which can be written as a finite union of intervals can be written as a finite union of disjoint intervals: if two intervals are not disjoint, then their union is an interval, and so they can be replaced by their union. Indeed, we can assume that the union is minimal in the sense that there is no pair of intervals in the union whose union is again an interval.

We also note the following:

Proposition 1.2.1
If A and B are elementary sets, then so are $A \cup B$, $A \cap B$ and A^c.

Proof:

Let $A = I_1 \cup \ldots \cup I_n$ and let $B = J_1 \cup \ldots \cup J_m$ be representations as finite unions of intervals. Then clearly $A \cup B = (I_1 \cup \ldots \cup I_n) \cup (J_1 \cup \ldots \cup J_m)$ is a finite union of intervals.

We note that $I_k \cap J_l$ is always an interval (although it may often be the empty set), and using the distributive law for unions and intersections, we have

$$A \cap B = \bigcup_{k=1}^{n} \bigcup_{l=1}^{m} I_k \cap J_l.$$

Hence $A \cap B$ is a finite union of intervals.

Finally, note that if I_k has left endpoint a_k and right endpoint b_k, then I_k^c is the union of two intervals: one of the form $(-\infty, a_k)$, $(-\infty, a_k]$ or the empty set (if $a_k = -\infty$); and the other of the form (b_k, ∞), $[b_k, \infty)$ or $\emptyset$ (if $b_k = \infty$). Now DeMorgan's law tells us that

$$A^c = (I_1 \cup \ldots \cup I_n)^c = I_1^c \cap I_2^c \cap \ldots \cap I_n^c.$$

So A^c is a finite intersection of elementary sets. Repeated application of the previous part shows us that A^c must therefore be an elementary set. ■

In fact, you can prove the last part of this result directly because by permuting the order of the sets I_k, we can always ensure that if the endpoints of I_k are a_k and b_k then

$$-\infty \leq a_1 \leq b_1 \leq a_2 \leq b_2 \leq \ldots \leq a_n \leq b_n \leq \infty.$$

It is then easy to see that A^c is the union of the intervals that lie between one interval of A and the next. Once we know $A \cup B$ and A^c are elementary, we can show that $A \cap B$ is also elementary by noting that DeMorgan's law tells us

$$A \cap B = (A^c \cup B^c)^c.$$

So in looking for an axiomatisation of things like $\mathcal{E}$, we really only need unions and complements:

Definition 1.2.2
A family $\mathcal{A}$ of subsets of $\mathbb{R}$ is an ***algebra*** *if $\mathbb{R} \in \mathcal{A}$, and whenever A, $B \in \mathcal{A}$, then $A \cup B$ and A^c are both in $\mathcal{A}$.*

Note: *There's no reason we need to restrict ourselves to subsets of $\mathbb{R}$, as we will see later.*

So Proposition 1.2.1 showed that $\mathcal{E}$ is an algebra of sets. We can also see that intersections of sets in an algebra must be in the algebra for the same reason it happens in $\mathcal{E}$. Indeed it is easy to show that if A and B are sets in an algebra, then $A \setminus B$ and $A \bigtriangleup B$ are in the algebra as well (see Exercise 1.2.1). These facts can often simplify arguments. To summarise:

Note: *when we need to distinguish an algebra in this sense from the linear algebra definition of an algebra (a vector space which has a consistent ring structure), we will call this sort of algebra a* set algebra. *Some texts call an algebra a* ***field*** *of sets. Also, some references work with a slightly more general object, called a* ***ring*** *which doesn't require that $\mathbb{R}$ is in the ring, and instead of requiring that A^c is in the ring, requires that $A \setminus B$ is in the ring.*

Proposition 1.2.2
If $\mathcal{A}$ is an algebra, and A, $B \in \mathcal{A}$, then $\emptyset$, $A \cap B$, $A \setminus B$ and $A \bigtriangleup B \in \mathcal{A}$.

Since we want to be able to deal with countable unions of intervals, rather than finite ones, we are led to the following definition.

Definition 1.2.3
A family of sets $\mathcal{M}$ is an ***σ-algebra*** *if $\mathbb{R} \in \mathcal{M}$, if whenever $A_i \in \mathcal{M}$, for $i \in \mathbb{N}$, then*

$$\bigcup_{k=1}^{\infty} A_k \in \mathcal{M},$$

and if $A \in \mathcal{M}$, then $A^c \in \mathcal{M}$.

Two examples of σ-algebras are $\mathcal{P}(\mathbb{R})$, the power set of $\mathbb{R}$, and $\{\emptyset, \mathbb{R}\}$, the trivial σ-algebra. $\mathcal{E}$, however, is not a σ-algebra.

Since a σ-algebra is an algebra, Proposition 1.2.2 still holds. We can go a little further with intersections, however.

Proposition 1.2.3
Let $\mathcal{A}$ be a σ-algebra. If $A_k \in \mathcal{A}$ for $k \in \mathbb{N}$, then

$$\bigcap_{k=1}^{\infty} A_k$$

is in $\mathcal{A}$.

Proof:

Since DeMorgan's law holds for countable intersections and unions,

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

Since $\mathcal{A}$ is an algebra, $A_k^c \in \mathcal{A}$, and so we have expressed the intersection as the complement of a countable union of elements of $\mathcal{A}$. Therefore the set is an element of the σ-algebra. ■

At this point it is worthwhile noting that although the definition of a σ-algebra requires closure of the family under arbitrary countable unions, we only need closure under countable disjoint unions.

Lemma 1.2.4

Let $\mathcal{A}$ be an algebra which is closed under disjoint countable unions, ie. given any family of sets $A_k \in \mathcal{A}$ such that $A_k \cap A_l = \emptyset$ for $k \neq l$, the union of these sets is in the algebra. Then $\mathcal{A}$ is a σ-algebra.

Proof:

Strategy: *we are creating a disjoint union from an arbitrary union by letting B_k be the "new" points added by A_k to the union. We need to show that these sets are in the σ-algebra, and that their union is the same as the union of the A_k.*

Assume $A_k \in \mathcal{A}$ be a countable collection of sets in the algebra which are not necessarily disjoint. Given the union of A_1 through A_{k-1}, let B_k be the set of elements of A_k which are not in the union, ie.

$$B_k = A_k \setminus \bigcup_{i=1}^{k-1} A_i = A_k \cap \left(\bigcup_{i=1}^{k-1} A_i \right)^c .$$

So B_k lies in $\mathcal{A}$, since it is built from complements, finite intersections and finite unions of sets in the algebra $\mathcal{A}$. Also, by construction, the sets B_k and B_l are disjoint. We also note that because $X \cup Y = (X \setminus Y) \cup Y$, we have that

$$\bigcup_{l=1}^{k} B_k = \bigcup_{l=1}^{k} A_k.$$

Finally, given any point x in the union of all the A_k, we must have $x \in A_k$ for some k. Hence $x \in B_l$ for some $l \leq k$, and so

$$\bigcup_{l=1}^{\infty} A_k \subseteq \bigcup_{l=1}^{\infty} B_k.$$

Similarly, any point x in the union of all the B_k must be in at least one set B_k, and so $x \in A_k$, which means

$$\bigcup_{l=1}^{\infty} B_k \subseteq \bigcup_{l=1}^{\infty} A_k.$$

Therefore
$$\bigcup_{l=1}^{\infty} A_k = \bigcup_{l=1}^{\infty} B_k.$$
So we have written an arbitrary countable union as a countable disjoint union, and so the countable union must be an element of the algebra. Hence $\mathcal{A}$ is a σ-algebra. ■

This little lemma is very useful for proving that a family of sets is in fact a σ-algebra. The strategy will be to first show that the family is an algebra, then to show that it is closed under countable disjoint unions and then use the lemma to get the final result.

We can now begin to define our volume function. Given any $E \in \mathcal{E}$, we have $E = I_1 \cup I_2 \cup \ldots \cup I_n$ where the I_k are intervals which are pairwise disjoint. If the left and right endpoints a_k and b_k (respectively), we define the function $m : \mathcal{E} \to [0, \infty]$ by
$$m(E) = \sum_{k=1}^{n} b_k - a_k.$$
Clealy, if I is an interval with left and right endpoints a and b (respectively), we have $m(I) = b - a$ (ie. the length of the interval I), and so we could write
$$m(E) = \sum_{k=1}^{n} m(I_k)$$
instead.

There is one technical consideration that we need to worry about with this definition. A set $E \in \mathcal{E}$ can potentially be given by many different unions of disjoint intervals. It ought to be the case that the value of m does not depend upon the choice of intervals.

Lemma 1.2.5
The function $m : \mathcal{E} \to [0, \infty]$ is well-defined. That is, if $E \in \mathcal{E}$ with $E = I_1 \cup I_2 \cup \ldots \cup I_n = J_1 \cup J_2 \cup \ldots \cup J_p$, then
$$\sum_{k=1}^{n} m(I_k) = \sum_{l=1}^{p} m(J_k).$$

This fact is intuitively obvious, but a full proof is fairly technical and picky, and not particularly enlightening. For this reason, the proof is given in Section 1.11.

Our ultimate objective is to extend our function m to some σ-algebra containing $\mathcal{E}$. Again, we want to give some axiomatic definitions. We observe that this function has two key properties:

Proposition 1.2.6
If A and $B \in \mathcal{E}$ are two disjoint sets, then
$$m(A \cup B) = m(A) + m(B).$$

Also $m(\emptyset) = 0$.

Proof:
Let $A = I_1 \cup I_2 \cup \ldots \cup I_n$ be a disjoint union, and $B = J_1 \cup J_2 \cup \ldots \cup J_m$ be a disjoint union. Then

$$A \cup B = (I_1 \cup I_2 \cup \ldots \cup I_n) \cup (J_1 \cup J_2 \cup \ldots \cup J_m),$$

and since A and B are disjoint, I_k and J_l are disjoint for all k and l, so this is a disjoint union of intervals. Therefore

$$\begin{aligned} m(A \cup B) &= m(I_1) + m(I_2) + \cdots + m(I_n) + m(J_1) + m(J_2) + \cdots + m(J_m) \\ &= m(A) + m(B). \end{aligned}$$

We also have $\emptyset = (a, a)$, so $m(\emptyset) = a - a = 0$. ■

This leads us to the following definitions:

Definition 1.2.4
A function whose domain is an algebra is called a ***set function****. A set function* $\mu : \mathcal{A} \to [0, \infty]$ *is* ***additive*** *if given two disjoint sets* A, $B \in \mathcal{A}$,

$$\mu(A \cup B) = \mu(A) + \mu(B).$$

A set function on a σ*-algebra* $\mathcal{A}$ *is* σ***-additive*** *(or* ***countably additive****) if given* $A_i \in \mathcal{A}$, *for* $i \in \mathbb{N}$, *with the* A_i *disjoint,*

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$$

A σ*-additive set function for which* $\mu(\emptyset) = 0$ *is called a* ***measure****.*

Note: *A measure satisfies conditions 1, 2 and 3 from Section 1.1. The set function* m *is defined by condition 7. The remaining properties will be consequences of these assumptions.*

So from our previous discussion, $m : \mathcal{E} \to [0, \infty]$ is additive, and satisfies $m(\emptyset) = 0$. Clearly, what we want is a measure based on m, but we don't have a candidate σ-algebra yet, nor if we did is there a clear way to extend m.

Exercises

1.2.1. (†) If $\mathcal{A}$ is an algebra of sets, and A, $B \in \mathcal{A}$, show that $\emptyset \in A$, $A \cap B$, $A \setminus B \in \mathcal{A}$ and $A \triangle B \in \mathcal{A}$. If $\mathcal{A}$ is a σ-algebra, and $A_k \in \mathcal{A}$ for all $k \in \mathbb{N}$, show that

$$\bigcap_{k=1}^{\infty} A_k \in \mathcal{A}.$$

1.2.2. A subset of $\mathbb{R}$ is ***cofinite*** if its compliment is a finite set. Let $\mathcal{F}$ be the family of all subsets of $\mathbb{R}$ which are either countable or co-countable. Show that $\mathcal{F}$ is an algebra.

1.2.3. (†) A subset of $\mathbb{R}$ is ***co-countable*** if its compliment is a countable set. Let $\mathcal{C}$ be the family of all subsets of $\mathbb{R}$ which are either countable or co-countable. Show that $\mathcal{C}$ is a σ-algebra. (Hint: countable unions of countable sets are still countable.)

1.2.4. Let $\mathcal{C}$ be as in the previous exercise. Show that the set function $\mu : \mathcal{C} \to [0, \infty]$ defined by

$$\mu(X) = \begin{cases} 0 & \text{if } X \text{ countable} \\ \infty & \text{if } X \text{ co-countable} \end{cases}$$

is a measure.

1.2.5. (†) Show that $\mathcal{P}(\mathbb{R})$ is a σ-algebra. Show that the set function

$$c(X) = \begin{cases} |X| & \text{for } X \text{ finite,} \\ \infty & \text{for } X \text{ infinite} \end{cases}$$

is a measure on $\mathcal{P}(\mathbb{R})$.

1.2.6. An algebra $\mathcal{A}$ in $\mathbb{R}^n$ satisfies the same axioms as an algebra in $\mathbb{R}$, except that we insist that $\mathbb{R}^n \in \mathcal{A}$. A ***box*** in $\mathbb{R}^n$ is a Cartesian product of intervals $I_1 \times I_2 \times \ldots \times I_n$.

Show that the family $\mathcal{E}^n$ of all finite unions of boxes is an algebra.

Let m_N be the set function $m_n : \mathcal{E}^n \to [0, \infty]$ given by $m_n(I_1 \times I_2 \times \ldots \times I_n) = m(I_1)m(I_2) \ldots m(I_n)$, and $m_N(E) = m(B_1)+m(B_2)+\ldots+m(B_n)$, where E is a disjoint union of the boxes B_1, B_2, $\ldots B_n$. Show that m_n is additive on $\mathcal{E}$.

1.2.7. Let $\mathcal{A}$ and $\mathcal{B}$ be algebras. Prove that $\mathcal{A} \cap \mathcal{B}$ is also an algebra.

Let $\mathcal{A}$ and $\mathcal{B}$ be σ-algebras. Prove that $\mathcal{A} \cap \mathcal{B}$ is also a σ-algebra.

Let I be an aribrary index set, and let $\mathcal{A}_\alpha$ be a σ-algebra for all $\alpha \in I$. Prove that

$$\bigcap_{\alpha \in I} \mathcal{A}_\alpha$$

is also a σ-algebra.

1.2.8.

1.3 Properties of Measures

In Section 1.1 we gave a list of conditions which a reasonable measure of volume should satisfy. By definition, a measure satisfies the first three of these conditions, but we can use these three to prove most of the remaining conditions without any more theory.

We start with the fifth of these conditions, as it is the easiest to prove, and is used in the proof of the others:

Proposition 1.3.1
If $\mathcal{A}$ is an algebra, and μ is an additive set function on $\mathcal{A}$, then if A and $B \in \mathcal{A}$, and $A \subseteq B$,

$$\mu(A) \leq \mu(B).$$

Proof:
If $A \subseteq B$, then we can write

Strategy: *The basic strategy here, and in many proofs with measures, is to break the sets we are interested in into disjoint unions of other sets, so we can use additivity of the measure.*

$$B = A \cup (B \setminus A).$$

By Exercise 1.2.1, we know that $B \setminus A \in \mathcal{A}$ and A and $B \setminus A$ are disjoint. So

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

Since m is always at least 0, we have that $\mu(A \setminus B) \geq 0$, and so

$$\mu(B) \geq \mu(A).$$

■

The fourth condition follows easily from the previous result.

Proposition 1.3.2
If $\mathcal{A}$ is an algebra, and μ is an additive set function on $\mathcal{A}$, then if A and $B \in \mathcal{A}$, and $\mu(A \cap B) < \infty$,

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B).$$

Proof:
For simplicity, assume that $\mu(A)$, $\mu(B) < \infty$ (see Exercise 1.3.1 for the infinite case). Clearly, we can write $A \cup B$ as a disjoint union:

$$A \cup B = (A \setminus B) \cup (A \cap B) \cup (B \setminus A).$$

By Exercise 1.2.1, we know that $A \setminus B$ and $B \setminus A \in \mathcal{A}$, so that

$$\mu(A \cup B) = \mu(A \setminus B) + \mu(A \cap B) + (B \setminus A).$$

However, we also have that $A = (A \setminus B) \cup (A \cap B)$ and $B = (B \setminus A) \cup (A \cap B)$, and both unions are disjoint. Therefore

$$\mu(A) = \mu(A \setminus B) + \mu(A \cap B) \qquad \text{and} \qquad \mu(B) = \mu(B \setminus A) + \mu(A \cap B),$$

and therefore

$$\begin{aligned} \mu(A) + \mu(B) - \mu(A \cap B) &= \mu(A \setminus B) + \mu(A \cap B) + \\ &\quad \mu(B \setminus A) + \mu(A \cap B) - \mu(A \cap B) \\ &= \mu(A \setminus B) + \mu(A \cap B) + \mu(B \setminus A) \\ &= \mu(A \cup B), \end{aligned}$$

as required. ■

There is another condition which was not listed earlier, but which is useful on occasion. You can think of this as a continuity condition which tells us that the measure of the union of an increasing sequence of sets is equal to the limit of their measures. The proof uses the fact that we can break a union of sets into a disjoint union of sets in the same way that we did in Lemma 1.2.4.

Proposition 1.3.3
Let $\mathcal{A}$ be a σ-algebra, and μ a σ-additive set function on $\mathcal{A}$. If A_k, $k \in \mathbb{N}$, is a sequence of sets in $\mathcal{A}$ with $A_{k-1} \subseteq A_k$, then

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \lim_{k\to\infty} \mu(A_k).$$

Proof:
Let

$$B_n = A_n \setminus \left(\bigcup_{k=1}^{n-1} A_k\right).$$

Strategy: *this uses the same strategy as Lemma 1.2.4, of looking at sets of new points added by each A_k.*

Again, by Exercise 1.2.1, we have that $B_n \in \mathcal{A}$, and by following the strategy of the proof of Lemma 1.2.4,

$$\bigcup_{k=1}^{n} A_k = \bigcup_{k=1}^{n} B_k.$$

and

$$\bigcup_{k=1}^{\infty} A_k = \bigcup_{k=1}^{\infty} B_k.$$

Now since $A_k \subseteq A_n$ for all $k \leq n$,

$$A_n = \bigcup_{k=1}^{n} B_k.$$

Furthermore, the B_k are disjoint, so we have

$$\begin{aligned}\mu\left(\bigcup_{k=1}^{\infty} A_k\right) &= \sum_{k=1}^{\infty} \mu(B_k) \\ &= \lim_{n\to\infty} \sum_{k=1}^{n} \mu(B_k) \\ &= \lim_{n\to\infty} \mu(A_n),\end{aligned}$$

as required. ■

You will notice that these results in no way used the fact that we are considering any particular measure, or an algebra living in any particular set. This fact will be very important when we start talking about general measures. The remaining two conditions which we discussed, however, are very specific to modelling a measure on $\mathbb{R}^n$ which gives the volume of sets. We cannot prove (or even state) those in a general setting, and must instead construct a particular measure for which we will prove the results.

Exercises

1.3.1. Complete the proof of Proposition 1.3.2 for the cases where one or both of $\mu(A)$ and $\mu(B)$ are infinite.

1.3.2. Show that if $\mathcal{A}$ is a σ-algebra and μ is a measure on A, then $\mu(A \triangle B) = \mu(A \setminus B) + \mu(B \setminus A)$ for all A, $B \in \mathcal{A}$.

1.3.3. (†) Show that if $\mathcal{A}$ is a σ-algebra, μ is a measure on $\mathcal{A}$, and A_k, $k \in \mathbb{N}$, is a sequence of sets in $\mathcal{A}$ with $A_k \subseteq A_{k-1}$, and $\mu(A_k) < \infty$ eventually, then

$$\mu\left(\bigcap_{k=1}^{\infty} A_k\right) = \lim_{k\to\infty} \mu(A_k).$$

1.3.4. (†) Let m be a measure on a σ-algebra $\mathcal{A}$. A ***null set*** is a set $A \in \mathcal{A}$ for which $m(A) = 0$. Show that if $A = \bigcup_{i=1}^{\infty} A_i$ and A_i are all null then so is A.

1.4 Outer Measure

Although we can't extend m to a σ-additive measure on $\mathcal{P}(\mathbb{R})$, we can do something a little weaker, but nearly as useful. Our strategy is to extend m to something which is *sub*-additive on $\mathcal{P}(\mathbb{R})$, and then try to identify sets where it is actually σ-additive.

Definition 1.4.1
*A set function $\mu^* : \mathcal{P}(\mathbb{R}) \to [0, \infty]$ is an **outer measure** if $\mu^*(\emptyset) = 0$, $\mu^*(A) \leq \mu^*(B)$ for all $A \subseteq B$, and*

$$\mu^*\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mu^*(A_i).$$

In other words, if we can't get σ-additivity, we can try to get at least *sub*-additivity.

Definition 1.4.2
For any $A \subset \mathbb{R}$, let

$$m^*(A) = \inf\left\{\sum_{k=1}^{\infty} m(A_k) : A_k \in \mathcal{E}, A \subseteq \bigcup_{k=1}^{\infty} A_k\right\}.$$

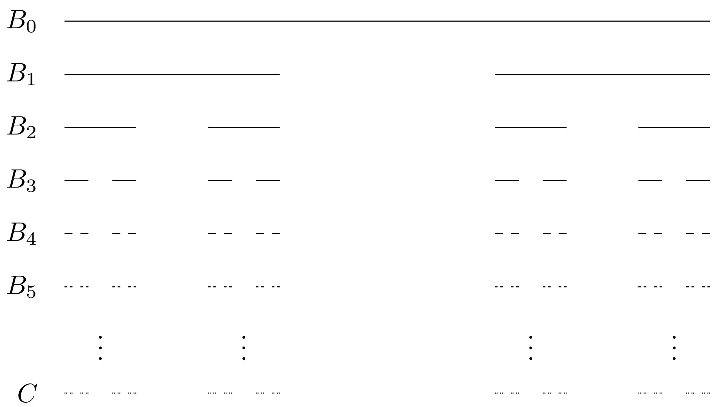

Figure 1.3: Constructing the Cantor Set

This set function is called the ***Lebesgue outer measure****.*

In other words, to calculate the Lebesgue outer measure of a set A, we cover A by elements of $\mathcal{E}$ and take the infimum of the total volumes of these covers. Intuitively, we are approximating the set A by countable unions of elements of $\mathcal{E}$ and use these to find an upper bound for the volume of A.

We note immediately, that since each A_k is a finite union of intervals, we have

$$m^*(A) = \inf \left\{ \sum_{k=1}^{\infty} m(I_k) : I_k \in \mathcal{I}, A \subseteq \bigcup_{k=1}^{\infty} I_k \right\}.$$

Indeed, since the intersection of two intervals is again an interval, the intervals I_k can be chosen to be disjoint. The reason for the more general definition, is that it is the way we will define outer measures in more general settings.

Example 1.4.1

The following is an example of how you can calculate m^* of a non-trivial set from first principles.

The Cantor set C is the set obtained by taking the interval $[0, 1]$, removing the open middle third of the interval, ie. the set $(1/3, 2/3)$, to obtain $[0, 1/3] \cup [2/3, 1]$, removing the open middle third of of each of these two intervals to get

$$[0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1],$$

and so on and so forth. The Cantor set C is the set obtained in the limit.

More formally, let $B_0 = [0, 1]$, $B_1 = [0, 1/3] \cup [2/3, 1]$, $B_3 = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1]$, and so on. Then

$$C = \bigcap_{k=1}^{\infty} B_k.$$

Another way of thinking of this set is by recognising that every number in the interval $[0,1]$ can be represented by a sum $\sum_{i=1}^{\infty} a_i 3^{-i}$ where $a_i \in \{0,1,2\}$. C is the set of numbers which can be represented by a sequence where $a_i \neq 1$ for all i. The equivalence between these two ways of thinking of C follows because the first middle third, $(1/3, 2/3)$, is the set of all numbers for which $a_1 = 1$; the next middle thirds, $(1/9, 2/9)$ and $(7/9, 8/9)$ are the remaining numbers for which $a_2 = 1$; and so on.

Note: *This is representing number in $[0,1]$ as a base 3 "decimal".*

The Cantor set is a good starting point for certain types of counter examples. It can be shown that it is a closed and bounded (ie. compact) subset of $\mathbb{R}$ whose interior is empty. In other words, it is a set which is "all boundary."

Note: *We'll examine these topological properties more closely in Chapter 2*

The Cantor set is uncountable, since we can define a function $\psi : C \to [0,1]$ by taking a point $x = \sum_{i=1}^{\infty} a_i 3^{-i}$ to

$$\psi(x) = \sum_{i=1}^{\infty} \frac{a_i}{2} 2^{-i}.$$

We claim the function ψ is a surjection. Since every number y in $[0,1]$ can be represented as a sum of the form $y = \sum_{i=1}^{\infty} b_i 2^{-i}$, where $b_i \in \{0,1\}$, we note that the point $x = \sum_{i=1}^{\infty} (2b_i) 3^{-i}$ lies in the cantor set, and that $\psi(x) = y$. Since ψ is a surjection onto an uncountable set, C is uncountable.

We will now calculate $m^*(C)$. Notice that $C \subset B_k$ for all k, and that B_k is a union of intervals, and hence a cover of C of the form we wish to consider. More concretely,

$$\begin{aligned} m^*(C) &= \inf \left\{ \sum_{k=1}^{\infty} m(I_k) : I_k \in \mathcal{I}, C \subseteq \bigcup_{k=1}^{\infty} I_k \right\} \\ &\leq \inf \{ m(B_k) : k \in \mathbb{N} \} \end{aligned}$$

Now the set B_k is a disjoint union of 2^k intervals, each of length 3^{-k}, so $m(B_k) = (2/3)^k$. Since $m^*(C)$ is an infimum taken over a set of non-negative numbers,

$$0 \leq m^*(C) \leq \inf \{ m(B_k) : k \in \mathbb{N} \} = 0.$$

Hence $m^*(C) = 0$. ◇

We now justify calling m^* the Lebesgue outer measure.

Proposition 1.4.1
The function m^* is an outer measure.

Proof:
First note that $\emptyset \in \mathcal{E}$, so $0 \leq m^*(\emptyset) \leq m(\emptyset) = 0$.

Strategy: *We use two standard "tricks" of proofs in analysis:*
1. if for each n and any $\varepsilon_n > 0$ we can find $x_n \leq y_n + \varepsilon_n$, then by choosing $\varepsilon_n = 2^{-n}\varepsilon$ we get $\sum x_n \leq \sum y_n + \varepsilon$.
2. if $x \leq y + \varepsilon$ for all $\varepsilon > 0$, then $x \leq y$.

Next observe that if $A \subseteq B$, then whenever $B \subseteq \bigcup_{k=1}^{\infty} B_k$ we automatically have $A \subseteq \bigcup_{k=1}^{\infty} B_k$, and hence

$$\left\{ \sum_{k=1}^{\infty} m(A_k) : A_k \in \mathcal{E}, A \subseteq \bigcup_{k=1}^{\infty} A_k \right\} \supseteq \left\{ \sum_{k=1}^{\infty} m(B_k) : B_k \in \mathcal{E}, B \subseteq \bigcup_{k=1}^{\infty} B_k \right\}.$$

Therefore

$$\begin{aligned}
m^*(A) &= \inf\left\{\sum_{k=1}^{\infty} m(A_k) : A_k \in \mathcal{E}, A \subseteq \bigcup_{k=1}^{\infty} A_k\right\} \\
&\leq \inf\left\{\sum_{k=1}^{\infty} m(B_k) : B_k \in \mathcal{E}, B \subseteq \bigcup_{k=1}^{\infty} B_k\right\} = m^*(B).
\end{aligned}$$

Let $\{A_k\}_{k=1}^{\infty} \subseteq \mathcal{P}(\mathbb{R})$. From the definition of m^* and basic facts about infima, we have that for any $\varepsilon > 0$ we can find $B_{k,j} \in \mathcal{E}$ (with the choice of sets depending on choice of ε) such that

$$A_k \subseteq \bigcup_{j=1}^{\infty} B_{k,j}$$

and $\sum_{j=1}^{\infty} m(B_{k,j}) \leq m^*(A_k) + 2^{-k}\varepsilon$. Now

$$\bigcup_{k=1}^{\infty} A_k \subseteq \bigcup_{k,j=1}^{\infty} B_{k,j}$$

which means that

$$\begin{aligned}
m^*\left(\bigcup_{k=1}^{\infty} A_k\right) &\leq \sum_{k,j=1}^{\infty} m(B_{k,j}) \\
&\leq \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} m(B_{k,j}) \\
&\leq \sum_{k=1}^{\infty} \big(m^*(A_k) + 2^{-k}\varepsilon\big) \\
&= \left(\sum_{k=1}^{\infty} m^*(A_k)\right) + \varepsilon.
\end{aligned}$$

Note: *The order of summation can safely be changed since all the term are non-negative and so the series converges absolutely or diverges to $+\infty$.*

But since we can do this for any $\varepsilon > 0$ (albeit with possibly different $B_{k,j}$), we have

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

as required. ■

For elementary sets $E \in \mathcal{E}$, $m^*(E) = m(E)$. This relies on the intuitively plausible, but somewhat technical proposition that if you partition an interval into a countable collection of subintervals, the length of the interval is the sum of the lengths of the subintervals.

Lemma 1.4.2
If $I \in \mathcal{I}$ and $I = \bigcup_{k=1}^{\infty} I_k$ where the I_k are pairwise disjoint, then

$$m(I) = \sum_{k=1}^{\infty} m(I_k).$$

This lemma is formally proved in Section 1.11.

Example 1.4.2
The outer measure $m^*(I)$ of an interval I is $m(I)$.

Clearly $m^*(I) \leq m(I)$, since I is an elementary set.

On the other hand, if $I \subseteq \bigcup_{k=1}^{\infty} I_k$, where the sets I_k are intervals, we let $E_1 = I_1 \cap I$, and

$$E_k = \left(I_k \setminus \bigcup_{j=1}^{k=1} I_j \right) \cap I.$$

Therefore $m(E_k) \leq m(I_k)$, E_k are pairwise disjoint, and

$$\bigcup_{k=1}^{\infty} E_k = \left(\bigcup_{k=1}^{\infty} I_k \right) \cap I = I$$

Furthermore, each set E_k can be written as a finite union of intervals $J_{k,l}$, where these intervals are pairwise disjoint. So

$$\sum_{k=1}^{\infty} m(I_k) \geq \sum_{k=1}^{\infty} m(E_k) \geq \sum_{k=1}^{\infty} \sum_{l=1}^{n_k} m(J_{k,l})$$

and

$$\bigcup_{k=1}^{\infty} \bigcup_{j=1}^{n_k} J_k = I.$$

So by Lemma **??**, we have

$$\sum_{k=1}^{\infty} m(I_k) \geq \sum_{k=1}^{\infty} \sum_{l=1}^{n_k} m(J_{k,l}) = m(I).$$

and we are done. $\diamond$

The fact that $m^*(E) = m(E)$ follows from this example and Carathéodory's Theorem (Theorem 1.4.3).

We now want to identify the family of sets for which this outer measure behaves like a measure, with the aim that restricting the outer measure to these "good" sets will make it a measure. In particular, we want to rule out sets like the set N discussed earlier.

Definition 1.4.3
*If μ^* is an outer measure, a set A is **μ^*-measurable** if*

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c).$$

for every subset E of $\mathbb{R}$.

From the definition of outer measure we automatically have

$$\mu^*(E) \leq \mu^*(E \cap A) + \mu^*(E \cap A^c)$$

so we only need check that

$$\mu^*(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c) \tag{1.2}$$

to prove a set is μ^*-measurable.

Theorem 1.4.3 (Carathéodory)
Let μ^* be an outer measure. Then the family $\mathcal{M}$ of μ^*-measurable sets is a σ-algebra, and the restriction of μ^* to $\mathcal{M}$ is a measure.

Proof:
First we will show that $\mathcal{M}$ is an algebra. Immediately we notice that if A is μ^*-measurable then so is A^c, since the definition of μ^*-measurability is symmetric. Now if A and B are in $\mathcal{M}$, for any E we have

Strategy: *we want to use Lemma 1.2.4 to prove that $\mathcal{M}$ is a σ-algebra, so we need to show that it is an algebra by checking the axioms, and we need to show that it is closed under disjoint unions.*

$$\begin{aligned}\mu^*(E) &= \mu^*(E \cap A) + \mu^*(E \cap A^c) \\ &= \mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c) \\ &\quad + \mu^*(E \cap A^c \cap B) + \mu^*(E \cap A^c \cap B^c).\end{aligned}$$

Now $(E \cap A \cap B) \cup (E \cap A \cap B^c) \cup (E \cap A^c \cap B) = E \cap (A \cup B)$, so

$$\mu^*(E \cap (A \cup B)) \leq \mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c) + \mu^*(E \cap A^c \cap B)$$

and therefore

$$\mu^*(E) \geq \mu^*(E \cap (A \cup B)) + \mu^*(E \cap (A \cup B)^c).$$

and hence $A \cup B \in \mathcal{M}$. Hence $\mathcal{M}$ is an algebra.

Now assume that we have a countable collection A_k of disjoint sets in $\mathcal{M}$. Let $B_n = \bigcup_{k=1}^n A_k$ and $B = \bigcup_{k=1}^\infty A_k$. Given any set E we have

$$\begin{aligned}\mu^*(E \cap B_n) &= \mu^*(E \cap B_n \cap A_n) + \mu^*(E \cap B_n \cap A_n^c) \\ &= \mu^*(E \cap A_n) + \mu^*(E \cap B_{n-1})\end{aligned}$$

and so by induction

$$\mu^*(E \cap B_n) = \sum_{k=1}^n \mu^*(E \cap A_k).$$

Therefore

$$\begin{aligned}\mu^*(E) &= \mu^*(E\cap B_n) + \mu^*(E\cap B_n^c)\\ &= \sum_{k=1}^{n}\mu^*(E\cap A_k) + \mu^*(E\cap B_n^c)\\ &\geq \sum_{k=1}^{n}\mu^*(E\cap A_k) + \mu^*(E\cap B^c)\end{aligned}$$

since $B^c \subseteq B_n^c$, and taking the limit as $n\to\infty$, we have

$$\begin{aligned}\mu^*(E) &\geq \sum_{k=1}^{\infty}\mu^*(E\cap A_k) + \mu^*(E\cap B^c)\\ &\geq \mu^*\left(\bigcup_{k=1}^{\infty}(E\cap A_k)\right) + \mu^*(E\cap B^c)\\ &= \mu^*(E\cap B) + \mu^*(E\cap B^c)\end{aligned}$$

so B is μ^*-measurable.

So $\mathcal{M}$ is an algebra where countable unions of disjoint sets in $\mathcal{M}$ are in $\mathcal{M}$. Therefore $\mathcal{M}$ is a σ algebra by Lemma 1.2.4.

Note that this last set of inequalities means that in the special case that $E = B$,

$$\mu^*(B) \geq \sum_{k=1}^{\infty}\mu^*(A_k) \geq \mu^*(B),$$

and hence

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

So μ^* is in fact a measure when restricted to $\mathcal{M}$. ■

We are almost done at this point, since we now have a measure. We just need to check that everything is well-behaved for our original sets $\mathcal{E}$.

Proposition 1.4.4
Every elementary set is m^*-measurable.

Proof:
Let A be any elementary set, and E any set, then for any $\varepsilon > 0$ we can find a family of sets $B_i \in \mathcal{E}$ such that $E \subseteq \bigcap_{i=1}^{\infty} B_i$ and $\sum_{i=1}^{\infty} m(B_i) \leq m^*(E) + \varepsilon$.

Hence

$$\begin{aligned} m^*(E) + \varepsilon &\geq \sum_{i=1}^{\infty} m(B_i) \\ &= \sum_{i=1}^{\infty} (m(B_i \cap A) + m(B_i \cap A^c)) \\ &= \sum_{i=1}^{\infty} m(B_i \cap A) + \sum_{i=1}^{\infty} m(B_i \cap A^c) \\ &\geq m^*(E \cap A) + m^*(E \cap A^c) \end{aligned}$$

since $E \cap A = \bigcup_{i=1}^{\infty}(B_i \cap A)$ and $E \cap A^c = \bigcup_{i=1}^{\infty}(B_i \cap A^c)$. So A satisfies Equation 1.2, hence A is m^*-measurable. ■

The following fact is also noteworthy:

Proposition 1.4.5
If $A \subset B$ and $B \in \mathcal{M}$ with $\mu^*(B) = 0$, then $A \in \mathcal{M}$ with $\mu^*(A) = 0$.

Proof:
We first note that if $\mu^*(B) = 0$, then the outer measure of any subset must be 0 as well, so we need only check that A is μ^*-measurable. For any set E, we have $\mu^*(E \cap A) \leq \mu^*(A) = 0$, so

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

and hence A is μ^*-measurable. ■

Measures with this property are called ***complete***.

Example 1.4.3
Every subset X of the Cantor set C is m^*-measureable and has $m^*(X) = 0$. ◇

Exercises

1.4.1. Show that the Lebesgue outer measure of $\mathbb{Q}$ is 0.

1.4.2. (†) Let N be the set described in Example 1.1.2. Show that the Lebesgue outer measure of N is not 0. Show that the Lebesgue outer measure of $[0, 1) \setminus N$ is 1. Conclude that N is not m^*-measureable.

1.4.3. (†) Show that any countable set has Lebesgue outer measure 0.

1.4.4. Show that any measure μ on $\mathcal{P}(\mathbb{R})$ is also an outer measure.

1.4.5. (†) Let f be a monotone increasing function on $\mathbb{R}$. If I is an interval with endpoints a and b, define

$$m_f(I) = f(b) - f(a),$$

and extend m_f to $\mathcal{E}$ in the obvious way. Now define

$$m_f^*(A) = \inf\left\{\sum_{k=1}^{\infty} m_f(A_k) : A_k \in \mathcal{E}, A \subseteq \bigcup_{k=1}^{\infty} A_k\right\}$$

for any $A \subseteq \mathbb{R}$.

Show that m_f^* is an outer measure.

1.4.6. Refer to Exercise 1.2.6. When working with concepts from this section in $\mathbb{R}^n$, replace references to $\mathbb{R}$ by $\mathbb{R}^n$ throughout. Define, for any set $A \subseteq \mathbb{R}^n$,

$$m_n^*(A) = \inf\left\{\sum_{k=1}^{\infty} m_n(A_k) : A_k \in \mathcal{E}^n, A \subseteq \bigcup_{k=1}^{\infty} A_k\right\}.$$

Show that m_n^* is an outer measure. Show that every set in $\mathcal{E}^n$ is m_n^*-measurable. Show that m_n^* is complete.

1.5 Lebesgue Measure

We have found a measure which is defined on a σ-algebra which includes $\mathcal{E}$ and, as we observed in the previous section, is equal to m on $\mathcal{E}$. Since this measure extends m we will also denote it by m. We call this measure the ***Lebesgue measure*** on $\mathbb{R}$, and the m^*-measurable sets ***Lebesgue measurable*** sets.

Proposition 1.5.1
If A is open then A is Lebesgue measurable. If A is closed then A is Lebesgue measurable.

If A is open and not empty, then $m(A) > 0$.

Proof:

By Proposition 1.11.2, every open set is a countable disjoint union of open intervals. All intervals are in $\mathcal{E}$ and hence $\mathcal{M}$ and since $\mathcal{M}$ is a σ-algebra, $A \in \mathcal{M}$.

If A is closed then A^c is open and so $A^c \in \mathcal{M}$. Therefore $A \in \mathcal{M}$.

Any open set A which is not empty contains a non-trivial open interval (a, b). Hence $0 < b - a = m((a, b)) \leq m(A)$. ■

As discussed earlier, we really want this measure to be invariant under length-preserving transformations. Additionally we would like the measure of a set to scale under compression and expansion. Fortunately, this is the case. For any set $X \subset \mathbb{R}$, and number $r \in \mathbb{R}$, define $X + r = \{x + r : x \in X\}$ (translation by r units) and $rX = \{rx : x \in X\}$ (expansion or contraction, and reflection by a factor r centered at the origin).

Theorem 1.5.2
If A is Lebesgue measurable, so are $A+r$ and rA for any $r \in \mathbb{R}$ and, moreover, $m(A+r) = m(A)$ and $m(rA) = |r|m(A)$.

Proof:
First note that for any interval $m(r + I) = m(I)$, since the length of an interval is invariant under translation. We can also easily verify that $m(rI) = |r|m(I)$. Given $A \in \mathcal{E}$ we clearly have $m(A + r) = m(A)$ and $m(rA) = |r|m(A)$ by applying the result for intervals to the intervals which make up A.

Now for any $A \subseteq \mathbb{R}$, $m^*(A + r) = m^*(A)$, since $A \subseteq \bigcup_{i=1}^\infty B_i$ iff $A + r \subseteq \bigcup_{i=1}^\infty (B_i + r)$, and so

$$\begin{aligned} m^*(A) &= \inf\{\sum_{i=1}^\infty m(B_i) : A \subseteq \bigcup_{i=1}^\infty B_i\} \\ &= \inf\{\sum_{i=1}^\infty m(B_i + r) : A + r \subseteq \bigcup_{i=1}^\infty (B_i + r)\} \\ &= m^*(A + r). \end{aligned}$$

A similar argument shows that $m^*(rA) = |r|m^*(A)$.

Now for any Lebesgue measurable set A, and any other set E,

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

Hence $A + r$ is Lebesgue measurable. A similar argument shows that rA is Lebesgue measurable.

Finally we simply need to note that of A is Lebesgue measurable, $m(A+r) = m^*(A + r) = m^*(A) = m(A)$ and $m(rA) = m^*(rA) = |r|m^*(A) = |r|m(A)$. ■

It is not hard to see that any countable set has Lebesgue measure 0, but it turns out that there are *uncountable* sets which have Lebesgue measure 0. The most famous of these is the Cantor set.

Example 1.5.1 (The Cantor Set)
In the previous section we showed from first principles that the Cantor set has $m^*(C) = 0$, and hence is m^*-measureable, and has $m(C) = 0$.

However, using the tools from this section we can show that $m(C) = 0$ without needing to use outer measures at all. Recall that the sets B_k are elementary sets with

$$m(B_k) = (2/3)^k.$$

By Exercise 1.3.3, we have

$$m(C) = m\left(\bigcap_{k=1}^\infty B_k\right) = \lim_{k\to\infty} m(B_k) = \lim_{k\to\infty} (2/3)^k = 0.$$

So the Cantor set is an uncountable set whose Lebesgue measure is 0. ◇

One can modify the construction of the Cantor set in such a way that the resulting set has any measure in the range $[0, 1)$. See the exercises for this chapter.

Exercises

Several of the following exercises refer to so-called "thick" Cantor sets, which are a generalization of Example 1.4.1. When we need to distinguish between a general Cantor set and the particular set of Example 1.4.1, we will refer to the latter as the Cantor middle third set.

Example 1.5.2 ("Thick" Cantor Sets)
Let ξ_n be a sequence of positive number such that $\xi_0 = 1$ and $\xi_n > 2\xi_{n+1}$ for all n. Remove from the interval $[0, 1]$ the open middle interval of length $\xi_0 - 2\xi_1$ to get the set $B_1 = [0, \xi_1] \cup [1 - \xi_1, 1]$. As in the construction of the Cantor set (Example 1.5.1), remove the open middle interval of length $\xi_1 - 2\xi_2$ from each of the subintervals of B_1 to get a set B_2. Repeat this construction inductively to get a sequence of sets B_n. Let $B = \bigcap_n B_n$.

These sets are compact and have empty interior, just like the usual Cantor set. ◇

1.5.1. (†) Let B be as in Example 1.5.2. Show that $m(B) = \lim_{n\to\infty} 2^n \xi_n$.

Show that for any θ with $0 < \theta < 1$, the sequence

$$\xi_n = \left(\frac{1-\theta}{2}\right)^n$$

gives $m(B) = 0$.

Show that the sequence $\xi_n = \alpha^{n/n+1} 2^{-n}$ for $0 < \alpha < 1$ gives $m(B) = \alpha$.

Note: *This shows that no thick Cantor set from the previous example can have measure 1.*

1.5.2. (†) Show that there is no nowhere dense set $A \subseteq [0, 1]$ such that $m(A) = 1$. (Recall: A is nowhere dense if the closure of A has empty interior, ie. $\overline{A}^o = \emptyset$).

Hint: use Proposition 1.5.1.

1.5.3. (†) Let q_n be an enumeration of the rational numbers. Consider intervals $A_n = (q_n - 2^{-n}, q_n + 2^{-n})$, and let $A = \bigcup_{i=1}^{\infty} A_n$. Show that A is an open, dense set with $m(A) \leq 2$.

Hint: *refine the previous example so the sets A_n are always disjoint, and so the lengths sum to the value you need; or use the fact that the complements of open, dense sets are nowhere dense sets, together with 1.5.1.*

1.5.4. (†) Show that you can construct an open, dense subset A of $\mathbb{R}$ whose measure $m(A)$ is any value greater than 0.

1.5.5. Let N be the set described in Example 1.1.2 and let C be the Cantor set of Example 1.5.1. Show that $m(N \cap C) = 0$.

1.5.6. Let N be the set described in Example 1.1.2 and let C be the Cantor set of Example 1.5.1. Let $x \in N$ be written as

$$x = \sum_{k=1}^{\infty} a_k 2^{-k},$$

where $a_k \in \{0, 1\}$, and a_k is not eventually 1 (ie. represent x in base 2), and let $\varphi : N \to C$ be given by

$$\varphi(x) = \sum_{k=1}^{\infty} (2a_k) 3^{-k}.$$

Show that $m(\varphi(N)) = 0$.

1.5.7. Let E be any Lebesgue measurable set. Given any $\alpha \in [0, m(E)]$ show that there is a Lebesgue measureable set F with $m(F) = \alpha$.

1.5.8. Refer to Exercises 1.2.6 and 1.4.6. Let m_n be the restriction of $\mathbb{R}^n$ to the m_n^*-measurable sets, and call this Lebesgure measure on $\mathbb{R}^n$.

Show that all open and closed sets are m_n measurable. Show that if A is m_n^*-measurable, then $m_n(x + A) = m_n(A)$ for all $x \in \mathbb{R}^n$ and $m_n(cA) = |x|^n m_n(A)$ for all $c \in \mathbb{R}$.

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. Show that for each T there is some number $D(T)$ so that $m_n(T(A)) = D(T) m_n(A)$. In particular, show that an orthogonal (ie. distance-preserving) transformation leaves the measure invariant.

Show that in fact the number $D(T) = |\det T|$, where $\det T$ is the determinant of a matrix which represents the linear transformation with respect to some basis of $\mathbb{R}^n$.

(These last two parts are quite hard.)

1.6 Measurable Functions

Now that we have a notion of the "volume" of a subset of the real numbers we want to apply this to define an integration theory based on this. One consequence of Example 1.1.2 is that not every function can be integrated. In particular, the characteristic function χ_N of the set described in that example is not a good candidate for being integrated. What we want are a class of functions that behave nicely on measurable sets.

Definition 1.6.1
Let X be any subset of $\mathbb{R}$. A function $f : X \to \mathbb{R}$ is ***(Lebesgue) measurable*** *on X if given any interval $I \subseteq \mathbb{R}$, the set $f^{-1}(I)$ is Lebesgue measurable.*

If f is measurable on $\mathbb{R}$, we will simply say that it is measurable.

We will denote the set of all measurable functions on X by $\mathcal{L}(X)$.

Note: *this definition is very similar to the definition or theorem which says that a continuous function has the property that the inverse images of open sets are open.*

Clearly if f is measurable on X, then X must be a measurable set, for $f^{-1}(\mathbb{R}) = X$. Similarly, it is not hard to see that if $X \subseteq Y$ with X a measureable set, and f is measurable on Y, $f|_X$ is measurable on X. For this reason, we will be slightly sloppy and refer to f being measurable on X, even when it is defined on a larger set.

Example 1.6.1

Let X be any measurable subset of $\mathbb{R}$. A characteristic function $\chi_A : X \to \mathbb{R}$ of a set $A \in \mathbb{R}$ is measurable on X if and only if $A \cap X$ is Lebesgue measurable.

Since the range of the characteristic function is 0 and 1, we need only ask whether or not these two points lie in the interval I to calculate $f^{-1}(I)$. If neither lie in I, $f^{-1}(I) = \emptyset$, which is Lebesgue measurable. If both lie in I, $f^{-1}(I) = X$, which is also Lebesgue measurable. If $1 \in I$, but 0 is not in I, then $f^{-1}(I) = X \cap A$, which is measurable. Finally, if $0 \in I$, but $0 \notin I$, then $f^{-1}(I) = X \cap A^c$, which is Lebesgue measurable, since $X \cap A$ and X are Lebesgue measurable. ◇

If f is measurable on X, the product $f\chi_X$ is measurable on any measurable set Y containing X. In particular this means that if $f : X \to \mathbb{R}$, we can extend it to a measurable function on all of $\mathbb{R}$ by simply setting $f(x) = 0$ for all $x \notin X$.

A ***simple function*** is a function of the form

$$\varphi(x) = \sum_{k=1}^{n} c_k \chi_{E_k}(x)$$

for some numbers c_k and sets E_k. Simple functions are precisely the functions whose range is a finite set. It is worthwhile noting that we can always write φ so that the sets E_k are the sets $\varphi^{-1}(c_k)$, for c_k in the range of φ, and we will say that this way of writing φ is the ***standard representation***.

An argument similar to Example 1.6.1 shows that a simple function φ is measurable on X if and only if the sets $E_k \cap X$ of the standard representation are Lebesgue measurable.

There are a number of equivalent ways of verifying that a more sophisticated function is measurable. In particular, we need not check the inverse image of every interval.

Proposition 1.6.1

Let $f : X \to \mathbb{R}$. Then the following are equivalent:

(i) $f \in \mathcal{L}(X)$.

(ii) $f^{-1}((a, \infty))$ *is measurable for all* $a \in [-\infty, \infty)$

(iii) $f^{-1}([a, \infty))$ *is measurable for all* $a \in (-\infty, \infty)$

(iv) $f^{-1}((-\infty, b))$ *is measurable for all* $b \in (-\infty, \infty]$

(v) $f^{-1}((-\infty, b])$ *is measurable for all* $b \in (-\infty, \infty)$

(vi) $f^{-1}((a,b))$ *is measurable for all* a, $b \in [-\infty,\infty]$

(vii) $f^{-1}([a,b])$ *is measurable for all* a, $b \in (-\infty,\infty)$.

(viii) $f^{-1}(U)$ *is measurable for every open subset* U *of* $\mathbb{R}$.

(ix) $f^{-1}(F)$ *is measurable for every closed subset* F *of* $\mathbb{R}$.

The proof of each of these cases is very similar, and in each case revolves around the fact that you can construct arbitrary intervals out of unions, intersections and complements of these sets, and the inverse images of these unions, intersections and complements must therefore be measurable. We will prove some of these cases and leave the rest as an exercise.

Proof:

Clearly (i) ⇒ (ii), (iii), (iv), (v), (vi) and (vii).

We also have (viii) ⇒ (ii), (iv) and (vi) since the intervals in these cases are all open sets, while (ix) ⇒ (iii), (v) and (vii) since the intervals in these cases are closed sets.

(ii) ⇒ (iv): Given any b, we note that

$$\bigcup_{k=1}^{\infty}(b-1/k,\infty)^c = \bigcup_{k=1}^{\infty}(-\infty, b-1/k] = (-\infty, b),$$

and so

$$f^{-1}((-\infty,b)) = f^{-1}\left(\bigcup_{k=1}^{\infty}(b-1/k,\infty)^c\right) = \bigcup_{k=1}^{\infty} f^{-1}((b-1/k,\infty))^c,$$

which is a countable union of Lebesgue measurable sets; and hence is a Lebesgue measurable set.

(ii) ⇒ (vi): Given any a and b,

$$(a,b) = (-\infty,b)\cap(a,\infty),$$

so

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

and we know that the second set is Lebesgue measurable by (ii), and the first by the fact that (ii) ⇒ (vi). Therefore $f^{-1}((a,b))$ is Lebesgue measurable by Exercise 1.2.1.

(vi) ⇒ (i): Given any interval I, with left and right endpoints a and b, we are done if I is open. If $I = [a,b]$, we have

$$[a,b] = \bigcap_{k=1}^{\infty}(a-1/k, b+1/k),$$

and so

$$f^{-1}([a,b]) = \bigcap_{k=1}^{\infty} f^{-1}((a-1/k, b+1/k))$$

which is Lebesgue measurable by Exercise 1.3.3. If $I = (a, b]$, then we can write $I = [a, b] \cap (a, b+1)$, and similarly if $I = [a, b)$ we can write $I = [a, b] \cap (a-1, b)$, and taking inverse images, we can easily see that $f^{-1}(I)$ is Lebesgue measurable in both cases.

(vi) ⇒ (viii): Every open set U in $\mathbb{R}$ is a countable union of bounded open intervals by Corollary 1.11.3. In other words we have a_k, $b_k \in \mathbb{R}$ such that

$$U = \bigcup_{k=1}^{\infty} (a_k, b_k).$$

But then

$$f^{-1}(U) = \bigcup_{k=1}^{\infty} f^{-1}((a_k, b_k)),$$

and so it is a countable union of Lebesgue measurable sets, and hence is Lebesgue measureable.

(viii) ⇒ (ix): If F is closed, then F^c is open, and so $f^{-1}(F^c)$ is Lebesgue measurable. But then

$$f^{-1}(F) = f^{-1}((F^c)^c) = f^{-1}(F^c)^c$$

is measurable, since complements of measurable sets are measurable. ■

Using these equivalent ways of describing measurability, we can easily get important examples of measurable functions.

Example 1.6.2

Recall that a function $f : X \to \mathbb{R}$ is continuous if given any open interval U, $f^{-1}(U)$ is relatively open (ie. there is some open set $V \subseteq \mathbb{R}$ with $f^{-1}(U) = V \cap X$).

Every continuous function is therefore measurable, since by Proposition 1.6.1 we only need check the inverse images of open intervals to determine measurability, and the inverse images of open intervals are relatively open sets, which are measurable by Proposition 1.5.1 and Exercise 1.2.1, since $f^{-1}(U) = V \cap X$ where V is open and X is measurable. ◇

In fact the set of measurable functions is fairly nice from an algebraic and analytic point of view:

Theorem 1.6.2

If f and $g \in \mathcal{L}(X)$, h is a continuous function on $f(X)$, and c is any constant, then:

(i) cf is measurable;

(ii) $f + g$ is measurable;

(iii) fg is measurable;

(iv) $h \circ f$ is measurable.

Proof:

(i) If $c = 0$, the result is trivial because the zero function is continuous. If $c \neq 0$ we have that $(cf)^{-1}(I) = c^{-1}(f^{-1}(I))$, and a constant multiple of a measurable set is measurable by Theorem 1.5.2.

(ii) Let $\Psi : \mathbb{R}^2 \to \mathbb{R}$ be the function given by addition, $\Psi(x, y) = x + y$, and $u : \mathbb{R} \to \mathbb{R}^2$ be given by $u(x) = (f(x), g(x))$, so that $f + g(x) = \Psi(f(x), g(x)) = \Psi(u(x))$.

Now for any interval of the form $I = (a, \infty)$, $(f + g)^{-1}(I) = u^{-1}(\Psi^{-1}(I))$, and $\Psi^{-1}(I) = \{(x, y) : x + y \geq a\}$. This region can be written as a countable union of open rectangles $R_k = I_k \times J_k$, where I_k and J_k are open intervals. For example, we could take a countable union of all rectangles of the form $(n2^{-j}, (n+1)2^{-j}) \times (m2^{-j}, (m+1)2^{-j})$ which are contained in $\Psi^{-1}(I)$ for some n, $m \in \mathbb{Z}$ and $j \in \mathbb{N}$.

We observe that $u^{-1}(R_k) = f^{-1}(I_k) \cap g^{-1}(J_k)$, which is Lebesgue measurable since f and g are Lebesgue measurable, and so

$$(f + g)^{-1}(I) = u^{-1}(\varphi^{-1}(I)) = u^{-1}\left(\bigcup_{k=1}^{\infty} R_k\right) = \bigcup_{k=1}^{\infty} f^{-1}(I_k) \cap g^{-1}(J_k)$$

which is Lebesgue measurable.

(iii) The proof of this is essentially the same as (ii), except that we use the multiplication function $\Phi(x, y) = xy$ in place of Ψ.

(iv) For any open interval I, $h^{-1}(I)$ is a relatively open set in X, because h is continuous on X. So there is some open set U with $h^{-1}(I) = U \cap f(X)$, and therefore $(h \circ f)^{-1}(I) = f^{-1}(U \cap f(X)) = f^{-1}(U)$ is measurable by Proposition 1.6.1, part (vii). ■

This result says that the collection of measurable functions is a vector space. In fact it is also an algebra. Also note that this means that $f - g$ is measurable if f and g are measurable.

Unfortunately we cannot guarantee that $f \circ g$ is measurable if f and g are measurable.

Example 1.6.3

Let X be the set $[0, 1)$. Let N be the unmeasurable set described Example 1.1.2 (with the additional restriction that the one rational number in N is not a diadic rational number) and let C be the Cantor set of Example 1.5.1.

Note: *a diadic rational number is a number of the form* $\frac{k}{2^n}$.

Let $x \in [0, 1)$ be written as

$$x = \sum_{k=1}^{\infty} a_k 2^{-k},$$

where $a_k \in \{0, 1\}$, and a_k is not eventually 1 (ie. represent x in base 2), and let $\varphi : [0, 1) \to C$ be given by

$$\varphi(x) = \sum_{k=1}^{\infty} (2a_k) 3^{-k}.$$

One can show that φ is measurable, since it is monotone increasing (see Exercise 1.6.3), and it is one-to-one for points in $[0,1)$ which are not diadic rationals.

Now $\varphi(N) \subseteq C$ as in Exercise 1.5.6, and since it has the cardinality of the continuum there is a bijection $\theta : \varphi(N) \to (0,\infty)$, so let

$$\psi(x) = \begin{cases} \theta(x) & x \in \varphi(N), \\ 0 & x \notin \varphi(N). \end{cases}$$

Again, this function ψ is measurable since $\psi^{-1}(I)$ is either contained in a set of measure 0, if $0 \notin I$, and so is measurable; or its complement is contained in a set of measure 0 if $0 \in I$, and so it is measurable.

However $\psi \circ \varphi$ is not measurable, since

$$(\psi \circ \varphi)^{-1}((0,\infty)) = \varphi^{-1}(\psi^{-1}((0,\infty))) = \varphi^{-1}(\varphi(N)) = N$$

is an unmeasurable set. ◇

The class of measurable functions behave well under limiting constructs as well.

Theorem 1.6.3
If f_k is a sequence of functions in $\mathcal{L}(X)$, then

$$g(x) = \sup_k f_k(x) \qquad \text{and} \qquad h(x) = \inf_k f_k(x)$$

are measurable functions on X.

Proof:
Let $I = (a,\infty)$. Then $x \in g^{-1}(I)$ iff $g(x) > a$ iff $f_k(x) > a$ for some k, and so

$$g^{-1}(I) = \bigcup_{k=1}^{\infty} f_k^{-1}(I),$$

which is measurable. Hence g is measurable.

The same argument works for h with intervals $I = (-\infty, b)$. ■

Corollary 1.6.4
If f_k is a sequence in $\mathcal{L}(X)$, then

$$g(x) = \limsup_k f_k(x) \qquad \text{and} \qquad h(x) = \liminf_k f_k(x)$$

are measurable functions on X.

Proof:
This follows immediately from Theorem 1.6.3 since

$$\limsup_k f_k(x) = \inf_k \sup_{n \geq k} f_n(x)$$

and

$$\liminf_k f_k(x) = \sup_k \inf_{n \geq k} f_n(x).$$

■

Corollary 1.6.5
If f_k is a sequence in $\mathcal{L}(X)$ which converge pointwise to a function f, then $f \in \mathcal{L}(X)$.

Proof:
If $f_k \to f$ pointwise, then

$$f(x) = \lim_k f_k(x) = \limsup_k f_k(x).$$

So f is measurable by Corollary 1.6.4. ■

Since uniform convergence implies pointwise convergence, the above Corollary also implies that uniform limits of measurable functions are measurable.
Given any function f, we define functions f^+ and f^- by

$$f^+(x) = \max\{f(x), 0\} \qquad f^-(x) = \max\{-f(x), 0\}.$$

From these definitions it is immediate that $f = f^+ - f^-$. We also have that $|f| = f^+ + f^-$.

Corollary 1.6.6
Let $f : X \to \mathbb{R}$. Then the following are equivalent:

(i) $f \in \mathcal{L}(X)$.

(ii) f^+ *and* f^- *are both in* $\mathcal{L}(X)$.

Proof:
(i) $\Rightarrow$ (ii): Since the maximum of a finite set is the same as the supremum, and f and 0 and $-f$ are all measurable, Theorem 1.6.3 gives the result.
(ii) $\Rightarrow$ (i): If f^+ and f^- are measurable, then Theorem 1.6.2 tells us that $f = f^+ - f^-$ is measurable. ■

Note that this also tells us that $|f|$ is measurable if f is measurable.

Corollary 1.6.7
Let $f : X \to \mathbb{R}$ be measurable. Then $|f|$ is measurable.

Proof:
We observe that $|f| = f^+ + f^-$. Therefore it is measurable by Corollary 1.6.6 and Theorem 1.6.2. ■

In proofs it is often convenient to work with the positive and negative portions of functions, or with the absolute value, so these corollaries are very important from a theoretical standpoint. An example of this can be seen in the proof of the next proposition.

Recall that a measurable set E is a null set if $m(E) = 0$. We say that two functions f and g from X to $\mathbb{R}$ are equal ***almost everywhere*** if the set where they differ,

$$\{x \in X : f(x) \neq g(x)\}$$

is a null set. We often abbreviate "almost everywhere" as "a.e." and we will write "$f = g$ almost everywhere" or "$f = g$ a.e." when this condition holds.

Proposition 1.6.8
If $f \in \mathcal{L}(X)$, and $f = g$ almost everywhere, then $g \in \mathcal{L}(X)$.

Proof:
Let $E = \{x \in \mathbb{R} : f(x) \neq g(x)\}$ and let $h = f - g$, so that $h = 0$ almost everywhere.

Then h^+ is measurable, since $(h^+)^{-1}((a, \infty))$ is X if $a < 0$, or a subset of E if $a \geq 0$. But E is null, so any subset of it is measurable by Proposition 1.4.5. So h^+ is measurable on X. The same argument shows that h^- is measurable in X.

Since h^+ and h^- are measurable on X, so is h, and $g = f + h$, and therefore g is measurable on X. ■

Finally, and perhaps most importantly, given any non-negative measurable function $f : X \to \mathbb{R}$, we can find a sequence φ_n of measurable simple functions which increase pointwise to f.

Theorem 1.6.9
Let $f \in \mathcal{L}(X)$ with $f \geq 0$. Then there is a sequence of simple functions $\varphi_n \in \mathcal{L}(X)$ with $0 \leq \varphi_n \leq \varphi_{n+1} \leq f$ for all n, and $\varphi_n \to f$ pointwise on X.

Strategy: *We are subdividing the part of the range up into smaller and smaller subintervals, while increasing the range that is covered by these subintervals.*

Proof:
For each n, let

$$E_{n,k} = f^{-1}([(k-1)2^{-n}, k2^{-n}))$$

for $k = 1, 2, \dots, 2^{2n}$, and let

$$\varphi_n = \sum_{k=1}^{2^n} (k-1)\chi_{E_{n,k}}.$$

Clearly the sets $E_{n,k}$ are measurable, so each simple function φ_n is measurable. For a fixed $x \in X$, then, we have that

$$\varphi_n(x) = \min\{p_n 2^{-n}, 2^n\}$$

where $p_n \in \mathbb{N}$ with $p_n 2^{-n} \leq f(x) < (p_n + 1)2^{-n}$. Hence $\varphi_n(x) \leq \varphi_{n+1}(x)$, and as $n \to \infty$, $\varphi_n(x) \to f(x)$. ■

This theorem is a key part of many other results since, in conjunction with the monotone convergence theorem (Theorem 1.8.1) of Section 1.8 it helps us to extend results from measurable simple functions to measurable functions.

Indeed, positive measurable functions play such an important role in the theory we define $\mathcal{L}^+(X)$ to be the set of all non-negative measurable functions on X.

Exercises

1.6.1. Complete the proof of Proposition 1.6.1.

1.6.2. Show that piecewise continuous functions are measurable.

1.6.3. (†) A function is ***monotone increasing*** if $f(x) \geq f(y)$ whenever $x \geq y$ and ***monotone decreasing*** if $f(x) \leq f(y)$ whenever $x \geq y$.

Show that monotone increasing and montone decreasing functions are measurable on any measurable set X.

1.6.4. Carefully verify the claims made in Example 1.6.3.

1.6.5. Show that there are functions f and g which are not measurable, but $f + g$ is measurable.

1.6.6. Show that if h is continuous on a Lebesgue measureable set $X \subseteq R$, and f is measurable on $h(X)$, then $f \circ h$ is measurable on X.

1.6.7. Give an example of a sequence of measurable simple functions $(\varphi_n)_{n=1}^{\infty}$ with $\varphi_n \leq \varphi_{n+1}$, and $\varphi_n(x) \to x^2$ for all $x \in \mathbb{R}$.

1.6.8. (†) Let X be a measurable set. Define a relation on the set of measurable functions on X by $f \sim g$ when $f = g$ almost everywhere. Prove that $\sim$ is an equivalence relation.

1.6.9. Refer to Exercises 1.2.6, 1.4.6 and 1.5.8. Let $X \subseteq \mathbb{R}^n$. A function $f : X \to \mathbb{R}$ is (Lebesgue) measurable if given any interval $I \subseteq \mathbb{R}$, $f^{-1}(I)$ is Lebesgue measurable in $\mathbb{R}^n$.

Prove versions of all the results of this section in this setting.

1.7 Lebesgue Integration

In developing our integration theory, we want to start with the simplest cases, and move up to the more general cases. The strategy is to start with non-negative measurable simple functions, use these to define the integral of general non-negative measurable functions, and then use these to define the integral of general measurable functions. Unfortunately, although we can define the general integral fairly easily, we need some deep results about integration of non-negative measurable functions before we can prove even something basic like the additivity of integration.

As a result, we'll spend some time looking just at integration of non-negative functions. We start with the simple functions, as we already have an idea of how to integrate these.

Note: *when working with integrals of simple functions, remember that $0 \cdot \infty = 0$ by definition.*

Definition 1.7.1
If E is any Lebesgue measurable set, and $\varphi \in \mathcal{L}^+(E)$ is a simple function with standard representation

$$\varphi(x) = \sum_{k=1}^{n} c_k \chi_{E_k}(x),$$

*we define the **Lebesgue integral** of φ to be*

$$\int_E \varphi \; dm = \sum_{k=1}^{n} c_k m(E_k \cap E).$$

This is a natural definition, since for each component $c_k \chi_{E_k}$ of the simple function we are adding the "volume" of E_k in E times the height of the function above it. Note that if φ is a step function and E is the interval $[a, b]$, then the Lebesgue and Riemann integrals agree.

Example 1.7.1
The characteristic function of the set of rational numbers, $\chi_{\mathbb{Q}}$ is a trivial example of a simple function.

$$\int_E \chi_{\mathbb{Q}} \; dm = m(\mathbb{Q} \cap E) = 0$$

for any measurable set E. This is what we hoped would be the case from the discussion in Section 1.1. ◇

As a matter of notation, observe that we have dropped the variable from the integral. This is to emphasise that the integral really operates on *functions*, rather than expressions in a variable. If we wish to emphasise the variable we are integrating over, we use the notation

$$\int_E \varphi(x) \; dm(x).$$

We immediately can observe that, for a non-negative, simple, measurable function φ,

$$\int_E \varphi \; dm \in [0, \infty].$$

Also

$$\int_E 0 \; dm = 0.$$

The following facts follow from the definition. The proofs are not particularly enlightening, so we relegate them to Section 1.11.

Proposition 1.7.1
If E is any Lebesgue measurable set, φ and ψ are simple functions in $\mathcal{L}^+(E)$, and $c \geq 0$ is any constant, then

(i) if φ has a representation $\varphi = \sum_{k=1}^{l} a_k \chi_{G_k}$, with $a_k \geq 0$, then

$$\int_E \varphi \, dm = \sum_{k=1}^{l} a_k m(G_k \cap E).$$

(ii) $\int_E c\varphi \, dm = c \int_E \varphi \, dm.$

(iii) $\int_E \varphi + \psi \, dm = \int_E \varphi \, dm + \int_E \psi \, dm.$

(iv) if $\varphi \leq \psi$, then $\int_E \varphi \, dm \leq \int_E \varphi \, dm.$

(v) if $F \subseteq E$ is measurable, $\int_E \varphi \chi_F \, dm = \int_F \varphi \, dm.$

(vi) if E is null, then $\int_E \varphi \, dm = 0.$

This proposition tells us that integration of measurable simple functions behaves pretty much the way we expect. Part (i) of this last theorem is useful from the standpoint of calculating integrals, since it means that we don't need to be careful about how we represent our simple function.

We can now meaningfully make the following definition:

Definition 1.7.2
*If E is any Lebesgue measurable set and $f \in \mathcal{L}^+(E)$, then we define the **Lebesgue integral** of f to be*

$$\int_E f \, dm = \sup \left\{ \int_E \varphi \, dm : 0 \leq \varphi \leq f \text{ and } \varphi \text{ simple, measurable} \right\}$$

If $f = \psi$ is a simple function, this general definition of the Lebesgue integral and the previous one agree, since part (iv) of the previous Lemma tells us that the integral of ψ is an upper bound for the set

$$\left\{ \int_E \varphi \, dm : 0 \leq \varphi \leq \psi \text{ and } \varphi \text{ simple, measurable} \right\},$$

but the integral is also a member of the set, so therefore it is the supremum.

We may as well state the definition of the integral of arbitrary measurable functions at this point, although we will immediately set it aside and go back to considering just non-negative functions. The key to the definition is that we can write $f = f^+ - f^-$ and integrate the positive and negative parts, since f^+ and f^- are non-negative measurable functions if f is measurable. The only thing we need to be wary of is situations where we have $\infty - \infty$ for the integral.

Definition 1.7.3
If E is any measurable set, and $f \in \mathcal{L}(E)$ with one of

$$\int_E f^+ \, dm < \infty \qquad \text{or} \qquad \int_E f^- \, dm < \infty,$$

*then we define the **Lebesgue integral** of f on E to be*

$$\int_E f \, dm = \int_E f^+ \, dm - \int_E f^- \, dm.$$

This agrees with the Lebesgue integral defined for non-negative measurable simple functions and non-negative measurable functions, since in both those cases $f^- = 0$.

We turn back to the properties of the integral for non-negative functions.

Proposition 1.7.2
Let E be any measurable set, f and $g \in \mathcal{L}^+(E)$, and $c \geq 0$ be any constant. Then

(i) $\displaystyle\int_E cf \, dm = c \int_E f \, dm.$

(ii) if $f \leq g$, then $\displaystyle\int_E f \, dm \leq \int_E g \, dm.$

(iii) if $F \subseteq E$ is measurable, $\displaystyle\int_E f\chi_F \, dm = \int_F f \, dm.$

(iv) if E is null, then $\displaystyle\int_E f \, dm = 0.$

Proof:

Strategy: *We use the facts about suprema from Section A.4 to extend the results of Proposition 1.7.1 to more general functions.*

(i) If $c = 0$ the result is trivial. If $c > 0$ and ψ is any non-negative, simple, measurable function with $\psi \leq f$, then $c\psi \leq cf$. Therefore,

$$\left\{ \int_E \varphi \, dm : 0 \leq \varphi \leq cf \text{ and } \varphi \text{ simple, measurable} \right\}$$
$$\supseteq \left\{ \int_E c\psi \, dm : 0 \leq \psi \leq f \text{ and } \psi \text{ simple, measurable} \right\}$$

Taking suprema of both sides,

$$\begin{aligned}\int_E cf\ dm &\geq \sup\left\{\int_E c\psi\ dm : 0 \leq \psi \leq f \text{ and } \psi \text{ simple, measurable}\right\}\\ &\geq c\sup\left\{\int_E \psi\ dm : 0 \leq \psi \leq f \text{ and } \psi \text{ simple, measurable}\right\}\\ &\geq c\int_E f\ dm.\end{aligned}$$

But this also means that

$$c\int_E f\ dm \geq cc^{-1}\int cf\ dm = \int_E cf\ dm,$$

so

$$\int_E cf\ dm = c\int_E f\ dm.$$

(ii) If φ is non-negative, simple, and measurable, with $\varphi \leq f$, then $\varphi \leq g$, so

$$\begin{aligned}&\left\{\int_E \varphi\ dm : 0 \leq \varphi \leq f \text{ and } \varphi \text{ simple, measurable}\right\}\\ &\qquad\subseteq \left\{\int_E \psi\ dm : 0 \leq \psi \leq g \text{ and } \varphi \text{ simple, measurable}\right\}\end{aligned}$$

and simply taking suprema of both sides gives,

$$\int_E f\ dm \leq \int_E g\ dm.$$

(iii) It is trivial that $0 \leq \varphi \leq f\chi_F$ on E if and only if $0 \leq \varphi(x) \leq f(x)$ for all $x \in F$. Combining this with the fact that $\varphi = \varphi\chi_F$ if this is the case and Proposition 1.7.1, part (iv), we have

$$\begin{aligned}&\left\{\int_E \varphi\ dm : 0 \leq \varphi \leq f\chi_F \text{ and } \varphi \text{ simple, measurable}\right\}\\ &\qquad= \left\{\int_F \varphi\ dm : 0 \leq \varphi \leq f \text{ and } \varphi \text{ simple, measurable}\right\}\end{aligned}$$

and so taking suprema we get the result.

(iv) by Proposition 1.7.1,

$$\begin{aligned}\int_E f\ dm &= \sup\left\{\int_E \varphi\ dm : 0 \leq \varphi \leq f \text{ and } \varphi \text{ simple, measurable}\right\}\\ &= \sup\{0\} = 0.\end{aligned}$$

■

Note that we haven't yet shown some key properties that we expect this integral to have, such as additivity. The proof of additivity requires the convergence theorem that will be proved in the next section.

We now show how an example of the difficulty of working with this definition of the integral.

Example 1.7.2

Note: *the method of this example can be generalized to any monotone increasing continuous function.*

Consider the function $f(x) = x$ on the set $[0, 1]$. f is measurable, since f is continuous.

We immediately note that $f(x) = x$ is Riemann integrable. Given any partition $\mathcal{P} = \{0 = a_0 < a_1 < \ldots < a_n = 1$ of $[0, 1]$, let $I_k = [a_{k-1}, a_k)$ for $k = 1, 2, \ldots n - 1$ and $I_n = [a_{n-1}, 1]$. Let

$$\psi = \sum_{k=1}^{n} a_{k-1}\chi_{I_k}.$$

It is straighforward that $\psi \leq f$, and that

$$\int_{[0,1]} \psi \; dm = \sum_{k=1}^{n} a_{k-1}(a_k - a_{k-1}) = L(f, \mathcal{P}).$$

Since these step functions ψ are a special subset of the general simple functions φ with $0 \leq \varphi \leq f$, we have that

$$\int_{[0,1]} f \; dm \geq \sup\left\{\int_{[0,1]} \psi \; dm : \psi \text{ as above}\right\} \tag{1.3}$$

$$= \sup\{L(f, \mathcal{P}) : \mathcal{P} \text{ a partition}\} = \int_0^1 x \; dx = 1/2. \tag{1.4}$$

If $0 \leq \varphi \leq f$ on $[0, 1]$ with $\varphi \in \mathcal{L}^+([0, 1))$, then if $\varphi = \sum_{k=1}^{n} c_k\chi_{E_k}$ is the standard representation we immediately have that the sets E_k are pairwise disjoint. Also

$$c_k \leq \inf E_k$$

for all k, otherwise we can find $x \in E_k$ with $c_k \geq x$, and then $\varphi(x) = c_k \geq x = f(x)$, which contradicts our assumption.

Without loss of generality, we can assume that $0 = c_1 < c_2 < c_3 < \ldots < c_n < 1$. Let $I_k = [c_k, c_k + 1)$ for $1 \leq k < n$, and $I_n = [c_n, 1]$. Then let

$$\psi = \sum_{k=1}^{n} c_k\chi_{I_k}.$$

We have $\varphi \leq \psi$, since if $\varphi(x) = c_k$, then $x \in E_k$, which means $x \in I_m$ for some $m \geq k$, as otherwise $x < c_k \leq \inf E_k \leq x$, which is a contradiction. Therefore $\psi(x) = c_m \geq c_k = \varphi(x)$. So

$$\int_{[0,1]} \varphi \; dm \leq \int_{[0,1]} \psi \; dm.$$

Also, clearly, $0 \leq \psi \leq f$, so this means that it is sufficient to take the suprema of functions like ψ, instead of general simple functions, to find the integral of f. But ψ is exactly the step function coming from the lower sum of f with the partition $\mathcal{P} = \{c_1, c_2, \ldots, c_n, 1\}$, ie.

$$\int_{[0,1]} \psi \; dm = L(f, \mathcal{P}) \leq \int_0^1 x \; dx = 1/2,$$

and so

$$\int_{[0,1]} f \; dm \leq 1/2.$$

Therefore

$$\int_{[0,1]} f \; dm = 1/2$$

using Lebesgue integration, which is very good, since it agrees with the Riemann integral. ◇

While this example demonstrates that you can find Lebesgue integrals from first principles, in practice it is usually much easier to approximate the function we want by functions which are easy to integrate, and then use the convergence theorems we will see in the next few sections.

The next result is important both for technical reasons and philosophical reasons when we come to generalise integration.

Proposition 1.7.3
If $f \in \mathcal{L}^+(\mathbb{R})$, then the set function

$$\mu_f(A) = \int_A f \; dm$$

defined for A in the σ-algebra of Lebesgue measurable sets is a measure.

Proof:
First note that

$$\mu_f(\emptyset) = \int_\emptyset f \; dm = 0$$

for every function f.

Let A be the disjoint union of a countable family of sets A_k.

Firstly we note that if $f = \chi_E$ for some measurable set E, then σ-additivity of μ_f follows directly from σ-additivity of m, for

$$\begin{aligned}\mu_f(A) = \int_A \chi_E \, dm &= m(A \cap E) \\ &= m\left(\bigcup_{k=1}^{\infty} A_k \cap E\right) \\ &= m\left(\bigcup_{k=1}^{\infty} (A_k \cap E)\right) \\ &= \sum_{k=1}^{\infty} m(A_k \cap E) = \sum_{k=1}^{\infty} \mu_f(A_k).\end{aligned}$$

Immediately from this we can see that if f is a simple function $\sum_{k=1}^{n} c_k \chi_{E_k}$, then μ_f is also σ-additive:

$$\begin{aligned}\mu_f(A) = \sum_{k=1}^{n} c_k m(A \cap E_k) &= \sum_{k=1}^{n} c_k \sum_{j=1}^{\infty} m(A_j \cap E_k) \\ &= \sum_{j=1}^{\infty} \sum_{k=1}^{n} c_k m(A_j \cap E_k) = \sum_{k=1}^{\infty} \mu_f(A_k).\end{aligned}$$

Now for any $f \in \mathcal{L}^+(\mathbb{R})$, if we have any simple function φ with $0 \leq \varphi \leq f$,

$$\int_A \varphi \, dm = \sum_{k=1}^{\infty} \int_{A_k} \varphi \, dm \leq \sum_{k=1}^{\infty} \mu_f(A_k),$$

and taking suprema over all such φ we then get

$$\mu_f(A) \leq \sum_{k=1}^{\infty} \mu_f(A_k).$$

If any of the A_k have $\mu_f(A_k) = \infty$, we are done, for then $\mu_f(A) \geq \mu_f(A_k) = \infty$. So we now assume that $\mu_f(A_k)$ is finite for all k.

Now, given any two disjoint measurable sets E and F with $\mu_f(E)$ and $\mu_f(F)$ finite, and any $\varepsilon > 0$, we can find a measurable simple function φ such that

$$\int_E \varphi \, dm \geq \int_E f \, dm - \varepsilon/2 \qquad \text{and} \qquad \int_F \varphi \, dm \geq \int_F f \, dm - \varepsilon/2$$

and so

$$\mu_f(E \cup F) \geq \int_{E \cup F} \varphi \, dm = \int_E \varphi \, dm + \int_E \varphi \, dm \geq \mu_f(E) + \mu_f(F) - \varepsilon.$$

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

$$\mu_f(E \cup F) \geq \mu_f(E) + \mu_f(F),$$

and by repeated application of this result, we have

$$\mu_f\left(\bigcup_{k=1}^n A_k\right) \geq \sum_{k=1}^n \mu_f(A_k).$$

Finally, since $A \supseteq \bigcup_{k=1}^n A_k$ for all n we have

$$\mu_f(A) \geq \sum_{k=1}^n \mu_f(A_k)$$

for all n, and so

$$\mu_f(A) \geq \sum_{k=1}^\infty \mu_f(A_k).$$

Therefore μ_f is σ-additive. ■

The philosophical importance of this result is that there are *lots* of measures on $\mathbb{R}$. It is also extremely useful as it allows us to break integrals up over measurable sets. Consider the following corollaries:

Corollary 1.7.4
If E and F are disjoint measurable sets and $f \in \mathcal{L}(E \cup F)$, then

$$\int_{E\cup F} f\ dm = \int_E f\ dm + \int_F f\ dm.$$

Proof:
This is just the previous result applied to $f\chi_{E\cup F} \in \mathcal{L}^+(\mathbb{R})$ and written as integrals. ■

Corollary 1.7.5
If E and F are measurable sets with $F \subset E$ and $m(E \setminus F) = 0$, then

$$\int_E f\ dm = \int_F f\ dm.$$

Proof:

$$\int_E f\ dm = \int_{F\cup(E\setminus F)} f\ dm = \int_F f\ dm + \int_{E\setminus F} f\ dm = \int_F f\ dm + 0.$$

■

Another way of thinking of this result is that if we alter the value of a function on a set of measure 0, the value of the integral is unaffected.

Corollary 1.7.6
If f is a measurable function and $f = g$ almost everywhere, and E is a measurable set, then

$$\int_E f\ dm = \int_E g\ dm.$$

Proof:
Let $F = \{x \in \mathbb{R} | f(x) = g(x)\}$, so $m(E \setminus F) = 0$. Then

$$\int_E f\ dm = \int_F f\ dm = \int_F g\ dm = \int_E g\ dm$$

since $f = g$ on F. ■

This fact means that it is often sufficient to consider the behaviour of functions almost everywhere. For example, in the next section we will

Exercises

1.7.1. Show that $\int_\emptyset f\ dm = 0$ for every $f \in \mathcal{L}(E)$.

1.7.2. (†) Let $m(E) < \infty$. Show that if $f_n \in \mathcal{L}^+(E)$ and $f_n \to f$ uniformly, then $f \in \mathcal{L}^+(E)$ and

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

1.7.3. Let $m(E) = \infty$. Find f_n and $f \in \mathcal{L}^+(E)$ with

$$\int_E f\ dm = 0,$$

and $f_n \to f$ uniformly, but

$$\lim_{n\to\infty} \int_E f_n\ dm \to \infty.$$

1.7.4. (†) Let $f \in \mathcal{L}^+(\mathbb{R})$ and $c \in \mathbb{R}$ and constant.
Show that

$$\int_{\mathbb{R}} f(x + c)\ dm(x) = \int_{\mathbb{R}} f(x)\ dm(x)$$

and if $c \neq 0$, then

$$\int_{\mathbb{R}} f(cx)\ dm(x) = |c|^{-1} \int_{\mathbb{R}} f(x)\ dm(x).$$

1.7.5. Using the method of Example 1.7.2, show that if $f(x) = x^2$,

$$\int_{[0,1]} f\ dm = \int_0^1 x^2\ dx = 1/3.$$

1.8 Monotone Convergence Theorem

The most powerful results of Lebesgue integration theory come from the fact that it is comparatively well-behaved under limits. There are two big convergence theorems, the monotone convergence theorem (which we'll often abbreviate as the MCT) and the dominated convergence theorem (or DCT). We are at a point where we can prove the first of these, which applies only to integrals non-negative functions. The dominated convergence theorem applies more generally, and so we'll present it once we've discussed integrals of more general functions.

Theorem 1.8.1 (Monotone Convergence Theorem)
Let f_n be a monotone increasing sequence of functions in $\mathcal{L}^+(E)$. That is, $0 \leq f_n(x) \leq f_{n+1}(x)$ for all $x \in E$ and $n \in \mathbb{N}$. If f is the pointwise limit of the f_n, then

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

Proof:
It is clear that, since f_n is increasing, so are the integrals, and so there is some $\alpha \in [0, \infty]$ such that

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

Moreover, since $\int_E f_n\ dm \leq \int_E f\ dm$, for all n,

$$\alpha \leq \int_E f\ dm.$$

Choose c with $0 < c < 1$. Now let φ be a simple function such that $0 \leq \varphi \leq f$, and let

$$E_n = \{x : f_n(x) \geq c\varphi(x)\}$$

Since the $f_n(x)$ are increasing for each x, $E_n \subseteq E_{n+1}$ for all n, and since $f_n(x) \to f(x)$, we have that every x is eventually in some E_n, so

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

Now

$$\int_E f_n\ dm \geq \int_{E_n} f_n\ dm \geq \int_{E_n} c\varphi\ dm = \mu_{c\varphi}(E_n),$$

and since $\mu_{c\varphi}$ is a measure by Proposition 1.7.3, Proposition 1.3.3 tells us that $\mu_{c\varphi}(E) = \lim_{n\to\infty} \mu_{c\varphi}(E_n)$, and so

$$\alpha = \lim_{n\to\infty} \int_E f_n\ dm \geq \int_E c\varphi\ dm = c \int_E \varphi\ dm$$

and therefore, letting $c \to 1$,

$$\alpha \geq \int_E \varphi \; dm.$$

Then using the definition of the integral to take suprema, we get

$$\alpha \geq \int_E f \; dm,$$

and so

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

■

Exercise 1.8.4 and Exercise 1.8.5 shows that is is necessary that the sequence be monotone increasing.

Part of the utility of this theorem is that it allows us to evaluate Lebesgue integrals with ease.

Example 1.8.1

Let $f(x) = x$ and consider the interval $[0, 1)$. Let φ_n be the simple function

$$\varphi_n = \sum_{k=0}^{2^n-1} k2^{-n} \chi_{[k2^{-n},(k+1)2^{-n})}.$$

It is not hard to see that $\varphi_n \leq \varphi_{n+1}$, and also that $\varphi_n(x) \to x$ as $n \to \infty$.

The monotone convergence theorem then tells us that

$$\begin{aligned}\int_{[0,1)} f \; dm &= \lim_{n\to\infty} \int_{[0,1)} \varphi_n \; dm \\ &= \lim_{n\to\infty} \sum_{k=0}^{2^n-1} k2^{-2n} \\ &= \lim_{n\to\infty} \frac{(2^n-1)2^n}{2} 2^{-2n} \\ &= \lim_{n\to\infty} \frac{2^{2n}-2^n}{2^{2n+1}} \\ &= 1/2.\end{aligned}$$

◇

This is clearly superior to the method used in Example 1.7.2.

In fact, it turns out that we don't need a monotone sequence f_n to converge to f at every point, all we need is convergence at almost every point, giving this slight refinement of the monotone convergence theorem.

Corollary 1.8.2
Let f_n be a monotone increasing sequence of functions in $\mathcal{L}^+(E)$, and f a function. If

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

almost everywhere in E, then

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

Proof:
Let

$$F = \{x \in E : f(x) = \lim_{n\to\infty} f_n(x)\}.$$

Then $m(E \setminus F) = 0$, and using the monotone convergence theorem on F, we have

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

■

In such cases as this we will say that the sequence f_n converges to f ***pointwise almost everywhere*** or ***pointwise a.e.*** Convergence pointwise a.e. is weaker than pointwise convergence, since if $f_n \to f$ pointwise, then automatically $f_n \to f$ pointwise a.e.

This monotone convergence theorem is vital in proving many useful results, not least of which is that integration is additive.

Theorem 1.8.3
Let f_k a (finite or infinite) sequence of functions in $\mathcal{L}^+(E)$, and let

$$f = \sum_k f_k.$$

Then

$$\int_E f\ dm = \sum_k \int_E f_k\ dm.$$

Proof:
Given two functions f_1 and f_2, we can use find monotone increasing sequences of simple functions φ_k and ψ_j which converge pointwise to f_1 and f_2 respectively, by Theorem 1.6.9. Then the sequence $\varphi_k + \psi_k$ converges pointwise to $f_1 + f_2$, and $\varphi_k + \psi_k \leq \varphi_{k+1} \leq \psi_{k+1}$. Therefore

$$\begin{aligned}\int_E f_1 + f_2\ dm &= \lim_{k\to\infty} \int_E \varphi_k + \psi_k\ dm \\ &= \lim_{k\to\infty} \int_E \varphi_k\ dm + \lim_{k\to\infty} \int_E \psi_k\ dm = \int_E f_1\ dm + \int_E f_2\ dm,\end{aligned}$$

using the monotone convergence theorem and additivity for simple functions. Induction gives the result for finite sums.

For infinite sums, note that since $f_k \geq 0$, the partial sums

$$s_n = \sum_{k=1}^{n} f_k$$

form a monotone increasing sequence which converges pointwise to f so, using the monotone convergence theorem again, we have

$$\int_E f \; dm = \lim_{n\to\infty} s_n \; dm = \lim_{n\to\infty} \sum_{k=1}^{n} \int_E f_k \; dm = \sum_{k=1}^{\infty} \int_E f_k \; dm.$$

■

In general, sequences of functions are not going to be monotone increasing, and it turns out that we can say a little bit about intergals of functions in such sequences, even if they do not converge at all. The relevant result is known as Fatou's Lemma, and it tells us that the limit infima behaves fairly nicely when integrated.

Proposition 1.8.4 (Fatou's Lemma)
If f_n is a sequence of functions in $\mathcal{L}^+(E)$, then

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

Proof:

Recall that

$$\liminf_{n\to\infty} f_n(x) = \lim_{n\to\infty} \inf\{f_k(x) : i \geq n\}.$$

Let $g_n(x) = \inf\{f_k(x) : i \geq n\}$. Then the functions g_n are non-negative, measurable and form a monotone increasing sequence. Hence by the monotone convergence theorem,

$$\begin{aligned}\int_E \liminf_{n\to\infty} f_n \; dm &= \int_E \lim_{n\to\infty} g_n \; dm \\ &= \lim_{n\to\infty} \int_E g_n \; dm\end{aligned}$$

Moreover, since $g_n \leq f_k$ for all $i \geq n$, we have

$$\int_E g_n \; dm \leq \int_E f_k \; dm$$

for all $i \geq n$, and taking infima of both sides, we get

$$\int_E g_n \; dm \leq \inf\{\int_E f_k \; dm : i \geq n\}.$$

Therefore

$$\int_E \liminf_{n\to\infty} f_n \; dm = \lim_{n\to\infty} \int_E g_n \; dm \leq \lim_{n\to\infty} \inf\{\int_E f_k \; dm : i \geq n\}$$

as required. ■

Notice that Fatou's Lemma makes no assumptions about the convergence properties of the sequence, and so is very general. The conclusion is correspondingly weak, however. But if f_n *does* converge, we have

$$\int_E \lim_{n\to\infty} f_n \; dm \leq \liminf_{n\to\infty} \int_E f_n \; dm, \tag{1.5}$$

which is often useful to know.

Exercises

1.8.1. (†) Use the Monotone Convergence Theorem, together with an appropriate sequence of simple functions to show that

$$\int_{[0,1]} x^2 \; dm(x) = 1/3.$$

Hint: *the sum of the first n squares is $n(n+1)(2n-1)/6$.*

1.8.2. Let $f(x) = \dfrac{1}{x^2+1}$ and show that

$$\int_{\mathbb{R}} f \; dm = \pi.$$

Note: *it is possible to prove Fatou's lemma without using MCT, so this exercise can give an alternative proof of MCT.*

1.8.3. (†) Show that the monotone convergence theorem can be proved as a corollary of Fatou's lemma.

1.8.4. (†) Find a sequence of non-negative measurable functions f_n with $f(x) \leq 1$ for all $x \in \mathbb{R}$, such that $f_n \to 0$ pointwise, but

$$\int_{\mathbb{R}} f_n \, dm = 1.$$

1.8.5. (†) Find a sequence of non-negative measurable functions f_n, with $f_n \to 0$ pointwise on $[0,1]$ and

$$\int_{[0,1]} f_n \, dm = 1.$$

Hint: *Find $\int_{(0,1)} f \, dm$ first.*

1.8.6. (†) Let $f(x) = x^{-1/2}$ on $(0,1)$, and $f(x) = 0$ otherwise. Let q_n be an enumeration of the rational numbers, and let $g_n = f(x - q_n)2^{-n}$. Then let

$$g = \sum_{i=0}^{\infty} g_n.$$

Let D be the domain of g.

Show that $g \in \mathcal{L}(D)$, and moreover that

$$\int_{\mathbb{R}} g \, dm = 4.$$

Show that g is and unbounded on every subinterval and hence g is not Riemann integrable.

Show that $g^2 < \infty$ almost everywhere, but

$$\int_I g^2 \, dm = \infty$$

for any interval $I \subseteq \mathbb{R}$.

1.9 The General Lebesgue Integral

Recall that if E is any measurable set, and f is a measurable function with one of

$$\int_E f^+ \, dm < \infty \qquad \text{or} \qquad \int_E f^- \, dm < \infty,$$

then we define the Lebesgue integral of f on E to be

$$\int_E f \, dm = \int_E f^+ \, dm - \int_E f^- \, dm.$$

Note that these integrals may be infinite. We say that f is ***Lebesgue integrable*** on E if the Lebesgue integrals of both f^+ and f^- are finite. We let $L^1(E)$ be the set of Lebesgue integrable functions on E.

Lemma 1.9.1
Let E be any measurable set. Then the following are equivalent:

(i) $f \in L^1(E)$.

(ii) $\int_E f \; dm$ *is finite.*

(iii) $\int_E |f| \; dm < \infty$.

Proof:
(i) $\Rightarrow$ (iiI): If $f \in L^1(E)$, then both

$$\int_E f^+ \; dm < \infty \qquad \text{and} \qquad \int_E f^- \; dm < \infty,$$

so

$$\int_E |f| \; dm = \int_E f^+ \; dm + \int_E f^- \; dm$$

must be a finite quantity.
(ii) $\Rightarrow$ (i): If

$$\int_E |f| \; dm = \int_E f^+ \; dm + \int_E f^- \; dm$$

is finite, then both parts of the sum must be finite. Hence $f \in L^1(E)$.
(i) $\Leftrightarrow$ (ii): This is left as an (easy) excercise. ■

The following are immediate consequences of the definition of the general integral, and are all reasonable properties that we expect an integral to have.

Proposition 1.9.2
Let E be any measurable set, and f, $g \in \mathcal{L}(E)$ be functions for which the Lebesgue integral exists, and let c be any constant. Then

(i) $\int_E cf \; dm = c \int_E f \; dm$.

(ii) if we do not have

$$\int_E f \; dm = +\infty \qquad \text{and} \qquad \int_E g \; dm = -\infty$$

or vice versa, then

$$\int_E f + g \; dm = \int_E f \; dm + \int_E g \; dm$$

(iii) if $f \leq g$, *then* $\int_E f \; dm \leq \int_E g \; dm$.

(iv) if $F \subseteq E$ is measurable, $\displaystyle\int_E f\chi_F \, dm = \int_F f \, dm.$

(v) if E is null, then $\displaystyle\int_E f \, dm = 0.$

(vi) if f is bounded on E and $m(E) < \infty$, then $f \in L^1(E)$. Moreover, if $\alpha \leq f(x) \leq \beta$ on E, then

$$\alpha m(E) \leq \int_E f \, dm \leq \beta m(E).$$

Proof:

Almost all of these follow directly from the corresponding result for non-negative measurable functions applied to f^+ and f^-.

(i) The result for $c = 0$ is trivial. If $c > 0$, $(cf)^+ = cf^+$ and $(cf)^- = cf^-$, so

$$\begin{aligned}\int_E cf \, dm &= \int_E cf^+ \, dm - \int_E cf^- \, dm \\ &= c\left(\int_E f^+ \, dm - \int_E f^- \, dm\right) = c\int_E f \, dm.\end{aligned}$$

On the other hand, $(-f)^+ = f^-$ and $(-f)^- = f^+$, so

$$\int_E -f \, dm = \int_E f^- \, dm - \int_E f^+ \, dm = -\int_E f \, dm.$$

Combining this with the $c > 0$ case, we are done.

(ii) Using the fact that $f + g = (f+g)^+ - (f+g)^- = f^+ - f^- + g^+ - g^-$, we have

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

and so by the Theorem 1.8.3,

$$\begin{aligned}&\int_E (f+g)^+ \, dm + \int_E f^- \, dm + \int_E g^- \, dm \\ &\qquad = \int_E (f+g)^- \, dm + \int_E f^+ \, dm + \int_E g^+ \, dm.\end{aligned}$$

Rearranging the terms gives

$$\begin{aligned}\int_E f + g \, dm &= \int_E (f+g)^+ \, dm - \int_E (f+g)^- \, dm \\ &= \int_E f^+ \, dm - \int_E f^- \, dm + \int_E g^+ \, dm - \int_E g^- \, dm \\ &= \int_E f \, dm + \int_E g \, dm.\end{aligned}$$

(iii) we have $0 \leq f^+ \leq g^+$ and $0 \leq g^- \leq f^-$, and so

$$0 \leq \int_E f^+ \, dm \leq \int_E g^+ \, dm \qquad \text{and} \qquad 0 \leq \int_E g^- \, dm \leq \int_E f^- \, dm$$

Hence

$$\int_E f^+ \, dm - \int_E f^- \, dm \leq \int_E g^+ \, dm - \int_E g^- \, dm,$$

as required.

(iv) we have that $(f\chi_F)^+ = f^+\chi_F$ and $(f\chi_F)^- = f^-\chi_F$, so

$$\int_E f\chi_F = \int f^+\chi_F \, dm - \int f^-\chi_F \, dm = \int_F f^+ \, dm - \int_F f^- \, dm = \int_F f \, dm.$$

(v) $\int_E f \, dm = \int_E f^+ \, dm - \int_E f^- \, dm = 0.$

(vi) For a bounded measurable function f, we have numbers α and β such that $\alpha \leq f(x) \leq \beta$ for all $x \in E$.

If $\alpha \geq 0$, then $f \in \mathcal{L}^+(E)$, and since $\alpha \leq f \leq \beta$ on E, we have

$$\alpha m(E) = \int_E \alpha \, dm \leq \int_E f \, dm \leq \int_E \beta \, dm = \beta m(E).$$

If $beta \leq 0$, then $-f \in \mathcal{L}^+(E)$, and $-\beta \leq -f \leq -\alpha$, so we can apply (i) and the previous part to get

$$-\beta m(E) \leq -\int_E f \, dm \leq -\alpha m(E),$$

and multiplying through by -1 gives the result.

Finally, if $\alpha < 0$ and $\beta > 0$ we have $0 \leq f^+ \leq \beta$ and $0 \leq f^- \leq -\alpha$, and so

$$0 \leq \int_E f^+ \, dm \leq \beta m(E) \qquad \text{and} \qquad \alpha m(E) \leq -\int_E f^- \, dm \leq 0.$$

Adding these inequalities together gives the result.

In all three cases, the integral of f is finite, and so $f \in L^1(E)$. ∎

Looking at the proof of these results, you will notice that there is a common technique which is to: (i) prove the result for measurable simple functions, (ii) use this result for measurable simple functions to prove the result for measurable $f(x) \geq 0$ (usually by either taking suprema, or using the MCT), and (iii) use the second result to prove for general measurable f. This is generally a good approach to proving general facts involving Lebesgue integrals.

The one additional component of our theory is a convergence theorem which applies in this more general setting. Unfortunately, the Monotone Convergence Theorem cannot apply for general integrable functions, because of the following example:

Example 1.9.1

Let $f_n(x) = -1/n$ for all $x \in \mathbb{R}$. Then $f_n(x)$ is pointwise monotone increasing to $f(x) = 0$. But

$$\int_{\mathbb{R}} f_n \, dm = -\infty, \qquad \text{and} \qquad \int_{\mathbb{R}} f \, dm = 0,$$

so

$$\int_{\mathbb{R}} f \, dm \neq \lim_{n\to\infty} \int_{\mathbb{R}} f_n \, dm.$$

◇

Fortunately, the convergence theorem which does hold is more powerful because it allows any sort of pointwise convergence, as long as we have some upper bound to the integrals.

Theorem 1.9.3 (Dominated Convergence Theorem)
Let $f_n \in \mathcal{L}(E)$ be a sequence of functions for which the Lebesgue integral exists on E, and let $f_n \to f$ pointwise. If there is some function $g \in L^1(E)$, for which

$$|f_n(x)| \leq g(x),$$

for all $x \in E$, then

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

Proof:

Note that for each n, the functions $g + f_n$ and $g - f_n$ are non-negative, so by Fatou's Lemma

$$\begin{aligned}
\int_E g \, dm + \int_E f \, dm &= \int_E g + f \, dm \\
&\leq \liminf_{n\to\infty} \int_E g + f_n \, dm = \int_E g \, dm + \liminf_{n\to\infty} \int_E f_n \, dm, \\
\int_E g \, dm - \int_E f \, dm &= \int_E g - f \, dm \\
&\leq \liminf_{n\to\infty} \int_E g - f_n \, dm = \int_E g \, dm + \liminf_{n\to\infty} -\int_E f_n \, dm.
\end{aligned}$$

Removing the common terms (which we can do because the integral of g is finite), and recognising the fact that $\liminf -c_n = -\limsup c_n$ for any sequence $(c_n)_{n=1}^{\infty}$, we have

$$\begin{aligned}
\int_E f \, dm &\leq \liminf_{n\to\infty} \int_E f_n \, dm, \\
\int_E f \, dm &\geq \limsup_{n\to\infty} \int_E f_n \, dm.
\end{aligned}$$

Therefore

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

■

As with the monotone convergence theorem, the result remains true if we replace pointwise convergence by pointwise a.e. convergence:

Corollary 1.9.4
Let $f_n \in \mathcal{L}(E)$ be a sequence of functions for which the Lebesgue integral exists on E, and let $f_n \to f$ pointwise almost everywhere. If there is some function $g \in L^1(E)$, for which

$$|f_n(x)| \leq g(x),$$

for all $x \in E$, then

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

The proof of this is similar to the proof for the corresponding version of the monotone convergence theorem, and so is left as an exercise.

Proposition 1.9.5
If $f_n \in L^1(E)$ and

$$\sum_{n=1}^{\infty} \int_E |f_n|\ dm < \infty,$$

then there is an $f \in L^1(E)$, such that

$$f = \sum_{n=1}^{\infty} f_n$$

almost everywhere on E, and

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

Proof:
We know from Theorem 1.8.3 for non-negative measurable functions, that

$$\int_E \sum_{n=1}^{\infty} |f_n|\ dm = \sum_{n=1}^{\infty} \int_E |f_n|\ dm < \infty,$$

and so $\sum_{n=1}^{\infty} |f_n|$ is finite almost everywhere (else it would not have a finite integral). Hence

$$\sum_{n=1}^{\infty} f_n(x)$$

converges absolutely almost everywhere, and we can define a function which is equal to the sum at the convergent points and 0 at the divergent points. Furthermore, $f \in L^1(E)$, since

$$\int_E |f| \, dm = \int_E \sum_{n=1}^{\infty} |f_n| \, dm < \infty.$$

Looking at the partial sums of the sequence, s_n, we see that

$$|s_n| = \left| \sum_{k=1}^{n} f_k \right| \le \sum_{k=1}^{n} |f_k| \le \sum_{k=1}^{\infty} |f_k|$$

and so, by the dominated convergence theorem,

$$\begin{aligned} \int_E f \, dm = \int_E \lim_{n\to\infty} s_n \, dm &= \lim_{n\to\infty} \int_E \sum_{k=1}^{n} f_k \, dm \\ &= \lim_{n\to\infty} \sum_{k=1}^{n} \int_E f_k \, dm = \sum_{k=1}^{\infty} \int_E f_k \, dm. \end{aligned}$$

■

Exercises

1.9.1. (†) Let $f : \mathbb{R} \to \mathbb{C}$. We define $\operatorname{Re} f$ and $\operatorname{Im} f$ to be the functions which give the real and imaginary parts of $f(x)$, so

$$f = \operatorname{Re} f + i \operatorname{Im} f,$$

and $\operatorname{Re} f$ and $\operatorname{Im} f$ are real-valued functions. If $\operatorname{Re} f$ and $\operatorname{Im} f \in L^1(E)$, we define

$$\int_E f \, dm = \int_E \operatorname{Re} f \, dm + i \int_E \operatorname{Im} f \, dm$$

and say that f is Lebesgue integrable.

Show that f is Lebesgue integrable if and only if

$$\int_E |f| \, dm < \infty.$$

1.9.2. (†) Using the definitions of Exercise 1.9.1, show that the dominated convergence theorem holds for complex-valued Lebesgue integrable functions.

Show that if f_n are complex-valued and Lebesgue integrable, and

$$\sum_{n=1}^{\infty} \int_E |f_n| \, dm < \infty,$$

then

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

converges a.e to a complex-valued, Lebesgue integrable function f, and

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

1.9.3. Complete the proof of Lemma 1.9.1

1.9.4. Prove Corollary 1.9.4.

1.9.5. Show that

$$\int_{[a,b]} x\ dm(x) = b^2/2 - a^2/2.$$

1.10 Comparison With the Riemann Integral

While the Lebesgue integral has very nice properties and can be used on a wide variety of functions, it can be difficult to calculate the value. While the convergence theorems allow us to find integrals by approximating by simple functions, we don't have an analogue of the fundamental theorem of calculus at present.

Additionally, we would hope that when a function is Riemann integrable that the Lebesgue integral and Riemann integrals ought to agree. If this is the case, it will allow us to use the standard techniques of calculus to calculate integrals, but we will also have available the powerful convergence theorems which allow us to evaluate many more integrals than we might otherwise be able to do.

Theorem 1.10.1
Let f be a bounded function on an interval $[a, b]$. If f is Riemann integrable, then f is Lebesgue integrable, and

$$\int_a^b f(x)\ dx = \int_{[a,b]} f\ dm.$$

Proof:

If f is Riemann integrable, then for any n, we can find partitions $\mathcal{P}_n$ such that $\mathcal{P}_{n+1}$ is a refinement of $\mathcal{P}_n$, that $\|\mathcal{P}_n\| \to 0$ and $L(f, \mathcal{P})$ and $U(f, \mathcal{P})$ converge to the Riemann integral of f.

Letting

$$l_{\mathcal{P}_n} = \sum_{k=1}^{n} m_k \chi_{(x_{k-1}, x_k]} \qquad u_{\mathcal{P}_n} = \sum_{k=1}^{n} M_k \chi_{(x_{k-1}, x_k]},$$

it is easy to see that

$$\int_{[a,b]} l_{\mathcal{P}_n}\ dm = L(f,\mathcal{P}_n) \qquad \text{and} \qquad \int_{[a,b]} u_{\mathcal{P}_n}\ dm = U(f,\mathcal{P}_n).$$

Now because each partition is a refinement of the previous one, $l_{\mathcal{P}_n}$ is a monotone increasing sequence and $u_{\mathcal{P}_n}$ is a monotone decreasing sequence, and they are bounded above and below by the bounds of f. Hence both sequences converge pointwise to functions l and u, respectively. By the dominated convergence theorem,

$$\int_{[a,b]} l\ dm = \lim_{n\to\infty} L(f,\mathcal{P}_n) = \int_a^b f(x)\ dx$$

and

$$\int_{[a,b]} u\ dm = \lim_{n\to\infty} U(f,\mathcal{P}_n) = \int_a^b f(x)\ dx.$$

Hence $\int_{[a,b]} u - l\ dm = 0$, so $u = l$ almost everywhere, and since $l \le f \le u$, $f = u = l$ almost everywhere as well. Hence f is Lebesgue integrable with

$$\int_{[a,b]} f\ dm = \int_{[a,b]} l\ dm = \int_a^b f(x)\ dx.$$

■

Indeed, one can show that the Riemann integrable functions are precisely those which are discontinuous on sets of measure zero.

Note also that these integrals are all proper integrals. We can easily extend the result to absolutely convergent improper integrals.

Theorem 1.10.2
Let f be an function on $[a,b)$ (where b may be $+\infty$) which has an absolutely convergent improper Riemann integral. Then f is Lebesgue integrable and

$$\int_{[a,b)} f\ dm = \int_a^b f(x)\ dx.$$

Proof:

Let t_n be any increasing sequence which converges to b. We then define $f_n = \chi_{[a,t_n]} f$. Then $|f_n|$ is a monotone increasing sequence of functions on $[a,b)$ which converges pointwise to $|f|$, and hence by the Monotone Convergence Theorem,

$$\int_{[a,b)} |f|\ dm = \lim_{n\to\infty} \int_{[a,b)} |f_n|\ dm = \lim_{n\to\infty} \int_a^{t_n} |f(x)|\ dx = \int_a^b |f(x)|\ dx < \infty.$$

Hence $f \in \mathcal{L}^1([a,b))$. But then f_n converges pointwise to f on $[a,b)$, and $|f_n| \le |f|$, so the Dominated Convergence Theorem tells us that f is Lebesgue integrable, and

$$\int_{[a,b)} f\ dm = \lim_{n\to\infty} \int_{[a,b)} f_n\ dm = \lim_{n\to\infty} \int_a^{t_n} f(x)\ dx = \int_a^b f(x)\ dx.$$

■

The proof for improper integrals on $(a, b]$ is similar.

Note that there are functions on $\mathbb{R}$ for which the improper Riemann integral is defined, but which are not Lebesgue integrable. This might seem like a flaw in the Lebesgue theory, except that we need to apply this sort of restriction to Riemann integrals as soon as we consider improper Riemann integrals in $\mathbb{R}^n$. In other words, the existence of such functions is more an accident of the topology of $\mathbb{R}$ and pales in significance compared to the power of the Lebesgue integral.

To conclude, the Lebesgue integral is more powerful in almost every way than the Riemann integral. Compared to the Riemann integral, by using the Lebesgue integral you gain:

- The ability to integrate a function over any measurable set, not just intervals.
- The ability to integrate a much larger class of functions.
- The Monotone and Dominated Convergence Theorems, which are powerful tools for calculating integrals.
- Agreement with the Riemann integral (and so the Fundamental Theorem of Calculus) in almost every case.
- A theory of integration which is based on simple axiomatic constructions which can easily be generalized to an abstract setting.

We will pursue this last advantage in Chapter 3, but to gain a proper understanding of this general setting, we need to cover some point-set topology first.

Exercises

1.10.1. Show that $\int_{[0,1]} x\ dm(x) = 1/2$.

1.10.2. (†) Given any interval $I = (a - R, a + R)$, consider the power series

$$f(x) = \sum_{n=1}^{\infty} c_n (x-a)^n.$$

Show that if

$$\sum_{n=1}^{\infty} |c_n| R^n < \infty,$$

then $f(x)$ is Lebesgue integrable on I.

1.10.3. Let f be a continuous, differentiable, monotone increasing function on $\mathbb{R}$. For any interval I with endpoints a and b, define

Note: *This exercise gives a fundamental theorem of ca... of sorts. It also foreshadow... Radon-Nikodym derivative... Chapter 3.*

$$m_f(I) = f(b) - f(a),$$

and for $E \in \mathcal{E}$, define $m_f(E)$ in the obvious way.

Show that

$$m_f^*(A) = \inf\left\{\sum_{k=1}^{\infty} m_f(A_k) : A_k \in \mathcal{E}, A \subseteq \bigcup_{k=1}^{\infty} A_k\right\}$$

is an outer measure.

Show that every Lebesgue measurable set is m_f^*-measurable, and let m_f be the restriction of m_f to the Lebesgue measurable sets.

Show that

$$m_f(A) = \mu_{f'}(A) = \int_A f' \, dm.$$

1.10.4. Let

$$f(x) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n} \chi_{[n,n+1)}.$$

Show that f is Riemann integrable on $[0,\infty)$, but is not Lebesgue integrable on the same integral.

1.11 Technical Details

In this section we detail some of the proofs which were glossed over earlier in the interest of a cleaner exposition.

Example 1.1.2

Recall from Example 1.1.2 of Section 1.1, that we had an equivalence relation defined by $x \sim y$ if $x - y$ is a rational number. We let N be a set of equivalence class representatives for this equivalence relation on $[0,1)$, and we defined sets

$$N_q = \{x + q : x \in N, 0 \leq x < 1 - q\} \cup \{x - (1-q) : x \in N, 1 - q \leq x < 1\},$$

for every rational number $q \in Q = [0,1) \cap \mathbb{Q}$.

The following claim was made, and although the proof is easy, it was too long to include in the eaxample.

Claim 1.11.1
Let q, $r \in Q$ with $q \neq r$. Then N_q and N_r are disjoint.

Proof:

Assume otherwise, so that $q \neq r$ and there is some $y \in N_q cap N_r$. Without loss of generality, we may assume that $q > r$. There are 3 cases that we need to consider: $y \geq q$, $q > y \geq r$ and $r > y$.

In the first case, let $x_1 = y - q$ and $x_2 = y - r$. Then we have that $0 \leq x_1 = y - q < 1 - q$ and $y = x_1 + q$, from which we conclude that $x_1 \in N$. Similarly we conclude that $x_2 \in N$. Also note that $x_2 - x_1 = (y - r) - (y - q) = q - r$ is rational, so $x_1 \sim x_2$.

In the second case we let $x_1 = y + (1 - q)$ and $x_2 = y - r$. Using the same argument as before, we have $x_2 \in N$. We also have $x_1 \in N$, since $1-q = x_1 - y \leq x_1 < 1$ and $x_1 - (1 - q) = y$. Also note that $x_2 - x_1 = (y - r) - (y + (1 - q)) = q - r - 1$ is rational, so $x_1 \sim x_2$.

In the third case, we let $x_1 = y+(1-q)$ and $x_2 = y+(1-r)$. Then the same argument that we used in ther previous case for x_1 applies to both x_1 and x_2, so both are elements of N. Also note that $x_2 - x_1 = (y+(1-r))-(y+(1-q)) = q-r$ is rational, so $x_1 \sim x_2$.

In other words, in each of the 3 cases, x_1 and x_2 are elements of N with $x_1 \sim x_2$. But our construction of N assumes that there is precisely one element from each equivalence class in N, so this is a contradiction.

Hence N_q and N_r are disjoint. ■

Lemma 1.2.5

Recall from Section 1.2 that the set function $m : \mathcal{E} \to [0, \infty]$ was defined by

$$m(E) = \sum_{k=1}^{n} m(I_k),$$

for $E \in \mathcal{E}$ where $E = I_1 \cup I_2 \cup \ldots \cup I_n$ is a disjoint union of intervals, and

$$m(I_k) = b_k - a_k$$

where the left and right endpoints of I_k are a_k and b_k (respectively). Note that the order of the intervals in the union does not affect the value of the sum, so without loss of generality, one can assume that the endpoints of the interval I_k satisfy a_k and b_k with $a_1 \leq b_1 \leq a_2 \leq b_2 \leq \ldots \leq a_n \leq b_n$.

Recall also, that E has a disjoint union of intervals which is minimal in the sense that the union of no pair of intervals is again an interval.

Proof (Lemma 1.2.5):

Let I be any interval with end-points a and b, and J_k, $k = 1, \ldots, n$ be intervals with end points $J_k = [a_k, b_k$ such that $a_1 \leq b_1 \leq a_2 \leq b_2 \leq \ldots \leq a_n \leq b_n$. If $I = J_1 \cup J_2 \cup \ldots \cup J_n$, then it must be the case that $a_{k+1} = b_k$ for $k = 1, \ldots, n - 1$, and $a_1 = a$, $b_n = b$, so

$$\sum_{k=1}^{n} m(J_k) = \sum_{k=1}^{n} (b_k - a_k) = b - a = m(I).$$

So we have just shown the intuitively obvious fact that if you partition an interval, the volume of the sub-intervals sums to the volume of the interval. Hence m is well-defined for intervals.

Letting $E = I_1 \cup I_2 \cup \ldots \cup I_n$ be the minimal disjoint union with I_k having end points a_k and b_k with $a_1 \leq b_1 \leq a_2 \leq b_2 \leq \ldots \leq a_n \leq b_n$, and $E = J_1 \cup J_2 \cup \ldots \cup J_p$ an arbitrary disjoint union, where each J_l has endpoints c_l and d_l and $c_1 \leq d_1 \leq c_2 \leq d_2 \leq \ldots \leq c_p \leq d_p$. Then there are numbers $0 = l_0 < l_1 < l_2 < \ldots < l_n = p$, so that $I_k = J_{l_{k-1}+1} \cup J_{l_{k-1}+2} \cup \ldots \cup J_{l_k}$. Therefore

$$\sum_{l=1}^{p} m(J_l) = \sum_{k=1}^{n} \sum_{l=l_{k-1}+1}^{l_k} m(J_l) = \sum_{k=1}^{n} m(I_k) = m(E).$$

Hence m is well-defined on $\mathcal{E}$. ■

Lemma 1.4.2

Proof (Lemma 1.4.2):
We know that for any n,

$$m(A) \geq m\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{\infty} m(A_k),$$

and taking limits we then have

$$m(A) \geq \sum_{k=1}^{\infty} m(A_k).$$

Now assume that A is a bounded interval, and let the endpoints of A be a and b and the endpoints of A_k be a_k and b_k. Then given any $\varepsilon > 0$ the open intervals, let $\varepsilon_k = 2^{-i-1}\varepsilon$ $B_k = (a_k - \varepsilon_k, b_k + \varepsilon_k)$ cover the compact set $[a + \varepsilon/2, b - \varepsilon/2]$. Hence we can find a finite subcover by open intervals B_{k_j}, $j = 1, \ldots, n$, and for which there is some N such that $i_j \leq N$ for all j. Moreover, we can choose the B_{k_j} is such a way that $a_{k_1} < a_{k_2} < a_{k_3} < \ldots < a_{k_n}$ and

$b_{k_j} + \varepsilon_{k_j} \in (a_{k_{j+1}} - \varepsilon_{k_{j+1}}, b_{k_j} + \varepsilon_{k_j})$. Now

$$\begin{aligned}
m(A) &\leq m([a + \varepsilon/2, b - \varepsilon/2]) + \varepsilon \\
&\leq m((a_{k_1} - \varepsilon_{k_1}, b_{k_n} + \varepsilon_{k_n})) + \varepsilon \\
&= b_{k_n} + \varepsilon_{k_n} - (a_{k_1} - \varepsilon_{k_1}) + \varepsilon \\
&= b_{k_n} + \varepsilon_{k_n} - (a_{k_n} - \varepsilon_{k_n}) + \sum_{j=1}^{n-1} (a_{k_{j+1}} - \varepsilon_{k_{j+1}}) - (a_{k_j} - \varepsilon_{k_j})) + \varepsilon \\
&\leq b_{k_n} - (a_{k_n} - \varepsilon_{k_n}) + \sum_{j=1}^{n-1} (b_{k_j} + \varepsilon_{k_j}) - (a_{k_j} - \varepsilon_{k_j})) + \varepsilon \\
&\leq \sum_{j=1}^{n} (b_{k_j} - a_{k_j}) + 2 \sum_{j=1}^{n} \varepsilon_{k_j} + \varepsilon \\
&\leq \sum_{j=1}^{n} m(A_{k_j}) + 2\varepsilon \\
&\leq \sum_{j=1}^{\infty} m(A_{k_j}) + 2\varepsilon.
\end{aligned}$$

Since ε can be any number, we are done in the case of a bounded interval. By restricting to an unbounded interval to $[-M, M]$ for some M, and using the above result, we get that

$$m(A \cap [-M, M]) \leq \sum_{j=1}^{\infty} m(A_{k_j})$$

for all M, and so letting $M \to \infty$, we get

$$m(A) \leq \sum_{j=1}^{\infty} m(A_{k_j}).$$

■

Proposition 1.11.2

Proposition 1.11.2
If $U \subset \mathbb{R}$ is an open set, then U is a countable union of disjoint open intervals.

Proof:
Any open set in $\mathbb{R}$ is a union of open intervals, say $A = \bigcup_{\alpha \in J} I_\alpha$. More concretely, bu the definition of an open set, for every $xinU$, we have some $\varepsilon_x > 0$ such that $N_{\varepsilon_x}(x) = (x - \varepsilon_x, x + \varepsilon_x) \subseteq U$, so that

$$U = \bigcup_{x \in U} (x - \varepsilon_x, x + \varepsilon_x).$$

However, without loss of generality, we can assume that they are disjoint, for if two open intervals are not disjoint then their union is an open interval. Now for each interval I_α we can find a rational number $q \in I_\alpha$, and since the intervals are disjoint, this number is only in the one interval. So we have an injective function $J \to \mathbb{Q}$ defined by this association. Hence J is countable and so A is a countable union of disjoint open intervals. ■

Related to this is the following fact:

Corollary 1.11.3
If $U \subset \mathbb{R}$ is an open set, then U is a countable union of bounded open intervals.

Proof:
If the union from the previous proposition does not contain an interval of the form (a, ∞) or $(-infty, b)$, then we are done. If not, we simlpy replace the interval (a, ∞) by the countable set of bounded intervals

$$(a, 2n+1), (2n, 2n+2), (2n+1, 2n+3), \ldots, (2n+k, 2n+k+2), \ldots$$

where $n \in \mathbb{Z}$ such that $2n > a$. While these intervals are no longer disjoint, they are bounded. We adapt $(-\infty, b)$ in a similar manner. ■

Proposition 1.7.1

Proof:
Let φ and ψ have standard representations

$$\varphi = \sum_{k=1}^{n} c_k \chi_{E_k} \qquad \text{and} \qquad \psi = \sum_{k=1}^{m} b_k \chi_{F_k}$$

respectively.

(i, part 1) We deal with the case where the sets G_k are pairwise disjoint first. In this case, each $a_k = c_j$ for some j, and we have a disjoint unions $E_j = G_{k_1} \cup G_{k_2} \cup \ldots \cup G_{k_{l_j}}$, and each G_k is in exactly one E_j. Therefore

$$\begin{aligned} \sum_{k=1}^{l} a_k m(G_k \cap E) &= \sum_{j=1}^{n} \sum_{p=1}^{l_j} c_j m(G_{k_p} \cap E) \\ &= \sum_{j=1}^{n} c_j m(E_j \cap E) = \int_E \varphi \, dm. \end{aligned}$$

We prove the general case after we have completed (iii).

(ii) If $c = 0$ the result is trivial. If $c \neq 0$ it is easy to see that, the standard representation of $c\varphi$ is $c\varphi = \sum_{k=1}^{n} c c_k \chi_{E_k}$, and so

$$\int_E c\varphi \, dm = \sum_{k=1}^{n} c c_k m(E_k \cap E) = c \sum_{k=1}^{n} c_k m(E_k \cap E) = c \int_E \varphi \, dm.$$

(iii) We can represent $\varphi + \psi$ by

$$\varphi + \psi = \sum_{j=1}^{n} \sum_{k=1}^{m} (c_j + b_k) \chi_{E_j \cap F_k}.$$

Since $E_j \cap F_k$ are pairwise disjoint, (i) tells us that

$$\begin{aligned}
\int_E \varphi + \psi \; dm &= \sum_{j=1}^{n} \sum_{k=1}^{m} (c_j + b_k) m(E_j \cap F_k \cap E) \\
&= \sum_{j=1}^{n} c_j \sum_{k=1}^{m} m(E_j \cap F_k \cap E) + \sum_{k=1}^{m} b_k \sum_{j=1}^{n} m(E_j \cap F_k \cap E) \\
&= \sum_{j=1}^{n} c_j m(E_j \cap E) + \sum_{k=1}^{m} b_k m(F_k \cap E) \\
&= \int_E \varphi \; dm + \int_E \psi \; dm.
\end{aligned}$$

(i, part 2) If G_k are not disjoint, by (ii) and (iii) used repeatedly,

$$\int_E \varphi \; dm = \sum_{k=1}^{l} a_k \int_E \chi_{G_k} \; dm = \sum_{k=1}^{l} a_k m(G_k \cap E).$$

(iv) We have that $c_j \leq b_k$ on $E_j \cap F_k$, so

$$\begin{aligned}
\sum_{j=1}^{n} c_j m(E_j \cap E) &= \sum_{j=1}^{n} \sum_{k=1}^{m} c_j m(E_j \cap F_k \cap E) \\
&\leq \sum_{j=1}^{n} \sum_{k=1}^{m} b_k m(E_j \cap F_k \cap E) = \sum_{k=1}^{m} b_k m(F_k \cap E),
\end{aligned}$$

and hence

$$\int_E \varphi \; dm \leq \int_E \varphi \; dm.$$

(v) we note that any subset of a null set is null, and so

$$\int_E \varphi \; dm = \sum_{k=1}^{n} c_k m(E_k \cap E) = 0.$$

∎

Sample Exam Questions

These sample questions are designed to give you an idea of what questions may be asked on the midterm.

1. State the Monotone Convergence Theorem.
2. State the definition of a Lebesgue measurable function.
3. State the definition of a σ-algebra.
4. If $\mathcal{A}$ is a σ-algebra, show that given any countable collection of sets $A_n \in \mathcal{A}$, $n = 1, 2, \ldots$, then
$$\bigcap_{n=1}^{\infty} A_n \in \mathcal{A}.$$
5. Consider the set function $c : \mathcal{P}(\mathbb{R}) \to [0, \infty]$ defined by
$$c(X) = \begin{cases} |X| & \text{if } X \text{ is finite} \\ \infty & \text{if } X \text{ is infinite.} \end{cases}$$
Show that c is a measure.
6. Show that if $f \in \mathcal{L}(E)$, then $cf \in \mathcal{L}(E)$.
7. Give an example of a Lebsgue measurable set X which contains no intervals, but for which $m(X) = 1/3$.
8. Find
$$\int_{[-1,2]} x \; dm(x),$$
carefully stating the results you use.
9. Let f be a Riemann integrable function on $[a, b]$. Prove that
$$\int_a^b f(x) \; dx \leq \int_{[a,b]} f \; dm.$$
10. Give an example of a sequence of functions f_n in $\mathcal{L}^+([0, 1])$ which converge pointwise to 0, but for which
$$\lim_{n\to\infty} \int_{[0,1]} f_n \; dm \neq 0.$$
Verify that your example is valid.
11. Let E be a measurable set with $m(E) < \infty$. Show that if f_n is a sequence in $\mathcal{L}(E)$, and f_n converges to some $f \in L^1(E)$ uniformly, then
$$\int_E f \; dm = \lim_{n\to\infty} \int_E f_n \; dm.$$

Chapter 2

General Topology

2.1 Topological Spaces

Topology generalises the notions of convergence and continuity which you have seen in the context of $\mathbb{R}^n$ and (hopefully) more general metric spaces. In these settings convergence is defined in terms of the distance between points, and continuity of functions by the preservation of limiting behaviour.

In general topology, we simply start with the notion of an open set.

Definition 2.1.1
A ***topological space*** *is a pair* (X, τ) *where* X *is any set and* τ *is a family of subsets of* X *which satisfy:*

1. X *and* $\emptyset$ *are in* τ.

2. *If* A_α *is a family of sets in* τ *indexed by* $\alpha \in I$, *then*

$$\bigcup_{\alpha \in I} A_\alpha \in \tau.$$

3. *If* $A_1, A_2, \ldots, A_n$ *is a finite family of sets in* τ, *then*

$$\bigcap_{k=1}^{n} A_k \in \tau.$$

The family of sets τ *is called a* ***topology*** *on* X. *The sets in* τ *are called* ***open sets***. *Sets whose complements are in* τ *are called* ***closed sets***

A concise way of describing the axioms of a topology is that a topology contains $\emptyset$ and X, and is closed under arbitrary unions and finite intersections. These definitions are reminsicent of the definitions of a σ-algebra from Chapter 1, but the details are distinctly different.

From the definition of topology and DeMorgan's laws, it is easy to see that closed sets have the following properties:

1. X and $\emptyset$ are closed.
2. If A_α is a family of closed sets indexed by $\alpha \in I$, then

$$\bigcap_{\alpha\in I} A_\alpha$$

is open.

3. If $A_1, A_2, \ldots, A_n$ is a finite family of closed sets, then

$$\bigcup_{k=1}^{n} A_k$$

is open.

Indeed, it is possible to define topological theory starting from closed sets and these three facts as axioms, but it is customary to use open sets as the principal objects.

Example 2.1.1 (Metric Space Topology)

Note: *A metric is axiomatizes the intuitive notion of "distance" between 2 points in a set.*

Let X be any set. A ***metric*** on X is a function $d : X \times X \to [0, \infty)$ such that $d(x, y) = 0$ if and only if $x = y$; and $d(x, z) \le d(x, y) + d(y, z)$ for all x, y, $z \in X$ (this property is called the ***triangle inequality***). We call the pair (X, d) a ***metric space***.

Let (X, d) be a metric space. An ***open ball*** in (X, d) is a set of the form

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

We say a set $U \subseteq X$ is open if for every $x \in U$, there is some $\varepsilon > 0$ so that $B(x, \varepsilon) \subseteq U$.

It is straightforward to see that this is a toplogy. The empty set and the whole metric space trivially satisfy the condition. If we have an arbitrary collection of such sets U_α for $\alpha \in I$, then for any $x \in \bigcup_{\alpha\in I} U_\alpha$, $x \in U_{\alpha'}$ for some particular $\alpha' \in I$, and so there is some $\varepsilon > 0$ with $B(x, \varepsilon) \subseteq U_{\alpha'}$, and so

$$B(x, \varepsilon) \subseteq \bigcup_{\alpha\in I} U_\alpha.$$

Finally, given open sets $U_1, U_2, \ldots U_n$, for every x in the intersection of these sets we can find $\varepsilon_1, \ldots, \varepsilon_n$ so that $B(x, \varepsilon_k) \subseteq U_k$ for $k = 1, \ldots, n$. But letting $\varepsilon = \min_k \varepsilon_k > 0$, we have $B(x, \varepsilon) \subseteq B(x, \varepsilon_k) \subseteq U_k$ for $k = 1, \ldots, n$. Hence

$$B(x, \varepsilon) \subseteq \bigcap_{k=1}^{n} U_k.$$

Equivalently, a set U is open if and only if it is a union of open balls. To see this, we first note that because of the triangle inequality, open balls are in fact open sets. For any $x \in X$ and $r > 0$, given any $y \in B(x,r)$ we can let $0 < \varepsilon < r - d(x,y)$, and then given any $z \in B(y,\varepsilon)$ we have

$$d(x,z) \leq d(x,y) + d(y,z) < d(x,y) + r - d(x,y) = r,$$

so $B(y,\varepsilon) \subseteq B(x,r)$ and so $B(x,r)$ is open. Therefore any union of open balls is an open set.

On the other hand, given any open set U. By definition, for every $x \in U$, there is a $\varepsilon_x > 0$ with $B(x,\varepsilon_x) \subseteq U$. But then

$$U = \bigcup_{x \in U} B(x,\varepsilon_x),$$

so every open set is a union of open balls.

We will denote the collection of all these open sets τ_d, and we call this topology the ***metric space topology***.

The standard topology on $\mathbb{R}^n$ is the metric topology where $d(x,y)$ is the standard Euclidean distance between the two points.

If V is a vector space with a norm $\|\cdot\|$ then $d(u,v) = \|u - v\|$ is a metric. ◇

Example 2.1.2

The power set of a set X trivially satisfies the axioms of a topology. In particular, every set of the form $\{x\}$ for some $x \in X$ is open, so this topology is called the ***discrete topology*** on X. It is also the metric topology given by the discrete metric

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

◇

Not all topologies come from metrics.

Example 2.1.3

The family of subsets $\tau = \{\emptyset, X\}$ of X also satisfies the axioms of a topology, and is called the ***trivial topology*** on X. ◇

Example 2.1.4

Consider sets in $\mathbb{R}^n$ of the form

Note: *The Zariski topology plays an important role in algebraic geometry.*

$$\bigcup_{\alpha \in I} p_\alpha^{-1}(\mathbb{R} \setminus \{0\})$$

where p_α for α in an index set I are all real polynomials in n variables. It can be shown that sets of this form are a topology on $\mathbb{R}^n$ and that this topology is not the collection of open sets for any metric. This topology is called the ***Zariski topology***. ◇

Example 2.1.5

Let X be any set, let (Y, τ) be a topological space, and let $F(X, Y)$ be the set of all functions from X to Y. A set of functions is open if it is of the form

$$\bigcup_{\alpha \in I} \{f \in F(X, Y) : f(x_\alpha) \in U_\alpha\}$$

where $x_\alpha \in X$ and $U_\alpha \in \tau$. $\diamond$

Example 2.1.6

If (X, τ) is a topological space and $Y \subseteq X$, then

$$\tau_Y = \{A \cap Y : A \in \tau\}$$

is a topology on Y called the ***relative topology***. $\diamond$

Clearly there may be more than one topology on a set. If τ_1 and τ_2 are two topologies on a set X with $\tau_1 \subset \tau_2$, then we say τ_1 is ***weaker*** or ***coarser*** than τ_2, and τ_2 is ***stronger*** or ***finer*** than τ_1. If τ_1 and τ_2 are both topologies on X, then so is $\tau_1 \cap \tau_2$. Indeed, if $\mathcal{T}$ is any collection of topologies on a set X, then the intersection of all these topologies,

$$\bigcap_{\tau \in \mathcal{T}} \tau,$$

is also a topology.

If A is any subset of X, the ***interior*** of A, denoted A° is the union of all open sets contained in A; while the ***closure*** of A, denoted $\overline{A}$, is the intersection of all closed sets containing A. These are, respectively, the largest open set contained in A and the smallest closed set containing A. Clearly, if A is open, $A^\circ = A$, and if A is closed $\overline{A} = A$.

If $x \in X$, then a ***neighbourhood*** of x is any set N which contains some open set U containing x, or equivalently $x \in N^\circ$. If N is open, then we say that it is an ***open neighbourhood*** of x.

Open neighbourhoods give a useful way of determining whether a set is open or not.

Lemma 2.1.1

Let (X, τ) be a topological space. Then a set $A \subseteq X$ is open if and only if for every $x \in A$ we can find an open neighbourhood U of x so that $U \subseteq A$.

Proof:

If A is open, then it is an open neighbourhood of any of its elements, so that direction is trivial.

On the other hand, given every $x \in A$, we can find an open neighbourhood of x, $U_x \subseteq A$. But then

$$A = \bigcup_{x \in A} U_x,$$

which is open. ■

If $\overline{A} = X$ we say A is ***dense***. A topological space (X, d) which has a countable dense set Q is said to be ***separable***.

A set A for which the interior of the closure is empty is said to be ***nowhere dense***.

If one can find two disjoint, closed sets E and F, with $X = E \cup F$, then (X, τ) is said to be ***disconnected***. If a set $Y \subseteq X$ is disconnected in the relative topology, then Y is said to be a ***disconnected set*** in X. A space or set which is not disconnected is ***connected***.

If given any distinct pair of points x and $y \in X$, there are disjoint closed sets E and F with $x \in E$ and $y \in F$ and $X = E \cup F$, then (X, τ) is said to be ***totally disconnected***.

Example 2.1.7

If C is the Cantor middle third set with the relative topology in $\mathbb{R}$, then C is totally disconnected. ◇

Exercises

2.1.1. Let $\tau = \{(a, \infty) : a \in [-\infty, \infty)\}$. Show that this is a topology on $\mathbb{R}$ which is different from the usual topology.

2.1.2. Show that $(A^\circ)^c = \overline{A^c}$ and that $(\overline{A})^c = (A^c)^\circ$.

2.1.3. Let A be a closed, nowhere dense set. Show that A^c is an open, dense set.

2.1.4. Show that a set with the discrete topology is totally disconnected.

2.2 Continuous Functions

Recall from undergraduate real analysis that one way to tell if a function was continuous was that the inverse image of open sets were open. In topology, since open sets play a fundamental role, it is this that we base the definition on, rather than one of the other equivalent versions of continuity.

Definition 2.2.1

*If (X, τ) and (Y, σ) are topological spaces then a function $f : X \to Y$ is **continuous** if given any open set $U \in Y$, the set $f^{-1}(U)$ is open in X. f is **continuous on a subset** A of X if its restriction to A is continuous, where A is given the relative topology.*

*f is **continuous at a point** $x \in X$ if given any neighbourhood U of $f(x)$, $f^{-1}(U)$ is a neighbourhood of x.*

It may seem counterintuitive that we want to look at inverse images of open sets, rather than the images of open sets. Functions which have the property that $f(U)$ is open in Y whenever U is open in X are called ***open functions***, and they may or not be continuous. Open functions have some importance in functional analysis.

The following facts about continuity are not difficult to verify:

Proposition 2.2.1
If (X, τ) and (Y, σ) are topological spaces and $f : X \to Y$, then:

(i) f is continuous on a subset A of X if and only if it is continuous at every point in A.

(ii) f is continuous if and only if it is continuous at every point in X.

Part (iii) of this proposition tells us that you only need to check the inverse images of sets in a base or sub-base to confirm continuity.

Example 2.2.1
If (X, τ) is a topological space, and (Y, d) is a metric space, then you only need to verify that the inverse images of open balls are open.

In this setting, (v) is the familiar ε-δ definition of continuity at a point when combined with the neighbourhood bases of open balls $\{B(f(x), \varepsilon) : \varepsilon > 0\}$ and $\{B(x, \delta) : \delta > 0\}$.

We have $d(f(x), f(y)) < \varepsilon$ for all y with $d(x, y) < \delta$, if and only $f(y) \in B(f(x), \varepsilon)$ for all $y \in B(x, \delta)$, if and only if $B(x, \delta) \subseteq f^{-1}(B(f(x), \varepsilon))$. So if for every set $B(f(x), \varepsilon)$ in the first nieghbourhood base, we have a set $B(x, \delta) \subseteq f^{-1}(B(f(x), \varepsilon))$ in the second neighbourhood base, then for every $\varepsilon > 0$, there is a $\delta > 0$ such that $d(f(x), f(y)) < \varepsilon$ for all y with $d(x, y) < \delta$, and vice-versa. $\diamond$

If we have three topological spaces (X, τ), (Y, σ) and (Z, ω), and $f : X \to Y$ is continuous and $g : Y \to Z$ is continuous, then $g \circ f : X \to Z$ is continuous.

If $f : X \to Y$ is bijective and both it and its inverse are continuous, then we say that f is a ***homeomorphism***, and that the two topological spaces are ***homeomorphic***. Two topological spaces which are homeomorphic are equivalent from the point of view of topology.

Example 2.2.2
Any open interval is homeomorphic with any other open interval. If $I = (a, b)$ and $J = (c, d)$, are two bounded intervals, then

$$f(x) = \frac{d - c}{b - a}(x - a) + c$$

is a homeomorphism which demonstrates this. One can find ecplicit examples for unbounded intervals with similar ease. $\diamond$

Example 2.2.3

A sphere, torus and coffee mug are given the a relative topology in $\mathbb{R}^3$. The torus and the coffe mug are homeomorphic, but the sphere is not homeomorphic to either one. Showing this requires a bit of work. $\diamond$

Continuous functions from a topological space (X, τ) to $\mathbb{R}$ or $\mathbb{C}$ with their usual topologies play an important role in analysis. We let $C(X, Y)$ be the set of all continuous functions from X to Y. We define $C_b(X, \mathbb{R})$ and $C_b(X, \mathbb{C})$ to be the set of all continuous, bounded functions from X to $\mathbb{R}$ and $\mathbb{C}$ respectively, and we will often simply write $C(X)$ and $C_b(X)$ when the codomain is implict, or where either can be used (usually we want the codomain to be $\mathbb{C}$ for full generality).

Exercises

2.2.1. Show that if (X, τ) is a topological space and $Y \subseteq X$, then the inclusion map $\iota : Y \hookrightarrow X$ is continuous when Y is given the relative topology.

2.2.2. Let τ be the usual topology on $\mathbb{R}$ and σ be the topology of Exercise 2.1.1. Show that a function $f : \mathbb{R} \to \mathbb{R}$ is continuous from $(\mathbb{R}, \tau)$ to $(\mathbb{R}, \sigma)$ if and only if it is lower semicontinuous.

Which topology should you use to get upper semicontinuous functions?

2.2.3. Show that the functions $+ : \mathbb{R}^2 \to \mathbb{R}$ and $\cdot : \mathbb{R}^2 \to \mathbb{R}$ are both continuous in the usual topology on $\mathbb{R}^2$ and $\mathbb{R}$.

Show that the function $x \mapsto x^{-1}$ is a homeomorphism on $\mathbb{R} \setminus \{0\}$ with the relative topology.

2.2.4. Using the previous exercise, show that if (X, τ) is any topological space, and f, $g \in C(X, \mathbb{R})$, then $f + g$ and fg are in $C(X, \mathbb{R})$.

Hint: *show* $u(x) = (f(x), g(x))$ *is continuous from* X *to* $\mathbb{R}^2$.

2.2.5. If (X, τ) is a connected topological space, (Y, σ) is any topological space, and $f \in C(X, Y)$, show that $f(X)$ is connected.

2.2.6. An ***arc***, or ***curve***, γ in a topological space (X, τ) is a continuous function $\gamma : [0, 1] \to X$ (where $[0, 1]$ has the usual topology). We say that a subset A of a topological space (X, τ) is ***path connected*** or ***arcwise connected*** if given any x and y in A, there is an arc γ such that $\gamma(0) = x$ and $\gamma(1) = y$, and $\gamma([0, 1]) \subseteq A$ (ie. there is a continuous curve inside A connecting the two points). If the space as a whole is path connected, then we say it is a ***path connected topological space***.

Show that $\mathbb{R}^n$ is path connected in the usual topology.

Show that if A is path connected, then it is connected.

Show that the set $A = \{(x, y) : x > 0, y = \sin x^{-1}\} \cup \{(0, 0)\}$ is connected as a subset of $\mathbb{R}^2$ with the usual topology, but is not path connected.

2.2.7. A ***loop*** γ in a topological space (X, τ) is an arc such that $\gamma(0) = \gamma(1)$, and we call this starting and finishing point the base point of the loop. Fix a point $x_0 \in X$, and assume that X is path connected. Two loops γ_0 and γ_1 with base point x_0 are ***homotopy equivalent***, written $\gamma_0 \sim \gamma_1$ if there are loops γ_t with base point x_0, for $0 < t < 1$, such that $t \mapsto \gamma_t(a)$ is continuous from $[0, 1]$ to X for all $a \in [0, 1]$ (in other words, we can *continuously deform* γ_0 to γ_1.

Note: *this example is the starting point of algebraic topology.*

Show that if (X, τ) and (Y, σ) are homeomorphic topological spaces, with homeomorphism $\phi : X \to Y$, then loops γ_0 and γ_1 in X are homotopy equivalent if and only if loops $\Phi \circ \gamma_0$ and $\Phi \circ \gamma_1$ are homotopy equivalent.

Show that every loop on a sphere is homotopy equivalent to a constant loop $\gamma(s) = x_0$. Show that the outer circumference of a torus is not homotopy equivalent to a constant loop. Conclude that a sphere and a torus are not homeomorphic.

2.2.8. A topological group G is a group together with a topology τ for which the binary operation is continuous from $G^2 \to G$, where G^2 has the product topology, and the inverse map is continuous from $G \to G$.

Show that G is a topological group if and only if the map

$$(x, y) \mapsto xy^{-1}$$

is continuous from $G^2 \to G$.

Show that $\mathcal{T} = \{z \in \mathbb{C} : |z| = 1\}$ is a topolgical group under multiplication with the relative topology.

Show that any group is a topological group when given the discrete topology.

Show that if $U \subseteq G$ is an open set, then $g \cdot U$ is an open set for any $g \in G$.

2.3 Nets

Topology can also be used to define convergence properties for sequences. We say that a sequence $(x_n)_{n=1}^{\infty}$ ***eventually*** satisfies a property P if there is some n_0 such that P is true for all x_n with $n > n_0$.

Definition 2.3.1
Let $(x_n)_{n=1}^{\infty}$ be a sequence in X and τ a topology on X. $(x_n)_{n=1}^{\infty}$ ***converges*** *to x in (X, τ) if and only if given any neighbourhood U of x, $(x_n)_{n=1}^{\infty}$ is eventually in U.*

In some sense it is convergence which we really care about in analysis, but sometimes (generally in spaces which are "big") we need a more general notion of convergence than sequences give us.

Recall from the appendix that a directed set is a set together with a partial order (ie. a reflexive, transitive relation) for which every pair of elements has an upper bound. The natural numbers with the usual order $\leq$ is a directed set, and we can regard a sequence in X as a function from $\mathbb{N}$ to X.

Definition 2.3.2
*A **net** in a set X is a function $\lambda \mapsto x_\lambda$ from a directed set Λ to X. We will usually denote such a net by $(x_\lambda)_{\lambda\in\Lambda}$, and say (x_λ) is **indexed** by Λ.*

$\mathbb{N}$ with the relation $\leq$ is a directed set, so all sequences are nets. However there are other notions of convergence and limits that you have encountered, even as far back as your calculus courses.

Example 2.3.1
The set of all partitions $\mathcal{P}$ of an interval $[a, b]$ with the relation $\mathcal{P} \preceq \mathcal{Q}$ iff $\|\mathcal{P}\| \leq \|\mathcal{Q}\|$, where

$$\|\mathcal{P}\| = \max_{k=1,\ldots,n}(p_k - p_{k-1})$$

is the mesh of $\mathcal{P}$, is a directed set.

If $f : [a, b] \to \mathbb{R}$ is a function, then

$$\ell_{\mathcal{P}} = \sum_{k=1}^{n} (\inf\{f(x) : x \in [p_{k-1}, p_k]\})\,(p_k - p_{k-1})$$

is a net of lower Riemann sums which should be familiar from Riemann integration. ◇

Example 2.3.2
If x is any point in a topological space (X, τ), then the set $\mathcal{N}$ of all open neighbourhoods of x, with the relation by $U \preceq V$ iff $U \supseteq V$ (ie. *reverse inclusion*) is a directed set. ◇

Example 2.3.3
If Λ and Γ are two directed sets, then the set $\Lambda \times \Gamma$ is directed by the relation $(\lambda, \gamma) \preceq (\lambda', \gamma')$ iff $\lambda \preceq \lambda'$ and $\gamma \preceq \gamma'$. ◇

We say that a net $(x_\lambda)_{\lambda\in\Lambda}$ satisfies a property P ***eventually*** if there is some $\lambda_0 \in \Lambda$ such that x_λ satisfies P for all $\lambda \succeq \lambda_0$; and we say that it satisfies P ***frequently*** if for every $\lambda \in \Lambda$ there is a $\mu \succeq \lambda$ with x_μ satisfying P.

A ***subnet*** of a net $(x_\lambda)_{\lambda\in\Lambda}$ is a net $(y_\gamma)_{\gamma\in\Gamma}$ along with a map $\gamma \mapsto \lambda_\gamma$ from $\Gamma \to \Lambda$ such that for every $\lambda \in \Lambda$, eventually $\lambda_\gamma \succeq \lambda$, and $y_\gamma = x_{\lambda_\gamma}$ for all $\gamma \in \Gamma$. Subsequences of sequences are subnets.

We can define convergence for nets in exactly the same way we do for sequences.

Definition 2.3.3
*Let $(x_\lambda)_{\lambda\in\Lambda}$ be a net in X and τ a topology on X. $(x_\lambda)_{\lambda\in\Lambda}$ **converges** to x in X if and only if given any open neighbourhood U of x, x_λ is eventually in U.*

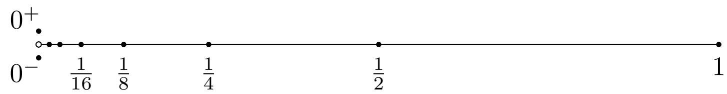

Figure 2.1: A Sequence with Two Limits

*In this case we say that x is a **limit** of $(x_\lambda)_{\lambda\in\Lambda}$, and we write $x_\lambda \to x$ for $\lambda \in \Lambda$, or more commonly*

$$\lim_{\lambda\in\Lambda} x_\lambda = \lambda.$$

For example, it is this sort of convergence that we are talking about when we consider a limit like

$$\lim_{\|\mathcal{P}\|\to 0} \sum_{k=1}^{n} f(x_k^*)(x_k - x_{k-1})$$

Indeed, you are really dealing with nets when considering limits of the form

$$\lim_{x\to a} f(x)$$

as well:

Example 2.3.4
Let $a \in \mathbb{R}$. Then $\mathbb{R}$ is directed by the relation $x \le y$ if $|x-a| \ge |y-a|$. Any function $f : \mathbb{R} \to \mathbb{R}$ is a net.

We have that $f_x(= f(x))$ converges to L if and only if

$$\lim_{x\to a} f(x) = L$$

in the usual sense of this limit. ◇

Note that a net may converge to multiple points if the topology is unusual enough.

Example 2.3.5
Give $(0,1]$ the usual topology $\tau_{(0,1]}$ considering it as a subset of $\mathbb{R}$ with its usual topology $\tau_\mathbb{R}$, and let $X = \{0^+, 0^-\} \cup (0,1]$, where 0^+ and 0^- are abstract points. Give X the topology

$$\tau = \tau_{(0,1]} \cup \{\{0^+, 0^-\} \cup (U \cap (0,1]) : 0 \in U \in \tau_\mathbb{R}\}.$$

You can think of this as splitting 0 into two points 0^+ and 0^- (as illustrated in Figure 2.1). τ is a topology, but every open set which contains $0^\pm$ must contain the other point $0^\mp$.

Consider the sequence $x_n = 2^{-n}$. Then $x_n \to 0^+$ and $x_n \to 0^-$. ◇

This sort of "bad" behaviour can be avoided if it is known that the space satisfies the separation axiom T_2 (see below).

Knowing which nets converge is very closely related to the topology of a topological space. To explore this connection, we introduce two new definitions. Let (X, τ) be a topological space. If A is any subset of X, we say that $x \in X$ is a ***limit point*** of A if there is a net $(x_\lambda)_{\lambda\in\Lambda}$ with $x_\lambda \in A$ which converges to x. It is an ***accumulation point*** if there is a net $(x_\lambda)_{\lambda\in\Lambda}$ with $x_{\lambda_1} \neq x$ which converges to x. We immediately notice that all accumulation points are limit points.

Example 2.3.6

The set of limit points of $(0, 1]$ in $(\mathbb{R}, d)$ is $[0, 1]$. This is also the set of accumulation points. ◇

Example 2.3.7

The set of limit points of $\{1/n \ : n \in \mathbb{N}\}$ in $(\mathbb{R}, d)$ is $\{0\} \cup \{1/n \ : n \in \mathbb{N}\}$. However the only accumulation point is 0, since any sequence x_k from the set which converges to $1/n$ must eventually have $x_k = 1/n$. ◇

These concepts are closely related to closed sets. We start with the following concrete way of thinking of the closure:

Proposition 2.3.1

Let A be any subset of a topological space (X, τ), and let A' be the set of all limit points of A. Then $\overline{A} = A'$.

Proof:

Assume that there is some limit point x of A which is not in $\overline{A}$ ie. there is a net with $x_\lambda \in A$ which converges to x, but $x \notin \overline{A}$. Then $x \in \overline{A}^c$ which is an open set, and hence an open neighbourhood of x, so x_λ is eventually in $\overline{A}^c$, which is a contradiction. Therefore the limit of the net must lie in $\overline{A}$, and so $A' \subseteq \overline{A}$.

Conversely, assume that there is some $x \in \overline{A} \setminus A'$. This means that $A \subseteq A' \subset \overline{A}$, so A' is not closed (else it would be the closure, by definition). Then A'^c is not open, so there must be some point $x \in A'^c$ so that every open neighbourhood of x fails to be contained in A'^c (otherwise it would satisfy the hypotheses of Lemma 2.1.1, and so would be open). Let Λ_x be the set of open neighbourhoods of x. Then Λ_x is a directed set when ordered by reverse inclusion, ie. $U \leq V$ iff $V \subseteq U$. For each $U \in \Lambda_x$, we can find an element $y_U \in A' \cap U$, and hence an $x_U \in A \cap U$ since there is some net in A converging to y_U, and so it must eventually be in U. Then $(x_U)_{U\in\Lambda_x}$ is a net, and clearly $x_U \to x$. But we have just shown that $x \in A'$, contradicting our initial assumption. ■

This immediately gives us the following characterisation of closed sets:

Corollary 2.3.2
Let (X, τ) be a topological space. A subset F of X is closed if and only if F contains all its limit points.

In other words, if we know which nets converge, we can identify which sets are closed, and hence which sets are open, so we can recover the topology. Indeed, we could have started with the notion of convergence and defined the core ideas of topology in terms of convergence, but this approach is not typical. It does tell us, however, that whenever we define some notion of convergence in analysis, such as pointwise convergence, or uniform convergence, that there is likely some sort of topology involved.

Most arguments dealing with sequences transfer directly to nets. The following proposition should be familiar from metric space theory.

Proposition 2.3.3
If X and Y are topological spaces then a function $f : X \to Y$ is continuous at $x \in X$ if and only if for every net $(x_\lambda)_{\lambda\in\Lambda}$ which converges to x, the net $(f(x_\lambda))_{\lambda\in\Lambda}$ converges to $f(x)$.

Proof:

If f is continuous, then given any open neighbourhood V of $f(x)$, we know that $f^{-1}(V)$ is a neighbourhood of x, and so there is some open neighbourhood U of x, with $U \subseteq f^{-1}(V)$. But since $x_\lambda \to x$, we know that x_λ must eventually lie in U, and so $f(x_\lambda)$ must eventually lie in $f(U) \subseteq V$. Hence $f(x_\lambda)$ is eventually in V, so $f(x_\lambda) \to f(x)$.

On the other hand, assume that $x_\lambda \to x$ implies that $f(x_\lambda) \to f(x)$ for all convergent nets. If f is not continuous, then there is some neighbourhood V of $f(x)$, such that $f^{-1}(V)$ is not a neighbourhood of x, or in other words it does not contain any open neighbourhoods of x. Hence given any open neighbourhood U of x, there is a point $x_U \in U \cap f^{-1}(V)^c$. As in Proposition 2.3.1, $(x_U)_{U\in\Lambda_x}$ is a net, and x_U converges to x. Hence $f(x_U)$ converges to $f(x)$, by assumption, but also $f(x_U) \notin V$ for all U, so $f(x_U)$ cannot converge to $f(x)$. Therefore f must be continuous. ■

Note that we require nets in this proposition, as it is possible to find topological spaces and functions for which every sequence converges as above, but for which some more general nets do not. Fortunately it is usually not any more difficult to consider nets in the place of sequences when proving theorems, as most sequence arguments translate fairly directly to the more general setting, as in the proposition above.

An obvious question which should arise is the question of what is the analogue of a subsequence? If $(x_\lambda)_{\lambda\in\Lambda}$ is a net, then a ***subnet*** $(y_\theta)_{\theta\in\Theta}$ is a net, together with a map $\alpha : \Theta \to \Lambda$ with $y_\theta = x_{\alpha(\theta)}$ is increasing, in the sense that if $\theta_1 \leq \theta_2$, then $\alpha(\theta_1) \leq \alpha(\theta_2)$, and given any $\lambda_0 \in \Lambda$, $\alpha(\theta) \geq \lambda$ eventually.

This is a slightly weaker definition than a subsequence, since we're allowed to go backwards or stay stationary, as long as we eventually get big. Also, our directed set Θ could potentially be very different from the directed set Λ. In particular, subnets of sequences may not be subsequences, even if $\Theta = \mathbb{N}$ with the usual order. Never the less, subnets play the same role as subsequences.

Lemma 2.3.4
If (X, τ) is a topological space, and $(x_\lambda)_{\lambda \in \Lambda}$ is a net in X, then $x_\lambda \to x$ if and only every subnet converges to x.

Proof:
If $x_\lambda \to x$, then given any open neighbourhood U of x, there is some λ_0 with $x_\lambda \in U$ for all $\lambda \geq \lambda_0$. But then given any subnet $y_\theta = x_{\alpha(\theta)}$, there is some θ_0 so that $\alpha(\theta) \geq \lambda_0$ for all $\theta \geq \theta_0$. So $y_\theta = x_{\alpha(\theta)} \in U$ for all $\theta \geq \theta_0$, and so $y_\theta \to x$.

On the other hand, if every subnet converges to x, but x_λ does not then there is some open neighbourhood U of x for which given any λ_0 there is some $\lambda \geq \lambda_0$ with $x_\lambda \notin U$. Let $\Theta = \{\lambda \in \Lambda : x_\lambda \notin U\} \subseteq \Lambda$, and give it the order from Λ. Then Θ is directed, since if we have any θ_1, $\theta_2 \in \Theta$, there is some $\lambda_0 \in \Lambda$ with $\lambda_0 \geq \theta_1$ and $\lambda_0 \geq \theta_2$, since Λ is directed. But then we have some $\lambda \geq \lambda_0$ with $x_\lambda \notin U$, so $\lambda \in \Theta$, and $\lambda \geq \theta_1$, $\lambda \geq \theta_2$. Moreover, $(x_\theta)_{\theta \in \Theta}$ is a subnet with the trivial map $\alpha(\theta) = \theta$, and clearly x_θ does not converge to x, contradicting our hypothesis. Hence $x_\lambda \to x$ as well. ■

A subset Θ of a directed set Λ is ***cofinal*** if for each $\lambda \in \Lambda$, there is a $\theta \in \Theta$ with $\theta \geq \lambda$. For example, the set Θ in the proof of the above lemma is cofinal. It is not hard to see, as in the above lemma, that a cofinal subset determines a subnet of a net on Λ with α being the inclusion map. We can use cofinal sets to give us the following result, which is sometimes useful:

Proposition 2.3.5
If (X, τ) is a topological space, and $(x_\lambda)_{\lambda \in \Lambda}$ is a net in X, then $x_\lambda \to x$ if and only if for every cofinal $\Theta \subseteq \Lambda$, there is a cofinal $\Gamma \subseteq \Theta$ such that $(x_\gamma)_{\gamma \in \Gamma} \to x$.

Using this proposition to show convergence can be simpler than using the previous lemma, because the collection of subnets of a particular net can be quite wild.

Exercises

2.3.1. Let x be a point in a topological space (X, τ). If $\mathcal{N}_x$ is the set of all open neighbourhoods of x, show that it is a directed set with order $U \leq V$ iff $V \subseteq U$.

Show that if for every $U \in \mathcal{N}_x$ you choose any element x_U of U, that the net $(x_U)_{U \in \mathcal{N}_x}$ must converge to x.

2.3.2. Prove Proposition 2.3.5.

2.4 Bases and Sub-bases

As we saw with the example of the metric space topology, somtimes it is easier to deal with a smaller or nicer subset of a topology, such as the set of open balls. The hope is that we can prove things that we need to know by using this smaller collection of sets. To know just how small we can go, we introduce the concept of the base and sub-base of a topology.

Given any set $\mathcal{E} \subseteq \mathcal{P}(X)$, there is a minimal topology $\tau_{\mathcal{E}}$ such that $\mathcal{E} \subseteq \tau_{\mathcal{E}}$, namely the topology that is the intersection of all the topologies containing $\mathcal{E}$. We say that $\mathcal{E}$ ***generates*** τ, or sometimes, that $\mathcal{E}$ is a ***sub-base*** of τ. If $\mathcal{E} \subseteq \mathcal{F}$, then $\tau_{\mathcal{E}} \subseteq \tau_{\mathcal{F}}$ by simple set theoretic considerations.

Note: *Some references call a "base" a "basis" and a "sub-base" a "subbasis."*

Example 2.4.1

Let (X, d) be a metric space. The family of all open balls forms a sub-base for the metric topology, τ_d.

Let τ be the topology generated by the open balls. As we saw in Example 2.1.1, every open ball is in τ_d, so $\tau \subseteq \tau_d$. On the other hand, every set in τ_d is a union of open balls, so every topology containing the open balls must contain τ_d, and so τ_d is contained in the intersection. Thus $\tau_d \subseteq \tau$, and so $\tau = \tau_d$. ◇

Example 2.4.2

Let X be a set and $\mathbb{C}(X)$ the set of functions $f : X \to \mathbb{C}$. The topology of pointwise convergence is the topology generated by sets of the form

$$\{g \in \mathbb{C}(X) : |f(x) - g(x)| < \varepsilon\}$$

where $f \in \mathbb{C}(X)$, $x \in X$ and $\varepsilon > 0$. ◇

A ***neighbourhood base*** at x is a family $\mathcal{N}$ of open neighbourhoods of x, and if A is a neighbourhood of x, then there is some $B \in \mathcal{N}$ with $B \subseteq A$.

Example 2.4.3

If (X, d) is a metric space, the open balls $B(x, r)$ are a neighbourhood base for every $x \in X$. ◇

Example 2.4.4

If (X, d) is a metric space, the open balls $B(x, 1/n)$, where $n \in \mathbb{N}$ are a neighbourhood base for every $x \in X$. ◇

A topological space is called ***first countable*** if it has a countable neighbourhood base. The above example shows that every metric space is first countable. First countable spaces have the nice property that we can usually use sequences instead of nets, as show in Exercise 2.4.5 below.

A collection of open sets $\mathcal{B} \subseteq \tau$ is a ***base*** for τ if $\mathcal{B}$ contains a neighbourhood base for every point $x \in X$. A topological space is ***second countable*** if it has a countable base.

Example 2.4.5
If (X, d) is a metric space, the collection of all open balls is a base. $\diamond$

Example 2.4.6
In particular, in $\mathbb{R}^n$ using the metrics

$$d_2(x, y) = \sqrt{\sum_k (x_k - y_k)^2}$$

and

$$d_\infty(x, y) = \max_k |x_k - y_k|,$$

which both give the usual topology on $\mathbb{R}^n$, we get that the collection of all open spheres, and the collection of all open rectangles are bases for the topology on $\mathbb{R}^n$. $\diamond$

Bases for topologies can be made to play similar roles to those of open balls in metric spaces. Compare the proof of the following proposition to the discussion of open sets in Example 2.1.1 and Lemma 2.1.1.

Proposition 2.4.1
Given a topological space (X, τ), $\mathcal{B}$ is a base iff every open set is a union of elements of $\mathcal{B}$.

Proof:
If $\mathcal{B}$ is a collection of sets such that every open set is a union of elements of $\mathcal{B}$, then given $x \in X$ and anu neighbourhood N of x, there is some open neighbourhood U of x with $U \subseteq N$. We can write U as a union of sets B_α for $\alpha \in I$, and so $x \in B_{\alpha'}$ for some particular $\alpha' \in I$. But then $B_{\alpha'}$ is an open neighbourhood of x contained in U, and therefore N. Hence $\mathcal{B}$ is a neighbourhood base for x, and since x was arbitrary, $\mathcal{B}$ is a base.

Let $\mathcal{B}$ be a base of (X, τ), and let U be any open set. If $x \in U$, then there is some $B_x \in \mathcal{B}$ so that $B_x \subseteq U$. But then

$$U = \bigcup_{x \in U} B_x,$$

and so every open set is a union of elements of the base. ■

Note that we consider the empty set to be the trivial union of an empty collection of sets.

We note that a base is automatically a sub-base.

Corollary 2.4.2
Let (X, τ) be a topological space and $\mathcal{B}$ a base for this topology. Then the topology generated by $\mathcal{B}$ is τ.

Proof:
Let $\tau_{\mathcal{B}}$ be the topology generated by $\mathcal{B}$.

By the previous proposition, every element of τ is a union of elements of $\mathcal{B}$, so $\tau \subseteq \tau_{\mathcal{B}}$. On the other hand, by definition, the elements of $\mathcal{B}$ are all elements of τ, so τ is a topology containing $\mathcal{B}$ and so $\tau_{\mathcal{B}} \subseteq \tau$.

Therefore $\tau = \tau_{\mathcal{B}}$. ■

We might ask when a collection of sets is a base, rather than a sub-base for the topology it generates.

Lemma 2.4.3
Let X be a set. Given a collection $\mathcal{B}$ of susbsets of X which $\mathcal{B}$ satisfy the conditions:

(i) if every $x \in X$ is an element of some $B \in \mathcal{B}$,

(ii) if A and $B \in \mathcal{B}$ and $x \in A \cap B$, then there is some $C \in \mathcal{B}$ with $x \in C \subseteq A \cap B$,

then the collection of all unions (including the empty union) of elements of $\mathcal{B}$ is a topology, and $\mathcal{B}$ is a base for that topology.

Proof:
To verify that τ is a topology, we only need to verifiy that finite intersections of sets in τ are also in τ, and that $X \in \tau$. It is easy to see that (i) implies that $X \in \tau$, since if B_x is the lement of $\mathcal{B}$ given by (i),

$$X = \bigcup_{x \in X} B_x \in \tau.$$

We let U_1, U_2 be sets in τ. Then for any x in the intersection of these sets, $x \in U_k$ for $k = 1, 2$, and since these are unions of elements of $\mathcal{B}$, we can find $B_k \subseteq U_k$ with $x \in B_k$. But then there is some $C_x \in \mathcal{B}$ with $x \in C_x \subseteq B_1 \cap B_2 \subseteq U_1 \cap U_2$. We then have that

$$U_1 \cap U_2 = \bigcup_{x \in U_1 \cap U_2} C_x \in \tau.$$

Repeated application of this result gives that τ is closed under finite intersections.

Now if $x \in X$ and N is any neighbourhood of x, there is some $U \in \tau$ with $x \in U \subseteq N$. But U is a union of elements of $\mathcal{B}$, so there is some $B \in \mathcal{B}$ with $x \in B \subseteq U$. Hence $\mathcal{B}$ is a neighbourhood base at x, and since x was arbitrary, it is a base for τ. ■

Note that since $\mathcal{B}$ is a base for τ, the previous corollary shows that τ is generated by $\mathcal{B}$.

We can now precisely characterise the topology generated by a family of sets:

Proposition 2.4.4
Given a topological set X, and $\mathcal{E} \subseteq \mathcal{P}(X)$, the topology generated by $\mathcal{E}$ is the collection consisting of $\emptyset$, X, and all unions of finite intersections of sets in $\mathcal{E}$.

Proof:
Let $\mathcal{B}$ be the collection of all finite intersections of elements of $\mathcal{E}$, together with $\emptyset$ and X. Then $\mathcal{B}$ satisfies the conditions of the previous lemma. (i) follows trivially from the fact that $X \in \mathcal{B}$. If A and $B \in \mathcal{B}$ have non-trivial intersection, then either A or B is X, in which case (ii) follows trivially, or $A = E_1 \cap \ldots \cap E_n$ and $B = F_1 \cap \ldots \cap F_m$, with E_k, $F_j \in \mathcal{E}$, $k = 1, \ldots, n$, $j = 1, \ldots, m$. But then

$$C = (E_1 \cap \ldots \cap E_n) \cap (F_1 \cap \ldots \cap F_m) \in \mathcal{B}.$$

So the topology, $\tau_{\mathcal{B}}$, generated by $\mathcal{B}$, has $\mathcal{B}$ as a base.

If τ is the topology generated by $\mathcal{E}$, it is straightforward that $\mathcal{B} \subseteq \tau$, and so $\tau_{\mathcal{B}} \subseteq \tau$. On the other hand, $\mathcal{E} \subseteq \mathcal{B}$, and so $\tau \subseteq \tau_{\mathcal{B}}$. Hence $\tau = \tau_{\mathcal{B}}$, and we are done. ■

Now that we have established these facts about bases and sub-bases, we can use them to simplify what we need to check when considering continuity and convergence.

Proposition 2.4.5
Let (X, τ_X) and (Y, τ_Y) be topological spaces, $f : X \to Y$ and $\mathcal{B}$ is a sub-base for Y. Then f is continuous if and only if $f^{-1}(U)$ is open for every $U \in \mathcal{B}$.

Proof:
Clearly if f is continuous, then the inverse image of every element of the sub-base is open in X because every element of the sub-base is open in Y.

On the other hand, if U is any open set in Y, we can write U as a union of finite intersection of sets in $\mathcal{B}$, ie.

$$U = \bigcup_{\alpha \in I} \bigcap_{k=1}^{n_\alpha} V_{\alpha,k}.$$

But then

$$f^{-1}(U) = \bigcup_{\alpha \in I} \bigcap_{k=1}^{n_\alpha} f^{-1}(V_{\alpha,k}),$$

and $f^{-1}(V_{\alpha,k})$ is open for all α and k by the hypothesis, and so $f^{-1}(U)$ is open. ■

Proposition 2.4.6
Let (X, τ_X) and (Y, τ_Y) be topological spaces, $f : X \to Y$ and $\mathcal{N}_x$ an open neighbourhood base for $x \in X$, and $\mathcal{N}_{f(x)}$ an open neigbourhood base for $f(x)$. The f is continuous at x if and only if for every $U \in \mathcal{N}_{f(x)}$ there is some $V \in \mathcal{N}_x$ such that $V \subseteq f^{-1}(U)$.

Proof:

If f is continuous at $x \in X$, then $f^{-1}(U)$ is a neigbourhood of x, so we can find some element $V \in \mathcal{N}_x$ so that $V \subseteq f^{-1}(U)$ by the definition of a neighbourhood base.

On the other hand, let U be any open neighbourhood of x. We have some $U' \subseteq U$ with in $U' \in \mathcal{N}_{f(x)}$. This means that there is some $V \subseteq f^{-1}(U')$ with $V \in \mathcal{N}_x$, by hypothesis. But then $V \subseteq f^{-1}(U)$, so that $f^{-1}(U)$ is a neighbourhood of x, and so f is continuous at x. ∎

These two propositions are very useful for proving continuity, since it is far easier to work with some nice sub-base or neighbourhood base in Y than with some arbitrary open set. Indeed, it is this way that we see that the definition of continuity that we know from undergraduate real analysis corresponds to the definition given above.

Example 2.4.7

The above proposition leads to the classical ε-δ definition of continuity. We know that for a metric space, the open balls $B(x, \varepsilon)$ are a neighbourhood base at x, so the above proposition tells us that if (X, d_X) and (Y, d_Y) are metric spaces, and $f : X \to Y$, then f is continuous at x if and only if for all $\varepsilon > 0$, there exists a $\delta > 0$, such that

$$B(x, \delta) \subseteq f^{-1}(B(f(x), \varepsilon)),$$

or when you unravel the definitions, $d(f(x), f(y)) < \varepsilon$ for all y with $d(x, y) < \delta$.

The definition for continuity on X follows from Proposition 2.2.1. ◇

Example 2.4.8

This connection between continuity ans bases leads us to be able to define topologies in terms of functions. If X is any set and $f_\alpha : X \to Y_\alpha$ is a family of maps into topological spaces Y_α for all α in an index set I, the ***weak topology*** on X generated by this family is the weakest topology which makes all the maps f_α continuous. This is simply the topology generated by the sets $f_\alpha^{-1}(A)$ where A is any open set in Y_α, and α is any element of I.

In the opposite direction if you have functions $f_\alpha : Y_\alpha \to X$, then the ***strong topology*** generated by this family of functions is the strongest topology for which all these functions are continuous. This is the topology generated by the sets A where $f_\alpha^{-1}(A)$ is open for all $\alpha \in I$. ◇

If (X, τ) is a topological space, and $\sim$ is an equivalence relation on X, then the ***quotient topology*** on $X/\sim$ is the strong topology given by the quotient map $q : X \to X/\sim$.

If X_α is a family of topological spaces for α in some index set I, the ***product topology*** is the topology on the cross product of the X_α, $\prod_{\alpha \in I} X_\alpha$, which is generated by the projection maps

$$\pi_\alpha : \prod_{\beta \in I} X_\beta \to X_\alpha$$

given by $\pi_\alpha(x) = x_\alpha$ where $x = (x_\beta)_{\beta \in I} \in \prod_{i \in \beta} X_\beta$.

For finite products, the family of sets $\prod_{i=1}^n U_i$ where U_i is an open set in X_i forms a base for the product topology, because every such set is a finite intersection of the geenrating sets.

Example 2.4.9

The usual metric space topology on $\mathbb{R}^n$ is the same as the topology that you get by considering it as the product of n copies of $\mathbb{R}$ with the usual topology.

In particular, open boxes are a base for the usual topology on $\mathbb{R}^n$. ◇

Neighbourhood bases are also useful for proving that a net converges.

Proposition 2.4.7

Let (X, τ) be a topological space, $(x_\lambda)_{\lambda \in \Lambda}$ be a net in X, and $\mathcal{N}$ an open neighbourhood base of x. Then x_λ converges to x if and only if given any $U \in \mathcal{N}_x$ we have x_λ in U eventually.

Proof:

The "if" part of this proposition follows directly from the definition of convergence.

On the other hand, if U is any open neighbourhood of x, then there is some $V \subseteq U$, with $V \in \mathcal{N}_x$, and by hypothesis, $x_\lambda \in V$ eventually. Therefore $x_\lambda \in U$ eventually, and so $x_\lambda \to x$ for $\lambda \in \Lambda$. ■

Again, this has particular utility for showing that the definition of convergence for sequences in metric spaces corresponds to the definition given above.

The previous proposition in particular is useful for nailing down the topologies that correspond to different notions of convergence for sequences of functions.

Example 2.4.10

Let X be any set, and (Y, τ_Y) any topological space. Let $F(X, Y)$ be the set of all functions $f : X \to Y$. We say that a net $(f_\lambda)_{\lambda \in \Lambda}$ of functions in $F(X, Y)$ converges ***pointwise*** to $f \in F(X, Y)$ if and only if

$$\lim_{\lambda \in \Lambda} f_\lambda(x) = f(x)$$

for all $x \in X$.

Consider the collection of subsets of $F(X, Y)$

$$U_{x,V} = \{f \in F(X, Y) : f(x) \in V\}$$

where $x \in X$, and $V \in \tau_Y$. Let τ_{pw} be the topology generated by the sets $U_{x,V}$.

The $(f_\lambda)_{\lambda \in \Lambda}$ converges to f in τ_{pw} if and only if $f_\lambda \to f$ pointwise.

This is the same topology that you get by identifying $F(X, Y)$ with the cartesian product $\prod_{x \in X} Y$ by regarding $f : X \to Y$ as the element $f = (f_x)_{x \in X}$, where $f_x = f(x)$, and giving both spaces the product topology. ◇

Example 2.4.11

Let X be any set, and (Y, d) a metric space. We say that a net $(f_\lambda)_{\lambda\in\Lambda}$ in $F(X, Y)$ converges uniformly to $f \in F(X, Y)$ if

$$\lim_{\lambda\in\Lambda} \sup |f_\lambda - f| \to 0.$$

This is precisely convergence in the topology τ_u generated by the collection of sets

$$U_{f,r} = \{g \in F(X, Y) : \sup |g - f| < r\}$$

where $f \in F(X, Y)$ and $r > 0$. In fact the collection of such $U_{f,r}$ for fixed f is a neighbourhood base for f, so this collection of sets is in fact a base.

The identity function $\mathrm{id} : F(X, Y) \to F(X, Y)$ is continuous from the topology of uniform convergence to the topology of pointwise convergence, because given any $x \in X$, and $V \in \tau_d$, and $f \in U_{x,V}$, we can find an $r_f > 0$ such that $B(f(x), r) \subseteq V$, and then $U_{f,r_f} \subseteq U_{x,V}$. Hence

$$\mathrm{id}^{-1}(U_{x,V}) = \bigcap_{f\in U_{x,V}} U_{f,r_f},$$

which is open.

Note: *It's pretty easy to prove this fact directly.*

As a corollary, if $(f_\lambda)_{\lambda\in\Lambda} \in F(X, Y)$ converges in the uniform topology to f, then it converges pointwise. ◇

Exercises

2.4.1. Show that the family of sets

$$\mathcal{E} = \{(a, \infty) : a \in \mathbb{R}\} \cup \{(-\infty, b) : b \in \mathbb{R}\}$$

is a sub-base of the usual topology on $\mathbb{R}$.

2.4.2. Show that every second countable topological space is separable.

2.4.3. Show that a metric space (X, d) is second countable if and only if it is separable.

2.4.4. Let (X, τ) be a first countable topological space. If $x \in X$, show that the countable neighbourhood base $N_x = \{U_n : n \in \mathbb{N}\}$ of x may be chosen in such a way that $U_n \supset U_{n+1}$.

Note: *this exercise shows that we can usually use sequences instead of nets in first countable topological spaces.*

2.4.5. Let (X, τ) be a first countable topological space. If A is any subset of X, show that x is a limit point of A if any only if there is a sequence $x_n \in A$ which converges to x.

Let (Y, τ_Y) be any topological space. Show that $f : X \to Y$ is continuous at x if and only if for every sequence $x_n \to x$, $f(x_n) \to x_n$.

2.5 Compact Sets

An important notion in topology and analysis is the concept of a compact set. In $\mathbb{R}^n$ these are the closed, bounded sets, and so theorems which involve closed bounded sets are likely to generalize to theorems involving compact sets. For example, a real-valued, continuous function on a closed, bounded subset of $\mathbb{R}$ attains its maximum and minimum values. Also, if we have a sequence in a closed, bounded set in $\mathbb{R}^n$, the Bolzano-Weierstrauss theorem tells us that there must be a convergent subsequence. This second fact is key in one way of looking at compact sets in general topological spaces.

Compact sets can be defined in a number of equivalent ways, so we first introduce some terminology.

Definition 2.5.1
A given a set X and some set $K \subseteq X$, we say that a family of sets $\mathcal{E}$ ***covers*** *K if*

$$K \subseteq \bigcup_{E \in \mathcal{E}} E.$$

A ***subcover*** *of a cover $\mathcal{E}$ is any family of sets $\mathcal{F} \subseteq \mathcal{E}$ which is still a cover of K. A cover is* ***finite*** *if it contains a finite number of sets.*

If (X, τ) is a topological space, and $K \subseteq X$, an ***open cover*** *of K is any cover $\mathcal{U}$ of K which consists of open sets.*

Example 2.5.1
The family of sets $\mathcal{U} = \{(n-1, n+1) : n \in \mathbb{N}\}$ is an open cover of $\mathbb{R}$. ◇

Definition 2.5.2
If (X, τ) is a topological space, and $K \subseteq X$, we say that K has the ***finite intersection property*** *if whenever we have a family, $\mathcal{F}$, of closed sets with the property that*

$$(F_1 \cap F_2 \cap \ldots \cap F_n) \cap K \neq \emptyset \tag{2.1}$$

for any $F_1, F_2, \ldots F_n \in \mathcal{F}$, then

$$\left(\bigcap_{F \in \mathcal{F}} F \right) \cap K \neq \emptyset.$$

Example 2.5.2
$\mathbb{R}$ does not have the finite intersection property, because the family of closed sets

$$\mathcal{F} = \{[a, \infty) : a \in \mathbb{R}\}$$

satisfies (2.1), because

$$[a_1, \infty) \cap [a_2, \infty) \cap \ldots \cap [a_n, \infty) = [\max\{a_1, a_2, \ldots, a_n\}, \infty) \neq \emptyset,$$

but

$$\bigcup_{a\in\mathbb{R}}[a,\infty) = \emptyset.$$

◇

It turns out that these two ideas are closely related, and moreover, are linked to convergence properties of nets.

Theorem 2.5.1
Let (X, τ) be a topological space, and $K \subseteq X$. Then the following are equivalent:

Note: *if (X, τ) is first countable, we can replace nets with sequences in this theorem.*

(i) Every open cover of K has a finite subcover.

(ii) K has the finite intersection property

(iii) every net in K has a subnet which converges in K.

(iv) every infinite subset of K has an accumulation point.

Definition 2.5.3
If (X, τ) is a topological space, a set K satisfying any one of the above equivalent conditions is called a ***compact*** *set. If the X itself is compact, we sat that it is a* ***compact topological space****.*

A set whose closure is compact is called ***precompact****.*

Proof:

Theorem 2.5.1 (i) ⇔ (ii): This follows essentially from DeMorgan's laws. Assume that every open cover of K has a finite subcover, but that K does not satisfy the finite intersection property. So there is a family, $\mathcal{F}$, of closed sets with the property that

$$(F_1 \cap F_2 \cap \ldots \cap F_n) \cap K \neq \emptyset$$

for any $F_1, F_2, \ldots F_n \in \mathcal{F}$, but

$$\left(\bigcap_{F\in\mathcal{F}} F\right) \cap K = \emptyset,$$

or in other words, K is contained in the complement of the intersection. But letting $\mathcal{E} = \{F^c : f \in \mathcal{F}\}$, and applying DeMorgan's laws, we have that

$$K \subseteq \left(\bigcup_{F\in\mathcal{F}} F^c\right),$$

and so $\mathcal{E}$ is an open cover of K. However given any finite subset $F_1^c, F_2^c, \ldots, F_n^c$ of $\mathcal{E}$, we from our assumption, and DeMorgan's laws, that

$$K \not\subseteq (F_1 \cap F_2 \cap \ldots \cap F_n)^c = F_1^c \cup F_2^c \cup \ldots \cup F_n^c.$$

There are no finite subcovers of K, which contradicts our original assumption. Hence K satisfies the finite intersection property.

Conversely, assume that F satisfies the finite intersection property, but that there is some open cover $\mathcal{E}$ that does not have a finite subcover. This means that if we let $\mathcal{F} = \{E^c : E \in \mathcal{E}\}$, then given any finite collection $E_1^c, \ldots E_n^c \in \mathcal{F}$, we have

$$(E_1^c \cap E_2^c \cap \ldots \cap E_n^c) \cap K = (E_1 \cup E_2 \cup \ldots \cup E_n)^c \cap K \neq \emptyset,$$

since the E_k, $k = 1, \ldots, n$ cannot be a cover. So by the finite intersection property,

$$\left(\bigcap_{E \in \mathcal{E}} E^c \right) \cap K = \emptyset,$$

but this means that

$$\left(\bigcup_{E \in \mathcal{E}} E \right)^c \cap K = \emptyset,$$

meaning our original set was not a cover, which contradicts our assumption. Hence every open cover of K has a finite subcover.

The remaining results will be omitted for the time being. In fact, I should probably put this into a technical details section.

■

The following two facts follow almost immediately from the above theorem and definition.

Lemma 2.5.2
Let F be a closed subset of a compact set K in a topological space (X, τ). Then F is compact.

In a metric space, there is another, equivalent way of talking about compact sets. We would like to say that a set is compact if it is closed and bounded, as is the case in $\mathbb{R}^n$, but this is not true in general for metric spaces. We do know that metric spaces are Hausdorff, and so the following lemma tells us that compact sets in metric spaces are closed.

Lemma 2.5.3
A compact subset K of a Hausdorff (see the next section) topological space (X, τ) is closed.

The following example shows us why boundedness does not suffice.

Example 2.5.3

Let X be an infinite set with the discrete metric. Then X is a bounded set, since every point is at most distance 1 from any other. However, it is not compact, since the collection of singleton sets $\mathcal{E} = \{\{x\} : x \in X\}$ is an open cover of X, but it has no finite cover. ◇

To replace boundedness, we have to use a slightmly more general concept.

Definition 2.5.4
*A is **totally bounded** if for every $\varepsilon > 0$, there is a finite collection of sets $B_1, \ldots, B_n \subset A$ (where both n and the sets B_j may depend on ε) such that*

$$A = \bigcup_{j=1}^{n} B_j$$

and $\text{diam}(B_j) < \varepsilon$ for all $j = 1, \ldots, n$.

Theorem 2.5.4
Let (X, d) be a metric space, and $K \subseteq X$. Then K is compact if and only if it is closed and totally bounded.

The following theorem is a special case of the above, which tells us that compact sets to indeed generalise the closed bounded sets in $\mathbb{R}^n$.

Theorem 2.5.5 (Heine-Borel Theorem)
Let K be a subset of $\mathbb{R}^n$ with the usual metric topology. K is compact if and only if it is closed and bounded.

Compact sets have quite a number of useful properties. In particular, it makes sense to look at versions of facts we know involving closed and bounded sets in $\mathbb{R}^n$.

Proposition 2.5.6
Let (X, τ_X) and (Y, τ_Y) be topological spaces, and let $f : X \to Y$ be a continuous function. If $K \subseteq X$ is a compact set, $f(K) \subseteq Y$ is also compact.

Corollary 2.5.7
Let (X, τ_X) be a topological space, and $f : X \to \mathbb{R}$ be continuous with the usual topology on $\mathbb{R}$. If K is a compact set in X, then there are points a and $b \in K$ such that

$$f(a) = \sup\{f(x) : x \in K\} \qquad \text{and} \qquad f(b) = \sup\{f(x) : x \in K\}.$$

That is, f achieves its maximum and minimum values on K.

Corollary 2.5.8
Let (X, τ_X) and (Y, τ_Y) be topological spaces, with X compact. If $f : X \to Y$ is a continuous bijection, then f^{-1} is continuous, ie. f is a homeomorphism.

We conclude with an important theorem:

Theorem 2.5.9 (Tychonoff's Theorem)
Let (X_α, τ_α) for $\alpha \in I$ be a family of compact topological spaces. Then the product space

$$\prod_{\alpha \in I} X_\alpha$$

with the product topology is compact.

The proof of this fact relies on the Axiom of Choice, and in fact one can show that Tychonoff's Theorem is equivalent to the Axiom of Choice.

Exercises

2.5.1. Let (X, τ_X) be a compact space, and let $(f_\lambda)_{\lambda\in\Lambda}$ be a net of functions in $C(X, \mathbb{R})$ which is increasing in the sense that if $\lambda_1 \leq \lambda_2$, then $f_{\lambda_1}(x) \leq f_{\lambda_2}(x)$ for all $x \in X$, and which converges pointwise to a continuous function f. Show that f_λ converges uniformly to f.

2.5.2. Let $(X, \leq)$ be a totally ordered set. The ***order topology*** is the topology on X generated by the sets $\{y : x < y\}$ and $\{y : y < x\}$ for all $x \in X$. A subset A of X is order bounded if there are x and $z \in X$ such that $x \leq y \leq z$ for all $y \in A$. Recall that X is (order) complete if every order bounded set has an infimum and a supremum.

Show that X is order complete if and only if every closed, order bounded subset of X is compact.

2.5.3. Let X be any set, and $\mathbb{T}$ the unit circle of $\mathbb{C}$. Show that $\mathbb{T}$ with multiplication is a compact topological group. Show that

$$\mathbb{T}^X = \prod_X \mathbb{T}$$

is a compact topological group.

If G is a group with the discrete topology, let $\hat{G} = \hom(G, \mathbb{T})$ be the dual group of all group homomorphisms from $G \to \mathbb{T}$, with the product and inverse $(\alpha_1\alpha_2)(g) = \alpha_1(g)\alpha_2(g)$ and $\alpha^{-1}(g) = (\alpha(g))^{-1}$. Show that $\hat{G}$ is a compact topological group.

If G is a compact topological group, show that the dual group $\hat{G}$ of all continuous group homomorphisms $\alpha : G \to \mathbb{T}$ is a topological group with the discrete topology.

Show that $\hat{\mathbb{T}} = \mathbb{Z}$ and $\hat{\mathbb{Z}} = \mathbb{T}$.

2.6 Separation and Extension

As you may have guessed, particularly from examples like Example 2.3.5, topologies can have radically different propoerties that what you would expect from your knowledge of convergence and continuity in $\mathbb{R}^n$, or even your understanding of metric space topology. Fortunately, the "bad" behaviour exhibited by some examples does not often occur in the topologies which arise from examples in analysis.

Most of this odd behaviour occurs because the topology cannot distinguish points. In Example 2.3.5, we get two limits two a sequence simply because we cannot separate the two points 0^+ and 0^- into their own open sets. To understand and regulate these pathologies, we introduce some new conditions on topological spaces, called the spearation axioms.

Definition 2.6.1 (Separation Axioms)
If (X, d) is a topological space, we say that it is:

Note: *Some references do not require regular or normal spaces to be T_1.*

- T_0 *if for every pair of distinct points* x, $y \in X$, *there is an open set* U *with either* $x \in U$ *and* $y \notin U$, *or* $y \in U$ *and* $x \notin U$.

- T_1 *if for every pair of distinct points* x, $y \in X$, *there is an open set* U *with* $x \in U$ *and* $y \notin U$.

- T_2, *or* ***Hausdorff***, *if for every pair of distinct points* x, $y \in X$, *there are disjoint open sets* U *and* V *with* $x \in U$ *and* $y \in V$.

- T_3, *or* ***regular***, *if it is* T_1 *and for every* $x \in X$ *and closed set* $F \subset X$ *with* $x \notin F$, *there are disjoint open sets* U *and* V *with* $x \in U$ *and* $F \subseteq V$.

- T_4, *or* ***normal***, *if it is* T_1 *and for every pair of disjoint closed sets* E, $F \subset X$, *there are disjoint open sets* U *and* V *with* $E \subseteq U$ *and* $F \subseteq V$.

An alternative way of considering axiom T_1 is given by the following lemma.

Lemma 2.6.1
(X, d) *is* T_1 *if and only if every singleton set* $\{x\}$, *for* $x \in X$, *is closed.*

Proof:
If (X, d) is T_1, then the set $\{x\}^c$ is open because if $y \neq x$, then there is an open set U_y with $y \in U_y$ and $x \notin U_y$, and so

$$\{x\}^c = \bigcup_{y \in \{x\}^c} U_y.$$

Hence $\{x\}$ is closed.

On the other hand if every singleton set is closed, then given distinct points x and $y \in X$, the set $U = \{y\}^c$ is open and $x \in U$, $y \notin U$, and so (X, d) is T_1. ■

With this lemma in hand it is easy to see that $T_4 \implies T_3 \implies T_2 \implies T_1 \implies T_0$. One can come up with examples of topological spaces which show that the reverse implications do not hold.

Example 2.6.1
If (X, d) is a metric space, then the metric topology is easily seet to be Hausdorff: given distinct points x and y, let $r = d(x, y)/3$ and let $U = B(x, r)$, $V = B(y, r)$. Clearly U and V are disjoint by the triangle inequality, and we know that they are open.

In fact, metric spaces are always normal, but we need some additional concepts before we can easily show this fact. ◇

Example 2.6.2
If X is a set with at least 2 distinct points, the trivial topology on X satisfies none of the separation axioms. ◇

Example 2.6.3
If $X = \{a, b\}$ and $\tau = \{\emptyset, \{a\}, X\}$, then (X, τ) is a topological space which is T_0, but not T_1. ◇

We have to go to infinite sets to get an example which is T_1 but not T_0.

Example 2.6.4
If X is any infinite set, and τ is the cofinite topology, ie. the collection of all sets whose complement is finite, then (X, τ) is a topological space which is T_1, but not T_2. ◇

The following lemma tells us that if a space is T_2, then that is sufficient to ensure that if a net converges, it converges to only one point, so something like Example 2.3.5 cannot occur.

Lemma 2.6.2
If (X, τ) is a T_2 topological space, and a net $(x_\lambda)_{\lambda\in\Lambda}$ converges to both x and y in X, then $x = y$.

Proof:
Assume that the net $(x_\lambda)_{\lambda\in\Lambda}$ converges to both x and y in X, but $x \neq y$. Then we can find disjoint open sets U and V such that $x \in U$, $y \in V$. But then x_λ is eventuall in U, and is also eventually in V, so it must eventually be in $U \cap V$, but this is the empty set, so we have a contradiction. ■

Note that we can have bad convergence properties even in a T_1 space:

Example 2.6.5
Let (X, τ) be as in Example 2.6.4, and let $(x_n)_{n=1}^{\infty}n$ be any sequence in X such that $x_n \neq x_m$ for $n \neq m$. Then x_n is eventually in *any* open set (because every open set excludes at most a finite number of points), so x_n converges to every point of X. ◇

Hopefully some of these examples illustrate that topologies can, in general, be very different from the intuition that you may have developed from metric space theory or $\mathbb{R}^n$. However, the topologies which occur in analysis tend not to be too strange. Most are at least Hausdorff, and at worst they are T_1.

In certain circumstances we want to know when we can find continuous functions with certain properties, such as taking certain values on certain subsets of our topological space. For concrete spaces, such as $\mathbb{R}^n$, this is not usually an issue because we can usually explicitly construct the function if we have to. However, if we are trying to prove general results we usually don't have this luxury.

There are two typical sorts of situations. The first is that we want to find a continuous function which takes value a on a set A, and b on a set B. The second sort of situation is that we have a function defined on a subset of out topological space which is continuous in the relative topology. Can we extend this function to a continuous function on the whole space?

A little thought should tell you that the first situation for certain sets is closely related to the separation axioms discussed in the first section. For example, if we know that for any closed sets A and B we can find a continuous function f with $f|_A = a$ and $f|_B = b$, then $f^{-1}(B(a,\varepsilon))$ and $f^{-1}(B(b,\varepsilon))$, where $0 < \varepsilon \leq |b-a|/2$ are two disjoint open sets containing A and B respectively. In other words, we should only expect to be able to do this if the topological space is normal.

The first result shows that this example in fact characterizes normal topological spaces.

Theorem 2.6.3 (Urysohn's Lemma)
Let (X,τ) be a normal space, and E, F disjoint closed sets in X. Then there is a continuous function $f : X \to \mathbb{R}$ where $f(E) = \{0\}$ and $f(F) = \{1\}$.

Once one has Urysohn's lemma for normal spaces, it is not difficult to show that one can extend a function from a closed subset to the whole set.

Theorem 2.6.4 (Tietze Extension Theorem)
Let (X,τ) be a normal space. If F is a closed subset of X, and $f \in C(F,[a,b])$, then there is a function $\overline{f} \in C(X,[a,b])$ such that $\overline{f}\big|_F = f$.

Corollary 2.6.5
Let (X,τ) be a normal space. If F is a closed subset of X, and $f \in C(F)$, then there is a function $\overline{f} \in C(X)$ such that $\overline{f}\big|_F = f$.

If instead of considering all closed subsets, we restrict our focus to certain classes of closed subsets, we get some additional results. The most natural restrictions are to make one of the closed sets either a compact set, or a single point.

A topological space (X,τ) is called ***completely regular***, $T_{3.5}$, or ***Tychanoff*** if it is T_1 and for each closed subset F of X, and every $x \notin F$, there is a function $f \in C(X,[0,1])$ such that $f(x) = 1$ and $f = 0$ on F. Every $T_{31/2}$ space is T_3, and Urysohn's Lemma shows that every T_4 space is $T_{31/2}$.

A topological space which is locally compact and Hausdorff is often called an ***LCH space***.

Theorem 2.6.6 (Urysohn's Lemma)
Let (X,τ) be a LCH space, and $K \subseteq U \subseteq X$, where K is compact and U is open. Then there is a continuous function $f : X \to \mathbb{R}$ where $f(K) = \{1\}$ and $f = 0$ outside of some compact subset of U.

This version of Urysohn's lemma allows us to conclude that every LCH space is automatically completely regular, and hence regular, since given a closed set F and a point $x \notin F$, we let $K = \{x\}$ and $U = F^c$.

There is also an extension-type theorem in this setting.

Theorem 2.6.7 (Tietze Extension Theorem)
Let (X, τ) be an LCH space. If K is a compact subset of X, and $f \in C(K)$, then there is a function $\overline{f} \in C(X)$ such that $\overline{f}\big|_F = f$, and moreover, one can also guarantee that $\overline{(}f) = 0$ outside some compact set.

Exercises

1. Show that the topology of Exercise 2.1.1 is T_0 but not T_1.
2. Let (X, τ) be a topological space where X is a finite set. Show that if (X, τ) is T_1, then X must have the discrete topology.
3. Show that a topological space is normal if and only if it satisfies the conclusion
4. Let (X, d) be a metric space. Given a set A and $B \subset X$, define
$$d_1(x, A) = \inf\{d(x, a) : a \in A\}$$
and
$$d_2(A, B) = \sup\{d(a, B) : a \in A\}.$$
Show that d_2 is a metric on the closed sets of the topology.
Let F be a closed set. Show that
$$\rho_F(x) = d_1(x, F)$$
is a continuous function. Use this to show that any metric space is completely regular.
Show that (X, d) is normal.

2.7 Function Spaces

We will be concerned with certain spaces of functions later on, and in many cases the continuous functions will be an important subset. For example, we have already seen that Lebesgue measurable functions include all continuous functions. From undergraduate real analysis you should know that there are several ways of topologising spaces of continuous functions. We'll look at two important theorems about sets of continuous functions. In both cases we want to consider $C(X)$ with the uniform topology, and in both cases the theorem applies if X is a compact Hausdorff space (so $C(X) = C_b(X)$). Also
$$d_u(f, g) = \sup_{x \in X} |f(x) - g(x)|$$
is in fact a metric, since it can never be infinite.

The first tells us what the uniformly compact sets of functions are. A set of functions $f \in \mathcal{F} \subseteq C(X)$ is ***equicontinuous at*** x if for every $\varepsilon > 0$, there is an open neighbourhood U of x so that $|f(x) - f(y)| < \varepsilon$ for all $y \in U$ and $f \in \mathcal{F}$. It is ***equicontinuous*** if it is equicontinuous at each $x \in X$. A set of functions is ***pointwise bounded*** if $\{f(x) : f \in \mathcal{F}\}$ is a bounded susbset of $\mathbb{R}$.

Theorem 2.7.1 (Arzela-Ascoli)
If X is a compact Hausdorff space, and $\mathcal{F}$ is a pointwise bounded, equicontinuous subset of $C(X)$, then $\mathcal{F}$ is totally bounded in the uniform metric, and so the closure of $\mathcal{F}$ is compact.

The second theorem tells us about dense sets in $C(X)$. A subset $\mathcal{A}$ of $C(X)$ ***separates points*** if given any x, $y \in X$, there is some $f \in \mathcal{A}$ such that $f(x) \neq f(y)$. It ***vanishes nowhere*** if for every $x \in X$ there is some $f \in \mathcal{A}$ such that $f(x) \neq 0$. A subset of $\mathcal{A}$ is a ***subalgebra*** of $C(X)$ if it is a vector subspace of $C(X)$ which is closed under multiplication of functions. If we are considering complex-valued functions, $\mathcal{A}$ is ***self-adjoint*** if $\overline{f} \in \mathcal{A}$ for all $f \in \mathcal{A}$.

Theorem 2.7.2 (Stone-Weierstrauss)
If X is a compact Hausdorff space, and $\mathcal{A}$ is a subalgebra of $C(X, \mathbb{R})$ (or a self-adjoint subalgebra of $C(X, \mathbb{C})$), and it separates points and vanishes nowhere, then $\mathcal{A}$ is uniformly dense in $C(X)$.

This has the following very useful corollary (which was actually discovered first):

Corollary 2.7.3 (Weierstrauss Approximation Theorem)
If X is a compact subset of $\mathbb{R}^n$ with the usual metric topology, then the restriction to X of all real polynomial functions in n unknowns is uniformly dense in $C(X, \mathbb{R})$.

In other words, we can find polynomials which are as close as we like (uniformly) to any continuous function on X.

Exercises

2.7.1. Let $\mathbb{T}$ be the unit circle in $\mathbb{C}$. A ***complex trigonometric polynomial*** on $\mathbb{T}$ is a function of the form

$$f(z) = \sum_{k=-n}^{n} c_k z^k.$$

where c_k are complex constants. Show that the set of all trigonometric polynomials is dense in $C(\mathbb{T}, \mathbb{C})$.

Hint: *relate these functions to functions on $\mathbb{T}$ and use the previous exercise.*

2.7.2. Show that the set of all ***real trigonometric polynomials***, ie. functions of the form

$$f(x) = a_0 + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx),$$

where a_k, b_k are real constants, is dense in the set of all functions $f \in C([0, 2\pi])$ with $f(0) = f(2\pi)$.

Chapter 3

Measure Theory

3.1 General Measure Spaces

Now that we have some topology under our belts, our aim is to generalise the ideas of Chapter 1 to arbitrary situations. Just as with the Lebesgue measure on $\mathbb{R}$, our starting point is the concept of a σ-algebra

Definition 3.1.1
Let X be any set. A family $\mathcal{A}$ of subsets of X is an ***algebra*** *if $X \in \mathcal{A}$, and whenever A, $B \in \mathcal{A}$, then $A \cup B$ and A^c are both in $\mathcal{A}$. A family $\mathcal{M}$ of subsets of X is an* ***σ-algebra*** *if $X \in \mathcal{M}$, if whenever $A_k \in \mathcal{M}$, for $k \in \mathbb{N}$, then*

$$\bigcup_{k=1}^{\infty} A_k \in \mathcal{M},$$

and if $A \in \mathcal{M}$, then $A^c \in \mathcal{M}$.

A pair $(X, \mathcal{M})$ where X is a set, and $\mathcal{M}$ is a σ-algebra of sets in X is called a ***measurable space****. The sets in $\mathcal{M}$ are called* ***measurable sets****.*

In other words, an algebra is closed under finite unions and complements, while a σ-algebra is closed under countable unions and complements.

Example 3.1.1
Let X be any set. Then $\mathcal{P}(X)$ is a σ-algebra. $\diamond$

Example 3.1.2
Let X be any set. Then $\{\emptyset, X\}$ is a σ-algebra. $\diamond$

The basic results we had for algebras and σ-algebras in $\mathbb{R}$ all hold for general algebras and σ-algebras.

Lemma 3.1.1
Let $\mathcal{A}$ be an algebra which is closed under disjoint countable unions, ie. given any family of sets $A_k \in \mathcal{A}$ such that $A_k \cap A_l = \emptyset$ for $k \neq l$, the union of these sets ins in the algebra. Then $\mathcal{A}$ is a σ-algebra.

Proof:
See Lemma 1.2.4. ■

Proposition 3.1.2
If $\mathcal{A}$ is an algebra of sets, and A, $B \in \mathcal{A}$, then $\emptyset \in A$, $A \cap B$, $A \setminus B \in \mathcal{A}$ and $A \triangle B \in \mathcal{A}$.

If $\mathcal{A}$ is a σ-algebra, and $A_k \in \mathcal{A}$ for all $k \in \mathbb{N}$, then

$$\bigcap_{k=1}^{\infty} A_k \in \mathcal{A}.$$

Proof:
See Exercise 1.2.1. ■

On occasion we will want to extend some collection of sets, often an algebra of sets, to a full-blown σ-algebra. Inspired by our discussion of bases and subbases of topological spaces, we note the following fact:

Lemma 3.1.3
Let X be any set, and $\mathcal{M}_\alpha$, $\alpha \in I$, any collection of σ-algebras on X. Then

$$\mathcal{M} = \bigcap_{\alpha \in I} \mathcal{M}_\alpha$$

is a σ-algebra.

Proof:
Clearly $X \in \mathcal{M}_\alpha$ for all $\alpha \in I$, so $X \in \mathcal{M}$.
If A_k, $k \in \mathbb{N}$ is a countable collection of sets in $\mathcal{M}$, then

$$\bigcup_{k=1}^{\infty} A_k \in \mathcal{M}_\alpha$$

for all $\alpha \in I$, and hence the union is in $\mathcal{M}$.
Similarly, if $A \in \mathcal{M}$, then $A^c \in \mathcal{M}_\alpha$ for all $\alpha \in I$, and so $A^c \in \mathcal{M}$. ■

Definition 3.1.2
Let X be any set, and $\mathcal{E}$ any collection of subsets of X. The σ-algebra $\mathcal{M}_\mathcal{E}$ ***generated by*** *$\mathcal{E}$ is the smallest σ-algebra which contains $\mathcal{E}$, or equivalently, the intersection of all σ-algebras containing $\mathcal{E}$.*

If (X, τ) is a topological space, then the σ-algebra $\mathcal{B}_X$ generated by τ is called the ***Borel σ-algebra*** *of the topological space. An element of $\mathcal{B}$ is called a* ***Borel set***

We note that if $\mathcal{E}$ and $\mathcal{F}$ are both families of subsets of X, and $\mathcal{E} \subseteq \mathcal{F}$, then $\mathcal{E} \subseteq \mathcal{F}$. This observation will be useful in a number of proofs.

The Borel σ-algebras are important as a class of natural σ-algebras to consider when we want to think about integrating continuous functions on X. A little thought tells us that the Borel σ-algebra must contain all open sets, all closed sets, all countable intersections of open sets, all countable unions of closed sets, all countable unions of countable intersections of open sets, etc. In the literature the various permutations of countable intersections and unions are labelled as follows: countable intersections of open sets are G_δ sets, countable unions of closed sets are F_σ sets, countable unions of G_δ sets are $G_{\delta\sigma}$ sets, countable intersections of F_σ sets are $F_{\sigma\delta}$ sets, etc.

Note: *δ comes from "durschnitt," the German for intersection; and σ comes from "summe," the German word for union.*

Example 3.1.3

Of particular interest is the Borel σ-algebra $\mathcal{B}_\mathbb{R}$ on $\mathbb{R}$ with its usual topology, as this will be key in defining the functions that we can attempt to integrate.

If $\mathcal{L}$ is the set of Lebesgue measurable sets on $\mathbb{R}$, then we note that since every open set is Lebesgue measurable, $\mathcal{B}_\mathbb{R} \subseteq \mathcal{L}$.

In fact there are sets which are Lebesgue measurable, but not Borel. ◇

However, just as with topology, it can be a hassle to have to work with arbitrary open sets in $\mathbb{R}$. The following lemma, which is closely related to Proposition 1.6.1 makes our life somewhat simpler.

Lemma 3.1.4

The Borel σ-algebra on $\mathbb{R}$ is generated by the following collections of sets:

(i) $\mathcal{E}_1 = \{(a, \infty) : a \in \mathbb{R}\}$

(ii) $\mathcal{E}_2 = \{[a, \infty) : a \in \mathbb{R}\}$

(iii) $\mathcal{E}_3 = \{(-\infty, b) : a \in \mathbb{R}\}$

(iv) $\mathcal{E}_4 = \{(-\infty, b] : a \in \mathbb{R}\}$

(v) $\mathcal{E}_5 = \{(a, b) : a, b \in \mathbb{R}\}$

(vi) $\mathcal{E}_6 = \{[a, b] : a, b \in \mathbb{R}\}$

(vii) $\mathcal{E}_7 = \{(a, b] : a, b \in \mathbb{R}\}$

(viii) $\mathcal{E}_8 = \{[a, b) : a, b \in \mathbb{R}\}$

(ix) $\mathcal{E}_9$*, the collection of all closed sets in* $\mathbb{R}$.

The proof of this lemma is very closely related to Proposition 1.6.1, and so will be left as an exercise.

The second ingredient in the integration theory of Chapter 1 was the concept of a measure. Again, this transfers painlessly across to the general setting:

Definition 3.1.3
A function whose domain is an algebra is called a ***set function****. A set function* $\mu : \mathcal{A} \to [0, \infty]$ *is* ***additive*** *if given two disjoint sets* A*,* $B \in \mathcal{A}$*,*

$$\mu(A \cup B) = \mu(A) + \mu(B).$$

A set function on a σ-algebra $\mathcal{A}$ *is* σ***-additive*** *(or* ***countably additive****) if given* $A_k \in \mathcal{A}$*, for* $k \in \mathbb{N}$*, with the* A_k *disjoint,*

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k)$$

A σ*-additive set function for which* $\mu(\emptyset) = 0$ *is called a* ***measure****.*

We will call a triple $(X, \mathcal{M}, \mu)$*, where* X *is a set,* $\mathcal{M}$ *a* σ-algebra*, and* μ *a measure a* ***measure space****.*

We will often regard the σ-algebra as a secondary object implicit in the definition of the measure μ. In this case we call the sets in $\mathcal{M}$ ***μ-measurable sets***. When the measure is also implict, we will simply say that these sets are measurable.

Example 3.1.4
If X is any set and $f : X \to [0, \infty]$ is any function, then the set function $\mu(A) = \sum_{x \in A} f(x)$ (where if A is uncountable, and f is non-zero on uncountably many elements of A, the sum is infinite) gives us a measure on $\mathcal{P}(X)$.

Two particular examples of this are of note. First if $f(x) = 1$ for all $x \in X$, then $\mu(E) = c(E)$, the ***counting measure***. Secondly, if there is some $x_0 \in X$ with $f = \chi_{\{x_0\}}$, then $\mu(A) = \delta_{x_0}(A)$, the ***unit point mass*** or ***Dirac measure***.

The Dirac measure can alternatively be defined by

$$\delta_{x_0}(A) = \begin{cases} 1 & x_0 \in A \\ 0 & x_0 \notin A. \end{cases}$$

◇

Example 3.1.5
If $(X, \mathcal{M}, \mu)$ is a measure space, and $A \in \mathcal{M}$, then $(A, \mathcal{M}_A, \mu|_{\mathcal{M}_A})$ is a measure space, where

$$\mathcal{M}_A = \{B \cap A : B \in \mathcal{M}\}.$$

◇

We use the names counting measure and Dirac measure for restrictions of those measures to smaller σ-algebras as well.

Let $(X, \mathcal{M}, \mu)$ be a measure space. A set $A \in \mathcal{M}$ is a ***((μ-) null set*** if $\mu(A) = 0$. A is a ***σ-finite set*** if there is a countable collection of sets $A_k \in \mathcal{M}$, $k \in \mathbb{N}$, with $\mu(A_k) < \infty$ and $A = \bigcup_{k=1}^{\infty} A_k$. If we have $\mu(X) < \infty$, we say that $(X, \mathcal{M}, \mu)$ is a ***finite measure space***, and μ is a ***finite measure***. If we have $\mu(X) = 1$, μ is often called a ***probability measure***. If X is a σ-finite set, we say that $(X, \mathcal{M}, \mu)$ is a ***σ-finite measure space*** and μ is a ***σ-finite measure***. If for every $A \in \mathcal{M}$ with $\mu(A) = \infty$, we can find a set $B \subseteq A$ with $0 < \mu(B) < \infty$, we say that $(X, \mathcal{M}, \mu)$ is a ***semifinite measure space*** and μ is a ***semifinite measure***.

As before, a property P holds ***(μ-) almost everywhere*** (or (μ-) a.e.) if it is true for every x except on a null set. We will omit the μ if there is no risk of confusion about the measure.

The facts we proved about measures on $\mathbb{R}$ also transfer without change to their proofs:

Proposition 3.1.5
If $\mathcal{A}$ is an algebra of sets in X, and μ is an additive set function on $\mathcal{A}$, then if A and $B \in \mathcal{A}$, and $A \subseteq B$,

$$\mu(A) \leq \mu(B).$$

Proof:
See Proposition 1.3.1. ■

Proposition 3.1.6
If $\mathcal{A}$ is an algebra of sets in X, and μ is an additive set function on $\mathcal{A}$, then if A and $B \in \mathcal{A}$, and $\mu(A \cap B) < \infty$,

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B).$$

Proof:
See Proposition 1.3.2. ■

Proposition 3.1.7
Let $\mathcal{A}$ be a σ-algebra of sets in X, and μ a σ-additive set function on $\mathcal{A}$. If A_k, $k \in \mathbb{N}$, is a sequence of sets in $\mathcal{A}$ with $A_{k-1} \subseteq A_k$, then

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \lim_{k\to\infty} \mu(A_k).$$

Proof:
See Proposition 1.3.3. ■

Proposition 3.1.8
If $\mathcal{A}$ is a σ-algebra of sets in X, μ is a measure on $\mathcal{A}$, and A_k, $k \in \mathbb{N}$, is a sequence of sets in $\mathcal{A}$ with $A_k \subseteq A_{k-1}$, and $\mu(A_k) < \infty$ eventually, then

$$\mu\left(\bigcap_{k=1}^{\infty} A_k\right) = \lim_{k\to\infty} \mu(A_k).$$

Proof:
See Exercise 1.3.3. ■

Exercises

3.1.1. Prove 3.1.4.

3.1.2. Let C be the Cantor set, and let $\varphi : [0,1] \to C$ be a bijection. Let N be the unmeasurable set of Example 1.1.2. Show that $\varphi(N)$ is a Lebesgue measurable set, but is not a Borel set.

3.1.3. Show that $(\mathbb{R}, \mathcal{L}, m)$ is σ-finite.

3.1.4. (†) Show that every σ-finite measure is semifinite.

3.1.5. (†) Let $\mu_1, \ldots, \mu_n$ be measures on $(X, \mathcal{M})$. Show that $\mu = \sum_{k=1}^{n} a_k \mu_k$, where $a_k \geq 0$ is also a measure on $(X, \mathcal{M})$.

3.1.6. (†) If μ is a semifinite measure, show that if $\mu(A) = \infty$ then for any $M > 0$ there is a subset B of F with $M < \mu(B) < \infty$.

3.2 Measurable Functions

Having transferred the basics of measure theory across to general sets, we need to identify the functions that we can hope to integrate. We start at an abstract level with the following definition:

Definition 3.2.1
*Let $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ be measurable spaces, and let $f : X \to Y$. We say that f is $(\mathcal{M}, \mathcal{N})$-**measurable** (or simply measurable if the σ-algebras are implicit), if $f^{-1}(A) \in \mathcal{M}$ for all $A \in \mathcal{N}$.*

*If $A \in \mathcal{M}$, we say that f is **measurable on** A if the restriction of f to A is $(\mathcal{M}_A, \mathcal{N})$-measurable.*

In other words, inverse images of measurable sets are measurable. In this sense, this definition is very much in the spirit of the definition of continuity that we discussed in the previous chapter.

It is immediate that if $(X, \mathcal{M})$, $(Y, \mathcal{N})$ and $(Z, \mathcal{O})$ are measurable spaces, and $f : X \to Y$ is $(\mathcal{M}, \mathcal{N})$-measurable and $g : Y \to Z$ is $(\mathcal{N}, \mathcal{O})$-measurable, then $f \circ g$ is $(\mathcal{M}, \mathcal{O})$-measurable, ie. compositions of measurable functions are measurable.

As with our discussion of sub-bases and continuity, we can make our lives easier when checking for measurability by only looking at a generating family of sets.

Note: *this lemma is a very generalised version of Proposition 1.6.1.*

Lemma 3.2.1
Let $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ be measurable spaces, and let $\mathcal{N}$ be generated by $\mathcal{E}$. Then $f : X \to Y$ is measurable if and only if $f^{-1}(E) \in \mathcal{M}$ for all $E \in \mathcal{E}$.

Proof:
If f is measurable, the result follows immediately from the definition.

The converse follows by observing that $\mathcal{O} = \{A \subseteq Y : f^{-1}(A) \in \mathcal{M}\}$ is a σ-algebra, and that the hypothesis tells us that $\mathcal{E} \subseteq \mathcal{O}$. But that implies that $\mathcal{N} \subseteq \mathcal{O}$, and so $f^{-1}(A) \in \mathcal{M}$ for all $A \in \mathcal{N}$, and we have that f is measurable. ■

An immediate corollary of this is the following fact.

Corollary 3.2.2
If (X, τ_X), (Y, τ_Y) are topological spaces, then every continuous function $f : X \to Y$ is $(\mathcal{B}_X, \mathcal{B}_Y)$-measurable.

Proof:
The open sets of Y generate $\mathcal{B}_Y$, and we know that $f^{-1}(U) \in \tau \subseteq \mathcal{B}_X$ by the definition of continuity, so the lemma gives the result immediately. ■

Our principle interest is with real- or complex- valued functions defined on a set X. In these case, we always assume that the codomains $\mathbb{R}$ and $\mathbb{C}$ have the Borel σ-algebras $\mathcal{B}_{\mathbb{R}}$ or $\mathcal{B}_{\mathbb{C}}$ respectively. If $(X, \mathcal{M})$ is a measurable space, and $f : X \to \mathbb{R}$ or $\mathbb{C}$, we say that f is ***$\mathcal{M}$-measurable*** (or simply measurable if $\mathcal{M}$ is implicit) if it is $(\mathcal{M}, \mathcal{B}_{\mathbb{R}})$- or $(\mathcal{M}, \mathcal{B}_{\mathbb{C}})$-measurable. If (X, τ) is a topological space, we say that the $\mathcal{B}_X$-measurable functions are ***Borel measurable***.

Example 3.2.1
If $\mathcal{L}$ are the Lebesgue measurable sets, then $f : \mathbb{R} \to \mathbb{R}$ is $\mathcal{L}$-measurable if and only if it is Lebesgue measurable in the sense discussed in Chapter 1.

Since $\mathcal{B}_{\mathbb{R}} \subset \mathcal{L}$, we have that every Borel measurable function is Lebesgue measurable, but not every Lebesgue measurable function is Borel Measurable. ◇

We would like to prove results analagous to those in Section 1.6, and some of them are straightforward given the above facts.

Proposition 3.2.3
Let $(X, \mathcal{M})$ be a measurable space, and let f and g be $\mathcal{M}$-measurable real- (or complex-) valued functions on X, and let h be a continuous real- (or complex-) valued function on $\mathbb{R}$ (or $\mathbb{C}$). Then:

(i) $h \circ f$ is $\mathcal{M}$-measurable,

(ii) cf is $\mathcal{M}$-measurable for every constant c,

(iii) $|f|$ is $\mathcal{M}$-measurable,

(iv) f^+ is $\mathcal{M}$-measurable,

(v) f^- is $\mathcal{M}$-measurable,

(vi) $f + g$ is $\mathcal{M}$-measurable,

(vii) fg is $\mathcal{M}$-measurable.

Proof:

All of these rely on Corollary 3.2.2. We will prove the real case, the complex case is similar.

(i) This follows simply from the fact that because h is continuous, h is $(\mathcal{B}_{\mathbb{R}}, \mathcal{B}_{\mathbb{R}})$-measurable, and f is $(\mathcal{M}, \mathcal{B}_{\mathbb{R}})$-measurable, and so the composition $h \circ f$ must be $(\mathcal{M}, \mathcal{B}_{\mathbb{R}})$-measurable, ie. it is $\mathcal{M}$-measurable.

(ii) This is just the special case of (i) where $h(x) = cx$.

(iii) This is just the special case of (i) where $h(x) = |x|$.

(iv) This is just the special case of (i) where $h(x) = \max\{x, 0\}$.

(v) This is just the special case of (i) where $h(x) = \max\{-x, 0\}$.

(vi) The map $\varphi : (x, y) \mapsto x + y$ is a continuous function from $\mathbb{R}^2$ to $\mathbb{R}$ and hence is $(\mathcal{B}_{\mathbb{R}^2}, \mathcal{B}_{\mathbb{R}})$-measurable, and the map $\psi : x \mapsto (f(x), g(x))$ is $(\mathcal{M}, \mathcal{B}_{\mathbb{R}^2})$-measurable, since one can show that $\mathcal{B}_{\mathbb{R}^2}$ is generated by open rectangles

$$\mathcal{R} = \{(a, b) \times (c, d) : a, b, c, d \in \mathbb{R}\},$$

and $\psi^{-1}((a, b) \times (c, d)) = f^{-1}((a, b)) \cap f^{-1}((c, d)) \in \mathcal{M}$. So $f + g = \varphi \circ \psi$ is $(\mathcal{M}, \mathcal{B}_{\mathbb{R}})$-measurable.

(vii) This is just like (vi), but using the fact that that $\varphi : (x, y) \mapsto xy$ is continuous. ■

As simple corollaries, we have that $f - g$ is $\mathcal{M}$-measurable and $1/f$ is $\mathcal{M}$-measurable if $f(x) \neq 0$. Also, if f^+ and f^- are both $\mathcal{M}$-measurable, then so is f.

In the case of complex-valued functions, we have the following as well:

Corollary 3.2.4

If $(X, \mathcal{M})$ is a measurable space, then $f : X \to \mathbb{C}$ is $\mathcal{M}$-measurable if and only if $\operatorname{Re} f$ and $\operatorname{Im} f$ are $\mathcal{M}$-measureable.

Also $\operatorname{sign} f$ defined by $\operatorname{sign} z = z/|z|$ (or 0 if $z = 0$) is $\mathcal{M}$-measurable.

Proof:

We know that $R : z \mapsto \operatorname{Re} z$ and $I : z \mapsto \operatorname{Im} z$ are both continuous functions from $\mathbb{C}$ to $\mathbb{R}$, and so $\operatorname{Re} f = R \circ f$ and $\operatorname{Im} f = I \circ f$ are both $\mathcal{M}$-measureable.

If $\operatorname{Re} f$ and $\operatorname{Im} f$ are $\mathcal{M}$-measurable, then $f = \operatorname{Re} f + i \operatorname{Im} f$ is $\mathcal{M}$-measurable by the previous result.

The function $\operatorname{sign} : z \mapsto z/|z|$ is a $\mathcal{B}_{\mathbb{C}}$-$\mathcal{B}_{\mathbb{C}}$-measurable function. This can be seen by the fact that it is continuous at every point except 0, and so if $U \subseteq \mathbb{C}$ does not contain 0, then $\operatorname{sign}^{-1}(U)$ is open in $\mathbb{C}$, and hence Borel measurable, and if U does contain 0 then $U \setminus \{0\}$ is open, and $\{0\}$ is closed, and $\operatorname{sign}^{-1}(U) = \operatorname{sign}^{-1}(U \setminus \{0\}) \cup \{0\}$, which is a union of an open set and a closed set, and hence is Borel measurable.

It follows immediately from this fact and the fact that composition of measurable functions between appropriate measureable spaces are measurable, that $\operatorname{sign} f = \operatorname{sign} \circ f$ is $\mathcal{M}$-measurable. ■

As in the case of Lebesgue measurable functions, measurability behaves well under limiting constructs:

Theorem 3.2.5
Let $(X, \mathcal{M})$ be a measurable space. If $f_k : X \to \mathbb{R}$ is a sequence $\mathcal{M}$-measurable functions, then

$$g_1(x) = \sup_k f_k(x) \qquad \text{and} \qquad g_2(x) = \inf_k f_k(x)$$

and

$$h_1(x) = \limsup_k f_k(x) \qquad \text{and} \qquad h_2(x) = \liminf_k f_k(x)$$

are all $\mathcal{M}$-measurable functions on X. Moreover, if f_k converges pointwise to some function $f : X \to \mathbb{R}$, then f us $\mathcal{M}$-measurable.

Proof:
The proof of these is exactly as in Theorem 1.6.3 and Corollaries 1.6.4 and 1.6.5, keeping in mind that Lemma 3.2.1 and Lemma 3.1.4 mean that we only need check that the inverse images of certain intervals I to prove measurability. ■

Although sup, inf, lim sup and lim inf do not make sense for complex-valued functions, limits do, and we can easily see that the limit of complex-valued measurable functions is again a measurable functions by recalling that f_n and f are complex-valued functions then $f_n \to f$ pointwise if and only if $\operatorname{Re} f_n \to \operatorname{Re} f$ and $\operatorname{Im} f_n \to \operatorname{Im} f$ pointwise in $\mathbb{R}$.

As before, we define a ***simple function*** on a set X to be a function of the form

$$f(x) = \sum_{k=1}^{n} c_k \chi_{E_k}$$

for some constants c_k and sets E_k. A simple function is in standard form if the values c_k are points in the range, and $E_k = f^{-1}(\{c_k\})$. A simple function is $\mathcal{M}$-measurable if and only if every E_k is measurable when the function is written in standard form.

Note that although we often assume that the constants c_k are real, or even non-negative, they can be chosen to be complex numbers.

As before, the key fact about simple functions is that we can approximate any measurable function by simple functions.

Theorem 3.2.6
Let $(X, \mathcal{M})$ be a measurable space. Let $f : X \to \mathbb{R}$ be $\mathcal{M}$-measurable with $f \geq 0$. Then there is a sequence of $\mathcal{M}$-measurable simple functions φ_n with $0 \leq \varphi_n \leq \varphi_{n+1} \leq f$ for all n, and $\varphi_n \to f$ pointwise on X.

Proof:
As you might expect, the proof is the same as the proof of Theorem 1.6.9 for Lebesgue measurable functions. ■

For simplicity, we will denote the set of non-negative $\mathcal{M}$-measurable functions on X by $\mathcal{M}^+(X)$.

Exercises

3.2.1. Verify that the proofs which refer to Chapter 1 are in fact valid.

3.2.2. (†) Show that the supremum of an uncountable family of measurable functions may not be measurable.

3.2.3. Show that any monotone function $f : \mathbb{R} \to \mathbb{R}$ is Borel measurable.

3.3 General Lebesgue Integrals

Now that we have general measurable functions, it is easy to define the integral in the general setting. This is done in exactly the same way as for the Lebesgue integral on $\mathbb{R}$: we start with simple functions, use these to define the integral for non-negative measurable functions, and then proceed to general measurable functions.

Definition 3.3.1
Let $(X, \mathcal{M}, \mu)$ be a measure space. If $\varphi \in \mathcal{M}^+(X)$ is simple, with standard representation

$$\varphi = \sum_{k=1}^{n} c_k \chi_{A_k}$$

then we define the Lebesgue integral of φ to be

$$\int \varphi \, d\mu = \sum_{k=1}^{n} c_k \mu(A_k).$$

If $f \in \mathcal{M}^+(X)$, then we define the Lebesgue integral of f to be

$$\int f \, d\mu = \sup\{\int \varphi \, d\mu : \varphi \in \mathcal{M}^+(X), \text{simple}, \varphi \leq f\}.$$

If $f : X \to \mathbb{R}$ is any $\mathcal{M}$-measurable function, and

$$\int f^+ \, d\mu < \infty \qquad \text{and} \qquad \int f^- \, d\mu < \infty,$$

then we define the Lebesgue integral of f to be

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu.$$

Note that we are taking the integral over the entire set X. If we want to integrate over a measurable subset $E \subseteq X$, we simply note that $\mathcal{M}_E = \{E \cap A : A \in \mathcal{M}\}$ is a σ-algebra on E, called the ***relative σ-algebra*** on E, and $(E, \mathcal{M}_E, \mu|_{\mathcal{M}_E})$ is a measure space, so we simply define

$$\int_E f \, d\mu = \int f|_E \, d\mu_E = \int f\chi_E \, d\mu.$$

For simplicity, we will work with integrals over the entire measure space in general, and any result can be easily converted to a result about integrals over a particular set with the above observations.

As before, the integral behaves as we expect:

Proposition 3.3.1
If $(X, \mathcal{M}, \mu)$ is a measure space, φ and $\psi \in \mathcal{M}^+(X)$ are simple, and $c \geq 0$, then

(i) *if φ has a representation $\varphi = \sum_{k=1}^{l} a_k \chi_{G_k}$, with $a_k \geq 0$, then*

$$\int \varphi \, d\mu = \sum_{k=1}^{l} a_k m(G_k).$$

(ii) $\int c\varphi \, d\mu = c \int \varphi \, d\mu.$

(iii) $\int \varphi + \psi \, d\mu = \int \varphi \, d\mu + \int \psi \, d\mu.$

(iv) *if $\varphi \leq \psi$, then* $\int \varphi \, d\mu \leq \int \varphi \, d\mu.$

(v) $\mu_\varphi(A) = \int_A \varphi \, d\mu$ *is a measure on $\mathcal{M}$.*

Proof:

This is the same as Proposition 1.7.1, except for the last part, which is Proposition 1.7.3 in the simple function case. ■

The integral for $\mathcal{M}^+(X)$ functions is similarly well-behaved, although we do not yet have additivity:

Proposition 3.3.2
Let $(X, \mathcal{M}, \mu)$ be any measurable set, f and $g \in \mathcal{M}^+(X)$, and $c \geq 0$ be any constant. Then

(i) $\int_E cf \, d\mu = c \int_E f \, d\mu.$

(ii) *if $f \leq g$, then* $\int_E f \, d\mu \leq \int_E g \, d\mu.$

(iii) $\mu_f(A) = \int_A f \, d\mu$ *is a measure on $\mathcal{M}$.*

(iv) $\int f \, d\mu = 0$ *iff $f = 0$ μ-almost everywhere.*

(v) *if* $f = g$ μ*-almost everywhere, then* $\int f \; d\mu = \int g \; d\mu$.

Proof:
(i) and (ii) are the same as Proposition 1.7.2. (iii) is the same as Proposition 1.7.3, and (iv) and (v) are the same as two of the Corollaries of that Proposition. ■

Since we know how the theory should progress, we may as well give the Monotone Convergence Theorem at this point.

Theorem 3.3.3 (Monotone Convergence Theorem)
Let $(X, \mathcal{M}, \mu)$ *be a measure space, and let* $f_n \in \mathcal{M}^+$ *be an increasing sequence which converges pointwise* μ*-almost everywhere to some function* $f : X \to \mathbb{R}$*. Then* $f \in \mathcal{M}^+(X)$*, and*

$$\int f \; d\mu = \lim_{n\to\infty} f_n \; d\mu.$$

Proof:
This is the same as Theorem 1.8.1 and the Corollary which allows us to relax to pointwise almost everywhere convergence. ■

As was the case before, we use the Monotone Convergence Theorem to prove additivity.

Theorem 3.3.4
Let $(X, \mathcal{M}, \mu)$ *be a measure space, and let* f_k *a (finite or infinite) sequence of functions in* $\mathcal{M}^+(X)$*, and let*

$$f = \sum_k f_k.$$

Then

$$\int f \; d\mu = \sum_k \int f_k \; d\mu.$$

Proof:
This is the same as Theorem 1.8.3. ■

Finally, we have Fatou's Lemma:

Proposition 3.3.5 (Fatou's Lemma)
If $(X, \mathcal{M}, \mu)$ *is a measure space, and* f_n *is a sequence of functions in* $\mathcal{M}^+(X)$*, then*

$$\int \liminf_{n\to\infty} f_n \; d\mu \leq \liminf_{n\to\infty} \int f_n \; d\mu.$$

For real and complex valued measurable functions, we have the following results. Note that if the integral of f is finite, we say that $f \in L^1(X, \mu)$ (or just $L^1(X)$ is μ is implicit).

Lemma 3.3.6
Let $(X, \mathcal{M}, \mu)$ be any measure space. Then the following are equivalent:

(i) $f \in L^1(X, \mu)$.

(ii) $\int |f| \, d\mu < \infty$.

Proposition 3.3.7
Let $(X, \mathcal{M}, \mu)$ be any measure space, and f, g be measurable functions for which the integral exists, and let c be any constant. Then

(i) $\int cf \, d\mu = c \int f \, d\mu$.

(ii) if we do not have

$$\int f \, d\mu = +\infty \qquad \text{and} \qquad \int g \, d\mu = -\infty$$

or vice versa, then

$$\int f + g \, d\mu = \int f \, d\mu + \int g \, d\mu$$

(iii) if $f \leq g$, then $\int f \, d\mu \leq \int g \, d\mu$.

(iv) if $F \subseteq E$ is measurable, $\int_E f\chi_F \, d\mu = \int_F f \, d\mu$.

(v) if E is null, then $\int_E f \, d\mu = 0$.

(vi) if f is bounded on X and $\mu(X) < \infty$ (ie. (X, μ) is a finite measure space), then $f \in L^1(X, \mu)$. Moreover, if $\alpha \leq f(x) \leq \beta$ on X, then

$$\alpha m(X) \leq \int f \, d\mu \leq \beta m(X).$$

Theorem 3.3.8 (Dominated Convergence Theorem)
Let $(X, \mathcal{M}, \mu)$ be a measure space, and f_n be a sequence of real- or complex-valued measurable functions for which the Lebesgue integral exists on X, and let $f_n \to f$ pointwise μ-almost everywhere. If there is some function $g \in L^1(X, \mu)$, for which

$$|f_n(x)| \leq g(x),$$

for all $x \in X$, then

$$\int f \, d\mu = \lim_{n\to\infty} \int f_n \, d\mu.$$

Proposition 3.3.9
Let $(X, \mathcal{M}, \mu)$ be a measure space. If $f_n \in L^1(X, \mu)$ and

$$\sum_{n=1}^{\infty} \int |f_n| \, d\mu < \infty,$$

then there is an $f \in L^1(X, \mu)$, such that

$$f = \sum_{n=1}^{\infty} f_n$$

almost everywhere on X, and

$$\int f \, d\mu = \sum_{n=1}^{\infty} \int f_n \, d\mu.$$

Having got this block of theory out of the way, we should have a look at what these general integrals can look like in particular examples.

Example 3.3.1
Consider a measure space of the form $(X, \mathcal{P}(X), \mu_\alpha)$ as in Example 3.1.4 with a "weighting function" α. If φ is a non-negative simple function (which means that it is non-zero at a finite number of points), then

$$\int \varphi \, d\mu_\alpha = \sum_{k=1}^{n} c_k \sum_{x \in E_k} \alpha(x) = \sum_{x \in X} \alpha(x)\varphi(x).$$

By letting simple functions φ_n increase pointwise to an arbitrary function f on X, we get that

$$\int f \, d\mu_\alpha = \lim_{n\to\infty} \int \varphi_n \, d\mu_\alpha = \lim_{n\to\infty} \sum_{x \in X} \alpha(x)\varphi_n(x) \tag{3.1}$$

$$= \sum_{x \in X} \lim_{n\to\infty} \alpha(x)\varphi_n(x) = \sum_{x \in X} \alpha(x) f(x). \tag{3.2}$$

(We can swap the sum and limit, because either the sums are absolutely convergent, or are eventually $+\infty$.)

In particular, counting measure (where $\alpha = 1$) gives us a simple sum:

$$\int f \, dc = \sum_{x \in X} f(x),$$

while the Dirac measure at x_0 gives us evaluation:

$$\int f \, d\delta_{x_0} = f(x_0).$$

◇

This last example shows that this general concept of integration is *very* flexible: sums and even evaluation of a function can be considered a type of integration.

Exercises

3.3.1. Verify the claims made in this section.

3.3.2. (†) Use the counting measure on $\mathbb{N}$ to state Fatou's Lemma and the convergence theorems as theorems about sums of series.

3.4 Types of Convergence

When considering a sequence (f_n) of $\mathbb{R}$ or $\mathbb{C}$ valued functions, we are already familiar with different types of convergence. The most familiar of these are uniform convergence and pointwise convergence. With the introduction of measures and integration, we have some new types of convergence to consider

In measure theory, null sets are usually negligible: it often suffices that some condition hold except on a set of measure 0. As we have seen, the monotone and dominated convergence theorems work when the functions converge pointwise almost everywhere. It is easy to see that uniform convergence implies pointwise convergence which in turn implies pointwise almost everywhere convergence. Pointwise almost everywhere convergence is only equivalent to pointwise convergence if the only null set is the empty set or, equivalently, all single point sets $\{x\} \subseteq X$ have positive measure.

For sequences of integrable functions we can naturally talk about a metric coming from the integral. We say $f_n \to f$ in L^1 if we have convergence in the pre-metric

$$d(f, g) = \int |f - g| \, d\mu.$$

This type of convergence does not automatically bear any relation to the other types of convergence. Examples can be found where there is uniform convergence, but not L^1 convergence, and vice-versa. If, however, the measure space is finite, ie. $\mu(X) < \infty$, then we have that uniform convergence implies L^1 convergence.

We say that a sequence of measurable functions ***converges in measure*** if for every $\varepsilon > 0$,

$$\mu(\{x \in X : |f_n(x) - f(x)| \geq \varepsilon\}) \to 0$$

as $n \to \infty$. We can easily see that if $f_n \to f$ in L^1 then $f_n \to f$ in measure, since if

$$E_{n,\varepsilon} = \{x \in X : |f_n(x) - f(x)| \geq \varepsilon\},$$

we have

$$\int |f_n - f| \, d\mu \geq \int_{E_{n,\varepsilon}} |f_n - f| \, d\mu \geq \varepsilon\mu(E_{n\varepsilon}),$$

and so $\mu(E_{n,\varepsilon}) \to 0$ as $n \to \infty$.

The final type of convergence we shall consider is called almost uniform convergence. A sequence of measurable functions converges to f almost uniformly if for every $\varepsilon > 0$, there is a set $E \subseteq X$ such that $\mu(E) < \varepsilon$ and $f_n \to f$

uniformly on E^c. Clearly, if $f_n \to f$ uniformly, then $f_n \to f$ almost uniformly. In addition, it is not hard to see that if $f_n \to f$ almost uniformly, then $f_n \to f$ almost everywhere, since the set of points where $f_n(x)$ does not converge to $f(x)$ is a subset of the intersection of all the sets E for all possible ε, and hence has measure 0. One can also show that almost uniform convergence implies convergence in measure.

Two slightly deeper theorems give us some partial results linking these types of convergence together.

Proposition 3.4.1
Let (f_n) be Cauchy in measure, ie. for every $\varepsilon > 0$, we have

$$\mu(\{x \in X : |f_n(x) - f_m(x)| \geq \varepsilon\}) \to 0$$

as m, $n \to \infty$. Then there is a measurable function f such that $f_n \to f$ in measure, and a subsequence f_{n_k} which converges pointwise a.e. Furthermore if $f_n \to g$ in measure, then $f = g$ a.e.

Proof:
Choose a subsequence $g_k = f_{n_k}$ so that the set $E_k = \{x \in X : |g_k(x) - g_{k+1}(x)| \geq 2^{-k}\}$ has $\mu(E_k) \leq 2^{-k}$. Letting $F_j = \bigcup_{k=j}^{\infty} E_k$, we have

$$\mu(F_j) \leq \sum_{j=k}^{\infty} 2^{-k} = 2^{1-j},$$

and for $x \notin F_j$ we have

$$|g_k(x) - g_l(x)| \leq \sum_{m=k}^{l} |g_{m+1}(x) - g_m(x)| \leq 2^{1-k}$$

so g_k is pointwise Cauchy for $x \notin F_j$. Moreover, the set of points where g_k is not pointwise Cauchy is contained in the intersection F of the F_j, and

$$\mu\left(\bigcap F_j\right) \leq 2^{1-j}$$

g_k is Cauchy almost everywhere. Therefore there is a measurable f such that $g_k \to f$ on F^c, and $f(x) = 0$ on F, i.e. $g_k \to f$ pointwise almost everywhere.

Moreover, $|g_k(x) - f(x)| \leq 2^{2-k}$ for $k \geq j$ and $x \notin F_j$, so $g_k \to f$ in measure, since $\mu(F_j) \to 0$. But then $f_n \to f$ in measure, since

$$\begin{aligned}\{x \in X : |f_n(x) - f(x)| \geq \varepsilon\} \subseteq &\{x \in X : |f_n(x) - g_k(x)| \geq \varepsilon/2\} \\ &\cup \{x \in X : |g_k(x) - f(x)| \geq \varepsilon/2\}\end{aligned}$$

and both sets on the right have measure converging to 0 as $n \to \infty$, so the measure of the set on the left converges to 0.

A similar argument shows that if $f_n \to g$, then $f = g$ a.e. ∎

An immediate corollary of this is that if $f_n \to f$ in L^1, then there is a subsequence f_{n_k} which converges almost everywhere.

Theorem 3.4.2 (Egoroff's Theorem)
If $\mu(X) < \infty$ and f_n, f are measurable functions such that $f_n \to f$ pointwise a.e. Then $f_n \to f$ almost uniformly.

Proof:
WLOG assume that $f_n \to f$ pointwise. Then let

$$E_n(k) = \bigcup_{m=n}^{\infty} \{x : |f_m(x) - f(x)| \geq k^{-1}\}$$

for n and $k \in \mathbb{N}$. For fixed k, $E_n(k)$ is a sequence of sets which decreases to $\emptyset$. Hence $\mu(E_n(k)) \to 0$ as $n \to \infty$, since $\mu(X) < \infty$.

Now given $\varepsilon > 0$ and $k \in \mathbb{N}$, let n_k sufficiently large that $\mu(E_{n_k}(k)) < \varepsilon 2^{-k}$ and so $E = \bigcup E_{n_k}(k)$ has $\mu(E) < \varepsilon$, and whenever $n > n_k$, $|f_n - f(x)| < k^{-1}$ on E^c, so $f_n \to f$ uniformly on E^c. ■

Example 3.4.1
The following sequences for Lebesgue measurable functions on $\mathbb{R}$ are useful to illustrate the differences between the various types of convergence:

1. $f_n = n^{-1}\chi_{(0,n)}$ converges to 0 uniformly, pointwise, pointwise a.e., almost uniformly and in measure, but not in L^1.

2. $f_n = \chi_{(n,n+1)}$ converges to 0 pointwise, pointwise a.e., but not uniformly, in L^1, in measure or almost uniformly.

3. $f_n = n\chi_{[0,1/n]}$ converges to 0 pointwise, pointwise a.e., almost uniformly and in measure, but not uniformly, or in L^1.

4. $f_n = \xi_{[j2^{-k},(j+1)2^{-k}]}$ for $0 \leq j < 2^k$ and $n = j + 2^k$. This converges in L^1 and in measure, but not uniformly, pointwise, pointwise a.e. or almost uniformly.

◇

In addition to situations where each type of convergence is the natural one to consider, there are also situations where we want one type of convergence, but it simpler to show that a related type of convergence occurs.

We will learn of further ways in which sequences of functions converge when we look at the L^p spaces (probably next semester).

Exercises

3.4.1. Verify the claims made in Example 3.4.1.

3.5 Outer Measure

Up to this point we have been taking the measure as an already known and understood quantity: given a measure space, we can replicate the theory of the Lebesgue integral in the general setting. However, if you recall the theory from Chapter 1, we had to construct the Lebesgue measure on $\mathbb{R}$ from a weaker object defined on the algebra of elementary sets via the concept of an outer measure. When set up appropriately, we can generalise this construction.

Definition 3.5.1
*Let X be any set. An **outer measure** is a set-function $\mu^* : \mathcal{P}(X) \to [0\infty]$ such that:*

1. $\mu^*(\emptyset) = 0$;
2. *if $A \subseteq B$, then $\mu^*(A) \leq \mu^*(B)$;*
3. *(subadditivity) if $A = \bigcup_{k=1}^{\infty} A_k$, then*

$$\mu^*(A) \leq \sum_{k=1}^{\infty} \mu^*(A_k).$$

As before, given an outer measure μ^* we say that a set E is ***measurable with respect to*** μ^* if for every set A we have

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

Again, to show that a set is measurable it suffices to verify that

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

since $\mu^*(A) \leq \mu^*(A \cap E) + \mu^*(A \setminus E)$ follows immediately from the subadditivity axiom.

As one would hope, these "measurable" sets do in fact form a σ-algebra.

Theorem 3.5.1 (Carathéodory)
Let μ^* be an outer measure. Then the family $\mathcal{M}$ of μ^*-measurable sets is a σ-algebra, and the restriction of μ^* to $\mathcal{M}$ is a measure.

Proof:
As with so much of this section, the proof is exactly the same as the proof given for the first version of Carathéodory's theorem (Theorem 1.4.3). ■

The measure produced by Carathéodory's theorem is always complete (that is, if $A \subseteq B$ and $\mu(B) = 0$, then $\mu(A) = 0$) for the same reasons as discussed in Chapter 1 (see Exercises).

With this behind us, we can now talk about generalizing the way that Lebesgue measure was defined. Recall that the definition started with the measure of an interval, and proceeded from those sets to elementary sets, and thence via Lebesgue outer measure and Carathèodory's theorem to Lebesgue measurable sets.

However, the collection of all intervals is not an algebra of sets. In fact it satisfies a weaker condition.

Definition 3.5.2
A collection $\mathcal{S}$ of subsets of a set X is called a ***semi-algebra*** *if*

1. $\emptyset \in \mathcal{S}$,
2. *given* A, $B \in \mathcal{S}$, *then* $A \cap B \in \mathcal{S}$,
3. *if* $A \in \mathcal{S}$, *then* A^c *is a finite disjoint union of elements of* $\mathcal{S}$.

It is not hard to prove that the collection of all finite unions of sets in a semi-algebra gives the algebra generated by the semi-algebra (see Exercises).

The next step was to define a measure-like function on the set of all intervals. The key property that this function had was that it was σ-additive wherever this made sense.

Definition 3.5.3
Let $\mathcal{S}$ be a semi-algebra. A ***pre-measure*** *is a set function $\mu : \mathcal{S} \to [0, \infty]$ such that*

1. $\mu(\emptyset) = 0$,
2. *if $A_k \in \mathcal{S}$ are disjoint, with*

$$\bigcup_{k=1}^{\infty} A_k \in \mathcal{S},$$

then

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k).$$

In other words, although countable unions of elements of the semi-algebra need not be contained in the algebra, we insist that in the cases where they are, the pre-measure must be σ-additive for those sets.

Example 3.5.1
The set function $m : \mathcal{I} \to [0, \infty]$ of Chapter 1 is a pre-measure on the semi-algebra of intervals sets. This is the content of Lemma 1.4.2.

Since every algebra is also a semi-algebra, the set function $m : \mathcal{E} \to [0, \infty]$ which extends m to elementary functions is also a pre-measure. ◇

Given a pre-measure μ on a semi-algebra $\mathcal{S}$ in a set X, we can define a set function on $\mathcal{P}(X)$ by

$$\mu^*(E) = \inf\{\sum_{k=1}^{\infty} \mu(A_k) : E \subseteq \bigcup_{k=1}^{\infty} A_k, A_k \in \mathcal{S}\}.$$

This is precisely analagous to the definition of the (Lebesgue) outer measure in Chapter 1, and this set function is an outer measure for the same reason.

Proposition 3.5.2
Let $\mu : \mathcal{S} \to [0, \infty]$ be a pre-measure on a semi-algebra of sets $\mathcal{S}$ in a set X. Then the set function μ^* is an outer measure with $\mu^*(A) = \mu(A)$ for all $A \in \mathcal{S}$, and every set in the semi-algebra $\mathcal{S}$ is μ^*-measureable.

Proof:
These first of these claims is proved similarly to Proposition 1.4.1, but with the elementary sets replaced by an arbitrary semi-algebra of sets.

The remaining parts are slightly more complex than the corresponding statements in chapter 1, since we are dealing with a semi-algebra rather than the algebra of elementary sets, but the basic approach is the same.

It is immediate from the definition of μ^* that if $A \in \mathcal{S}$, then $\mu^*(A) \leq \mu(A)$. On the other hand, if $A \subseteq \bigcup_{k=1}^{\infty} A_k$ with $A_k \in \mathcal{S}$, then for each k we can find disjoint sets $A_{k,j} \in \mathcal{S}$ such that $A_k^c = A_1 \cup \cdots \cup A_{n_k}$. We define $B_1 = A_1$ and

$$B_{k,j_1,j_2,\ldots,j_{k-1}} = A_k \cap A_{1,j_1} \cap \cdots \cap A_{k-1,j_{k-1}}$$

where $1 \leq j_l \leq n_l$. All such sets are elements of $\mathcal{S}$, are pairwise disjoint, and are chosen so that

$$A_n = \bigcup_{k,j_1,j_2,\ldots,j_{k-1}} B_{k,j_1,j_2,\ldots,j_{k-1}}$$

where the union is taken over $k = 1, 2, \ldots, n$, and $1 \leq j_l \leq n_l$. Clearly then,

$$A = A \cap \bigcup_{n=1}^{\infty} A_n = A \cap \bigcup_{k,j_1,j_2,\ldots,j_{k-1}} B_{k,j_1,j_2,\ldots,j_{k-1}} = \bigcup_{k,j_1,j_2,\ldots,j_{k-1}} (A \cap B_{k,j_1,j_2,\ldots,j_{k-1}}),$$

and hence by the σ-additivity of μ,

$$\mu(A) = \sum_{k,j_1,j_2,\ldots,j_{k-1}} \mu(A \cap B_{k,j_1,j_2,\ldots,j_{k-1}}) \leq \sum_{k,j_1,j_2,\ldots,j_{k-1}} \mu(B_{k,j_1,j_2,\ldots,j_{k-1}}) = \sum_{n=1}^{\infty} \mu(A_n).$$

Taking infima over all such unions of sets A_k, we have that

$$\mu(A) \leq \mu^*(A).$$

Let E be an arbitrary set, and $A \in \mathcal{S}$. Since $\mathcal{S}$ is a semi-algebra, there are disjoint sets $A_1, \ldots, A_n \in \mathcal{S}$ such that $A^c = \bigcup_{k=1}^{n} A_k$. Now, any $\varepsilon > 0$ we have that there is acollection of sets $B_k \in \mathcal{S}$ such that $E \subseteq \bigcup_{k=1}^{\infty} B_k$, and

$$\sum_{k=1}^{\infty} \mu^*(B_k) \leq \mu^*(E) + \varepsilon.$$

Therefore, noting that the sets $B_k \cap A_j \in \mathcal{S}$ are disjoint sets and that $\mu^* = \mu$ is σ-additive on such sets

$$\begin{aligned}
\mu^*(E) + \varepsilon &\geq \sum_{k=1}^{\infty} \mu^*(B_k) \\
&= \sum_{k=1}^{\infty} \mu^*((B_k \cap A) \cup (B_k \cap A^c)) \\
&= \sum_{k=1}^{\infty} \mu^*((B_k \cap A) \cup (B_k \cap (A_1 \cup A_2 \cup \cdots \cup A_n))) \\
&= \sum_{k=1}^{\infty} \mu^*((B_k \cap A) \cup (B_k \cap A_1) \cup (B_k \cap A_2) \cup \cdots \cup (B_k \cap A_n)) \\
&= \sum_{k=1}^{\infty} (\mu^*(B_k \cap A) + \mu^*(B_k \cap A_1) + \mu^*(B_k \cap A_2) + \cdots + \mu^*(B_k \cap A_n)) \\
&= \sum_{k=1}^{\infty} \mu^*(B_k \cap A) + \sum_{k=1}^{\infty} \mu^*(B_k \cap A_1) + \sum_{k=1}^{\infty} \mu^*(B_k \cap A_2) + \cdots + \sum_{k=1}^{\infty} \mu^*(B_k \cap A_n) \\
&\geq \mu^*\left(\bigcup_{k=1}^{\infty} B_k \cap A\right) + \mu^*\left(\bigcup_{k=1}^{\infty} B_k \cap A_1\right) + \mu^*\left(\bigcup_{k=1}^{\infty} B_k \cap A_2\right) + \cdots + \mu^*\left(\bigcup_{k=1}^{\infty} B_k \cap A_n\right) \\
&\geq \mu^*(E \cap A) + \mu^*(E \cap A_1) + \mu^*(E \cap A_2) + \cdots + \mu^*(E \cap A_n) \\
&= \mu^*(E \cap A) + \mu^*(E \cap A^c).
\end{aligned}$$

Since this holds for any $\varepsilon > 0$, we conclude that $\mu^*(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c)$, and hence that A is μ^*-measurable. ■

It is worthwhile noting that sometimes we do not want to consider the the entire σ-algebra of measurable sets generated by Carathéodory's theorem; in particular the σ-algebra generated by the algebra $\mathcal{A}$ is a natural σ-algebra to consider restricting the new measure to.

These ideas are crucial in the next section when we discuss product measures.

3.5.1. Prove that the measure produced by Carathéodory's theorem is complete.

3.5.2. Prove that the set of intervals $\mathcal{I}$ in $\mathbb{R}$ is a semi-algebra (remembering that $(a, a) = \emptyset$).

3.5.3. Show that the collection of all intervals of the forms $(p, q]$, $(-\infty, q]$ and (p, ∞), where p, $q \in \mathbb{Q}$, and $p \leq q$; together with the empty set, are a semi-algebra.

3.5.4. Let $\mathcal{S}$ be an arbitrary semi-algebra. Show that the collection of sets $\mathcal{A}$ consisting of and all finite unions of elements of $\mathcal{S}$ is an algebra; indeed it is the algebra generated by $\mathcal{S}$.

3.5.5. Let $f : \mathbb{R} \to \mathbb{R}$ be a monotone increasing, right-continuous function. Show that m_f defined by $m_f((a,b]) = f(b) - f(a)$, $m_f((-\infty, b]) = f(b) - \lim_{x\to-\infty} f(x)$, and $m_f((a,\infty)) = lim_{x\to\infty} f(x) - f(a)$ is a pre-measure.

We denote by m_f the measure generated by m_f via the construction described in this section, restricted to $\mathcal{B}_\mathbb{R}$. Show that if $g : \mathbb{R} \to \mathbb{R}$ is another monotone increasing, right-continuous function such that $m_f = m_g$, then $f - g$ is constant.

3.5.6. Let μ be a measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ such that $\mu(E) < \infty$ if E is a bounded Borel set. Define

$$F(x) = \begin{cases} \mu((0,x]) & x > 0 \\ 0 & x = 0 \\ -\mu((-x,0]) & x < 0. \end{cases}$$

Show that F is a monotone increasing, right-continuous function and that $\mu = m_F$ for all Borel sets (where m_F is defined as in the previous exercise).

3.6 Product Measures

One of the reasons for discussing general measure theory was to simplify the discussion of Lebesgue measure on $\mathbb{R}^n$. We could proceed by defining the measure of "boxes" and replicating the construction of Lebesgue measure for $\mathbb{R}$. Thanks to the discussion of general outer measure above, and the fact that any resonable definition of "boxes" give a semi-algebra, this is in fact not too onerous to do. However The ideas of constructing measures on product spaces are much more general, so we will work in this more general context.

We start by defining products of measurable spaces in much the same way that we define product topologies.

Definition 3.6.1
*Let $(X_\alpha, \mathcal{M}_\alpha)$ be a collection of measurable spaces indexed by $\alpha in I$. Then we define the **product σ-algebra** on the product X of the sets X_α to be the σ-algebra generated by*

$$\{\pi_\alpha^{-1}(A_\alpha) : A_\alpha \in \mathcal{M}_\alpha, \alpha \in I\},$$

where $\pi_\alpha : X \to X_\alpha$ are the coordinate functions $\pi_\alpha(x) = x_\alpha$.

We denote this σ-algebra by

$$\bigotimes_{\alpha\in I} \mathcal{M}_\alpha$$

or $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$ for finite products.

It is worthwhile noting that the product σ-algebra is the smallest σ-algebra such that all the coordinate functions π_α are $\left(\bigotimes_{\alpha\in I} \mathcal{M}_\alpha, \mathcal{M}_\alpha\right)$-measurable.

If $\mathcal{E}_\alpha$ generates $\mathcal{M}_\alpha$, then the product σ-algebra is generated by

$$\{\pi_\alpha^{-1}(E_\alpha) : E_\alpha \in \mathcal{E}_\alpha, \alpha \in I\}.$$

This can be seen by first observing that whatever σ-algebra, $\mathcal{O}$, is generated by this collection of sets, it must be contained in the product σ-algebra, since it is a smaller generating set. On the other hand, if we consider all sets of the form $\pi_\alpha^{-1}(A)$ in $\mathcal{O}$ where $A \subset X_\alpha$, we observe that it is also a σ-algebra, and that as a result, the collection of sets $\{A \subset X_\alpha : \pi_\alpha^{-1}(A) \in \mathcal{O}\}$ is a σ-algebra on X_α which contains all the sets in $\mathcal{E}$ Therefore it contains $\mathcal{M}_\alpha$, and so $\pi_\alpha^{-1}(A) \in \mathcal{O}$ for every $A \in \mathcal{M}_\alpha$. Therefore the product σ-algebra is contained in $\mathcal{O}$.

Although this is useful in abstract proofs, this collection of sets fails to even be a semi-algebra, in general. Therefore it is not that useful for defining measures on the product σ-algebra.

Lemma 3.6.1
Let $(X_\alpha, \mathcal{M}_\alpha)$ for $\alpha \in I$ be a collection of measurable spaces, where each $\mathcal{M}_\alpha$ is generated by a semi-algebra $\mathcal{S}_\alpha$. Then the collection of sets $\mathcal{S}$ consisting of finite intersections of sets of the form $\pi_\alpha^{-1}(A_)$ for $A \in \mathcal{S}_\alpha$, and $\alpha \in I$ is a semi-algebra which generates the product σ-algebra.

Proof:

That these sets generate is a consequence of the prior discussion. We need to prove that this collection of sets is a semi-algebra.

It is immediate that the empty set is in $\mathcal{S}$, since $\emptyset = \pi_\alpha^{-1}(\emptyset)$.

It is also immediate that intersections of sets in $\mathcal{S}$ are in $\mathcal{S}$, since an intersection of two sets which are finite intersections of sets of the form $\pi_\alpha^{-1}(A)$ for $A \in \mathcal{S}_\alpha$ is still a finite intersection of such sets.

Finally, if

$$A = \bigcap_{k=1}^{n} \pi_{\alpha_k}^{-1}(A_k)$$

and $A_k^c = \bigcup_{j=1}^{m_k} B_{k,j}$ with $B_{k,j} \in \mathcal{S}_\alpha$, then

$$A^c = \bigcup_{k=1}^{n} \pi_{\alpha_k}^{-1}\left(\bigcup_{j=1}^{m_k} B_{k,j}\right) = \bigcup_{k=1}^{n}\bigcup_{j=1}^{m_k} \pi_{\alpha_k}^{-1}(B_{k,j}),$$

which is a finite union of elements of $\mathcal{S}$, as required. ■

Example 3.6.1

This Lemma implies that $\mathcal{B}_{\mathbb{R}^n} = \mathcal{B}_\mathbb{R} \otimes \cdot \otimes \mathcal{B}_\mathbb{R}$, since $\mathcal{B}_\mathbb{R}$ is generated by the semi-algebra of intervals, and $\mathcal{B}_{\mathbb{R}^n}$ is generated by the semi-algebra of n-dimensional boxes, which is precisely all finite intersections of sets of the form $\pi_k^{-1}(I) = \mathbb{R} \times \mathbb{R} \times \cdots I \times \times \times \mathbb{R}$.

On the other hand, $\mathcal{L}_{\mathbb{R}^n} \neq \mathcal{L}_{\mathbb{R}} \otimes \cdot \otimes \mathcal{L}_{\mathbb{R}}$, since if N is a non $\mathcal{L}$-measurable subset of $\mathbb{R}$, then $N \times 0 \times \cdots \times 0$ is $\mathcal{L}_{\mathbb{R}^n}$-measurable (since it is a subset of a box with measure 0), but is not $\mathcal{L}_{\mathbb{R}} \otimes \cdot \otimes \mathcal{L}_{\mathbb{R}}$-measurable, as we will see below. $\diamond$

We now wish to consider measures on product σ-algebras. The principal area of interest will be finite products, since this avoids issues arising from infinite products of measures of sets. And for clarity of exposition, we will consider the simple case of a product of two measures, since we can inductively generate any finite product from this case.

Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be measure spaces. Then the set $\mathcal{S} = \{E \times F : E \in \mathcal{M}, F \in \mathcal{N}\}$ is precisely the semi-algebra from the previous lemma given by intersecting inverse images of sets from $\mathcal{M}$ and $\mathcal{N}$. In fact, one can show that it is an algebra.

We define a set function $\mu \otimes \nu$ on this semi-algebra by letting

$$(\mu \otimes \nu)(E \times F) = \mu(E)\nu(F).$$

Proposition 3.6.2
Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be measure spaces. Then the set function $\mu \otimes \nu$ defined on the algebra $\{E \times F : E \in \mathcal{M}, F \in \mathcal{N}\}$ is a pre-measure.

Proof:
If we have a set $E \times F$ which is itself a countable disjoint union of rectangles $E_k \times F_k$, then we observe that

$$\chi_{E\times F}(x, y) = \chi_E(x)\chi_F(y),$$

but also

$$\chi_{E\times F}(x, y) = \sum \chi_{E_k\times F_k}(x, y) = \sum \chi_{E_k}(x)\chi_{F_k}(y).$$

Holding y constant and integrating using the corollary of the monotone convergence theorem for sums, we get

$$\begin{aligned}\mu(E)\chi_F(y) &= \int \chi_E(x)\chi_F(y)\, d\mu(x)\\ &= \sum \int \chi_{E_k}(x)\chi_{F_k}(y)\, d\mu(x) = \sum \mu(E_k)\chi_{F_k}(y).\end{aligned}$$

Integrating again with the MCT, this time with respect to y, gives

$$\mu(E)\nu(F) = \sum \mu(E_k)\nu(F_k),$$

and hence that $\mu \times \nu$ is σ-additive on the semi-algebra.

The observation that $\mu \times \nu(\emptyset) = 0$ means that $\mu \times \nu$ is a premeasure. ■

Using the facts about pre-measures from the previous section, we can define an measure from $\mu \times \nu$, and the collection of $(\mu \times \nu)^*$-measurable sets includes all sets in the algebra, and therefore also contains all sets in $\mathcal{M} \otimes \mathcal{N}$. We define

$\mu \times \nu$ to be this outer measure restricted to $\mathcal{M} \otimes \mathcal{N}$. We call this measure the ***product measure*** of μ and ν.

Note that this means that the product measure will not be complete, in general.

Recall that a measure space $(X, \mathcal{M}, \mu)$ is ***σ-finite*** if there is a countable cover of X by measurable sets E_k such that $\mu(E_k) < \infty$. If both $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ-finite, then $(M \times N, \mathcal{M} \otimes \mathcal{N}, \mu \times \nu)$ is σ-finite, since if E_i cover X with $\mu(E_i) < \infty$ and F_j cover Y with $\nu(F_j) < \infty$, then $E_i \times F_j$ cover $X \times Y$ and have finite product measure.

It can be shown that if $\mu \times \nu$ is σ-finite, then it is the only measure on $\mathcal{M} \otimes \mathcal{N}$ such that

$$\mu \times \nu(E \times F) = \mu(E)\nu(F).$$

Given a set $E \subseteq X \times Y$ and two points $x \in X$ and $y \in Y$, we define the ***x-section*** and the ***y-section*** to be, respectively,

$$E_{1,x} = \{y \in Y : (x, y) \in E\} \qquad E_{2,y} = \{x \in X : (x, y) \in E\}.$$

Similarly for a function on $X \times Y$, we define

$$f_{1,x}(y) = f_{2,y}(x) = f(x, y)$$

to be the ***x-section*** and ***y-section*** of f.

Proposition 3.6.3
If $E \in \mathcal{M} \otimes \mathcal{N}$, then $E_{1,x} \in \mathcal{N}$ and $E_{2,y} \in \mathcal{M}$ for all x and y.

If f is $\mathcal{M} \otimes \mathcal{N}$-measurable, then $f_{1,x}(y)$ is $\mathcal{N}$-measurable and $f_{2,y}(x)$ is $\mathcal{M}$-measurable.

Proof:

Consider the collection of sets

$$\mathcal{R} = \{E \subseteq X \times Y | E_{1,x} \in \mathcal{N} \text{ and } E_{2,y} \in \mathcal{M} \text{ for all } x \in X, y \in Y\}.$$

Clearly all rectangles are in $\mathcal{R}$. Furthermore, since

$$\left(\bigcup_{k=1}^{n} E_k \right)_{i,z} = \bigcup_{k=1}^{n} (E_k)_{i,z} \qquad \text{and} \qquad (E^c)_{i,z} = (E_{i,z})^c$$

for $i = 1, 2$ and $z \in X$ or Y (depending on the value of i), the set $\mathcal{R}$ is a σ-algebra. Hence $\mathcal{M} \otimes \mathcal{N} \subseteq \mathcal{R}$.

The second part follows immediately from the fact that

$$(f_{i,z})^{-1}(E) = (f^{-1}(E))_{i,z}$$

for $i = 1, 2$ and $z \in X$ or Y, depending on the value of i. ■

Example 3.6.2

This proposition demonstrates that if $N \subseteq \mathbb{R}$ is not Lebsgue meaureable, then $E = N \times \{0\}$ is not $\mathcal{L}_{\mathbb{R}} \otimes \mathcal{L}_{\mathbb{R}}$-measurable, since the y-section at $y = 0$ is N, contradicting the conclusion of the proposition.

However, it is contained in $\mathbb{R} \times \{0\}$, and $(m \otimes m)^*(\mathbb{R} \times \{0\}) = 0$ so E is an $(m \otimes m)^*$-measurable set. Therefore it is $\mathcal{L}_{\mathbb{R}^2}$-measurable. ◇

The above example demonstrates that there are some definite trade-offs involved in restricting to the product σ-algebra, and also the earlier claim that the product measure will generally not be complete.

We need to obtain some way of finding the measure of sets which are not rectangles. Before we can do this, however, we need a technical lemma which gives an alternative characterisation of σ-algebras generated by algebras.

A ***monotone class*** $\mathcal{C}$ is a collection of subsets of a set X such that if $E_1 \subseteq E_2 \subseteq E_3 \subseteq \dots$ is an increasing sequence of sets in $\mathcal{C}$, then the union of the E_n lies in $\mathcal{C}$ and if $F_1 \supseteq F_2 \supseteq F_3 \supseteq \dots$ is a decreasing sequence of sets in $\mathcal{C}$, then the intersection of the F_n also lies in $\mathcal{C}$. Clearly all σ-algebras are monotone classes, but the converse is false.

Given a collection of monotone classes, their intersection is again a monotone class, so given a collection $\mathcal{E}$ of subsets of X, we define the monotone class generated by $\mathcal{E}$ to be the intersection of all monotone classes containing $\mathcal{E}$.

Lemma 3.6.4

If $\mathcal{A}$ is an algebra of sets in X, then the monotone class generated $\mathcal{C}$ by $\mathcal{A}$, and the σ-algebra $\mathcal{M}$ generated by $\mathcal{A}$ are identical.

Proof:

Firstly, since $\mathcal{M}$ is a monotone class containing $\mathcal{A}$, we have $\mathcal{C} \subseteq \mathcal{M}$.

Now for a set $E \in \mathcal{C}$, let

$$\mathcal{C}(E) = \{F \in \mathcal{C} | E \setminus F, F \setminus E \text{ and } E \cap F \in \mathcal{C}\}.$$

It is immediate by symmetry of the definition, that if $F \in \mathcal{C}(E)$, then $E \in \mathcal{C}(F)$. A little work shows that $\mathcal{C}(E)$ is a monotone class. For example, if F_n is an increasing sequence of sets in $\mathcal{C}(E)$,

$$E \setminus \left(\bigcup F_n\right) = E \cap \left(\bigcup F_n\right)^c = E \cap \left(\bigcap F_n^c\right) = \bigcap (E \cap F_n^c) = \bigcap (E \setminus F_n)$$

and $E \setminus F_n$ is a decreasing sequence of sets in $\mathcal{C}$, so the intersection is also in $\mathcal{C}$. Similar arguments work for all the other cases which need to be checked.

If $E \in \mathcal{A}$, then for any $F \in \mathcal{A}$ it is easy to see that $E \setminus F$, $F \setminus E$ and $E \cap F$ are all elements of $\mathcal{A}$, and hence $\mathcal{C}$. Therefore $\mathcal{C} \subseteq \mathcal{C}(E)$, since it is a monotone class containing $\mathcal{A}$. So for any $F \in \mathcal{C}$, $F \in \mathcal{C}(E)$ and so $E \in \mathcal{C}(F)$. So $\mathcal{A} \subseteq \mathcal{C}(F)$ for any $F \in \mathcal{C}$, and therefore $\mathcal{C} \subseteq \mathcal{C}(F)$. Hence $\mathcal{C}(F) = \mathcal{C}$.

Since $X \in \mathcal{A}$, $X \in \mathcal{C}$, and so for any $F \in \mathcal{C}$, $F^c = X \setminus F \in \mathcal{C}(F) = \mathcal{C}$. Also, if E_k is a sequence of sets in $\mathcal{C}$, let $F_n = \bigcup_{k=1}^n E_k$. F_n is then an increasing sequence of sets, and

$$\bigcap E_k = \bigcap F_n,$$

which is an element of $\mathcal{C}$. Therefore $\mathcal{C}$ is a σ-algebra which contains $\mathcal{A}$. Hence $\mathcal{M} \subseteq \mathcal{C}$.

Hence $\mathcal{C} = \mathcal{M}$. ■

Another way of thinking about this result is that it says that if $\mathcal{E}$ is a collection of sets and $\mathcal{A}$ is the algebra generated by $\mathcal{E}$, then the σ-algebra generated by $\mathcal{E}$ coincides with the monotone class generated by $\mathcal{A}$. The power of monotone classes come from the fact that, at least in finite measure spaces, increasing and decreasing sequences of sets give us monotone increasing and decreasing families of characteristic functions, and the union and intersection of the families correspond to the pointwise limit. The convergence theorems will then allow us conclude that certain classes of measurable sets are monotone classes, as we will see.

Theorem 3.6.5

If $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ-finite measure spaces, and $E \in \mathcal{M} \otimes \mathcal{N}$, then the functions

$$x \mapsto \nu(E_{1,x}) \qquad \text{and} \qquad y \mapsto \mu(E_{2,y})$$

are measurable on X and Y respectively. Furthermore,

$$\mu \times \nu(E) = \int \nu(E_{1,x})\ d\mu(x) = \int \nu(E_{2,y})\ d\nu(y).$$

Proof:

First assume that μ and ν are finite measures. Let $\mathcal{R}$ be the collection of sets $E \subseteq X \times Y$ such that

$$x \mapsto \nu(E_{1,x}) \qquad \text{and} \qquad y \mapsto \mu(E_{2,y})$$

are measurable on X and Y respectively, and that

$$\mu \times \nu(E) = \int \nu(E_{1,x})\ d\mu(x) = \int \nu(E_{2,y})\ d\nu(y).$$

We wish to show that this collection must include $\mathcal{M} \otimes \mathcal{N}$.

Clearly, all rectangles are in $\mathcal{R}$, for if $E = A \times B$, then

$$\nu(E_{1,x}) = \chi_A(x)\nu(B) \qquad \text{and} \qquad \mu(E_{2,y}) = \mu(A)\chi_B(y).$$

By additivity it follows that all finite disjoint unions of rectangles must also be in $\mathcal{R}$.

Now let E_n be a sequence of sets in $\mathcal{R}$ and let $E = \bigcup_n E_n$. Then the functions

$$f_n(y) = \mu((E_n)_{2,y})$$

form an increasing sequence which converges pointwise to $f(y) = \mu(E_{2,y})$, and so by the monotone convergence theorem, $f(y)$ is measurable, and

$$\int \mu(E_{2,y})\ d\nu(y) = \lim_{n\to\infty} \int f_n\ d\nu = \lim_{n\to\infty} \mu \times \nu(E_n) = \mu \times \nu(E).$$

Similarly, $\nu(E_{1,x})$ is measurable and

$$\int \mu(E_{1,x})\ d\mu(x) = \mu \times \nu(E).$$

Hence $E \in \mathcal{R}$.

So $\mathcal{R}$ is a monotone class containing the algebra of finite unions of rectangles. Hence, by the technical lemma above, $\mathcal{M} \otimes \mathcal{N} \subseteq \mathcal{R}$.

For the σ-finite case, we can write $X \times Y$ as the union of an increasing sequence of rectangles $X_j \times Y_j$ of finite measure. Given $E \in \mathcal{M} \otimes \mathcal{N}$, We can apply the result for finite measures to $E \cap X_j \times Y_j$, to get

$$\mu\times\nu(E\cap X_j\times Y_j) = \int \chi_{X_j}(x)\nu(E_{1,x}\cap Y_j)\ d\mu(x) = \int \chi_{Y_j}(y)\nu(E_{2,y}\cap X_j)\ d\nu(y).$$

Applying the monotone convergence theorem is then sufficient to give the general result. ■

Now that we have a good grasp on the nature of the product measure, at least in the σ-finite case, we can now proceed to calculate the integral of functions on the product space. As with the Riemann integral, it is most useful to be able to calculate the integral with respect to the product measure in terms of a double integral, and the following result tells us that this is OK.

Theorem 3.6.6 (Fubini-Tonelli)
Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ-finite measure spaces.

(Tonelli) If $f \in \mathcal{M}^+(X \times Y)$ then $g(x) = \int f_{1,x}\ d\nu$ and $h(y) = \int f_{2,y}\ d\mu$ are in $\mathcal{M}^+(X)$ and $\mathcal{M}^+(Y)$ respectively, and

$$\int f\ d(\mu \times \nu) = \int \left(\int f(x,y)\ d\nu(y)\right)\ d\mu(x) = \int \left(\int f(x,y)\ d\mu(x)\right)\ d\nu(y).$$

(Fubini) If $f \in L^1(\mu \times \nu)$ then $f_{1,x} \in L^1(\nu)$ for almost every x and $f_{2,y} \in L^1(\mu)$ for almost every y, and the functions $g(x) = \int f_{1,x}\ d\nu$ and $h(y) = \int f_{2,y}\ d\mu$, which are defined almost everywhere, are in $L^1(\mu)$ and $L^1(\nu)$ respectively. Moreover,

$$\int f\ d(\mu \times \nu) = \int \left(\int f(x,y)\ d\nu(y)\right)\ d\mu(x) = \int \left(\int f(x,y)\ d\mu(x)\right)\ d\nu(y).$$

We will normally drop the parentheses in the double integral.

Proof:

Tonelli's theorem is clearly true for characteristic functions by Theorem 3.6.5. By the linearity of the integral, it immediately extends to non-negative simple functions.

For general $f \in L^+(X\times Y)$, we can always find a sequence of simple functions φ_n which increase pointwise to f. For $x \in X$ and $y \in Y$, the section functions

$(\varphi_n)_{1,x}$ and $(\varphi_n)_{2,y}$ are increasing sequences of simple functions which converge pointwise to $f_{1,x}$ and $f_{2,y}$ respectively. Hence if

$$g_n(x) = \int (\varphi_n)_{1,x}\ d\nu \qquad \text{and} \qquad h_n(x) = \int (\varphi_n)_{2,y}\ d\mu$$

the monotone convergence theorem tells us that g_n converges to g pointwise, and h_n converges to h pointwise. Moreover, we know that each g_n and h_n is measurable, so g and h must be measurable. Furthermore, the sequences g_n and h_n are themselves monotone increasing, so applying the monotone convergence theorem again, gives

$$\int g\ d\mu = \lim \int g_n\ d\mu = \lim \int \varphi_n\ d(\mu \times \nu) = \int f\ d(\mu \times \nu)$$

and similarly for h and h_n, whence Tonelli's result.

Moreover, if $\int f\ d(\mu \times \nu) < \infty$, we must have $g < \infty$ and $h < \infty$ almost everywhere. This implies that if $f \in L^1$, then $f_{1,x} \in L^1$ for almost every x and $f_{2,y} \in L^1$ for almost every y. Fubini's result then follows by applying Tonelli's theorem to the positive and negative parts of the real and imaginary parts of f. ■

A common situation in which we want to use these theorems is when we want to swap the order of integration in a double integral, ie. to turn

$$\int \int f\ d\mu\ d\nu \qquad \text{into} \qquad \int \int f\ d\nu\ d\mu.$$

The typical strategy is to first show that $f \in L^1$ using Tonelli's theorem, ie. that

$$\int |f|\ d(\mu \times \nu) = \int \int |f|\ d\mu\ d\nu < \infty,$$

and then use Fubini's theorem to swap the integral. We can *only* swap the integral once we know that f is in L^1.

Example 3.6.3

The ***convolution*** of two functions plays an important role in the theory of the Fourier transform. We define the convolution of f and g to be

$$(f * g)(x) = \int f(x - y)g(y)\ dm(y)$$

If f, $g \in L^1$ we can conclude that $f * g \in L^1$ from the Fubini-Tonelli theorem. First, we note that $|f(x - y)g(y)|$ is measurable as a function of x and y, since the products and compositions of measurable functions are measurable, and $(x, y) \mapsto x - y$ is continuous and thus measurable. Therefore Tonelli's theorem applies and so

$$
\begin{aligned}
\int |f * g| \; dm &= \int \left| \int f(x-y)g(y) \; dm(y) \right| \; dm(x) \\
&\leq \int \int |f(x-y)g(y)| \; dm(y) \; dm(x) \\
&= \int \int |f(x-y)||g(y)| \; dm(x) \; dm(y) \qquad \text{(Tonelli)} \\
&= \int |g(y)| \int |f(x-y)| \; dm(x) \; dm(y) \\
&= \int |g(y)| \int |f(x)| \; dm(x) \; dm(y) \\
&= \int |f(x)| \; dm(x) \int |g(y)| \; dm(y) < \infty
\end{aligned}
$$

An example of the use of Fubini's theorem can be seen in showing that convolution is an associative operation, ie. $(f * g) * h = f * (g * h)$ almost everywhere for f, g and $h \in L^1$. We note that

$$
\begin{aligned}
((f * g) * h)(x) &= \int \int f((x-z)-y)g(y) \; dm(y)h(z) \; dm(z) \\
&= \int \int f(x-z-y)g(y)h(z) \; dm(y) \; dm(z) \\
(f * (g * h))(x) &= \int f(x-y) \int g(y-z)h(z) \; dm(z) \; dm(y) \\
&= \int \int f(x-y)g(y-z)h(z) \; dm(z) \; dm(y).
\end{aligned}
$$

Now since f, g and h are measurable, then for fixed x so is $|f(x-y)g(y-z)h(z)|$, regarded as a function from $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$ with y and z the variables. Tonelli's theorem then tells us that

$$
\begin{aligned}
&\int |f(x-y)g(y-z)h(z)| \; d(m \times m)(y,z) \\
&\qquad = \int \int |f(x-y)g(y-z)h(z)| \; dm(z) \; dm(y) \qquad \text{(Tonelli)} \\
&\qquad = \int \int |f(x-y)||g(y-z)||h(z)| \; dm(z) \; dm(y) \\
&\qquad = (|f| * (|g| * |h|))(x)
\end{aligned}
$$

which is finite for almost every x, since $|f| * (|g| * |h|)$ is an L^1 function by the prior discussion. Hence we have that $f(x-y)g(y-z)h(z)$ is in $L^1(\mathbb{R} \times \mathbb{R})$ for almost every x, and hence Fubini's theorem applies. So

$$\begin{aligned}
(f*(g*h))(x) &= \int\int f(x-y)g(y-z)h(z)\ dm(z)\ dm(y) \\
&= \int\int f(x-y)g(y-z)h(z)\ dm(y)\ dm(z) && \text{(Fubini)} \\
&= \int\int f(x-(y+z))g(y)h(z)\ dm(y)\ dm(z) && \text{(translation invariance)} \\
&= \int\int f(x-z-y)g(y)h(z)\ dm(y)\ dm(z) \\
&= ((f*g)*h)(x)
\end{aligned}$$

for almost every x. ◇

Example 3.6.4

The above is a particular example of a general situation involving what are known as integral kernels. Let $K(x,y)$ be $\mathcal{M}\otimes\mathcal{N}$-measurable on $X\times Y$ (assume everything is σ-finite). If there is a constant C with

$$\int |K(x,y)|\ d\mu(x) < C$$

for almost every $y \in Y$, and $f \in L^1(\nu)$, then the function

$$g(x) = \int K(x,y)f(y)\ d\nu(y)$$

is in L^1 and

$$\int |g|\ d\mu \leq C \int |f|\ d\nu.$$

◇

Exercises

3.6.1. Let $(X,\mathcal{M},\mu)$ be a σ-finite measure space, and let $f \in \mathcal{L}^+(X)$. Then let

$$G_f = \{(x,y) \in X \times [0,\infty] : 0 \leq y \leq f(x)\}.$$

This set can be theought of as the region underneath the graph of f.

Show that G_f is $\mathcal{M}\otimes\mathcal{B}_{\mathbb{R}^*}$-measurable and that

$$\mu \times m(G_f) = \int f\ d\mu.$$

Hint 1: the map $\varphi(x,y) = (f(x),y)$ is $(\mathcal{M}\otimes\mathcal{B}_{\mathbb{R}},\mathcal{B}_{\mathbb{R}^2})$-measurable, $\psi(z,y) = z-y$ is continuous, and show $G_f = (\psi\circ\varphi)^{-1}([0,\infty))$.

Hint 2: use Tonelli's theorem on χ_{G_f}.

Note: This example shows that the integral of f is the area under the graph in the most general sense.

3.6.2. (†) Use Tonelli's theorem to verify the claims of Example 3.6.4.

3.7 Signed and Complex Measures

Both in applications and in later theory, we want to be able to generalise the notion of a measure to allow a set to have negative or complex measure. We can do this, but we must be a little more cautious with infinities, just as we had to when moving from nonnegative measurable functions to real- and complex-valued functions.

Concept: *Intuitively, it may help to visualise signed measures using electrical charge. The net charge in a region of space gives a signed measure: sometimes it is positive, sometimes it is negative, and occasionally the positive and negative charges in a region balance out and give a net charge of* 0.

Let $(X, \mathcal{M})$ be a measurable space. A ***signed measure*** on $(X, \mathcal{M})$ is a function $\nu : \mathcal{M} \to [-\infty, \infty]$ such that

1. $\nu(\emptyset) = 0$,
2. If E_k is a sequence of disjoint sets in $\mathcal{M}$, then

$$\nu\left(\bigcup_k E_k\right) = \sum_k \nu(E_k)$$

   and if the measure on the left is finite, the sum converges absolutely,

3. ν assumes at most one of the values $\pm\infty$.

These conditions hold for measures, as defined before, so every measure is a signed measure. To prevent confusion, we will sometimes refer to measures as ***positive*** measures.

Terminology: *note that "positive," and "negative" are defined to allow* 0 *as a possible value, so null sets are also positive and negative. This definition is made largely to avoid having to say "non-negative sets" and "non-positive sets" frequently.*

If ν is a signed measure, we will say that a set $E \in \mathcal{M}$ is ***positive***, ***negative*** or ***null*** if given any measurable set $F \subseteq E$, respectively, $\nu(F) \geq 0$, ≤ 0 or $= 0$. If $\mu(E) \neq \pm\infty$ for all measurable sets E, then we say μ is a ***finite*** signed measure.

A ***complex measure*** on $(X, \mathcal{M})$ is a function $\nu : \mathcal{M} \to \mathbb{C}$ such that

1. $\nu(\emptyset) = 0$,
2. If E_k is a sequence of disjoint sets in $\mathcal{M}$, then

$$\nu\left(\bigcup_k E_k\right) = \sum_k \nu(E_k)$$

   and the sum converges absolutely.

We do not allow complex measures to take on infinite values. Again it is easy to see that finite measures, and signed measures which are not infinite on any set, are complex measures. Given a complex measure ν, we can immediately define signed measures from the real and imaginary parts of ν, ie.

$$\nu_r(E) = \operatorname{Re}\nu(E) \qquad \text{and} \qquad \nu_i(E) = \operatorname{Im}\nu(E).$$

We call these the real and imaginary parts of ν, and clearly

$$\nu(E) = \nu_r(E) + i\nu_i(E).$$

Just as with any function whose domain is $\mathbb{C}$, we can write this more simply as $\nu = \nu_r + i\nu_i$.

Example 3.7.1

If μ is a positive measure and $f : X \to \mathbb{R}$ is an L^1 function, then

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

is a finite signed measure. Similarly, if $f : X \to \mathbb{C}$ is in L^1, then

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

is a complex measure. $\diamond$

Example 3.7.2

It is easy to construct signed and complex measures from positive measures. If μ_1 and μ_2 are two positive measures, at least one of which is a finite measure, then $\nu = \mu_1 - \mu_2$ is a signed measure. If, on the other hand, ν_1 and ν_2 are finite signed measures, $\nu = \nu_1 + i\nu_2$ is a complex measure. $\diamond$

It turns out that these constructions are generic: every signed and complex measure can be represented in this way.

Proposition 3.7.1

Let ν be a signed or complex measure on $(X, \mathcal{M})$. If E_k is an increasing sequence of measurable sets with $\nu(E_k) > -\infty$ eventually if ν is a signed measure, then $\nu(\bigcup E_k) = \lim \nu(E_k)$.

On the other hand, if E_k is a decreasing sequence of measurable sets with $\nu(E_k) < \infty$ eventually if ν is a signed measure, then $\nu(\bigcap E_k) = \lim \nu(E_k)$.

The proof of this proposition is essentially the same as the proof of the corresponding fact for measures.

Corollary 3.7.2

If ν is a signed measure and P_k are positive sets for all k, then $\bigcup_k P_k$ is a positive set.

Proof:

Let $Q_n = P_n \setminus \bigcup_{k=1}^{n-1} P_k$. Then $Q_n \subseteq P_n$, so Q_n is positive. Then given any $E \subseteq \bigcup P_k$, we have $E = \bigcup (E \cap Q_k)$ and this is a disjoint union, so $\nu(E) = \sum \nu(E \cap Q_k) \geq 0$. ■

We can now show that a signed measure has distinct "regions" where it is positive and negative.

Theorem 3.7.3 (Hahn Decomposition Theorem)
If ν is a signed measure on $(X, \mathcal{M})$ then X can be split into disjoint positive and negative sets P and N. Furthermore, if P' and N' are another such pair, then $P\Delta P'$ (which equals $N\Delta N'$) is null.

Proof:
Without loss of generality, assume that $\nu(A) \neq +\infty$ for any $A \in \mathcal{M}$; otherwise use $-\nu$ instead and reverse the P and N that are obtained.

The first step is to observe that we can always find an "almost positive" subset of any set: given any measurable set A with $\nu(A) > -\infty$, and any $\varepsilon > 0$, we can find a set $B \subseteq A$ such that $\nu(B) \geq \nu(A)$ and $\mu(E) > -\varepsilon$ for all $E \subseteq B$. If this is not true, we can let $A_1 = A$, and find, inductively, $E_n \subset A_n \subseteq A$ such that $\nu(E_n) \leq -\varepsilon$ and $A_{n-1} = A_n \setminus E_n$. This follows from the simple observation that $\nu(A_{n-1}) \geq \nu(A_n) - \nu(E_n) > \nu(A_n) + \varepsilon \geq \nu(A_n)$. Now the sets E_n are disjoint, so if $E = \bigcup E_n$, $\nu(E) = -\infty$ and so, $\nu(A \setminus E) = \nu(A) - \nu(E) = \infty$, contradicting our initial assumption.

We now refine this result to produce a positive subset: given any measurable set A with $\nu(A) > -\infty$, there is a positive set $P \subseteq A$ with $\nu(P) \geq \nu(A)$. Letting $A_1 = A$, we again proceed inductively. Given $A_1, \ldots, A_{n-1}$, we can find $A_n \subset A_{n-1}$ such that $\nu(A_n) \geq \nu(A_{n-1})$ and $\nu(E) > -1/n$ for all $E \subseteq A_n$. Letting $P = \bigcap A_n$, we have that $\nu(P) = \lim_{n\to\infty} \nu(A_n) \geq \nu(A)$, and any subset E of P is contained in A_n for all n, so $\nu(E) > -1/n$ for all n, and hence $\nu(E) \geq 0$.

Now let $m = \sup\{\nu(A) : A \in \mathcal{M}\}$. Since $\nu(\emptyset) = 0$, $m \geq 0$ and so we can find $A_n \in \mathcal{M}$ with $\nu(A_n) \to m$. Letting $P_n \subseteq A_n$, with P_n positive and $\nu(P_n) \geq \nu(A_n)$, we now have a sequence of positive sets whose measures converge to m by the sandwich theorem. Letting $P = \bigcup P_n$, we have that P is positive and $\nu(P) = \lim \nu(P_n) = m$.

Now letting $N = P^c$, we observe that N is a negative set, since if $E \subseteq N$ with $\nu(E) > 0$, $\nu(P \cup E) = \nu(P) + \nu(E) > m$ which is a contradiction. This gives the result.

Given any other such pair, we simply observe that $P \setminus P' \subseteq P$ and $P \setminus P' \subseteq N'$, so $P \setminus P'$ is both positive and negative, and hence must be null. Similarly $P' \setminus P$ is null, so $P\Delta P' = (P \setminus P') \cup (P' \setminus P)$ is also null. ■

Such a decomposition into sets P and N is called a ***Hahn decomposition*** for the measure ν.

Example 3.7.3
Let μ be a positive measure on $(X, \mathcal{M})$, $f \in L^1(\mu)$ and $\nu(E) = \int_E f \, d\mu$ as in Example 3.7.1. Then if we let $P = \{x \in X : f(x) \geq 0\}$ and $N = \{x \in X : f(x) \leq 0\}$ then we have that $X = P \cup N$ and $P \cap N = \emptyset$, and moreover if $E \in \mathcal{M}$ with $E \subseteq P$, then $f \geq 0$ on E and so $\nu(E) = \int_E f \, d\mu \geq 0$. Similarly if $E \subseteq N$ we have that $\nu(E) \leq 0$.

Hence this P and N are a Hahn decomposition for ν. ◇

We say $\nu \geq \mu$ on a set A if $\nu(E) \geq \mu(E)$ for all $E \subseteq A$, or equivalently, if A is a positive set for the measure $\nu - \mu$.

Definition 3.7.1
If μ and ν are two positive measures on $(X, \mathcal{M})$, we say that μ and ν are ***mutually singular****, and write $\mu \perp \nu$, if there are sets E and $F \in \mathcal{M}$ which are disjoint, with $E \cup F = X$, and for which $\mu(E) = 0$ and $\nu(F) = 0$.*

More generally, if μ and ν are signed or complex measures, then we insist that E be μ-null and F be ν-null.

Sometimes we will say that ν is singular with respect to μ or vice-versa when $\mu \perp \nu$. Informally we think of mutual singularity as saying that μ and ν are supported on disjoint sets.

Example 3.7.4
The Lebesgue measure m and the Dirac point mass δ_{x_0} on $(\mathbb{R}, \mathcal{L})$ are mutually singular, with $A = \mathbb{R} \setminus \{x_0\}$ and $B = \{x_0\}$. $\diamond$

Example 3.7.5
Let μ be a positive measure on $(X, \mathcal{M})$, f, $g \in L^1(\mu)$ with $\nu(E) = \int_E f \, d\mu$, and $\rho(E) = \int_E g \, d\mu$ as in Example 3.7.1.

If f and g have the property that for all $x \in X$ either $f(x) = 0$ or $g(x) = 0$ (or both), then if we let $E = \{x \in X : f(x) \neq 0\}$ and $F = X \setminus E = \{x \in X : f(x) = 0\}$, then $g = 0$ on E, and so $\rho(E) = 0$ and $f = 0$ on F so that $\nu(F) = 0$. So in this case we have $\nu \perp \rho$. $\diamond$

Lemma 3.7.4
If μ and ν are finite measures on $(X, \mathcal{M})$, then either $\mu \perp \nu$ or there is some $E \in \mathcal{M}$ and $\varepsilon > 0$, such that $\mu(E) > 0$ and $\nu \geq \varepsilon\nu$ on E.

Proof:
Consider the measures $\nu - n^{-1}\mu$. These have Hahn decompositions P_n and N_n, and let $P = \bigcup P_n$ and $N = \bigcap N_n = P^c$. Then N is negative for all the measures $\nu - n^{-1}\mu$, and so $0 \leq \nu(N) \leq n^{-1}\mu(N)$ for all n. Hence $\nu(N) = 0$. If $\mu(P) = 0$, then $\mu \perp \nu$, as N and P are the sets required by the definition. On the other hand, if $\mu(P) > 0$ then $\mu(P_n) > 0$ for some n and P_n is a positive set for $\nu - n^{-1}\mu$, ie. $\nu \geq n^{-1}\mu$ on P_n. ■

The Hahn decomposition together with the concept of mutual signualrity allow us to decompose a signed measure into positive and negative parts.

Theorem 3.7.5 (Jordan Decomposition)
If ν is a signed measure, there are unique positive measures ν_+ and ν_- such that $\nu = \nu_+ - \nu_-$ and $\nu_+ \perp \nu_-$.

Proof:

Let $X = P \cup N$ be a Hahn decomposition for ν, and define

$$\nu_+(A) = \nu(A \cap P) \qquad \nu_- = -\nu(B \cap N).$$

Then $\nu(A) = \nu_+(A) - \nu_-(A)$, and P and N are sets which show that $\nu_+ \perp \nu_-$.

If μ_+ and μ_- are two other measures satisfying the conculsion, let E and F be disjoint sets so that μ_+ is null on E and μ_- is null on F, and $E \cup F = X$. Then E and F are a Hahn decomposition for ν, so $P \Delta N$ is ν-null and so

$$\mu_+(A) = \mu_+(A \cap E) = \nu(A \cap E) = \nu(A \cap P) = \nu_+(A),$$

and similarly for μ_- and ν_-. ■

The measures ν_+ and ν_- are the positive and negative variations (or parts) of ν, $\nu = \nu_+ - \nu_-$ is called the Jordan decomposition of ν, and we define the total variation of ν to be

$$|\nu| = \nu_+ + \nu_-.$$

Note that $|\nu|(E) \neq |\nu(E)|$ in general.

Example 3.7.6

Let μ be a positive measure on $(X, \mathcal{M})$, $f \in L^1(\mu)$ and $\nu(E) = \int_E f \; d\mu$ as in Example 3.7.1.

Then if we let $\nu_1(E) = \int_E f^+ \; d\mu$, and $\nu_2(E) = \int_E f^- \; d\mu$, it is immediate that $\int_E f \; d\mu = \int_E f^+ \; d\mu - \int_E f^- \; d\mu$, so that $\nu(E) = \nu_1(E) - \nu_2(E)$. But moreover for every $x \in X$ we have that $f^+(x) = 0$ or $f^-(x) = 0$, since f cannot be both positive and negative at the same time, we have $\nu_1 \perp \nu_2$ by Example 3.7.5.

So by the uniqueness of the Jordan decomposition, $\nu_1 = \nu_+$ and $\nu_2 = \nu_-$. The total variation is then

$$|\nu|(E) = \nu_+(E) + \nu_-(E) = \int_E f^+ + f^- \; d\mu = \int |f| \; d\mu.$$

◇

We can easily see that A is ν-null iff $|\nu|(A) = 0$, and $\nu \perp \mu$ iff $|\nu| \perp \mu$ iff $\nu_+ \perp \mu$ and $\nu_- \perp \mu$. Also ν is finite (ie. $\nu(A) \neq \pm\infty$ for any A) iff $|\nu|$ is finite iff ν_+ and ν_- ar finite.

Finally, the Jordan decomposition gives us an easy definition of integration with respect to a signed measure. We simply let

$$\int f \; d\nu = \int f \; d\nu_+ - \int f \; d\nu_-,$$

where f is in $L^1(\nu_+) \cap L^1(\nu_-) = L^1(\nu)$.

Similarly for a complex measure, we define

$$\int f \; d\nu = \int f \; d\nu_r + i \int f \; d\nu_i$$

where ν_r and ν_i are the real and imaginary parts of ν.

3.7.1. (†) In this chapter we used the fact that if μ and ν are positive measures on $(X, \mathcal{M})$ then so is the set function $\rho = \mu + \nu$ defined in the obvious way by $\rho(E) = \mu(E) + \nu(E)$.

Verify that ρ is in fact a measure.

3.7.2. (†) Show that if μ is a signed measure on X, then $\mu(X) \neq \pm\infty$ if and only if μ is a finite signed measure.

3.7.3. (†) Show that if ν is a signed measure on $(X, \mathcal{M})$ and μ_1 and μ_2 are positive measures on the same space such that $\nu = \nu_1 - \nu_2$, then $\nu_1 \geq \nu_+$ and $\nu_2 \geq \nu_-$.

3.8 The Radon-Nikodym Derivative

In classical calculus it is common to use substitution to change the variable of integration. For the Riemann integral this is typically stated something like

$$\int_a^b f(u(x))u'(x)\ dx = \int_{u(a)}^{u(b)} f(x)\ dx,$$

and is a consequence of the chain rule and the Fundamental Theorem of Calculus. Informally it is common to say "$du = u'dx$" in this situation, and it is the purpose of this section to try to generalize this sort of situation to measure theory.

Consider a differentiable function u on reals with $u' \geq 0$. Then we can define a measure using u by letting $\mu(I) = u(b) - u(a)$ on the pre-algebra of all intervals, and extending using outer measure to define a measure on Lebesgue measurable sets. Then we have

$$\mu([a, b]) = \int_{[u(a),u(b)]} d\mu = \int_{[a,b]} u'\ dm = m_{u'}([a, b]).$$

So we conclude that $\mu = m_{u'}$, but furthermore, if E is measurable,

$$\int \chi_E\ d\mu = \mu(E) = m_{u'}(E) = \int \chi_E u'\ dm,$$

from whence we get, via measurable simple functions and the monotone convergence theorem, that if f is a μ-integrable function, then

$$\int f\ d\mu = \int f u'\ dm.$$

In this situation it would seem very natural to write $d\mu = u'\ dm$, or $\frac{d\mu}{dm} = u'$. This is the prototypical example of what we will call a Radon-Nikodym derivative, but we can generalize this situation much further.

We cannot generalize this completely, however, since one can show that there

Note: *If u were to exist in this example, then u' would be the mythical "Dirac delta function" which has value 0 for $x \neq 0$ and has area 1 under the graph at 0.*

is no differentiable function u such that

$$\int f\, d\delta_0 = \int fu'\, dm$$

where δ_0 is the Dirac unit point mass measure at 0. The obstacle, as it turns out, is that δ_0 and m disagree about which sets are null.

Definition 3.8.1
Let $(X, \mathcal{M})$ be a measurable space, and let μ and ν be measures on the space.

*We say that ν is **absolutely continuous** with respect to μ, and write $\nu \ll \mu$, if for every $E \in \mathcal{M}$ such that $\mu(E) = 0$, we have $\nu(E) = 0$.*

The term "absolutely continuous" is derived from classical real analysis, and the connection can be seen via an ε-δ formulation of the concept.

Proposition 3.8.1
Let ν and μ be measures on $(X, \mathcal{M})$. Then $\nu \ll \mu$ iff for every $\varepsilon > 0$ there exists a $\delta > 0$ such that $\nu(E) < \varepsilon$ whenever $\mu(E) < \delta$.

Proof:

Assume that for every $\varepsilon > 0$ there exists a $\delta > 0$ such that $\nu(E) < \varepsilon$ whenever $\mu(E) < \delta$. Then if $\mu(E) = 0$, we have that for any $\varepsilon > 0$, $\mu(E) < \delta = 1$ and hence $\nu(E) < \varepsilon$. Hence $\nu(E) = 0$, and so $\nu \ll \mu$.

Conversely, if there exists a $\varepsilon > 0$ such that for every $n \in \mathbb{N}$, we can find $E_n \in \mathcal{M}$ with $\mu(E_n) < 2^{-n}$ and $\nu(E_n) \geq \varepsilon$, then if $F_n = \bigcup_{k=n}^{\infty} E_k$ and $F = \bigcap_{n=1}^{\infty} F_n$, then $\mu(F_n) \leq \sum_{k=n}^{\infty} \mu(E_k) \leq 2^{1-n}$ and so $\mu(F) = \lim_{n\to\infty} \mu(F_n) = 0$, but $\nu(F_n) \geq \varepsilon$, so $\nu(F) \geq \varepsilon$. Hence ν is not absolutely continuous with respect to μ. ∎

Mutual singularity and absolute continuity are almost completely exclusive defintions. As Exercise 3.8.1 shows, if $\nu \ll \mu$ and $\nu \perp \mu$, then $\nu(E) = 0$ for every measurable set E.

Example 3.8.1
Let $(X, \mathcal{M}, \mu)$ be a measure space and $f \in \mathcal{L}^+(X)$. Then

$$\mu_f(E) = \int_E f\, d\mu,$$

so if $\mu(E) = 0$, then $\mu_f(E) = 0$. Hence $\mu_f \ll \mu$. ◇

As in the above example, when we have two measures μ and ν with the relationship that $\nu(E) = \int_E\, d\nu = \int_E f\, d\mu$ it is common to abuse notation and express this relationship by saying that $d\nu = f\, d\mu$, or even refer to the measure as "$f\, d\mu$."

This situation is in fact typical of the case where ν is absolutely continuous with respect to μ. In fact we can say more: any σ-finite measure ν can be broken up into two parts, one part of the form $f\, d\mu$ which is absolutely continuous to

μ, and a second part which is mutually singular. This decomposition is unique up to modification of f on μ-null sets. This result is known as the Lebesgue-Radon-Nikodym theorem.

As is often the case, we must first prove this for finite measures, and then extend to σ-finite measures.

Theorem 3.8.2 (Lebesgue-Radon-Nikodym)
Let ν and μ be finite measures on $(X, \mathcal{M})$. Then there exist unique positive measures λ and ρ on $(X, \mathcal{M})$ such that $\lambda \perp \rho$, $\rho \ll \mu$ and $\nu = \lambda + \rho$. Furthermore, there is a μ-integrable function $f : X \to [0, \infty]$, unique up to modifications on a μ-null set, such that $d\rho = f\ d\mu$.

Proof:
Consider the collection of functions

$$\mathcal{F} = \left\{ g : X \to [0, \infty] : \int_E g\ d\mu \leq \nu(E) \text{ for all } E \in \mathcal{M} \right\}.$$

Strategy: *The aim of this proof is to first find a candidate for f and then use that to construct λ. To find f we look at all the functions which have the property that $g\ d\mu \leq d\nu$ and use pointwise limits and the MCT to find a candidate for f.*

If g_1 and $g_2 \in \mathcal{F}$, then $h = \max(g_1, g_2) \in \mathcal{F}$, since if $A = \{x : g_1(x) > g_2(x)\}$ is the set of points where $g_1 > g_2$, then for any $E \in \mathcal{M}$,

$$\int_E h\ d\mu = \int_{E \cap A} g_1\ d\mu + \int_{E \setminus A} g_2\ d\mu \leq \nu(E \cap A) + \nu(E \setminus A) = \nu(E).$$

Induction then tells us that $h = \max(g_1, \dots, g_n) \in \mathcal{F}$ if $g_1, \dots, g_n \in \mathcal{F}$.

Let

$$c = \sup \left\{ \int_X g\ d\mu : g \in \mathcal{F} \right\}.$$

Clearly $c \leq \nu(X) < \infty$, so we can find a sequence of functions $g_n \in \mathcal{F}$ whose integrals increase to c. Moreover, if $f_n = \max(g_1, \dots, g_n)$, then $f_n \in \mathcal{F}$, the integrals of f_n increase to c, and each $f_n(x)$ is an increasing sequence, so f_n converges pointwise some measurable function $f : X \to [0, \infty]$. The monotone convergence theorem then tells us that

Note: *because it is possible that $f_n(x) \to \infty$ for some points x, we have to allow f to take infty as a value. To integrate such a potentially infinite function, remember that our convention states that $0 \cdot \infty = 0$, so the integral of f is infinite if f is infinite on a set of positive measure, and is only finite if it is finite almost everywhere.*

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

so $f \in \mathcal{F}$, and moreover

$$c = \lim_{n \to \infty} \int f_n\ d\mu = \int f\ d\mu.$$

As a consequence of the fact that f has finite integral, we have that $f(x) < \infty$ μ-almost everywhere.

Let $\rho(E) = \int_E f\ d\mu$, and consider the measure λ defined by $\lambda(E) = \nu(E) - \rho(E)$. Since $f \in \mathcal{F}$ implies $\int_E f\ d\mu \leq \nu(E)$, so $0 \leq \lambda(E) < \infty$ for all $E \in \mathcal{M}$, so λ is a finite measure.

Additionally λ and μ are mutually singular, since otherwise by Lemma 3.7.4 we can find some $A \in \mathcal{M}$ and $\varepsilon > 0$ such that $\mu(A) > 0$ and $\lambda(E) \geq \varepsilon\mu(E)$ for any $E \in \mathcal{M}$ with $E \subseteq A$. But then we have

$$\int_E f + \varepsilon\chi_A \; d\mu = \rho(E) + \varepsilon\mu(E \cap A) \leq \rho(E) + \lambda(E) = \nu(E),$$

so $f + \varepsilon\chi_A \in \mathcal{F}$, but

$$\int_X f + \varepsilon\chi_A \; d\mu = c + \varepsilon\mu(A) > c$$

contradicting our construction of f.

If we have $\nu = \lambda' + f' \; d\mu$ also satisfying the theorem, then $\lambda - \lambda' = (f - f') \; d\mu$, but $\lambda - \lambda' \perp \mu$, and $(f - f') \; d\mu \ll \mu$, so $\lambda - \lambda' = (f - f') \; d\mu = 0$, and so $\lambda = \lambda'$ and

$$\int f - f' \; d\mu = 0$$

so $f = f'$ μ-almost everywhere. ■

Strategy: *The sequence of corollaries which follow are a very typical strategy for extending from finite positive measures to σ-finite measures, signed measures and complex measures. The σ-finite case is obtained by subdividing into a countable number of finite chunks. The signed case is obtained by applying the σ-finite case to the Jordan decomposition, and the complex case is obtained by applying the signed case to the real and imaginary parts.*

This immediately extends to the σ-finite measures.

Corollary 3.8.3 (Lebesgue-Radon-Nikodym)
Let ν and μ σ-finite positive measures on $(X, \mathcal{M})$. Then there exist unique σ-finite measures λ and ρ on $(X, \mathcal{M})$ such that $\lambda \perp \rho$, $\rho \ll \mu$ and $\nu = \lambda + \rho$. Furthermore, there is a μ-integrable function $f : X \to [0, \infty]$, unique up to modifications on a μ-null set, such that $d\rho = f \; d\mu$.

Proof:
We can divide X into a countable disjoint union of sets A_n with $\mu(A_n) < \infty$ and $\nu(A_n) < \infty$. Letting $\mu_n(E) = \mu(E \cap A_n)$ and $\nu_n(E) = \mu(E \cap A_n)$, we can break ν_n into measures λ_n and $\rho_n = f_n \; d\mu_n$ such that $\lambda_n \perp \mu_n$, and λ_n is null and f_n is 0 on the compliment of A. Letting $\lambda = \sum \lambda_n$ and $f = \sum f_n$, we have that $\nu = \lambda + f \; d\mu$ and $\lambda \perp \mu$. Uniqueness follows as for the finite case. ■

By applying the above to the positive and negative parts of a σ-finite signed measure, and then taking the difference of the measures so obtained, we can extend the result to signed measures.

Corollary 3.8.4 (Lebesgue-Radon-Nikodym)
Let ν be a σ-finite signed measure and μ a σ-finite measure on $(X, \mathcal{M})$. Then there exist unique σ-finite signed measures λ and ρ on $(X, \mathcal{M})$ such that $\lambda \perp \rho$, $\rho \ll \mu$ and $\nu = \lambda + \rho$. Furthermore, there is a μ-integrable function $f : X \to \mathbb{R}^{(+)}$, unique up to modifications on a μ-null set, such that $d\rho = f \; d\mu$.

Proof:
We take the Jordan decomposition $\nu = \nu_+ - \nu_-$ of ν. Then ν_+ and ν_- are a σ-finite measures (in fact at least one of them must be finite), and so

by the previous corollary we have that there are $\lambda_\pm$, $\rho_\pm$, and $f_\pm$, such that $\nu_\pm = \lambda_\pm + \rho_\pm$ with $\rho_\pm \ll \mu$, $\mu \perp \lambda_\pm$ and $d\rho_\pm = f_\pm \; d\mu$.

But then let $f = f_+ - f_-$ so that $\rho = \rho_+ - \rho_-$ satisfies $d\rho = f \; d\mu$, and let $\lambda = \lambda_+ - \lambda_-$. Then

$$\nu = \nu_+ - \nu_- = (\lambda_+ + \rho_+) - (\lambda_- + \rho_-) = \lambda + \rho.$$

Also $\rho \ll \mu$, since if $\mu(E) = 0$ then $\rho_\pm(E) = 0$, so $\rho(E) = 0$. Finally, $\lambda \perp \mu$, since there are disjoint sets $E_\pm$ and $F_\pm$ such that $E_\pm \cup F_\pm = X$, $\mu(E_\pm) = 0$ and $\lambda_\pm(F_\pm) = 0$, so if we let $E = E_+ \cup E_-$ and $F = F_+ \cap F_-$, then E and F are disjoint with $E \cup F = X$,

$$\mu(E) \leq \mu(E_+) + \mu(E_-) = 0$$

and if $A \subseteq F$, then

$$\lambda(A) = \lambda_+(A) - \lambda_-(A) = 0,$$

so F is λ-null. ■

Similarly, this may be extended to complex measures using the previous corollary on the real and imaginary parts of the measure.

Corollary 3.8.5 (Lebesgue-Radon-Nikodym)
Let ν be a complex measure and μ a σ-finite measure on $(X, \mathcal{M})$. Then there exist unique complex measures λ and ρ on $(X, \mathcal{M})$ such that $\lambda \perp \rho$, $\rho \ll \mu$ and $\nu = \lambda + \rho$. Furthermore, there is a μ-integrable function $f : X \to \mathbb{C}$, unique up to modifications on a μ-null set, such that $d\rho = f \; d\mu$.

We call $\nu = \rho + \lambda$ the ***Lebesgue decomposition*** of ρ with respect to λ. The function f is called the ***Radon-Nikodym derivative*** of ν with respect to μ, and we write

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

The sorts of things that you expect to hold for derivatives hold for the Radon-Nikodym derivative.

Proposition 3.8.6
Let ν, ν_1 and ν_2 be signed or complex measures, and μ and λ are measures, on $(X, \mathcal{M})$.

1. *If ν_1 and $\nu_2 \ll \mu$ and α is a scalar, then if $\nu = \nu_1 + \alpha\nu_2$, we have $\nu \ll \mu$ and*
$$\frac{d\nu}{d\mu} = \frac{d\nu_1}{d\mu} + \alpha\frac{d\nu_2}{d\mu}$$
μ-almost everywhere.

Note: *This is linearity of for the Radon-Nikodym derivative.*

2. *If $\nu \ll \mu$ then if $g \in L^1(\nu)$, $g\frac{d\nu}{d\mu} \in L^1(\mu)$, and*
$$\int g \; d\nu = \int g \frac{d\nu}{d\mu} \; d\mu.$$

Note: *This is the substitution rule for Lebesgue integration.*

Note: *This is the chain ru* *the Radon-Nikodym deriva*

3. *If $\nu \ll \mu$ and $\mu \ll \lambda$, then $\nu \ll \lambda$ and*

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda} \qquad \lambda\text{-almost everywhere.}$$

4. *If $\mu \ll \lambda$ and $\lambda \ll \mu$, then*

$$\frac{d\mu}{d\lambda} = 1/(d\lambda/d\mu) \qquad \lambda\text{-almost everywhere.}$$

Proof:

For the first part, note that

$$\nu(E) = \nu_1(E) + \alpha\nu_2(E) = \int_E \frac{d\nu_1}{d\mu}\ d\mu + \alpha \int_E \frac{d\nu_2}{d\mu}\ d\mu = \int_E \frac{d\nu_1}{d\mu} + \alpha \int_E \frac{d\nu_2}{d\mu}\ d\mu.$$

Hence $d\nu = \left(\frac{d\nu_1}{d\mu} + \alpha \int_E \frac{d\nu_2}{d\mu}\right)\ d\mu$, and so $\nu \ll \mu$ and by the uniqueness of the Radon-Nikodym derivative

$$\frac{d\nu}{d\mu} = \frac{d\nu_1}{d\mu} + \alpha\frac{d\nu_2}{d\mu}$$

μ-almost everywhere.

For the second part, assume that ν is a σ-finite positive measure.

If $g = \chi_E$ for E, then

$$\int g\ d\nu = \int_E\ d\nu = \int_E \frac{d\nu}{d\mu}\ d\mu = \int g\frac{d\nu}{d\mu}\ d\mu.$$

This can then be extended to simple functions

$$\int g\ d\nu = \int \sum c_k\chi_{E_k}\ d\nu = \sum c_k \int \chi_{E_k}\ d\nu = \sum c_k \int \chi_{E_k}\frac{d\nu}{d\mu}\ d\mu = \int g\frac{d\nu}{d\mu}\ d\mu,$$

and then since every non-negative measurable function is the pointwise limit of and increasing sequence of simple functions, we can use the monotone convergence theorem to extend to non-negative measurable functions: if f_n are an increasing sequence of simple functions which converge pointwise to g, then $f_n\frac{d\nu}{d\mu}$ is an increasing sequence of non-negative functions which converges pointwise to $g\frac{d\nu}{d\mu}$, so that

$$\int g\ d\nu = \lim_{n\to\infty} \int f_n\ d\nu = \lim_{n\to\infty} \int f_n\frac{d\nu}{d\mu}\ d\mu = \int g\frac{d\nu}{d\mu}.$$

Finally, if $g \in L^1(\nu)$, then

$$\int \left|g\frac{d\nu}{d\mu}\right|\ d\mu = \int |g|\frac{d\nu}{d\mu}\ d\mu = \int |g|\ d\nu < \infty,$$

so $g\frac{d\nu}{d\mu} \in L^1(\mu)$, and

$$\int g\ d\nu = \int g_+\ d\nu - \int g_-\ d\nu = \int g_+ \frac{d\nu}{d\mu}\ d\mu - \int g_- \frac{d\nu}{d\mu}\ d\mu = \int g \frac{d\nu}{d\mu}\ d\mu.$$

We then extend this to signed and complex measures in the usual way, by first applying the result for positive measures to the Jordan decomposition, and then applying the result for signed measures to the real and imaginary parts of a complex measure.

For the third part, replace ν by μ and μ by λ, and let $g = \chi_E \frac{d\nu}{d\mu}$, we get

$$\nu(E) = \int_E \frac{d\nu}{d\mu}\ d\mu = \int g\ d\mu = \int g \frac{d\mu}{d\lambda}\ d\lambda = \int_E \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda}\ d\lambda,$$

so $d\nu = \left(\frac{d\nu}{d\mu}\frac{d\mu}{d\lambda}\right)\ d\lambda$ and hence $\nu \ll \lambda$. By the uniqueness of the Radon-Nikodym derivative, we then conclude that

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda}$$

λ-almost everywhere.

For the last part, replacing ν by λ in the 3rd part, we get

$$\frac{d\lambda}{d\mu}\frac{d\mu}{d\lambda} = \frac{d\lambda}{d\lambda} = 1,$$

λ-almost everywhere, so that

$$\frac{d\mu}{d\lambda} = 1/(d\lambda/d\mu).$$

λ-almost everywhere. ■

Note that in general, products and quotients of measures are not measures (indeed, the quotient probably doesn't even make sense as a set function), so we would not expect any analogue of the product or quotient rules for the Radon-Nikodym derivative.

Exercises

3.8.1. Show that if $\nu \ll \mu$ and $\nu \perp \mu$, then $\nu(E) = 0$ for every measurable set E.

Chapter 4

Introduction to Functional Analysis

In earlier sections of these notes we have relied upon the fact that functions and measures have natural vector space structures. These vector spaces are infinite dimensional, in general, and so we have not been able to escape without using other techniques to get results. These extra techniques usually either come from identifying "positive" elements within the vector space, such as positive functions and measures; or from notions of convergence or topology.

For this reason, we abstract these situations and consider general vector spaces which have a topology.

4.1 Topological Vector Spaces

We start with the most general definition.

Definition 4.1.1
A ***topological vector space*** *is a real or complex vector space V together with a topology on V such that*

$$\begin{aligned} + : V \times V &\to V \\ (v, w) &\mapsto v + w, \end{aligned} \qquad \text{and} \qquad \begin{aligned} \cdot : \mathbb{F} \times V &\to V \\ (\lambda, v) &\mapsto \lambda V, \end{aligned}$$

are continuous.

This is typical of the way in which algebraic objects are given topological structure. One simply insists that the appropriate operations (in this case addition and scalar multiplication) are continuous.

General topological vector spaces tend not to have much structure. Fortunately most topological vector spaces that are encountered in practice satisfy much stronger conditions.

Perhaps the next most stringent condition to insist on is that there be some notion of the length or size of a vector. The axioms required for a length are encapsualted in the idea of a norm.

Definition 4.1.2
*A **seminorm** on a vector space V is a function $\rho : V \to [0, \infty)$ such that*

(i) $\rho(v + w) \leq \rho(v) + \rho(w)$ *(triangle inequality),*

(ii) $\rho(\lambda v) = |\lambda|\rho(v)$,

*for all v, $w \in V$ and scalars λ. If in addition $\rho(v) = 0$ implies $v = 0$, then the seminorm is called a **norm**. A norm is traditionally denoted by $\|\cdot\|$. A vector space V together with a norm is called a **normed vector space**.*

It is worthwhile noting that the triangle inequality implies that

$$\rho(v) \leq \rho(w) + \rho(v - w) \qquad \text{and} \qquad \rho(w) \leq \rho(v) + \rho(v - w),$$

which can be combined to give

$$|\rho(v) - \rho(w)| \leq \rho(v - w).$$

It is not hard to see that a norm immediately defines a metric on the vector space by

$$d(v, w) = \|v - w\|.$$

Recall that a metric space is called complete if every Cauchy sequence converges, that is if v_n is a sequence in V such that for every $\varepsilon > 0$ there is some N such that $d(v_n, v_m) < \varepsilon$ for all n, $m \geq N$ then v_n converges to something.

Proposition 4.1.1
A normed vector space V is a topological vector space with the metric topology coming from the norm.

Proof:
Let $v_n \to v$ and $w_n \to w$ in the metric topology on V. Then

$$d(v_n+w_n, v+w) = \|v_n+w_n-(v+w)\| \leq \|v_n-v\|+\|w_n-w\| = d(v_n, v)+d(w_n, w).$$

Hence $v_n + w_n \to v + w$, and so addition is continuous.
Similarly, if $\lambda_n \to \lambda$,

$$\begin{aligned} d(\lambda_n v_n, \lambda v) &= \|\lambda_n v_n - \lambda v\| \\ &= \|\lambda_n v_n - \lambda_n v + \lambda_n v - \lambda v\| \leq |\lambda_n|\|v_n - v\| + |\lambda_n - \lambda|\|v\| \\ &= |\lambda_n| d(v_n, v) + |\lambda_n - \lambda|\|v\|. \end{aligned}$$

Hence $\lambda_n v_n \to \lambda v$, and so scalar multiplication is continuous. ■

Example 4.1.1

Consider the vector space $C_b(X)$ of continuous, bounded functions on a locally compact Hausdorff topological space X. The ***uniform norm*** is defined by

$$\|f\|_u = \sup_{x \in X} |f(x)|.$$

This is a norm, and $f_n \to f$ in this vector space if and only if $f_n \to f$ uniformly. In fact, this norm makes $C_b(X)$ a complete metric space, since every uniformly Cauchy sequence converges. $\diamond$

Another common situation in analysis is that rather than a norm giving the topology, a collection of seminorms gives the topology.

Example 4.1.2

Consider $C_b(X)$ again, but this time define seminorms ρ_x for each $x \in X$ by

$$\rho_x(f) = |f(x)|.$$

Then $f_n \to f$ in the topology given by the seminorms iff $f_n(x) \to f(x)$ for all $x \in X$. In other words, this is the topology of pointwise convergence of functions. $\diamond$

Example 4.1.3

Let $C^\infty([0,1])$ be the vector space of all infinitely differentiable functions. Define seminorms ρ_n for $n \in \mathbb{N}$ by

$$\rho_n(f) = \|f^{(n)}\|_u.$$

Then $f_n \to f$ in the topology determined by these seminorms iff $f^{(n)} \to f^{(n)}$ uniformly for all $n \in \mathbb{N}$. $\diamond$

Sometimes we only have a seminorm around, but in such cases we may reduce to a normed vector space by taking quotients of vector spaces. Recall that if V is a vector space, and W is a vector subspace of V, then we can construct the quotient vector space V/W as follows. If we say that $u \sim v$ if $u - v \in W$, then this defines an equivalence relation on V, and one can see that the quotient space $V/\sim$ is a vector space, where $[u] + [v] = [u + v]$ and $\lambda[v] = [\lambda v]$. We note that if ρ is a seminorm, then $\ker \rho = \{w \in V : \rho(w) = 0\}$ is a vector subspace of V, and so we have a vector space $V/\ker\rho$. We can define a norm on this space by

$$\|[v]\|_\rho = \rho(v).$$

This is well-defined, since if $u \in [v]$, $\rho(u - v) = 0$, so

$$\rho(u) \leq \rho(u - v) + \rho(v) = \rho(v),$$

but similarly $\rho(v) \leq \rho(u)$. Hence it does not matter which equivalence class representative we choose to define the norm. One can easily check that the other norm properties hold, in particular that if $\|[v]\|_\rho = 0$, then $\rho(v) = 0$, and so $v \in \ker \rho$, and hence $[v] = [0] = 0$.

Example 4.1.4

Let $(X, \mathcal{M}, \mu)$ be a measure space. Then the set of functions in $L^1(X, \mu)$ is a vector space, since if f, $g \in L^1(X, \mu)$, and c any constant, then

$$\int |f + g| \; d\mu \leq \int |f| + |g| \; d\mu = \int |f| \; d\mu + \int |g| \; d\mu < \infty,$$

so $f + g \in L^1(X, \mu)$; and

$$\int |cf| \; d\mu = |c| \int |f| \; d\mu < \infty,$$

so $cf \in L^1(X, \mu)$.

Furthermore

$$\|f\|_1 = \int |f| \; d\mu$$

is a seminorm on $L^1(X, \mu)$, since

$$\|f + g\|_1 = \int |f + g| \; d\mu \leq \int |f| \; d\mu + \int |g| \; d\mu = \|f\|_1 + \|g\|_1.$$

and

$$\|cf\|_1 = \int |cf| \; d\mu = |c| \int |f| \; d\mu = |c|\|f\|_1.$$

However, it is not a norm, since $\|f\|_1 = 0$ if and only if $f = 0$ μ-almost everywhere. So we can take the quotient of this vector space with all functions which are equal to 0 almost everywhere and get a new vector space with a norm, as described above. This quotient can be concretely understood as saying that two functions are equivalent if they are equal almost everywhere. In other words, if two functions differ on a set of measure zero, we think of them as being the same.

It is common practice to write "f" for the equivalence class of f in this context, but when we want to make it clear that we are considering the equivalence class, not the function, we will use $\boldsymbol{f}$ for $[f]$. Hence

$$\|\boldsymbol{f}\|_1 = \int |f| \; d\mu$$

is a norm.

Similarly, it is customary to call the quotient vector space $L^1(X, \mu)$. $\diamond$

In some cases we are lucky enough to have an inner product on the vector space.

Definition 4.1.3

*An **inner product** on a complex vector space V is a function $\langle \cdot, \cdot \rangle : V \to \mathbb{R}$ (or $\mathbb{C}$) such that* *(i)* $\langle \lambda v + w, u\rangle = \lambda\langle v, u\rangle + \langle w, u\rangle$ *(linearity in first variable),*

(ii) $\langle w, v\rangle = \overline{\langle v, w\rangle}$ *(conjugate symmetry)*

(iii) $\langle v, v\rangle > 0$ *if* $v \neq 0$,

for all u, v, $w \in V$ *and scalars* $\lambda \in \mathbb{C}$.

The axioms for a real vector space are the same, except that one gets symmmetry instead of conjugate symmetry: $\langle w, v\rangle = \langle v, w\rangle$.

The conjugate symmetry and linearity conditions imply that the inner product is conjugate linear in the second variable:

$$\langle v, \lambda w + u\rangle = \overline{\lambda}\langle v, w\rangle + \langle v, u\rangle.$$

Example 4.1.5

The most familiar example of an inner product is the standard inner product (or dot product) on $\mathbb{C}^n$. Given $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n) \in \mathbb{C}^n$ we have

$$\langle x, y\rangle = \sum_{k=1}^{n} x_k \overline{y_k}.$$

$\diamond$

Example 4.1.6

Let M_n be the vector space of all n-by-n complex matrices. The adjoint A^* of such a matrix A is the conjugate transpose of the matrix, ie. the transpose of the matrix whose entries are the complex conjugates of the entries of A. The adjoint has the properties

$$(\lambda A)^* = \overline{\lambda} A^* \qquad (A+B)^* = A^* + B^* \qquad (AB)^* = B^* A^* \qquad (A^*)^* = A.$$

A matrix is called self-adjoint if $A^* = A$, and is called unitary if $U^*U = UU^* = I$, where I is the identity matrix.

Given any matrix A one can find an upper triangular matrix T and a unitary matrix U such that $A = U^*TU$. If A is self-adjoint, then T can be the diagonalization of A and U has columns which are eigenvectors of A.

The trace, $\mathrm{tr}A$, of a matrix is the sum of the diagonal entries from top-left to bottom-right. The trace has the properties

$$\mathrm{tr}(\lambda A) = \lambda \mathrm{tr}A \qquad \mathrm{tr}(A+B) = \mathrm{tr}A + \mathrm{tr}B \qquad \mathrm{tr}(AB) = \mathrm{tr}(BA) \qquad \mathrm{tr}(A^*) = \overline{\mathrm{tr}A}.$$

If A is self-adjoint, then $\mathrm{tr}(A) = \mathrm{tr}(U^*DU) = \mathrm{tr}(U^*UD) = \mathrm{tr}D$ is equal to the sum of the eigenvalues of A.

If we define

$$\langle A, B\rangle = \mathrm{tr}(B^*A),$$

then we observe that

$$\begin{aligned}\langle \lambda A + B, C\rangle &= \mathrm{tr}(C^*(\lambda A + B)) = \mathrm{tr}(\lambda C^* A + C^* B) = \lambda \mathrm{tr}(C^* A) + \mathrm{tr}(C^* B) \\ &= \lambda\langle A, C\rangle + \langle B, C\rangle,\end{aligned}$$

and

$$\langle B, A\rangle = \mathrm{tr}(A^* B) = \mathrm{tr}((A^* B)^*)^* = \overline{\mathrm{tr}(B^*(A^*)^*)} = \overline{\mathrm{tr}(B^* A)} = \overline{\langle A, B\rangle}.$$

Also if $\langle A, A\rangle = 0$, then $\mathrm{tr} A^* A = 0$, so

$$\begin{aligned}0 = \mathrm{tr}(A^* A) &= \mathrm{tr}((U^* T^* U)^* U^* T U) \\ &= \mathrm{tr}(U^* T^* U U^* T U) \\ &= \mathrm{tr}(U^* T^* T U) \\ &= \mathrm{tr}(U^* U T^* T) \\ &= \mathrm{tr}(T^* T).\end{aligned}$$

Now T^*T is a self-adjoint matrix with non-negative eigenvalues, and so if the trace is 0 then the sum of the eigenvalues is also 0, and since they are non-negative, every eigenvalue is 0 and hence $T^*T = 0$. A simple induction argument on n shows that this implies that $T = 0$ and hence $A = 0$.

Therefore this is an inner product. ◇

Given an inner product, one can define

$$\|v\| = \langle v, v\rangle^{1/2}.$$

As the notation would indicate, this is a norm, and it is easy to see that

$$\|v\| = 0 \iff v = 0,$$

and

$$\|\lambda v\| = (\lambda\overline{\lambda}\langle v, v\rangle)^{1/2} = |\lambda|\|v\|.$$

Showing that the triangle inequality holds is slightly harder, and requires the Shwarz inequality.

Lemma 4.1.2 (Shwarz Inequality)
If $\langle\cdot,\cdot\rangle$ is an inner product on V, then $|\langle v, w\rangle| \leq \|v\|\|w\|$ for all v and $w \in V$. We get equality if and only if v and w are linearly dependant.

Proof:

If $\langle v, w\rangle = 0$, then the result is immediate.

If $\langle v, w\rangle \geq 0$, then $v \neq 0$ and $w \neq 0$, and for any $t \in \mathbb{R}$,

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

But this is quadratic in t, and so has absolute minimum value when

$$t = \frac{\langle v, w \rangle}{\|w\|^2},$$

and hence if we use this value of t, we have

$$0 \leq \|v\|^2 - \frac{\langle v, w \rangle^2}{\|w\|^2},$$

and the inequality is immediate from this. We get equality if and only if $v - tw = 0$.

Finally, if $\langle v, w \rangle = c \neq 0$, then we can take the polar decomposition of $c = re^{i\theta}$. Then $\langle e^{-i\theta} v, w \rangle = r \geq 0$ and

$$|\langle v, w \rangle| = \langle e^{-i\theta} v, w \rangle \leq \|e^{i\theta} v\| \|w\| = \|v\| \|w\|,$$

with equality iff $e^{i\theta} v - tw = 0$. ■

Corollary 4.1.3
If $\langle \cdot, \cdot \rangle$ is an inner product on V, then $\|v\| = \langle v, v \rangle^{1/2}$ is a norm on V.

Proof:
We have shown everything except the triangle inequality. We have that

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

So the triangle inequality holds. ■

Every inner product space is therefore naturally a topological vector space in the metric topology coming from the norm given by the inner product.

Example 4.1.7
The norm corresponding to the standard inner product on $\mathbb{C}^n$ is the usual Euclidean 2-norm

$$\|x\|_2 = \left(\sum_{k=1}^{n} |x|^2 \right)^{1/2}.$$

◇

Example 4.1.8
The norm corresponding to the inner product of Example 4.1.6,

$$\|A\| = (\mathrm{tr} A^* A)^{1/2}$$

is called the Hilbert-Schmidt norm. ◇

Recalling metric space theory, we remember that complete metric spaces are particularly nice. As we will see, many important examples of normed vector spaces in analysis are complete. In fact, this property is of sufficient importance to give such spaces a special name.

Definition 4.1.4
If V is a normed vector space which is complete as a metric space (ie. every Cauchy sequence converges), we call it a ***Banach space****. A complete inner product space is called a* ***Hilbert space****.*

Hence every Hilbert space is a Banach space.

Testing for completeness can be a hassle when using the traditional definition of every Cauchy sequence converging. In the context of normed vector spaces, there is another test which is often more convenient.

Let V be a normed vector space. A series

$$\sum_{k=1}^{\infty} v_k$$

converges in V if the sequence of partial sums,

$$s_n = \sum_{k=1}^{n} v_k,$$

converge in V. A series is ***absolutely convergent*** if

$$\sum_{k=1}^{\infty} \|v_k\|$$

converges.

Proposition 4.1.4
If V is a normed vector space, V is complete if and only if every absolutely convergent series converges.

Proof:

First assume that V is complete. If we have an absolutely convergent series, then for all $\varepsilon > 0$, there is some N such that

$$\sum_{k=n+1}^{m} \|v_k\| = \sum_{k=1}^{m} \|v_k\| - \sum_{k=1}^{n} \|v_k\| < \varepsilon$$

for all m, $n \geq M$, since the sequence of partial sums of the norms converge in the scalar field, and so is a Cauchy sequence. But then for all m, $n \geq N$,

$$\|s_m - s_n\| = \| \sum_{k=n+1}^{m} v_k\| \leq \sum_{k=n+1}^{m} \|v_k\| < \varepsilon.$$

So the sequence of partial sums s_n is Cauchy in V, and hence converges.

Conversely, if every absolutely convergent series converges in V, let v_n be a Cauchy sequence in V. Then we can find N_k such that $N_1 < N_2 < \dots$ and

$$\|v_n - v_m\| < 2^{-k}$$

for all n, $m \geq N_k$. Let $w_1 = v_{N_1}$ and $w_k = v_{N_k} - v_{N_{k-1}}$. Then we have partial sums

$$\sum_{k=1}^{n} w_k = v_{N_n},$$

and for $k > 1$, $\|w_k\| = \|v_{N_k} - v_{N_{k-1}}\| < 2^{-k+1}$, so

$$\sum_{k=1}^{\infty} \|w_k\| < \|w_1\| + \sum_{k=2}^{\infty} 2^{-k+1} = \|w_1\| + 1 < \infty.$$

So this series is absolutely convergent, and so the series is convergent. But this means that the partial sums v_{N_n} converge to some $v \in V$. So we have a subsequence of our original Cauchy sequence which converges. But then given any $\varepsilon > 0$, we can find some k such that both $2^{-k} < \varepsilon/2$ and $\|v_{N_k} - v\| < \varepsilon/2$. Then for all $n \geq N_k$,

$$\|v_n - v\| \leq \|v_n - v_{N_k}\| + \|v_{N_k} - v\| < \varepsilon.$$

So $v_n \to v$. Hence every Cauchy sequence converges, and V is complete. ■

This result is particularly important when considering vector spaces of integrable functions.

Example 4.1.9

Let $(X, \mathcal{M}, \mu)$ be a measure space. We know that the vector space $L^1(X, \mu)$ of equivalence classes of L^1 functions has a norm on it. For clarity, we will use $\boldsymbol{f}$ to represent the equivalence class of all functions equal to f almost everywhere. Concretely, then, the norm is

$$\|\boldsymbol{f}\|_1 = \int |f| \; d\mu.$$

Let $\sum_{k=1}^{\infty} \boldsymbol{f_k}$ be an absolutely convergent series in $L^1(X, \mu)$. In other words,

$$\sum_{k=1}^{\infty} \|\boldsymbol{f_k}\|_1 = \sum_{k=1}^{\infty} \int |f_k| \; d\mu < \infty.$$

But Proposition 3.3.9 tells us that there is an L^1 function f such that

$$f = \sum_{k=1}^{\infty} f_k$$

almost everywhere. Hence

$$\boldsymbol{f} = \sum_{k=1}^{\infty} \boldsymbol{f_k}$$

and so any absolutely convergent series converges.

Therefore $L^1(X, \mu)$ is a Banach space. ◇

There are several constructions that give new vector spaces from other vector spaces. The most obvious are taking a subspace of a vector space, and taking a quotient by a subspace. Any subspace of a topological vector space is automatically a topological vector space in the relative topology. However, although if W is a vector subspace of a Banach space V it is a normed vector space, it may not be a Banach space.

Proposition 4.1.5
Let V be a Banach space. Then a vector subspace W is a Banach space if and only if it is closed.

Proof:
Let w_n be a Cauchy sequence in W. Then it is a Cauchy sequence in V and so it converges to some $v \in V$. But since W is closed, the limit of any sequence in W must be in W, and so $v \in W$. Hence W is complete. ■

4.2 Linear Operators

You will recall that in linear algebra, the most important functions were linear maps bewteen vector spaces. For topological vector spaces, we typically insist that these linear maps must be continuous as well as linear.

Definition 4.2.1
Let V and W be topological vector spaces. We define $L(V, W)$ to be the set of all continuous linear maps $T : V \to W$. We let $L(V) = L(V, V)$. We commonly call maps between topological vector spaces ***operators****, and write $T(v) = Tv$.*

If V and W are normed vector spaces, we say that a linear map $T : V \to W$ is ***bounded*** *if*

$$\sup\{\|Tv\| : \|v\| \leq 1\} < \infty,$$

or equivalently, there is some K such that $\|Tv\| \leq K\|v\|$ for all $v \in V$. We denote the set of all bounded operators by $B(V, W)$, and $B(V) = B(V, V)$.

Note that the definition of boundedness in this context is distinct from the idea of a bounded function. The only linear operator which can be bounded as a function is the zero operator $0v = 0$. Boundedness for linear operators is equivalent to being bounded as a function restricted to the unit ball.

Also note that $K = \|T\|$ works in the definition, ie.

$$\|Tv\| \leq \|T\|\|v\|.$$

A bounded linear map is an isometry if $\|Tv\| = \|v\|$ for all $v \in V$, and in this case $\|T\| = 1$. If $\|T\| < 1$, then T is a contraction, since it shrinks the length of every vector.

It is through these linear maps that we can identify when two topological vector spaces are the same.

Definition 4.2.2
Let V and W be two topological vector spaces. We say that they are isomorphic if there is a linear homeomorphism from V to W.

We say two Banach spaces V and W are isometrically isomorphic if there is a linear isomorphism which is an isometry (and so therefore has an inverse which is also isometric). We write $V \cong W$ if V and W are isometrically isomorphic.

If V and W are two Banach spaces, they are isomorphic if and only if there is a bounded linear isomorphism with bounded inverse from V to W. Banach spaces may be isomorphic, but not isometrically isomorphic.

Example 4.2.1
Consider $\mathbb{R}^d$ with the norms

$$\|x\|_p = \left(\sum_{k=1}^{d} |x_k|^p \right)^{1/p}$$

for $p = 1, 2$. One can verify that $\|x\|_1 \leq \|x\|_2 \leq d\|x\|_1$, and so the identity map is an isomorphism between the Banach spaces $(\mathbb{R}^d, \|\cdot\|_1)$ and $(\mathbb{R}^d, \|\cdot\|_2)$. However, it is not an isometric isomorphism. ◇

We note that both $L(V, W)$ and $B(V, W)$ are themselves vector spaces. Indeed, if we say that $T_n \to T$ iff $T_n v \to Tv$ for all $v \in V$, we have that $L(V, W)$ is a topological vector space.

Proposition 4.2.1
Let V and W be normed vector spaces, and let $T \in B(V, W)$. Then

$$\|T\| = \sup\{\|Tv\| : \|v\| \leq 1\}.$$

defines a norm on $B(V, W)$, called the ***operator norm****, and so $B(V, W)$ is itself a normed vector space.*

Proof:
We check that all 3 conditions required of a norm hold.

Given T and $S \in B(V, W)$ we have that

$$\begin{aligned}
\|T + S\| &= \sup\{\|Tv + Sv\| : \|v\| \leq 1\} \\
&\leq \sup\{\|Tv\| + \|Sv\| : \|v\| \leq 1\} \\
&\leq \sup\{\|Tv\| : \|v\| \leq 1\} + \sup\{\|Sv\| : \|v\| \leq 1\} \\
&\leq \|T\| + \|S\|.
\end{aligned}$$

Given $\lambda \in \mathbb{C}$, we have

$$\|\lambda T\| = \sup\{\|\lambda T v\| : \|v\| \le 1\} = |\lambda| \sup\{\|Tv\| : \|v\| \le 1\} = |\lambda|\|Tv\|.$$

Finally, if $\|T\| = 0$, then $\|Tv\| = 0$ for all $v \in V$ with $\|v\| \le 1$, and so $Tv = 0$. Now if $\|v\| \ge 1$, we have that

$$Tv = T\left(\|v\|\frac{1}{\|v\|}v\right) = \|v\|T\left(\frac{1}{\|v\|}v\right) = 0$$

since $\left\|\frac{1}{\|v\|}v\right\| = 1$.

Hence $\|\cdot\|$ is a norm. ∎

It turns out that for normed vector spaces, $L(V,W)$ and $B(V,W)$ are the same.

Proposition 4.2.2
If V and W are topological vector spaces, then $T : V \to W$ is continuous if and only if T is continuous at 0.

If V and W are normed vector spaces, then $T : V \to W$ is continuous if and only if it is bounded.

Proof:
Let V and W are topological vector spaces and $T : V \to W$. It is trivial that if T is continuous, then it is continuous at 0. On the other hand, if T is continuous at 0, then given any convergent sequence, $v_n \to v$, we have that $v_n - v \to 0$, and so

$$Tv_n - Tv = T(v_n - v) \to T(0) = 0.$$

So $Tv_n \to Tv$. So T is continuous at every point, and so it is continuous.

Now let V and W be normed spaces. If T is bounded, there is some K such that $\|Tv\| \le K\|v\|$ for all $v \in V$. Given any $\varepsilon > 0$, let $\delta = \varepsilon/K$ so if $\|v_1 - v_2\| < \delta$, then

$$\|Tv_1 - Tv_2\| = \|T(v_1 - v_2)\| \le K\|v_1 - v_2\| = \varepsilon.$$

So T is continuous.

On the other hand, if T is continuous, it is continuous at 0. Letting $\varepsilon = 1$, there is some $\delta > 0$ such that

$$\|Tv\| = \|T(v - 0)\| < 1$$

for all $v \in V$ with $\|v\| < \delta$. For general $v \in V$, $(\delta/\|v\|)v \in B(0,\delta)$, so

$$\|Tv\| = \left\|\frac{\|v\|}{\delta}T\left(\frac{\delta}{\|v\|}v\right)\right\| = \frac{\|v\|}{\delta}\left\|T\left(\frac{\delta}{\|v\|}v\right)\right\| < \frac{1}{\delta}\|v\|.$$

So T is bounded. ∎

The same linear operator may have quite different properties, depending on what topology we are dealing with.

Example 4.2.2

Consider the vector space $C^\infty([0,1])$, and the differentiation operator

$$D : C^\infty([0,1]) \to C^\infty([0,1])$$

given by $Df = f'$. This is a linear map by the usual rules of calculus. If we consider the topology on $C^\infty([0,1])$ given in Example 4.1.3, then D is continuous, since for all functions $f \in C^\infty([0,1])$, we have

$$\rho_n(Df) = \rho_n(f') = \|f^{(n+1)}\|_u = \rho_{n+1}(f).$$

Hence if $f_k \to f$ in $C^\infty([0,1])$, we have $\rho_n(f_k - f) \to 0$ for all n, and so

$$\rho_n(Df_k - Df) = \rho_n(D(f_k - f)) = \rho_{n+1}(f_k - f) \to 0.$$

Therefore $Df_k \to Df$, and so D is continuous.

On the other hand, if we consider the topology on $C^\infty([0,1])$ coming from the uniform norm alone, this is not a bounded operator, since for each $n \in \mathbb{N}$, let $f_n = \sin nx$. Then

$$\|\sin nx\|_u = 1,$$

but

$$\|D(\sin nx)\|_u = \|n \sin nx\| = n.$$

So there is no K such that $\|Df\|_u \leq K\|f\|_u$ for all $f \in C^\infty([0,1])$. Since it is not bounded, it is also not continuous in this topology. ◇

Example 4.2.3

Consider $C_b(X)$ with the uniform norm. The evaluation maps

$$\varphi_x(f) = f(x)$$

are bounded linear maps from $C_b(X)$ to $\mathbb{C}$. Moreover, since

$$|\varphi_x(f)| = |f(x)| \leq \sup_{x \in X} |f(x)| = \|f\|_u,$$

we have that $\|\varphi_x\| = 1$. ◇

Example 4.2.4

Consider $C_b(\mathbb{R})$ with the uniform norm. Let $\lambda_t : C_b(\mathbb{R}) \to C_b(\mathbb{R})$ be the translation map

$$(\lambda_t f)(x) = f(x + t).$$

It is easy to see that $\|\lambda_t f\|_u = \|f\|_u$, so this is an isometry, and so $\|\lambda_t\| = 1$. ◇

Example 4.2.5

Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be σ-finite measure spaces. We know that $L^1(X, \mu)$ and $L^1(Y, \nu)$ are Banach spaces as discussed in Example 4.1.9. Let $K : X \times Y \to \mathbb{C}$ be $\mathcal{M} \otimes \mathcal{N}$-measurable with some constant C such that

$$\int |K(x, y)| \; d\mu(x) \leq C$$

as in Example 3.6.4. Then the function $I_K : L^1(X, \mu) \to L^1(Y, \nu)$ defined by the integral kernel K,

$$(I_K f)(y) = \int K(x, y) f(x) \; d\mu(x)$$

is a linear operator, and a consequence of Example 3.6.4 is that

$$\|I_K\| \leq C.$$

$\diamond$

Not only is $B(V, W)$ a normed vector space, it is in fact a Banach space if W is a Banach space.

Proposition 4.2.3

If V and W are normed vector spaces, then if W is complete, so is $B(V, W)$.

Proof:

Let T_n be a Cauchy sequence in $B(V, W)$. Since

$$|\|T_n\| - \|T_m\|| \leq \|T_n - T_m\|,$$

we have that $\|T_n\|$ is a Cauchy sequence in $\mathbb{R}$ and so converges.

Also for each $v \in V$, $T_n v$ is a Cauchy sequence in W, since for any $\varepsilon > 0$, we can find N such that $\|T_n - T_m\| < \varepsilon/\|v\|$ for all n, $m > N$, and then $\|T_n v - T_m v\| \leq \|T_n - T_m\|\|v\| = \varepsilon$. So each sequence $T_n v$ converges.

Define a function $T : V \to W$ by

$$Tv = \lim_{n \to \infty} T_n v.$$

T is linear, since

$$T(\lambda v + w) = \lim_{n \to \infty} T_n(\lambda v + w) = \lambda \lim_{n \to \infty} T_n v + \lim_{n \to \infty} T_n w = \lambda T v + T w.$$

We observe that

$$|\|Tv\| - \|T_n v\|| \leq \|Tv - T_n v\|,$$

and so, since $\|Tv - T_n v\| \to 0$,

$$\|Tv\| = \lim_{n \to \infty} \|T_n v\|.$$

We then have that and T is bounded, since

$$\|Tv\| = \lim_{n\to\infty} \|T_n v\| \leq (\lim_{n\to\infty} \|T_n\|)\|v\|.$$

Finally, $\|T_n - T\| \to 0$, since for n sufficiently large and $\|v\| \leq 1$,

$$\|T_n v - Tv\| = \lim_{m\to\infty} \|T_n v - T_m v\| \leq \lim_{m\to\infty} \|T_n - T_m\|\|v\| \leq \varepsilon\|v\| \leq \varepsilon,$$

and hence

$$\|T_n - T\| = \sup\{\|T_n v - Tv\| : \|v\| \leq 1\} \leq \varepsilon.$$

■

In particular, this means that $B(V, \mathbb{C})$ (or $B(V, \mathbb{R})$) is a Banach space.

Since elements of $L(V, W)$ and $B(V, W)$ are functions, it is an obvious question to ask what happens when two such operators are composed. Since composition of continuous functions are continuous, and composition of linear functions are linear, it is immediate that if $T \in L(V, W)$ and $S \in L(W, U)$, then $ST = S \circ T \in L(V, U)$. Norms of bounded operators have a particularly nice relationship when composed

Proposition 4.2.4
Let U, V, and W be normed vector spaces, and let $T \in B(V, W)$, $S \in B(W, U)$. Then $\|ST\| \leq \|S\|\|T\|$.

Proof:
We have that

$$\begin{aligned}\|ST\| &= \sup\{\|S(Tv)\| : \|v\| \leq 1\} \\ &\leq \sup\{\|S\|\|Tv\| : \|v\| \leq 1\} \\ &= \|S\|\|T\|.\end{aligned}$$

■

A consequence of this is that $B(V) = B(V, V)$ has a natural multiplicative structure defined by composition.

Definition 4.2.3
*If V is a Banach space, the Banach space $V^* = B(V, \mathbb{C})$ is called the **dual** of V. If $V \cong (V^*)^*$, then we say that V is **reflexive**.*

Note that since $L(V, \mathbb{C})$ and $B(V, \mathbb{C})$ are the same as vector spaces, we have two different topologies of interest on V^*: the norm topology, and the topology coming from pointwise convergence. We call this second topology the weak-* topology. These two topologies are somewhat analagous to the uniform and pointwise topologies on functions.

4.3 Case Study: L^p Spaces

In this section we will develop an important class of examples of Banach spaces. Throughout, let $(X, \mathcal{M}, \mu)$ be a measure space, and $\mathcal{L}(X, \mu)$ be the vector space of measurable functions modulo the equivalence relation $f \sim g$ if $f = g$ μ-almost everywhere. As we did in Example **??**, we will write $\boldsymbol{f}$ for the equivalence class of f.

For $0 < p < \infty$, we define

$$\|\boldsymbol{f}\|_p = \|f\|_p = \left(\int |f|^p \; d\mu\right)^{1/p}.$$

Clearly this is not affected by chenges to f on null-sets, so this value is in fact well-defined.

For $p = \infty$, we also define

$$\|\boldsymbol{f}\|_\infty = \|f\|_\infty = \inf\{a \geq 0 : \mu(\{x : |f(x)| > a\}) = 0\},$$

where we assume $\inf \emptyset = \infty$ as usual. In other words $\|\boldsymbol{f}\|_\infty$ is the smallest value such that $\{x : |f(x)| > a\}$ is null. This quantity is called the **e**sential supremum of f. Again, if $g \in \boldsymbol{f}$ then $\{x : |g(x)| > a\} \subseteq \{x : |f(x)| > a\} \cup \{x : g(x) \neq f(x)\}$, and since the latter is a union of null sets, the former is null. Therefore $\|\boldsymbol{f}\|_\infty$ does not depend on choice of equivalence class representative. One can think of this as being like the uniform norm, but neglecting values on sets of measure 0. Indeed, an equivalent definition is that

$$\|\boldsymbol{f}\|_\infty = \inf\{\|g\|_u : g \in \boldsymbol{f}\}.$$

We let

$$L^p(X, \mu) = \{\boldsymbol{f} : f \in \mathcal{L}(X, \mu), \|\boldsymbol{f}\|_p < \infty\}.$$

Where μ is implict we will write $L^p(X)$.

4.3.1 L^p Spaces are Banach Spaces

We now set about showing that $\|.\|_p$ are norms in many cases, and that the spaces $L^p(X, \mu)$ are in fact Banach spaces. We start by showing that they are vector spaces.

Lemma 4.3.1
If $0 < p \leq \infty$, $L^p(X, \mu)$ is a vector space.

Proof:

We note that for $0 < p < \infty$,

Strategy: *we show here that $\|\boldsymbol{f}\|_p^p < \infty$, which immediately implies that $\|\boldsymbol{f}\|_p < \infty$. This is a very common strategy, as it avoids cumbersome p-th roots in our calculations.*

$$|f + g|^p \leq (2 \max |f|, |g|)^p = 2^p \max |f|^p, |g|^p \leq 2^p(|f|^p + |g|^p),$$

and so

$$\|\boldsymbol{f} + \boldsymbol{g}\|_p^p = \int |f + g|^p \; d\mu \leq 2^p \int |f|^p \; d\mu + 2^p \int |g|^p \; d\mu < \infty.$$

If $p = \infty$, then if a and $b > 0$ such that

$$\mu(\{x : |f(x)| > a\}) = 0 \qquad \text{and} \qquad \mu(\{x : |g(x)| > b\}) = 0,$$

then

$$\mu(\{x : |f(x) + g(x)| > a + b\}) = 0.$$

Hence

$$\begin{aligned}\|f + g\|_\infty &= \inf\{c : \mu(\{x : |f(x) + g(x)| > c\}) = 0\} \\ &\leq \inf\{a + b : \mu(\{x : |f(x)| > a\}) = 0, \mu(\{x : |g(x)| > b\}) = 0\} < \infty.\end{aligned}$$

The fact that $\lambda f \in L^p(X, \mu)$ if $f \in L^p(X, \mu)$ for $\lambda \in \mathbb{C}$ is straightforward. ■

As the notation suggests, we want $\|\cdot\|_p$ to be a norm. In fact this is only the case when $p \geq 1$, and the stumbling block is the triangle inequality.

Lemma 4.3.2
If $a \geq 0$, $b \geq 0$ and $0 < \lambda < 1$, then

$$a^\lambda b^{1-\lambda} \leq \lambda a + (1 - \lambda)b.$$

and these are equal iff $a = b$.

Proof:
If a or b is 0 the result is trivial. By rearranging and dividing by b, the inequality is equivalent to saying

$$a^\lambda / b^\lambda - \lambda a/b \leq 1 - \lambda,$$

or letting $t = a/b$,

$$t^\lambda - \lambda t \leq (1 - \lambda).$$

But by elementary calculus, $t^\lambda - \lambda t$ has an absolute maxiumum at $t = 1$, and the value of this absolute maximum is $1 - \lambda$. Hence the inequality holds, and we have equality iff $a/b = 1$. ■

The next result is vital for the theory of L^p spaces, and will be used repeatedly.

Theorem 4.3.3 (Hölder's Inequality)
Let $1 < p < \infty$ and $1/p + 1/q = 1$, and f, $g \in \mathcal{L}(X, \mu)$. Then

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

Proof:
If $\|f\|_p = 0$ or $\|g\|_q = 0$, then $f = 0$ or $g = 0$ μ-a.e., and so the result is trivial. If $\|f\|_p = \infty$ or $\|g\|_q = \infty$, then the result is also trivial.

For each $x \in X$, we may apply Lemma 4.3.2 with

$$a = \frac{|f(x)|^p}{\|f\|_p^p}, \qquad b = \frac{|g(x)|^q}{\|g\|_q^q},$$

and $\lambda = 1/p$ (so $1 - \lambda = 1/q$). After some simplification, we have

$$\frac{|f(x)g(x)|}{\|f\|_p \|g\|_q} \leq \frac{|f(x)|^p}{p\|f\|_p^p} + \frac{|g(x)|^q}{q\|g\|_q^q},$$

and we can then integrate both sides to get

$$\frac{\|fg\|_1}{\|f\|_p \|g\|_q} \leq \frac{\|f\|_p^p}{p\|f\|_p^p} + \frac{\|g\|_q^q}{q\|g\|_q^q} = \frac{1}{p} + \frac{1}{q} = 1,$$

which gives the result. ■

Note that we have equality in the above precisely when $a = b$ for almost every x, or in other words, if

$$\|g\|_q^q |f|^p = \|f\|_p^p |g|^q \quad \mu\text{-a.e.}$$

or equivalently, when $|f|^p$ and $|g|^q$ are scalar multiples of one another.

Pairs of numbers p and q which satisfy $1 < p, q < \infty$ and

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

are important in the theory of L^p spaces. Note that we can always find $q = p/(1-p)$. We call q the conjugate exponent of p. Thinking of $1/\infty$ as 0, we will also say that 1 is the conjugate exponent of ∞ and ∞ the conjugate exponent of 1. Also note that 2 is its own conjugate exponent.

Theorem 4.3.4 (Minkowski's Inequality)
If $1 \leq p \leq \infty$, and f, $g \in L^p(X)$, then

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

Proof:
The case $p = 1$ is easy.

Consider $1 < p < \infty$. If $f + g = 0$ μ-a.e., the result is trivial. If not, note that

$$|f+g|^p = |f+g||f+g|^{p-1} \leq (|f|+|g|)|f+g|^{p-1} = |f||f+g|^{p-1} + |g||f+g|^{p-1},$$

and integrate both sides:

$$\int |f+g|^p \, d\mu \leq \int |f||f+g|^{p-1} \, d\mu + \int |g||f+g|^{p-1} \, d\mu.$$

Then Hölder's inequality applied twice to the left-hand side tells us that

$$\begin{aligned}\int |f+g|^p \, d\mu &\leq \|f\|_p \||f+g|^{p-1}\|_q + \|g\|_p \||f+g|^{p-1}\|_q \\ &\leq (\|f\|_p + \|g\|_p) \left(\int |f+g|^{(p-1)q} \, d\mu \right)^{1/q} \\ &\leq (\|f\|_p + \|g\|_p) \left(\int |f+g|^p \, d\mu \right)^{1/q},\end{aligned}$$

since $p = (p-1)q$. But then

$$\|f\|_p + \|g\|_p \geq \left(\int |f+g|^p \, d\mu \right)^{1-1/q} = \|f+g\|_p,$$

as required.

Finally, for $p = \infty$, we observe that if x is such that $|f(x)+g(x)| > \|f\|_\infty + \|g\|_\infty$, then either $|f(x)| > \|f\|_\infty$ or $|g(x)| > \|g\|_\infty$, so

$$\{x : |f(x)+g(x)| > \|f\|_\infty + \|g\|_\infty\} \subseteq \{x : |f(x)| > \|f\|_\infty\} \cup \{x : |g(x)| > \|g\|_\infty\}.$$

But these are both μ-null sets, so the union is μ-null, and so

$$\|f+g\|_\infty = \inf\{a \geq 0 : \mu(\{x : |f(x)+g(x)| > a\}) = 0\} \leq \|f\|_\infty + \|g\|_\infty.$$

■

Corollary 4.3.5
For $1 \leq p \leq \infty$, $L^p(X)$ is a normed vector space.

Proof:
It is immediate that if $\|\boldsymbol{f}\|_p = 0$, then $f = 0$ μ-a.e. and so $\boldsymbol{f} = 0$.

It is immeduate from the definition that $\|\lambda \boldsymbol{f}\|_\infty = |\lambda| \|\boldsymbol{f}\|_\infty$. For $1 \leq p < \infty$, we have

$$\|\lambda \boldsymbol{f}\|_p = \left(\int |\lambda f\| \, d\mu \right)^{1/p} = |\lambda| \|\boldsymbol{f}\|_p.$$

Finally, Minkowski's inequality immediately implies the triangle inequality for equivalence classes. ■

Indeed, these are Banach spaces.

Theorem 4.3.6
For $1 \leq p \leq \infty$, $L^p(X)$ is a Banach space.

Proof:
For $1 \leq p < \infty$, we use Proposition 4.1.4. Assume that we have an absolutely convergent series in $L^p(X)$,

$$\sum_{k=1}^{\infty} \|\boldsymbol{f}_k\|_p = K < \infty.$$

Let s_n be the partial sum

$$s_n = \sum_{k=1}^{n} |f_k|,$$

so by Minkowski's inequality,

$$\|\boldsymbol{s}_n\|_p \le \sum_{k=1}^{n} \|\boldsymbol{f}_k\|_p = K < \infty,$$

and since s_n is a monotone sequence, the monotone convergence theorem tells us that s_n converges pointwise a.e. to some function s, and

$$\int s^p \; d\mu = \lim_{n\to\infty} \int s_n^p \; d\mu \le K^p < \infty.$$

Hence

$$\boldsymbol{s} = \sum_{k=1}^{\infty} |\boldsymbol{f}_k| \in L^p(X),$$

which implies in particular that this function is finite almost everywhere. But this means that Proposition 3.3.9 applies, so the series

$$\sum_{k=1}^{\infty} f_n$$

converges pointwise a.e. to some function f. Clearly $|f| \le s$, and so $f \in L^p(X)$, and furthermore

$$|f - \sum_{k=1}^{n} f_k|^p \le (|f| + \sum_{k=1}^{n} |f_k|)^p \le (2s)^p,$$

so these are L^1 functions, and

$$\left\| \boldsymbol{f} - \sum_{k=1}^{n} \boldsymbol{f}_k \right\|_p^p = \int |f - \sum_{k=1}^{n} f_k|^p \; d\mu \to 0$$

by the dominated convergence theorem. So the series converges in L^p, and so Proposition 4.1.4 tells us that this is a Banach space.

$p = \infty$ is an exercise. ■

It is perhaps worth considering why $\|\cdot\|_p$ is not a norm for $0 < p < 1$. For any a and $b > 0$, one can see that

$$a^p + b^p > (a+b)^p,$$

and if we can find any two disjoint sets A and B of finite measure, then

$$\|\chi_A + \chi_B\|_p = (\mu(A) + \mu(B))^{1/p} > \mu(A)^{1/p} + \mu(B)^{1/p} = \|\chi_A\|_p + \|\chi_B\|_p.$$

Hence the triangle inequality cannot hold.

The case of $L^2(X, \mu)$ is special, because it turns out that this is in fact a Hilbert space. We define an inner product on $L^2(X, \mu)$ by

$$\langle \boldsymbol{f}, \boldsymbol{g} \rangle = \int f\overline{g} \; d\mu.$$

The first thing to note is that this is actually finite for functions in $L^2(X, \mu)$, by Hölder's inequality with $p = q = 2$:

$$\begin{aligned} |\langle \boldsymbol{f}, \boldsymbol{f} \rangle| &\leq \int |f\overline{g}| \; d\mu \\ &= \|\boldsymbol{f}\overline{\boldsymbol{g}}\|_1 \\ &\leq \|\boldsymbol{f}\|_2 \|\overline{\boldsymbol{g}}\|_2 = \|\boldsymbol{f}\|_2 \|\boldsymbol{g}\|_2 < \infty. \end{aligned}$$

It is clear that this is linear in the first variable and conjugate linear in the second. The fact that it is antisymmetric is also immediate. Finally

$$\langle \boldsymbol{f}, \boldsymbol{f} \rangle = \int f\overline{f} \; d\mu = \int |f|^2 \; d\mu = \|\boldsymbol{f}\|_2^2,$$

so the inner product of $\boldsymbol{f}$ with itself is 0 if and only if $f = 0$ almost everywhere.

This equality also tells us that the norm given by the inner product is exactly the norm $\|.\|_2$, and the fact that $L^2(X, \mu)$ is a complete in this norm, means that it is in fact a Hilbert space.

We now investigate the relationship between L^p-spaces for different values of p in two special cases. The first is the case where the measure space $(X, \mathcal{M}, \mu)$ is finite.

Proposition 4.3.7
If $(X, \mathcal{M}, \mu)$ is a finite measure space, and $0 < p < q \leq \infty$, then $L^q(X, \mu) \subseteq L^p(X, \mu)$.

Proof:
As is often the case, we deal with $q = \infty$ separately. In this case we note that if $\boldsymbol{f} \in L^\infty(X)$, we have $|f|^p \leq \|f\|_q^p < \infty$ μ-a.e., and so

$$\|\boldsymbol{f}\|_p^p = \int |f|^p \; d\mu \leq \int \|\boldsymbol{f}\|_\infty^p \; d\mu \leq \|\boldsymbol{f}\|_\infty^p \mu(X) < \infty.$$

So $\boldsymbol{f} \in L^p(X)$.

If $q < \infty$, we note that q/p and $q/(p - q)$ are conjugate exponents (since $q/p > 1$), and so we can use Hölder's inequality with these exponents to turn a power of p into a power of q:

$$\begin{aligned}\|\boldsymbol{f}\|_p^p = \int |f|^p \, d\mu &= \||f|^p 1\|_1 \\ &\le \||f|^p\|_{q/p}\|1\|_{q/(p-q)} \\ &\le \left(\int |f|^q \, d\mu\right)^{p/q} \mu(X)^{q/(p-q)} \\ &\le \|\boldsymbol{f}\|_q^p \mu(X)^{q/(p-q)} < \infty,\end{aligned}$$

so if $\boldsymbol{f} \in L^q(X)$, then $\boldsymbol{f} \in L^p(X)$. ■

We note that in general, this is a proper inclusion.

Example 4.3.1

Consider the finite measure space $([0,1], m)$, where m is Lebesgure measure restricted to the interval $[0,1]$. Given any $0 < p < q \le \infty$, choose r so that $p < r < q$. Then the function

$$f(x) = x^{-1/r}$$

is in $L^p([0,1])$, because

$$\int |x^{-1/r}|^p \, dm = \int_0^1 x^{-p/r} \, dx = \left[\frac{1}{1-p/r} x^{1-p/r}\right]_0^1 = \frac{1}{1-p/r} x^{1-p/r},$$

since $p/r < 1$. However $f \notin L^q([0,1])$, because the same calculation with q has

$$\int |x^{-1/r}|^q \, dm = \int_0^1 x^{-q/r} \, dx = \infty,$$

because $q/r > 1$.

Hence in this case L^q is a proper subset of $L^p(X)$. ◇

The other special case where we have a relationship between the L^p spaces is when we have the counting measure on a set. This special case is significant enough that we have a special notation for such L^p spaces.

Definition 4.3.1

Let c be the counting measure on a set X. We define

$$\ell^p(X) = L^p(X, c).$$

We note that since the only null set for counting measure is the empty set, $\boldsymbol{f} = \{f\}$, ie. there is no real distinction between equivalence classes and the actual functions.

It turns out that for fixed p, the only thing which distinguishes these spaces is the cardinality of X.

Proposition 4.3.8
If X and Y are sets with $|X| = |Y|$, and $1 \leq p \leq \infty$, then $\ell^p(X)$ is isometrically isomorphic to $\ell^p(Y)$.

Proof:
Since X and Y have the same cardinality, we have a bijection $\varphi : X \to Y$, which in turn induces a map φ^* from functions on Y to functions on X by

$$\varphi^* f = f \circ \varphi.$$

This function φ^* is clearly linear, and furthermore, if $f \in \ell^p(Y)$, we have

$$\begin{aligned}
\|\varphi^* f\|_p^p &= \sum_{x \in X} |(\varphi^* f)(x)|^p \\
&= \sum_{x \in X} |f(\varphi(x))|^p \\
&= \sum_{y \in Y} |f(y)|^p = \|f\|_p^p,
\end{aligned}$$

since φ is bijective. Hence φ^* is an isometry. It is also a vector space isomorphism, since $(\varphi^{-1})^*$ is also a linear map, and $(\varphi^*)^{-1} = (\varphi^{-1})^*$. ■

For this reason, particularly if $|X|$ is finite, we sometimes use the notation

$$\ell^p_{|X|} = \ell^p(X).$$

For ℓ^p spaces, we get the reverse inclusion compared to finite measure spaces.

Proposition 4.3.9
Let X be a set, and $0 < p < q \leq \infty$. Then $\ell^p(X) \subseteq \ell^q(X)$.

To prove this, we use a little lemma which tells us that if a function is both in L^p and L^q, it is also in L^r for every r in between p and q.

Lemma 4.3.10
If $0 < p < r < q \leq \infty$, then $L^q(X) \cap L^p(X) \subseteq L^r(X)$, and moreover if $\boldsymbol{f} \in L^q(X) \cap L^p(X)$,

$$\|\boldsymbol{f}\|_r \leq \|\boldsymbol{f}\|_p^\lambda \|\boldsymbol{f}\|_q^{1-\lambda},$$

where λ is such that

$$\frac{1}{r} = \frac{\lambda}{p} + \frac{1-\lambda}{q},$$

or, if $q = \infty$, $1/r = \lambda/p$.

Proof:
If $q < \infty$, we can use Hölder's inequality with conjugate exponents

$$\frac{p}{\lambda r} \quad \text{and} \quad \frac{q}{(1-\lambda)r}$$

to obtain:

$$\begin{aligned}\|\boldsymbol{f}\|_r^r &= \int |f|^{\lambda r}|f|^{(1-\lambda)r}\ d\mu \\ &\leq \||f|^{\lambda r}\|_{\frac{p}{\lambda r}}\||f|^{(1-\lambda)r}\|_{\frac{q}{(1-\lambda)r}} \\ &= \|\boldsymbol{f}\|_p^{\lambda r}\|\boldsymbol{f}\|_q^{(1-\lambda)r},\end{aligned}$$

and so

$$\|\boldsymbol{f}\|_r \leq \|\boldsymbol{f}\|_p^{\lambda}\|\boldsymbol{f}\|_q^{1-\lambda} < \infty$$

as required.

If $q = \infty$, we have

$$\|\boldsymbol{f}\|_r^r = \int |f|^{r-p}|f|^p\ d\mu \leq \|\boldsymbol{f}\|_\infty^{r-p}\|\boldsymbol{f}\|_p^p,$$

so, since $\lambda = p/r$, taking rth roots gives

$$\|\boldsymbol{f}\|_r = \|\boldsymbol{f}\|_\infty^{1-\lambda}\|\boldsymbol{f}\|_p^{\lambda} < \infty,$$

as required. ■

Proof (Proposition 4.3.9):

If $q = \infty$, we have that

$$\|f\|_\infty^p \leq \sup_{x\in X}|f(x)|^p \leq \sum_{x\in X}|f(x)|^p = \|f\|_p^p < \infty.$$

If $q < \infty$, then we use the Lemma with $p < q < \infty$ as our three exponents, and the fact that we just showed that $\|f\|_\infty \leq \|f\|_p$, to get

$$\|f\|_q \leq \|f\|_p^{\lambda}\|f\|_\infty^{1-\lambda} \leq \|f\|_p,$$

where $\lambda = p/q$. ■

Once again, we expect the inclusion of ℓ^p spaces to be proper in general.

Example 4.3.2

Consider $\ell^p(\mathbb{N})$. This is essentially the space of series which are absolutely p-summable, ie. if $f : k \mapsto a_k$ is in $\ell^p(\mathbb{N})$

$$\sum_{k=1}^{\infty}|a_k|^p < \infty.$$

For $0 < p < r < q \leq \infty$, the function

$$f(k) = k^{-1/r}$$

is in $L^q(\mathbb{N})$, since

$$\|f\|_q^q = \sum_{k=1}^{\infty} k^{-q/r} < \infty,$$

since $q/r > 1$, or if $q = \infty$, since $\|f\|_\infty = 1$. However f is not in $L^p(\mathbb{N})$, since

$$\|f\|_p^p = \sum_{k=1}^{\infty} k^{-p/r} = \infty,$$

since $p/r < 1$.

So in particular, we have a proper inclusion $\ell^p(X) \subset \ell^q(X)$. ◇

Example 4.3.3

Let X be a finite set. Then $(X, \mathcal{P}(X), c)$ is a finite measure space, and so we have that $\ell^p(X)$ and $\ell^q(X)$ are isomorphic as vector spaces, since every vector of each is an element of the other by combining the two inclusion Propositions we have just proved.

Indeed, they are both isomorphic as vector spaces to $\mathbb{C}^{|X|}$, and one can show that they are isomorphic as topological vector spaces to $\mathbb{C}^{|X|}$ with the topology of pointwise (ie. coordinate-wise) convergence. However the norms are different in general, and so they are not isometrically isomorphic as Banach spaces. ◇

4.3.2 Duals of L^p Spaces

Having spent some time showing that L^p spaces are Banach spaces, we would like to now investigate some of their properties. Based on our previous discussion, the question of what the duals of these Banach spaces look like would seem to be of interest.

The first thing that we need to do in looking at the duals of these spaces, is to try to identify an interesting class of linear functionals.

Lemma 4.3.11

Let (X, m) be a measure space, and let p and q be conjugate exponents with $1 \leq q < \infty$. For each $\boldsymbol{g} \in L^q(X)$, we have that

$$\varphi_{\boldsymbol{g}}(\boldsymbol{f}) = \int fg \; d\mu$$

is a bounded linear functional on $L^p(X)$, with $\|\varphi_{\boldsymbol{g}}\| = \|\boldsymbol{g}\|_q$. The result holds for $q = \infty$ if m is semifinite.

Proof:

It is immediate that $\varphi_{\boldsymbol{g}}$ is linear. Hölder's inequality also immediately tells us that

$$|\varphi_{\boldsymbol{g}}(\boldsymbol{f})| \leq \|\boldsymbol{fg}\|_1 \leq \|\boldsymbol{f}\|_p \|\boldsymbol{g}\|_q,$$

so $\|\varphi_{\boldsymbol{g}}\| \leq \|\boldsymbol{g}\|_q$, and hence $\varphi_{\boldsymbol{g}} \in (L^p(X))^*$.

We trivially have $\|\varphi_{\boldsymbol{g}}\| = \|\boldsymbol{g}\|_q$ if $g = 0$ μ-a.e. If $\boldsymbol{g} > 0$ and $q < \infty$, we let

Strategy: *we need to find function f such that $\|\boldsymbol{f}\|_p =$ and $|\varphi_{\boldsymbol{g}}(\boldsymbol{f})| = \|\boldsymbol{g}\|_q$.*

$$f = \frac{g^{q-1}}{\|g\|_q^{q-1}}.$$

Then

$$\|\boldsymbol{f}\|_p^p = \frac{1}{\|\boldsymbol{g}\|_q^{(q-1)p}} \int g^{(q-1)p} \, d\mu = \frac{\|\boldsymbol{g}\|_q^q}{\|\boldsymbol{g}\|_q^q} = 1, \tag{4.1}$$

and

$$|\varphi_{\boldsymbol{g}}(\boldsymbol{f})| = \int fg \, d\mu = \frac{1}{\|\boldsymbol{g}\|_q^{q-1}} \int g^q \, d\mu = \|\boldsymbol{g}\|_q,$$

with the first equality holding since both f and g are non-negative functions. Hence we have that $\|\varphi_{\boldsymbol{g}}\| \geq \|\boldsymbol{g}\|_q$.

For arbitrary $g \neq 0$, we use

$$f = \frac{|g|^{q-1}\overline{\operatorname{sign} g}}{\|\boldsymbol{g}\|_q^{q-1}}$$

instead, so that $fg > 0$.

Finally, if $q = \infty$ and μ is seminfinite, we know that for every $\varepsilon > 0$, we can find a set $B = \{x : |g(x)| > \|\boldsymbol{g}\|_\infty - \varepsilon\}$ with $\mu(B) > 0$, and since μ is semifinite, we have an $A \subseteq B$ such that $0 < \mu(A) < \infty$. If $\boldsymbol{g} > 0$, then we let $f = \mu(A)^{-1}\chi_A$, so that

$$\|\boldsymbol{f}\|_1 = \mu(A)^{-1} \int \chi_A \, d\mu = 1,$$

and

$$|\varphi_{\boldsymbol{g}}(\boldsymbol{f})| = \int fg \, d\mu = \mu(A)^{-1} \int_A g \, d\mu \geq \mu(A)^{-1}(\mu(A)(\|\boldsymbol{g}\|_\infty - \varepsilon)) = \|\boldsymbol{g}\|_\infty - \varepsilon.$$

So $\|\varphi_{\boldsymbol{g}}\| \geq \|\boldsymbol{g}\|_\infty - \varepsilon$ for all $\varepsilon > 0$, and so $\|\varphi_{\boldsymbol{g}}\| \geq \|\boldsymbol{g}\|_\infty$. Again, for arbitrary $g \neq 0$, we use $f = \mu(A)^{-1}\chi_A\overline{\operatorname{sign} g}$ instead, so that $fg \geq 0$. ■

Notice that the map $\boldsymbol{g} \mapsto \varphi_{\boldsymbol{g}}$ is itself a linear map, so this fact, together with the above lemma, tells us that we have an isometric map from $L^q(X, \mu)$ to the dual space $(L^p(X, \mu))^*$. The obvious question to ask is whether or not this gives us all of the linear functionals in the dual space, ie. is it a surjection?

The first step to showing this is to rule out the most obvious case: that we can find linear functionals $\varphi_{\boldsymbol{g}}$ on $L^p(X)$ where $\boldsymbol{g}$ is not in $L^q(X)$. To do this, we prove a sort of converse of Hölder's inequality.

Theorem 4.3.12

Let $(X, \mathcal{M}, \mu)$ be a σ-finite measure space, and p and q are conjugate exponents. Let S be the set of all simple functions whose support is contained in a set of finite measure. If $g \in \mathcal{L}(X, \mu)$ is such that $fg \in L^1(X)$ for all $f \in S$, and

$$M_q = \sup\left\{\left|\int fg \, d\mu\right| : f \in S, \|f\|_p = 1\right\} < \infty,$$

then $\boldsymbol{g} \in L^q(X)$ *and* $\|\boldsymbol{g}\|_q = M_q(X)$.

Proof:

First assume that $q < \infty$. Since X is σ-finite, we can find sets E_n of finite measure, with $E_n \subseteq E_{n+1}$ and

$$X = \bigcup_{n=1}^{\infty} E_n.$$

Assume that $g > 0$. We know that we can find an increasing sequence of measurable simple functions ψ_n which converge pointwise to g, and by multiplying by χ_{E_n} we can assume without loss of generality that ψ_n is 0 off of E_n. Letting

$$f_n = \frac{\psi_n^{q-1}}{\|\psi_n\|_q^{q-1}},$$

we have $\|f_n\|_p = 1$, by the same reasoning as (4.1). Fatou's lemma then tells us that

$$\begin{aligned}\|g\|_q &\leq \liminf \|\psi_n\|_q \\ &= \liminf \int f_n \psi_n \, d\mu \\ &\leq \liminf \int f_n g \, d\mu \leq M_q(g).\end{aligned}$$

Given general g, we simply have $\psi_n \to g$ pointwise, and we use

$$f_n = \frac{\psi_n^{q-1} \overline{\operatorname{sign} \psi_n}}{\|\psi_n\|_q^{q-1}}.$$

If $q = \infty$, then given any $\varepsilon > 0$ we consider $A = \{x : |g(x)| \geq M_\infty(g) + \varepsilon\}$. If $\mu(A) > 0$, then there must be a subset B of A which has finite measure, and then there are measurable simple functions φ_n which converge pointwise to g. Letting $f_n = \mu(B)^{-1} \overline{\operatorname{sign} \varphi_n} \chi_B$, we have that

$$\|f_n\|_1 = \mu(B)^{-1} \int \chi_B \, d\mu = 1,$$

but

$$\int f_n \varphi_n \, d\mu = \mu(B)^{-1} \int_B |\varphi_n| \, d\mu \to \mu(B)^{-1} \int_B |g| \, d\mu \geq M_\infty(g) + \varepsilon.$$

But

Having shown that $g \in L^q$, we immediately get from Hölder's inequality that

$$\left| \int fg \, d\mu \right| \leq \int |fg| \, d\mu \leq \|f\|_p \|g\|_q = \|g\|_q,$$

for any $f \in S$ with $\|f\|_p = 1$, so that $M_q(g) \leq \|g\|_q$. ■

Indeed, this theorem holds even if μ is semifinite.

We are now in a position to prove the main duality result. Unfortunately, in the general case the L^∞ spaces are exceptional, and do not give us a nice duality theory. However, for all other values of p, we get that $(L^p(X))^*$ is isometrically isomorphic to $L^q(X)$.

Theorem 4.3.13
Let $(X, \mathcal{M}, \mu)$ be a σ-finite measure space, and p and q are conjugate exponents with $1 \leq p < \infty$. For all $\varphi \in (L^p(X))^*$ there is a $g \in L^q(X)$ such that $\varphi = \varphi_g$. For $1 < p < \infty$, this holds even if μ is not σ-finite.

Proof:

We first prove the case where μ is finite. In this case, every bounded function, and hence every simple function, in in L^p. Given some $\varphi \in (L^p)^*$, we define a complex measure ν_φ by $\nu_\varphi(E) = \varphi(\chi_E)$. That this is a measure follows from the fact that if $E = \bigcup_{n=1}^\infty E_n$ as a disjoint union, then $\chi_E = \sum_{n=1}^\infty \chi_{E_n}$, and the series converges in L^p, since

$$\left\| \chi_E - \sum_{k=1}^n \chi_{E_k} \right\|_p = \left(\int \sum_{k=n}^\infty \chi_{E_k} \, d\mu \right)^{1/p} = \mu(\bigcup_{k=n}^\infty E_k)^{1/p},$$

which converges to 0 as $n \to \infty$. Since φ is bounded, and hence continuous,

$$\nu_\varphi(\bigcup_{k=1}^n E_k) = \varphi(\sum_{k=1}^\infty \chi_{E_k}) = \sum_{k=1}^\infty \varphi(\chi_{E_k}) = \sum_{k=1}^\infty \nu_\varphi(E_k).$$

The other properties of measures are immediate from the definition.

Also $\nu_\varphi \ll \mu$, since if $\mu(E) = 0$, then $\chi_E = 0$ in L^p, and so $\nu_\varphi(E) = \varphi(\chi_E) = \varphi(0) = 0$. So the Radon-Nikodym theorem for complex measures tells us that there is a Radon-Nikodym derivative $g = \frac{d\nu_\varphi}{d\mu} \in L^1(\mu)$, such that

$$\varphi(\chi_E) = \nu(E) = \int_E g \, d\mu,$$

or, for any simple function f,

$$\varphi(f) = \int fg \, d\mu.$$

Furthermore,

$$\left| \int fg \, d\mu \right| = |\varphi(f)| \leq \|\varphi\| \|f\|_p,$$

so by the converse of Hölder's inequality, $g \in L^q$.

Finally, if μ is finite, then the simple functions are dense in L^p, since we can find simple functions ψ_n which converge pointwise to any $f \in L^p$ with

$|\psi_n| \leq |f|$, and then one can see that $|f - \psi_n|^p$ converges pointwise to 0, and is bounded above by $2^p|f|^p$, which is in L^1, and so the dominated convergence theorem tells us that

$$\lim_{n\to\infty} \|f - \psi_n\|_p^p = \lim_{n\to\infty} \int |f - \psi_n|^p \, d\mu = 0,$$

so $\psi_n \to f$ in L^p. But then it is immediate from the dominated convergence theorem that

$$\varphi(f) = \lim_{n\to\infty} \varphi(\psi_n) = \lim_{n\to\infty} \int \psi_n g \, d\mu = \int fg \, d\mu.$$

Hence $\varphi = \varphi_g$ for some $g \in L^q$.

Now we consider the case where μ is σ-finite. We can find an increasing sequence of sets E_n such that $X = \bigcup_{n=1}^\infty E_n$ and $0 < \mu(E_n) < \infty$. Let $L^p(E_n)$ and $L^q(E_n)$ be identified with the subspaces of $L^p(X)$ and $L^q(X)$ of functions which are 0 off E_n. The arguement for finite measures tells us that for each n, we can find a $g_n \in L^q(E_n)$ such that

$$\varphi(f) = \int f g_n \, d\mu$$

for all $f \in L^p(E_n)$. Moreover $\|g_n\|_q = \|\varphi|_{L^p(E_n)}\| \leq \|\varphi\|$, and this is an increasing sequence. Now g_n is unique up to modification on a null set, so $g_n = g_m$ on E_n for $n \leq m$. So we can define a pointwise limit

$$g = \lim_{n\to\infty} g_n,$$

and in fact $|g_n|$ is increasing pointwise to $|g|$, so the monotone convergence theorem tells us that

$$\|g\|_q^q = \lim_{n\to\infty} \int |g_n|^q \, d\mu = \lim_{n\to\infty} \|g_n\|_q \leq \|\varphi\| < \infty,$$

so $g \in L^q$.

Finally, if $f \in L^p$, $f\chi_{E_n} \in L^p(E_n)$, and $f\chi_{E_n} \to f$ in L^p by the dominated convergence theorem. So

$$\varphi(f) = \lim_{n\to\infty} \varphi(f\chi_{E_n}) = \lim_{n\to\infty} \int f g_n \, d\mu = \int fg \, d\mu,$$

again using the dominated convergence theorem. ∎

Corollary 4.3.14
If $1 < p < \infty$, then $L^p(X)$ is reflexive.

4.4 Hilbert Spaces

Recall that a Hilbert space is a vector space equipped with an inner product, and which is complete (ie. a Banach space) in the norm

$$\|v\| = \langle x, x \rangle^{1/2}$$

derived from the inner product. Hilbert spaces are particularly nice examples of Banach spaces, and their theory is important in many areas of analysis.

Example 4.4.1

If (X, μ) is a measure space, then $L^2(X, \mu)$ is a Hilbert space with inner product

$$\langle f, g \rangle = \int f\overline{g}\, d\mu.$$

Since $f\overline{f} = |f|^2$, it is immediate that the norm one gets from this inner product is just the usual L^2 norm $\|.\|_2$.

Indeed, if μ is the counting measure, we get

$$\langle f, g \rangle = \sum_{x \in X} f(x)\overline{g(x)},$$

which is very reminiscent of the inner products you will have seen from undergraduate linear algebra. ◇

This example is particularly instructive, since we will see that every Hilbert space can be written in this form for an appropriate set.

We observing that the parallelogram law holds for vectors in a Hilbert space.

Proposition 4.4.1 (Parallelogram Law)
If H is a Hilbert space, then for all u, $v \in H$,

$$\|u + v\|^2 + \|u - v\|^2 = 2(\|u\|^2 + \|v\|^2).$$

Proof:

We have

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

■

The inner product has an important continuity condition.

Lemma 4.4.2
If $u_n \to u$ and $v_n \to v$ in a Hilbert space H, then

$$\lim_{n\to\infty} \langle u_n, v_n\rangle = \langle u, v\rangle.$$

Proof:
We observe that

$$\begin{aligned}
|\langle u_n, v_n\rangle - \langle u, v\rangle| &= |\langle u_n, v_n\rangle - \langle u_n, v\rangle + \langle u_n, v\rangle - \langle u, v\rangle| \\
&\le |\langle u_n, v_n - v\rangle| + |\langle u_n - u, v\rangle| \\
&\le \|u_n\|\|v_n - v\| + \|u_n - u\|\|v\|,
\end{aligned}$$

which converges to 0 as $n \to \infty$ since $\|v_n - v\| \to 0$, $\|u_n - u\| \to 0$. ■

Hilbert spaces have the advantage over Banach spaces that the concept of orthogonal vectors makes sense. We say that two vectors u and v in a Hilbert space H are orthogonal, and write $u \perp v$, if

$$\langle u, v\rangle = 0.$$

This is clearly symmetric: $u \perp v \iff v \perp u$. Given any subset X of H, we define

$$X^\perp = \{v : v \perp x, \forall x \in X\}.$$

Lemma 4.4.3
For any subset X of a Hilbert space H, the set $X^\perp$ is a closed subspace of H.

Proof:
If u, $v \in X^\perp$, then given any $x \in X$,

$$\langle \lambda u + v, x\rangle = \lambda\langle u, x\rangle + \langle v, x\rangle = 0,$$

so $\lambda u + v \in X^\perp$, and so $X^\perp$ is a subspace.
Furthermore, if $u_n \in X^\perp$, and $u_n \to u$ in H, then for any $x \in X$,

$$\langle u, x\rangle = \lim_{n\to\infty} \langle u_n, x\rangle = 0.$$

So $X^\perp$ is closed. ■

In fact given a closed subspace K of a Hilbert space H, we have that K and $K^\perp$ span H. More precisely,

Theorem 4.4.4
If K is a closed subspace of a Hilbert space H, then every $v \in H$ can be expessed uniquely as a sum $v = u + w$ with $u \in K$ and $w \in K^\perp$, where u and w are the unique elements with minimal the distance from v in K and $K^\perp$, respectively.

Pythagoras' theorem holds for mutually orthogonal sets of vectors.

Theorem 4.4.5
If H is a Hilbert space, and $v_1, \ldots, v_n \in H$ are pairwise orthogonal, ie. $v_k \perp v_m$ if $k \neq m$, then

$$\left\| \sum_{k=1}^{n} v_k \right\|^2 = \sum_{k=1}^{n} \|v_k\|^2.$$

Proof:
We have

$$\begin{aligned} \left\| \sum_{k=1}^{n} v_k \right\|^2 &= \left\langle \sum_{k=1}^{n} v_k, \sum_{m=1}^{n} v_m \right\rangle \\ &= \sum_{m,k=1}^{n} \langle v_k, v_m \rangle \\ &= \sum_{k=1}^{n} \langle v_k, v_k \rangle \\ &= \sum_{k=1}^{n} \|v_k\|^2, \end{aligned}$$

since $\langle v_k, v_m \rangle = 0$ if $k \neq m$. ■

Mutually orthogonal sets of vectors are important in a number of areas of analysis.

Example 4.4.2
Consider functions $f_k(\theta) = e^{ik\theta} \in L^2([0, 2\pi])$ for $k \in \mathbb{Z}$. f_k and f_n are orthogonal if $k \neq n$, since

$$\begin{aligned} \langle f_k, f_n \rangle &= \int f_k \overline{f_n} \; dm \\ &= \int_0^{2\pi} e^{ik\theta} e^{-in\theta} \; d\theta \\ &= \int_0^{2\pi} e^{i(k-n)\theta} \; d\theta \\ &= \left[\frac{1}{i(k-n)} e^{i(k-n)\theta} \right]_0^{2\pi} = 0. \end{aligned}$$

This example plays a key role in Fourier analysis. ◇

Example 4.4.3
Consider a set X with counting measure. Let

$$e_y(x) = \begin{cases} 1 & x = y \\ 0 & \text{otherwise} \end{cases}.$$

Then $e_y \perp e_z$ if $y \neq z$. Indeed, in this example we have the additional fact that $\|e_y\| = 1$. $\diamond$

The case where we have mutually orthogonal vectors whose norm is 1 is of particular importance. We say that a set of vectors in a Hilbert space H, $\{u_\alpha\}_{\alpha \in I}$ where I is some index set, is orthonormal if $u_\alpha \perp u_\beta$ for all α, $\beta \in I$ where $\alpha \neq \beta$, and $\|u_\alpha\|_\alpha = 1$ for all $\alpha \in I$.

Orthonormal sets are common, since if we are given any countable collection of vectors $v_1, v_2, \ldots, v_n, \ldots \in H$ which are linearly independent, ie.

$$v_n \notin \operatorname{span}\{v_1, \ldots, v_{n-1}\},$$

then we can obtain an orthonormal set via the Gram-Schmidt process. We work inductively, setting

$$\|u_1\| = \frac{1}{\|v_1\|} v_1,$$

and then, given $u_1, \ldots, u_{n-1}$, letting

$$u'_n = v_n - \sum_{k=1}^{n-1} \langle v_n, u_k \rangle u_k,$$

and

$$u_n = \frac{1}{\|u'_n\|} u'_n.$$

It is easy to see that these vectors are orthonormal. That $\|u_n\| = 1$ is immediate from the definition, and we can show that $u_k \perp u_n$ for all k and n with $k \neq n$ by induction. Assume that $u_k \perp u_m$ for all k, $m \leq n - 1$ with $k \neq m$. Then

$$\begin{aligned}
\langle u_n, u_k \rangle = \left\langle \frac{1}{\|u'_n\|} u'_n, u_k \right\rangle &= \frac{1}{\|u'_n\|} \left\langle v_n - \sum_{m=1}^{n-1} \langle v_n, u_m \rangle u_m, u_k \right\rangle \\
&= \frac{1}{\|u'_n\|} \left(\langle v_n, u_k \rangle - \sum_{m=1}^{n-1} \langle v_n, u_m \rangle \langle u_m, u_k \rangle \right) \\
&= \frac{1}{\|u'_n\|} \left(\langle v_n, u_k \rangle - \langle v_n, u_k \rangle \|u_k\|^2 \right) = 0.
\end{aligned}$$

Hence $u_k \perp u_m$ for all k, $m \leq n$ and $k \neq m$.

The Gram-Schmidt process has the clear disadvantage that it only works for countable collections of linearly independent vectors. There are Hilbert spaces which are not spanned by countable collections of linearly independent vectors, such as $\ell^2(X)$ where X is uncountable.

We now want to show that all Hilbert spaces have an orthonormal basis. The first step in this process is Bessel's inequality.

Theorem 4.4.6 (Bessel's Inequality)
If $\{u_\alpha\}_{\alpha\in I}$ is an orthonormal set in a Hilbert space H, then for any $v \in H$,

$$\sum_{\alpha\in I} |\langle v, u_\alpha\rangle|^2 \le \|v\|^2.$$

Note that this sum may have uncountably many terms, and is unordered. It's value is therefore

$$\sup\left\{\sum_{\alpha\in F} |\langle v, u_\alpha\rangle|^2 : F \subseteq I, F \text{ finite}\right\}.$$

and so the fact that the sum is finite means that at most a countable number of them are non-zero.
Proof:
Let $F \subset I$ be finite. Then, by Pythagoras,

$$\begin{aligned}
0 &\le \|v - \sum_{\alpha\in F}\langle v, u_\alpha\rangle u_\alpha\|^2 \\
&\le \|v\|^2 - 2\operatorname{Re}\left\langle v, \sum_{\alpha\in F}\langle v, u_\alpha\rangle u_\alpha\right\rangle + \left\|\sum_{\alpha\in F}\langle v, u_\alpha\rangle u_\alpha\right\|^2 \\
&= \|v\|^2 - 2\operatorname{Re}\sum_{\alpha\in F}\overline{\langle v, u_\alpha\rangle}\langle v, u_\alpha\rangle + \sum_{\alpha\in F}\|\langle v, u_\alpha\rangle u_\alpha\|^2 \\
&= \|v\|^2 - 2\sum_{\alpha\in F}|\langle v, u_\alpha\rangle| + \sum_{\alpha\in F}|\langle v, u_\alpha\rangle|^2\,\|u_\alpha\|^2 \\
&\le \|v\|^2 - 2\sum_{\alpha\in F}|\langle v, u_\alpha\rangle| + \sum_{\alpha\in F}|\langle v, u_\alpha\rangle|^2.
\end{aligned}$$

Hence

$$\sum_{\alpha\in F} |\langle v, u_\alpha\rangle|^2 \le \|v\|^2.$$

Taking suprema gives the result. ■

We can now show that every Hilbert space has an orthonormal basis.

Theorem 4.4.7
Let $\{u_\alpha\}_{\alpha\in I}$ is an orthonormal set in a Hilbert space H. Then the following are equivalent:

1. *for all $v \in H$, $\sum_{\alpha\in I} |\langle v, u_\alpha\rangle|^2 = \|v\|^2$.*
2. *for all $v \in H$, $v = \sum_{\alpha\in I}\langle v, u_\alpha\rangle u_\alpha$*
3. *if $\langle v, u_\alpha = 0$ for all $\alpha \in I$, then $v = 0$.*

Proof:

(3) ⇒ (2): Given any vector v, define u by

$$u = \sum_{\alpha \in I} \langle v, u_\alpha \rangle u_\alpha.$$

Then $\langle u, u_\alpha \rangle = \langle v, u_\alpha \rangle$ for all $\alpha in I$, and so

$$\langle u - v, u_\alpha \rangle = \langle u, u_\alpha \rangle - \langle v, u_\alpha \rangle = 0$$

for all $\alpha \in I$, and so by (3) we conclude that $u - v = 0$ and hence that

$$v = \sum_{\alpha \in I} \langle v, u_\alpha \rangle u_\alpha$$

(1) ⇒ (3): Assume that $\langle v, u_\alpha = 0$ for all $\alpha \in I$. Then

$$\|v\|^2 = \sum_{\alpha \in I} |\langle v, u_\alpha \rangle|^2 = 0$$

and hence $\|v\| = 0$. But this then implies that $v = 0$ from the axioms for a norm.

(2) ⇒ (1): By the definition of the Hilbert space norm, we have that

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

■

Proposition 4.4.8
Every Hilbert space has an orthonormal basis.

Proof:

This proof relies on a standard application of Zorn's Lemma. Let $\mathcal{O}$ be the set of all orthonormal sets in the Hilbert space H, with the partial order given by set inclusion. Let $\mathcal{C}$ be a totally ordered subset of $\mathcal{O}$, and let

$$U = \bigcup_{C \in \mathcal{C}} C.$$

Then if v, $w \in U$ we have that $v \in C_1$ and $u \in C_2$ for some C_1, $C_2 \in \mathcal{C}$. Since $\mathcal{C}$ is totally ordered, one of these orthonormal sets includes the other, so without loss of generality assume that $C_1 \subseteq C_2$. But then $v \in C_2$ and so

$$\langle v, w \rangle = \begin{cases} 1 & v = w \\ 0 & v \neq w. \end{cases}$$

Therefore U is an orthonormal set and it is an upper bound of $\mathcal{C}$ so the conditions of Zorn's lemma are satisfied. Thus we conclude that $\mathcal{O}$ has a maximal element M.

Assume that M is not complete. That is there is some $v \in H$ such that $\langle v, u \rangle = 0$ for all $u \in M$, but $v \neq 0$. But if this is the case, then let $v' = v/\|v\|$ so that $\langle v', u \rangle = 0$ and $\|v'\| = 1$. Then the set $M \cup \{v'\}$ is an orthonormal set which is strictly larget than the maximal orthonormal set M, contradicting the maximality of M. Therefore M must be complete, and so it is an orthonormal basis. ■

Recall that a topological space is separable if it has a countable dense subset.

Proposition 4.4.9
A Hilbert space is separable if and only if it has a countable orthonormal basis.

Proof:
If a Hilbert space H is separable then there is a countable set $X = \{x_n\}$ which is dense. By discarding those x_n which are linearly dependent on $x_1, x_2, \ldots, x_{n-1}$ we get a subset x_{n_k} which is linearly independent and whose linear span includes every element of X, and so is dense. Applying Gram-Schmidt to this subset we get an orthonormal set u_k whose linear span V is also dense. But if $v \in H$ with $\langle v, u_k \rangle = 0$ but $v \neq 0$, then $v \in V^\perp$ and so the distance from v to $\overline{V} = (V^\perp)^\perp$ is $\|v\| > 0$, which implies that V is not dense. Therefore the orthonormal set must be complete and so is a basis.

Conversely, if $\{u_k\}$ is a countable orthonormal set, then consider the countable dense subset Q of $\mathbb{C}$ consisting of those numbers with rational real and imaginary parts. Let X be all finite linear combinations of elements of the orthonormal set with coefficients in Q. This set is dense, since given any $v \in H$ and any $\varepsilon > 0$ we have that there is some n such that

$$v_n = \sum_{k=1}^{n} \langle v, u_k \rangle u_k$$

has $\|v - v_n\| < \varepsilon/2$, since $\{u_k\}$ is a basis, and for each k we can find $q_k \in Q$ such that $\|\langle v, u_k \rangle - q_k|^2 \leq \varepsilon^2/2^{2k+2}$. Then

$$x = \sum_{k=1}^{n} q_k u_k \in X$$

and $\|v_n - x\|^2 \leq \sum_{k=1}^{n} \varepsilon^2/2^{2k+2} < \varepsilon^2/2^2$. Therefore by the triangle inequality $\|v - x\| < \varepsilon$ and so X is dense in H. ■

Most Hilbert spaces that are encountered in applications are separable.

When we consider bounded linear operators between two Hilbert spaces H and K, the operators which preserve the inner product are of particular interest. We say that a linear operator $U : H \to K$ is unitary if

$$\langle Uv, Uw \rangle_K = \langle v, w \rangle_H.$$

A unitary operator is automatically bounded: indeed it must be an isometry and so $\|U\| = 1$. It is straightforward that compositions of unitary operators are again unitary.

If, in addition U has an inverse, then

$$\langle U^{-1}v, U^{-1}w\rangle_H = \langle UU^{-1}v, UU^{-1}w\rangle_K = \langle v, w\rangle_K$$

and so U^{-1} is unitary, and hence an isometry. Such an invertible unitary is therefore an isometric isomorphism, and we will call such invertible unitary operators unitary isomorphisms. From a category theory standpoint, if there is a unitary isomorphism between two Hilbert spaces then they are essentially the same Hilbert space.

Proposition 4.4.10
Every Hilbert space H is unitarily isomorphic to $\ell^2(X)$ for some set X.

Proof:
Let $\{u_\alpha\}_{\alpha\in X}$ be an orthonormal basis for H. Given $v \in H$, we define a function $f_v : X \to \mathbb{C}$ by

$$f_v(\alpha) = \langle v, u_\alpha\rangle.$$

Then

$$\|f_v\|_2^2 = sum_{\alpha\in X}|\langle v, u_\alpha\rangle|^2 = \|v\|_2 < \infty,$$

so $f_v \in \ell^2(X)$. Therefore we can define a function $U : H \to \ell^2(X)$ by $Uv = f_v$. It is straightforward to verify that U is linear. Moreover

$$\begin{aligned}\langle Uv, Uw\rangle &= \sum_{\alpha\in X} f_v(\alpha)\overline{f_w(\alpha)}\\ &= \sum_{\alpha\in X}\langle v, u_\alpha\rangle\overline{\langle w, u_\alpha\rangle}\end{aligned}$$

and

$$\begin{aligned}\langle v, w\rangle &= \left\langle \sum_{\alpha\in X}\langle v, u_\alpha\rangle u_\alpha, \langle \sum_{\beta\in X}\langle w, u_\beta\rangle u_\beta \right\rangle\\ &= \sum_{\alpha\in X}\sum_{\beta\in X}\langle v, u_\alpha\rangle\overline{\langle w, u_\beta\rangle}\langle u_\alpha, u_\beta\rangle\\ &= \sum_{\alpha\in X}\langle v, u_\alpha\rangle\overline{\langle w, u_\alpha\rangle}.\end{aligned}$$

So U is unitary. Furthermore, given any $f \in \ell^2(X)$ we let

$$v_f = \sum_{\alpha\in X} f(\alpha)u_\alpha,$$

and then

$$\|v_f\|_2 = \sum_{\alpha\in X}|f(\alpha)|^2 = \|f\|_2^2,$$

so that v_f in ℓ^2 and so $Uv_f = f$ which means that U is surjective; and since it is an isometry it is injective. So U has an inverse map $U^{-1}f = v_f$ and this is easily seen to be linear. Hence U is a unitary isomorphism. ■

Corollary 4.4.11
Every orthonormal basis of a Hilbert space has the same cardinality.

Proof:
Let $\{u_\alpha\}_{\alpha\in X}$ and $\{v_\beta\}_{\beta\in Y}$ be two orthonormal bases in a Hilbert space H. Then H is unitarily isomorphic to both $\ell^2(X)$ and $\ell^2(Y)$. But this means that $\ell^2(X)$ and $\ell^2(Y)$ are unitarily isomorphic, and so in particular this means that they are isometrically isomorphic. But from our work on ℓ^p spaces, we know this happens if and only if $|X| = |Y|$. ∎

We define the dimension of a Hilbert space to be the cardinality of its orthonormal basis.

Corollary 4.4.12
If two Hilbert spaces H and K have the same dimension then they are unitarily isomorphic.

Corollary 4.4.13
If a Hilbert space is separable it is unitarily isomorphic to one of $\ell^2(\mathbb{N})$ or $\ell^2(\{1,\ldots,n\}) \cong \mathbb{C}^n$.

We now turn to consider the duals of Hilbert spaces. Since every Hilbert space is unitarily isomorphic to $\ell^2(X)$ and $\ell^2(X)$ is reflexive, it follows that every Hilbert space is reflexive. In fact, we know that $(\ell^2(X))^*$ is isometrically isomorphic to $\ell^2(X)$, so in addition the same will be true for general Hilbert spaces. However, we would like to push this slightly further to see what happens with unitary isomorphisms instead of isometric isomorphisms.

Given a Hilbert space H we define the conjugate space $\overline{H}$ of H to be the vector space which as a set equals H, and with the same vector addition, but with the scalar multiplication

$$(\lambda, v) \mapsto \overline{\lambda} v$$

where this scalar multiplication in the definition is the original scalar multiplication on H. One can verify that $\overline{H}$ is in fact a vector space, and moreover if we define

$$\langle v, w\rangle_{\overline{H}} = \langle w, v\rangle_H$$

then $\overline{H}$ is in fact a Hilbert space. The identity map $I : H \to \overline{H}$ defined by $Iv = v$ is a bijective isometry, since

$$\|v\|_H^2 = \langle v, v\rangle_H = \langle v, v\rangle_{\overline{H}} = \|v\|_{\overline{H}}^2,$$

but it is conjugate linear (ie. $I(v + w) = Iv + Iw$ and $I(\lambda v) = \overline{\lambda} v$) rather than linear.

We will show that H^* is isometrically isomorphic to $\overline{H}$. We start by defining linear functionals on H by

$$\varphi_w(v) = \langle v, w\rangle$$

for any $w \in H$. It is straightforward that this is a linear functional, but we need the following lemma to give the norm.

Lemma 4.4.14
Let H be a Hilbert space and $\varphi_w : H \to \mathbb{C}$ as above. Then $\varphi_w \in H^*$ and $\|\varphi_w\| = \|w\|$.

Proof:
By Cauchy-Schwarz, if $\|v\| \leq 1$, then

$$|\varphi_w(v)| = |\langle v, w\rangle| \leq \|v\|\|w\| \leq \|w\|,$$

so $\|\varphi_w\| \leq \|w\|$, and $\varphi_w \in H^*$.
On the other hand, if we let $v = w/\|w\|$, so $\|v\| = 1$, then

$$|\varphi(w)| = |\langle v, w\rangle| = |\langle w, w\rangle|/\|w\| = \|w\|^2/\|w\| = \|w\|.$$

Hence $\|\varphi_w\| = \|w\|$. ■

Proposition 4.4.15
If H is a Hilbert space and $\varphi \in H^*$ then there is some $w \in H$ such that $\varphi = \varphi_w$.

Proof:
Let $\ker\varphi = \{v \in H : \varphi(v) = 0\}$. It is easy to see that $\ker\varphi$ is a linear subspace of H, and it is closed, since if $v_n \in \ker\varphi$ with $v_n \to v \in H$, then

$$\varphi(v) = \lim_{n\to\infty} \varphi(v_n) = 0$$

so $v \in \ker\varphi$.
Now if $\ker\varphi = H$ we have $\varphi(v) = 0$ for all v and so $\varphi = \varphi_0$.
Otherwise let $V = (\ker\varphi)^\perp$. Let $w \in V$ such that $\|w\| = 1$. There must be such a w since if $v \in V$ with $\varphi(v) \neq 0$ and so let $w = v/\|v\|$ which lies in V since V is a closed linear subspace of H. We note that $V = \{\lambda w : \lambda \in \mathbb{C}\}$, since if $v \in V$ then $\varphi(v) = \lambda/\|\varphi\|$ for some $\lambda \in \mathbb{C}$, and that $\varphi(v - \lambda w) = 0$, so $v - \lambda w \in \ker\varphi \cap V = \{0\}$.
Now given any $v \in H$, let $u = \varphi(v)w - \varphi(w)v$. It is immediate that

$$\varphi(u) = \varphi(v)\varphi(w) - \varphi(w)\varphi(v)$$

and so $u \in \ker\varphi$. Therefore

$$0 = \langle u, w\rangle = \varphi(v)\|w\|^2 - \varphi(w)\langle v, w\rangle = \varphi(v) - \langle v, \overline{\varphi(w)}w\rangle,$$

or equivalently

$$\varphi(v) = \langle v, \overline{\varphi(w)}w\rangle = \varphi_{\overline{\varphi(w)}w}(v).$$

■

Corollary 4.4.16
The operator $\Phi : w \mapsto \varphi_w$ is an isometric isomorphism from $\overline{H}$ to H^*.

Proof:

We first observe that it is linear:

$$\Phi(w + \lambda \cdot u)(v) = \langle v, w + \overline{\lambda} u\rangle = \langle v, w\rangle + \lambda\langle v, u\rangle = (\Phi(w) + \lambda\Phi(u))(v).$$

and so $\Phi(w + \lambda \cdot u) = \Phi(w) + \lambda\Phi(u)$. That it is an isometry follows form the lemma, that it is onto follows from the proposition. Hence the map is an isometric isomorphism. ■

Proposition 4.4.17 (Polarization Identity)
For any u and v in a Hilbert space H, we have

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

In other words, the inner product can be recovered completely from the norm.

4.5 The Hahn-Banach Theorem

We now turn to consider the dual space of a Banach space as an abstract space. Although we have, until this point, considered the dual as a well-populated vector space, there is no a priori reason why there need by any bounded linear functionals other than the zero functional.

More generally, we would like to be able to extend a linear functional defined on a closed subspace of a Banach space to the whole of the Banach space. This sort of "extension" or "lifting" problem is common in categorical approaches to analysis, and can be seen again and again in different settings.

The approach to this problem uses the standard strategy of proving things first in the case of $\mathbb{R}$ Banach spaces, and then using that result to extend to general $\mathbb{C}$ Banach spaces.

Definition 4.5.1
*A **Minkowski functional** on a real vector space V is a function $\rho : V \to \mathbb{R}$ such that $\rho(u + v) \leq \rho(u) + \rho(v)$ and $\rho(\lambda v) = \lambda\rho(v)$ for all u, $v \in V$ and all $\lambda \geq 0$.*

Every seminorm is a Minkowski functional, but the converse is not true, since Minkowski functionals can take negative values. Minkowski functionals allow us to state the Hahn-Banach theorem in its most general form. In practice when applying the Hahn-Banach theorem, the Minkowski functional used will be a multiple of a norm or an appropriate seminorm.

Theorem 4.5.1 (Hahn-Banach Theorem)
Let V be a real vector space, ρ a Minkowski functional on V, and W a subspace of V. If $\varphi : W \to \mathbb{R}$ is a linear functional on W such that $\varphi(w) \leq \rho(w)$ for all $w \in W$, then there is a linear functional $\psi : V \to \mathbb{R}$ such that $\psi|_W = \varphi$ and $\psi(v) \leq \rho(v)$ for all $v \in V$.

In fact, typically there will be many such ψ which extend φ. The proof of Hahn-Banach relies on an application of Zorn's Lemma in a typical sort of way: we will look at all functionals which extend φ and then showw the existence of a maximal one.

Proof:

We will call a linear functional $\psi : W' \to \mathbb{R}$ a linear extension of φ if W' is a subspace of V containing W, $\psi|_W = \varphi$ and $\psi(w) \leq \rho(w)$ for all $w \in W'$. Let $\mathcal{F}$ be the set of all such linear extensions, and we will define a partial order on the set by saying that given $\psi_1 : W_1 \to \mathbb{R}$ and $\psi_2 : W_2 \to \mathbb{R}$ that $\psi_1 \leq \psi_2$ if $W_1 \subseteq W_2$ and $\psi_2|_{W_1} = \psi_1$ (in other words the domain of the first is contained in the domain of the second, and the two functionals agree on the common domain).

Now consider a totally ordered subset $\mathcal{O} = \{\psi_\alpha : \alpha \in I\}$ of $\mathcal{F}$. Let

$$W' = \bigcup_{\alpha \in I} W_\alpha$$

and note that W' is a subspace of V. We then define a new linear functional $\psi : W' \to \mathbb{R}$ by letting $\psi(w) = \psi_\alpha(w)$ for $w \in W_\alpha$. This is a well-defined function, since if w is an element of both W_α and W_β then since $\mathcal{O}$ is totally ordered we must have $\psi_\alpha \leq \psi_\beta$ or vice-versa. Assuming $\psi_\alpha \leq \psi_\beta$, then $w \in W_\alpha$ and so $\psi_\beta(w) = \psi_\alpha(w)$.

Furthermore ψ is a linear extension of φ. It is linear, since given u, $w \in W'$ there is some W_α which contains both and so given any $\lambda \in \mathbb{R}$,

$$\psi(u + \lambda w) = \psi_\alpha(u + \lambda w) = \psi_\alpha(u) + \lambda\psi_\alpha(w) = \psi(u) + \lambda\psi(w).$$

That $\psi|_W = \varphi$ is trivial. Finally if $w \in W'$, then $w \in W_\alpha$ for some α and so $\psi(w) = \psi_\alpha(w) \leq \rho(w)$. In other words $\psi \in \mathcal{F}$.

Finally, ψ is an upper bound for the totally ordered set $\mathcal{O}$ by the way it was defined.

So the conditions of Zorn's Lemma are satisfied, and therefore there is some maximal linear extension $\psi : W' \to \mathbb{R}$. If we can show that the domain of this linear extension is all of V, then we are done.

Assume otherwise, so that W' is a proper subspace of V. So there is some vector $u \in V \setminus W'$. Consider the subspace $W'' = W' + \mathbb{R}u$. We will obtain a contradiction by finding a linear extension on this subspace which extending ψ.

Note that for w_1, $w_2 \in W'$ we have that

$$\psi(w_1) + \psi(w_2) = \psi(w_1 + w_2) \leq \rho(w_1 + w_2) \leq \rho(w_1 - u) + \rho(u + w_2),$$

or equivalently

$$\psi(w_1) - \rho(w_1 - u) \leq \rho(u + w_2) - \psi(w_2).$$

Taking suprema of the left hand side and infema of the right, we obtain

$$\sup\{\psi(w) - \rho(w - u) : w \in W'\} \leq \inf\{\rho(u + w) - \psi(w) : w \in W'\}$$

and so there must be some number α such that

$$\sup\{\psi(w) - \rho(w-u) : w \in W'\} \leq \alpha \leq \inf\{\rho(u+w) - \psi(w) : w \in W'\}.$$

In other words we have that

$$\psi(w) + \alpha \leq \rho(w+u)$$

and

$$\psi(w) - \alpha \leq \rho(w-u)$$

for all $w \in W'$.

Now define $\psi' : W'' \to \mathbb{R}$ by $\psi'(w + \lambda u) = \psi(w) + \lambda\alpha$. It is trivial that ψ' is a linear and that $\psi \leq \psi'$. We only need show that $\psi'(w+\lambda u) \leq \rho(w + \lambda u)$ for all $w \in W'$ and $\lambda \in \mathbb{R}$.

Now if $\lambda = 0$ this is true since $\psi'(w + 0u) = \psi(w) \leq \rho(w)$. If $\lambda > 0$ then

$$\begin{aligned}\psi(w+\lambda u) &= \lambda\left(\psi\left(\frac{1}{\lambda}w\right) + \alpha\right) \\ &\leq \lambda\rho\left(\frac{1}{\lambda}w + u\right) \\ &= \rho(w+\lambda u),\end{aligned}$$

and if $\lambda < 0$ then

$$\begin{aligned}\psi(w+\lambda u) &= |\lambda|\left(\psi\left(\frac{1}{|\lambda|}w\right) - \alpha\right) \\ &\leq |\lambda|\rho\left(\frac{1}{|\lambda|}w - u\right) \\ &= \rho(w+\lambda u).\end{aligned}$$

So we have a contradiction of the maximality of ψ, and so our assumption that W' is a proper subspace of V is incorrect.

Therefore any maximal linear extension of φ satisfies the theorem. ■

Note that if ρ is a seminorm then the condition $\varphi(v) \leq \rho(v)$ is equivalent to $|\varphi(v)| \leq \rho(v)$, since

$$-\varphi(v) = \varphi(-v) \leq \rho(-v) = \rho(v).$$

This gives the following statement for seminorms:

Corollary 4.5.2 (Complex Hahn-Banach Theorem)
Let V be a real vector space, ρ a seminorm on V, and W a subspace of V. If $\varphi : W \to \mathbb{R}$ is a linear functional on W such that $|\varphi(w)| \leq \rho(w)$ for all $w \in W$, then there is a linear functional $\psi : V \to \mathbb{R}$ such that $\psi|_W = \varphi$ and $|\psi(v)| \leq \rho(v)$ for all $v \in V$.

To extend this result to complex vector spaces, we need to note that every vector space over the complex numbers is automatically a vector space over the real numbers. Furthermore, there is a direct correspondence between complex linear functionals over the complex vector space, and real linear functionals over the same vector space regarded as a real vector space.

Given V a complex vector space and $\varphi : V \to \mathbb{C}$ a complex linear functional, the functional $\psi(v) = \operatorname{Re}\varphi(v)$ is a real linear functional since if $\lambda in \mathbb{R}$ and v, $w \in V$ then

$$\psi(v + \lambda w) = \operatorname{Re}(\varphi(v) + \lambda\varphi(w)) = \operatorname{Re}\varphi(v) + \lambda \operatorname{Re}\varphi(w).$$

Furthermore, we can recover φ from ψ by noting that

$$\psi(v) - i\psi(iv) = \operatorname{Re}\varphi(v) - i \operatorname{Re} i\varphi(v) = \operatorname{Re}\varphi(v) + i \operatorname{Im}\varphi(v) = \varphi(v).$$

Indeed, given any real linear functional $\psi : V \to \mathbb{R}$, we can define a new functional $\varphi : V \to \mathbb{C}$ by

$$\varphi(v) = \psi(v) - i\psi(iv).$$

This new functional is linear, since if v, $w \in V$ and $\lambda = x + iy \in \mathbb{C}$ then

$$\varphi(v+\lambda w) = \psi(v+xw+yiw)-i\psi(iv+xiw-yw) = \psi(v)+x\psi(w)+y\psi(iw)-i\psi(iv)-xi\psi(iw)+yi\psi(w) = \varphi(v)+x\varphi(w)+iy$$

In other words, there is a bijection between $\mathcal{L}(V, \mathbb{C})$ and $\mathcal{L}(V, \mathbb{R})$.

Corollary 4.5.3 (Complex Hahn-Banach Theorem)
Let V be a complex vector space, ρ a seminorm on V, and W a subspace of V. If $\varphi : W \to \mathbb{C}$ is a linear functional on W such that $|\varphi(w)| \leq \rho(w)$ for all $w \in W$, then there is a linear functional $\psi : V \to \mathbb{C}$ such that $\psi|_W = \varphi$ and $|\psi(v)| \leq \rho(v)$ for all $v \in V$.

Proof:
Let $\sigma = \operatorname{Re}\varphi$ as in the previous discussion. Then we have that

$$|\sigma(w)| \leq |\varphi(w)| \leq \rho(w)$$

for all $w \in W$ and so by the real Hahn-Banach theorem there is a real linear functional $\tau : V \to \mathbb{R}$ which extends σ, so in particular $|\tau(v)| \leq \rho(v)$ for all $v \in V$.

Let $\psi(v) = \tau(v) - i\tau(iv)$ as in the previous discussion. It is an immediate consequence of the fact that $\tau|_W = \sigma$ that $\psi|_W = \varphi$. We also have that if $\lambda = \overline{\operatorname{sign}\psi(v)}$ then

$$|\psi(v)| = \alpha\psi(v) = \psi(\alpha v) = \tau(\alpha v) - i\tau(i\alpha v) = \tau(\alpha v)$$

since the left-hand side is a real number. But then

$$|\psi(v)| = \tau(\alpha v) \leq \rho(\alpha v) = |\alpha|\rho(v) = \rho(v).$$

So ψ is a complex linear extension of φ. ∎

The Hahn-Banach theorem has a number of fundamental consequences.

Theorem 4.5.4
Let V be a (complex) normed vector space.

1. *If $v \neq 0$ then there is a $\varphi \in V^*$ such that $\|\varphi\| = 1$ and $\varphi(x) = \|x\|$.*
2. *V^* separates points of V.*
3. *If $W \subset V$ is a proper closed subspace of V and $v \in V \setminus W$ then there exists $\varphi \in V^*$ such that $\varphi(w) = 0$ for all $w \in W$ but $\varphi(v) \neq 0$; in fact φ can be found so that $\|\varphi\| = 1$ and*
$$\varphi(v) = \inf\{\|v - w\| : w \in W\}.$$
4. *For any $v \in V$ there is a linear functional $\hat{v} : V^* \to \mathbb{C}$ given by $\hat{v}(\varphi) = \varphi(v)$. Then the function $v \mapsto \hat{v}$ is a linear isometry from V into V^{**}.*

Proof:
(i) Consider the linear functional $\psi : \mathbb{C}v \to \mathbb{C}$ defined by $\psi(\lambda v) = \lambda\|v\|$. Since $|\psi(\lambda v)| = |\lambda|\|v\| = \|\lambda v\|$, the Hahn-Banach theorem tells us there is an extension φ to V with $|\varphi(v)| \leq \|v\|$ for all $v \in V$, so that $\|\varphi\| \leq 1$. But since $|\varphi(v)| = |\psi(v)| = \|v\|$ we have that $\|\varphi\| \geq 1$, from which the result follows.

(ii) If $v \neq w$ then by (i) there is a linear functional φ such that $\varphi(v-w) \neq 0$. But that means that $\varphi(v) - \varphi(w) \neq 0$ and hence $\varphi(v) \neq \varphi(w)$. So V^* separates points of V.

(iii) In a similar manner to (i), consider the linear functional ψ on $W + \mathbb{C}v$ given by $\psi(w + \lambda v) = \lambda\delta$ where $\delta = \inf\{\|v - w\| : w \in W\}$. Then we have that
$$|\psi(w + \lambda v)| = |\lambda|\delta \leq |\lambda| \left\| v - \frac{-1}{\lambda} w \right\| = \|w + \lambda v\|,$$
and so the Hahn-Banach theorem applies, given an extension φ which has the required properties.

(iv) It is simple to verify that $\hat{v}$ is a linear functional and that $\widehat{v + \lambda w} = \hat{v} + \lambda\hat{w}$, so that the function is linear. Now we have that
$$|\hat{v}(\varphi)| = |\varphi(v)| \leq \|\varphi\|\|v\|$$
so that $\|\hat{v}\| \leq \|v\|$. On the other hand, from (i) we know that there is a linear functional φ such that $\|\varphi\| = 1$ and $\varphi(v) = \|v\|$, so that $|\hat{v}(\varphi)| = |\varphi(v)| = \|v\|$, so we can conclude that $\|\hat{v}\| = \|v\|$. So the function is an isometry. ■

One consequence of the last part is that, since dual spaces are always complete, the closure of the image of V under $\hat{}$ is a Banach space which contains an isometric image of V as a dense subspace. We call this Banach space $\overline{\hat{V}} \subseteq V^{**}$ the completion of V. Indeed, if V is already a Banach space, this Banach space is isometrically isomorphic to V.

The dual space of a normed vector space V determines a topology on the vector space V called the weak topology. This topology is the topology on V where a net $(v_\lambda)_{\lambda\in\Lambda}$ converges to v if and only if
$$\lim_{\lambda\in\Lambda} \varphi(v_\lambda) = \varphi(v)$$

for every $\varphi \in V^*$. Another way of thinking of this is that it is the topology of pointwise convergence when we identify V with $\hat{V}$; that is $(v_\lambda)_{\lambda\in\Lambda}$ converges to v if and only if $(\widehat{v_\lambda})_{\lambda\in\Lambda}$ converges to $\hat{v}$ pointwise.

Proposition 4.5.5
If V is a normed vector space, then the weak topology on V is weaker than the norm topology on V. In other words if $(v_\lambda)_{\lambda\in\Lambda}$ converges to v in the norm topology, then it converges in the weak topology.

More generally, for any subset of V^* one can define a weak topolgy by insisting that the net converge for just those linear functionals in the set. Clearly convergence in the overall weak topology implies convergence in these restricted weak topologies.

Since V^* itself has a dual, V^{**} there is a weak topology on V^* coming from this duality. However the weak topology generated by $\hat{V} \subseteq V^{**}$ turns out to be of particular importance and is called the weak-* topology. This is the topology where a net $(\varphi_\lambda)_{\lambda\in\Lambda}$ converges to φ if and only

$$\lim_{\lambda\in\Lambda} \hat{v}(\varphi_\lambda) = \hat{v}(\varphi)$$

for all $v \in V$. But this is the same as saying

$$\lim_{\lambda\in\Lambda} \varphi_\lambda(v) = \varphi(v)$$

for all $v \in V$, so the weak-* topology is simply the topology of pointwise convergence.

Proposition 4.5.6
If V is a normed vector space, then the weak- topology on V^* is weaker than the norm topology on V^*. In other words if $(\varphi_\lambda)_{\lambda\in\Lambda}$ converges to φ in the norm topology, then it converges in the weak-* topology.*

One of the particularly nice properties of the weak-* topology is that the unit ball of the dual is compact in this topology.

Theorem 4.5.7 (Alaoglu's Theorem)
If V is a normed vector space then the closed unit ball $B = \{\varphi \in V^* : \|\varphi\| \leq 1\}$ of V^* is compact in the weak- topology.*

Proof:
Let

$$X = \prod_{v\in V} \{z \in \mathbb{C} : |z| \leq \|v\|\}.$$

This set is a compact set in the product topology by Tychonoff's theorem (Theorem 2.5.9). Recall that elements of the product set are really functions from the index set into each of the sets in the product, we can identify certain elements of V^* as elements of X, namely those linear functionals φ such that $|\varphi(v)| \leq \|v\|$ for all $v \in V$, i.e. the φ with $\|\varphi\| \leq 1$.

In other words $B \subseteq X$, and the topology on X is just the topology of pointwise convergence, so we will have that B is compact in the weak-* topology if and only it is closed when regarded as a subset of X in the product topology.

Let $(\varphi_\lambda)_{\lambda\in\Lambda}$ be a net in B which converges pointwise to some $f \in X$. In other words

$$\lim_{\lambda\in\Lambda} \varphi_\lambda(v) = f(v)$$

for all $v \in V$. But then

$$f(v + \alpha w) = \lim_{\lambda\in\Lambda} \varphi_\lambda(v + \alpha w) = \lim_{\lambda\in\Lambda} \varphi_\lambda(v) + \alpha\varphi_\lambda(w) = f(v) + \alpha f(w).$$

So f is linear, and $\|f\| \leq 1$, so $f \in B$. Hence B is closed in the weak-* topology. ■

It is worthwhile mentioning again here that Tychonoff's theorem is equivalent to the axiom of choice, so this theorem, like the Hahn-Banach theorem, relies on that axiom.

4.6 More on Bounded Operators

In this section we will investigate the structure of bounded operators between Banach spaces in more detail. The key result that will drive the results in this section is a theorem from metric space theory.

Recall that a set E in a topological space X is ***nowhere dense*** if the interior of the closure is empty, i.e. $(\overline{E})^o = \emptyset$, or equivalently it's complement contains an open, dense set. A subset E of X is ***meager*** or ***of the first category*** if it is a countable union of nowhere dense sets. The complement of an meager set is called ***residual***. A set which is not of the first category is called ***of the second category***.

The Baire category theorem says that a complete metric space is always of the second category as a subset of itself. More precisely:

Theorem 4.6.1 (Baire Category Theorem)
Let X be a complete metric space. Then X is not a countable union of nowhere dense sets.

Equivalently, if $(U_n)_{n\in\mathbb{N}}$ is a countable sequence of open, dense subsets of X, then

$$\bigcap_{n=1}^{\infty} U_n$$

is dense in X.

Typical uses of the Baire Category Theorem are in existence proofs: if you can show that the set of elements which fail to have some property is meager, then there must be some elements (in fact lots, in most cases) which have the property.

Since Banach spaces are complete metric spaces, it should not be suprising that the Baire Category Theorem can be used to derive some significant results.

We say that if X and Y are topological spaces, then a function $f : X \to Y$ is ***open*** if $f(U)$ is open in Y for every open set $U \subset X$. In other words, the image of open sets are open. If X and Y are a metric spaces, then a function is open if and only if given any ball $B_{x,r} = \{y \in X : d(x, y) < r\}$ in X centred at x, then $f(B_{x,r})$ contains a ball $B_{f(x),s}$ centred at $f(x)$. If X and Y are normed linear spaces and f is linear, then f is open if and only if $f(B)$ contains a ball in Y centred at the origin, where B is the unit ball of X.

Note: *compare with continuous maps where inverse images of open sets are open*

Theorem 4.6.2 (Open Mapping Theorem)
If V and W are Banach spaces and $T \in B(V, W)$ is surjective, then T is open.

Proof:
Let B_r be the ball of radius r in V centred at 0. We have that

$$V = \bigcup_{n=1}^{\infty} B_n$$

and so since T is surjective,

$$W = T(V) = \bigcup_{n=1}^{\infty} T(B_n).$$

Assume that $T(B_1)$ is nowhere dense. Since the dilation maps $D_n : W \to W$ defined by $D_n w = nw$ are bounded and the inverses $D_n^{-1} = D_{1/n}$ are also bounded, these are homeomorphisms, and so $T(B_n) = T(nB_1) = nT(B_1) = D_n T(B_1)$ must also be nowhere dense. This implies that the complete metric space W is a countable union of nowhere dense sets, which cannot happen by the Baire Category Theorem. So $T(B_1)$ is not nowhere dense.

In other words, we can find some $w_0 \in W$ and $r > 0$ such that $B_{w_0,4r} \subset \overline{T(B_1)}$. Let $w_1 \in T(B_1)$ such that $\|w_0 - w_1\| < 2r$, and since T is surjective there is some $v_1 \in B_1$ such that $w_1 = Tv_1$. By the triangle inequality then we have that $B_{w_1,2r} \subset \overline{T(B_1)}$, and so given any $w \in W$ with $\|w\| < 2r$ we have

$$w = w_1 + (w - w_1) = Tv_1 + (w - w_1) \in \overline{T(v_1 + B_1)} \subseteq \overline{T(B_2)}.$$

Therefore, dividing throughout by 2, we have that if $\|w\| < r$ then $w \in \overline{T(B_1)}$. Indeed, if $\|w\| < r2^{-n}$ then $w \in \overline{T(B_{2^{-n}})}$.

Starting with $w_1 \in W$ such that $\|w_1\| < r/2$ we have that there is a $v_1 \in B_{1/2}$ such that $\|w_1 - Tv_1\| < r/4$, since $w_1 \in \overline{T(B_{1/2})}$ and $T(B_{1/2})$ is dense in that set. Letting $w_2 = w_1 - Tv_1$. We can then proceed inductively to find v_n and w_n such that $v_n \in B_{2^{-n}}$ and $w_n = w_{n-1} - Tv_{n-1}$ with $\|w_n\| < r2^{-n}$.

Then

$$\sum_{n=1}^{\infty} v_n$$

is absolutely convergent, and hence convergent, and

$$Tv = T\left(\sum_{n=1}^{\infty} v_n\right) = \sum_{n=1}^{\infty} T(v_n) = w_1.$$

Since

$$\|v\| \leq \sum_{n=1}^{\infty} \|v_n\| < \sum_{n=1}^{\infty} 2^{-n} = 1$$

we are done. ■

Corollary 4.6.3
If V and W are Banach spaces, and $T \in B(V, W)$ is a bijection, then $T^{-1} \in B(W, V)$, or equivalently, T is an isomorphism.

The graph of a function $f : X \to Y$ is the set

$$G(f) = \{(x, f(x)) \in X \times Y : x \in X\}.$$

If $T : V \to W$ is a linear map between vector spaces, then the graph is a vector subspace of the vector space $V \times W$. If V and W are topological vector spaces and T is continuous, then $G(T)$ is a closed subspace of the topological vector space $V \times W$, since if $(v_\lambda, Tv_\lambda) \to (v, w) \in V \times W$, then $v_\lambda \to v$ in V, and so by continuity

$$w = \lim_{\lambda \in \Lambda} Tv_\lambda = T\left(\lim_{\lambda \in \Lambda} v_\lambda\right) = Tv.$$

So $(v, w) = (v, Tv) \in G(T)$.

If V and W are Banach spaces, then the converse is true. We call a linear map from V to W ***closed*** if $G(T)$ is closed in $V \times W$.

Theorem 4.6.4 (Closed Graph Theorem)
Let V and W be Banach spaces and $T : V \to W$ is a closed linear map then T is bounded.

Proof:
Let $\pi_V : G(T) \to V$ and $\pi_W : G(T) \to W$ be the projections onto V and W respectively, ie. $\pi_V(v, Tv) = v$ and $\pi_W(v, Tv) = Tv$. By the definition of the product topology π_V and π_W are continuous, and they are easily seen to be linear, so $\pi_V \in B(G(T), V)$ and $\pi_W \in B(G(T), W)$.

Now since V and W are Banach spaces, so is $V \times W$, and since $G(T)$ is closed it is also a Banach space. But π_V is a bijection from $G(T)$ to V and so by the corollary of the open mapping theorem, $\pi_V^{-1} \in B(V, G(T))$. Now $\pi_W(\pi_V^{-1}(v)) = \pi_W(v, Tv) = Tv$, so $T = \pi_W \circ \pi_V^{-1}$ and so T is bounded since it is the composition of two bounded maps. ■

The significance of the closed graph theorem is that it can save some effort when proving that a linear map is continuous. Rather than showing that if

$v_n \to v$ then $Tv_n \to Tv$, which requires first confirming that Tv_n convereges to something, and then showing that the something is the right thing, it is sufficient to show that the graph is closed. In other words we only need to check that if $v_n \to v$ and $Tv_n \to w$ then $w = Tv$.

It is worth mentioning that when considering unbounded linear maps defined on a dense subset of a Banach space, knowing that the graph of the linear map is closed often makes the linear map amenable to analysis. With appropriate domains, differentiation and integration operators, although not bounded, are closed.

The final result in the trio of theorems that are consequences of the Baire category theorem is a result which allows you to infer facts about the norms of families of bounded linear maps from knowledge of what the maps do to vectors.

Theorem 4.6.5 (Uniform Boundedness Principle)
Let V and W be normed vector spaces and $\mathcal{F}$ a family of bounded linear maps from V to W, ie. $\mathcal{F} \subseteq B(V, W)$.

1. *If* $\sup\{\|Tv\| : T \in \mathcal{F}\} < \infty$ *for every v in a non-meager subset F of V, then* $\sup\{\|T\| : T \in \mathcal{F}\} < \infty$.
2. *if V is a Banach space and* $\sup\{\|Tv\| : T \in \mathcal{F}\} < \infty$ *for every $v \in V$, then* $\sup\{\|T\| : T \in \mathcal{F}\} < \infty$.

Proof:
(i) Let

$$F_n = \{v \in V : \sup\{\|Tv\| : T \in \mathcal{F}\} \leq n\} = \bigcap_{T \in \mathcal{F}} \{v \in V : \|Tv\| \leq n\}.$$

Since the T is continuous, F_n is an intersection of closed sets, and so is itself closed. But F is the union of the sets F_n and since F is non-meager, at least one of the sets F_n is not nowhere dense. But since this F_n is closed, failure to be nowhere dense means that the interior fo F_n is not empty, and so there is some $v_0 \in F_n^o$ and $r > 0$ such that $B_{v_0,r} \subseteq F_n^o$. But then taking closures we have that $\overline{B_{v_0,r}} \subseteq F_n$.

But this means that $\overline{B_{0,r}} \subseteq F_{2n}$, since if $\|v\| \leq r$ then $v - v_0 \in F_n$ and so for any $T \in \mathcal{F}$,

$$\|Tv\| \leq \|T(v - v_0)\| + \|Tv_0\| \leq n + n = 2n.$$

Thefore if $\|v\| \leq 1$ we have that $\|rv\| \leq r$ and so for any $T \in \mathcal{F}$

$$\|Tv\| = \left\| \frac{1}{r} T(rv) \right\| = \frac{1}{r} \|T(rv)\| \leq \frac{1}{r} 2n.$$

So $\|T\| \leq 2n/r$, and so

$$\sup\{\|T\| : T \in \mathcal{F}\} \leq \frac{2n}{r} < \infty.$$

(ii) Follows immediately from (i) and the Baire category theorem, since V is a Banach space and therefore not meager. ■

4.7 Banach Algebras

We observe that if $T : V \to W$ is bounded, and $S : W \to X$ is bounded, then

$$\|(S \circ T)v\| = \|S(Tv)\| \leq \|S\|\|Tv\| \leq \|S\|\|T\|\|v\|.$$

Hence $\|S \circ T\| \leq \|S\|\|T\|$. We normally simply write ST for $S \circ T$. In particular, this means that composition of bounded operators in $B(V)$ gives another bounded operator on $B(V)$, so composition gives a product,

$$\circ : B(V) \times B(V) \to B(V),$$

and if $T_n \to T$ and $S_n \to S$ in $B(V)$, then

$$\|S_nT_n - ST\| = \|S_nT_n - S_nT + S_nT - ST\| \tag{4.2}$$
$$\leq \|S_n(T_n - T)\| + \|(S_n - S)T\| \tag{4.3}$$
$$\leq \|S_n\|\|T_n - T\| + \|S_n - S\|\|T\| \to 0 \tag{4.4}$$

as $n \to \infty$. So this product is continuous in the topology on $B(V)$. This is the prototypical example of a ***Banach algebra***.

Recall that an algebra $(A, +, \cdot, \circ, \mathbb{F})$ over a field $\mathbb{F}$ is a vector space $(A, +, \cdot, \mathbb{F})$ over $\mathbb{F}$, along with a product $\circ : V \to V$ on the vector space which is associative:

$$a \circ (b \circ c) = (a \circ b) \circ c,$$

left distributive:

$$a \circ (b + c) = a \circ b + a \circ c,$$

right distributive:

$$(a + b) \circ c = a \circ c + b \circ c,$$

and commutes with scalar multiplication:

$$\lambda \cdot (a \circ b) = (\lambda \cdot a) \circ b = a \circ (\lambda \cdot b).$$

If there is a multiplictive identity e, we say that the algebra is unital, and if the multiplication is commutative, we say that the algebra is commutative. We will typically omit the multiplication symbols $\cdot$ and $\circ$.

An involution $* : A \to A$ on an algebra A over $\mathbb{C}$ is a conjugate linear, antimultiplicative map $a \mapsto a^*$, ie.

$$(\lambda a + b)^* = \overline{\lambda} a^* + b^*$$

and

$$(ab)^* = b^* a^*.$$

If A is unital, we insist that $e^* = e$.

Definition 4.7.1
A topological algebra is an algebra which is a topological vector space, and for which the multiplication is continuous.

A Banach algebra is an algebra together with a norm which makes it a Banach space, and for which

$$\|a \circ b\| \leq \|a\|\|b\|.$$

A C-algebra is a Banach algebra together with an involution* $*$*, which is isometric, so*

$$\|a^*\| = \|a\|,$$

and satisfies the C^* *identity*

$$\|a^* a\| = \|a\|^2.$$

So every C*-algebra is a Banach algebra, and every Banach algebra is a topological algebra. As we have seen, $B(V)$ is a Banach algebra, but there are other interesting examples.

Example 4.7.1
The Banach space $C_b(X)$ of bounded, continuous functions with the uniform norm is a unital algebra with the usual multiplication of functions. Furthermore,

$$\|fg\|_u \leq \|f\|_u \|g\|_u,$$

so this is in fact a Banach algebra. Moreover if we define $f^*(x) = \overline{f(x)}$, then

$$\|f^*\|_u = \|f\|_u,$$

and since $f^*(x)f(x) = |f(x)|^2$,

$$\|f^* f\|_u = \sup_{x \in X} \|f(x)\|^2 = \|f\|^2.$$

So $C_b(X)$ is a C*-algebra. ◇

Example 4.7.2
Let G be a finite group, and let $\mathbb{C}(G)$ be the finite dimensional vector space of all formal linear combinations of elements of G, ie. elements of the form

$$v = \sum_{g \in G} v_g g,$$

for $v_g \in \mathbb{C}$. We give $\mathbb{C}(G)$ the norm

$$\|v\| = \sum_{g \in G} |v_g|,$$

which makes it a vector space. We define a product, $*$, on this vector space by extending the group operation to $\mathbb{C}(G)$ by letting

$$v * w = \sum_{g \in G} \sum_{h \in G} v_g w_h gh.$$

By replacing h by $g^{-1}h'$, we see that

$$v * w = \sum_{h' \in G} \left(\sum_{g \in G} v_g w_{g^{-1}h'} \right) h',$$

and so the coefficients are given by the convolution of v and w, and using this fact one can see that this is a Banach algebra.

We can define an isometric involution $*$ by

$$v^* = \sum_{g \in G} \overline{\lambda_g} g^{-1},$$

but this does not make $\mathbb{C}(G)$ a C*-algebra in general. ◇

Sample Exam Questions

These sample questions are designed to give you an idea of what questions may be asked on the midterm.

1. State the Monotone Convergence Theorem.

2. State the definition of a Lebesgue measurable function.

3. State the definition of a σ-algebra.

4. If $\mathcal{A}$ is a σ-algebra, show that given any countable collection of sets $A_n \in \mathcal{A}$, $n = 1, 2, \ldots$, then
$$\bigcap_{n=1}^{\infty} A_n \in \mathcal{A}.$$

5. Show that if f is a measurable function, then cf is a measurable function.

6. Give an example of a Lebsgue measurable set X which contains no intervals, but for which $m(X) = 1/3$.

7. Find
$$\int_{[-1,2]} x \; dm(x),$$
carefully stating the results you use.

8. Let f be a Riemann integrable function on $[a, b]$. Prove that
$$\int_a^b f(x) \; dx \leq \int_{[a,b]} f \; dm.$$

9. Give an example of a sequence of functions f_n in $\mathcal{L}^+([0, 1])$ which converge pointwise to 0, does not converge in $L^1([0, 1], m)$. Verify that your example is valid.

10. Give an example of a measure space (X, μ) a sequence of measurable functions on that space such that $f_n \to f$ in L^1, but f_n does not converge to f pointwise almost everywhere.

11. Let (X, μ) be a finite measure space. Show that if f_n is a sequence of measurable functions, and f_n converges to some $f \in L^1(X, \mu)$ uniformly, then
$$\int f \; d\mu = \lim_{n\to\infty} \int f_n \; d\mu.$$

12. If $(X, \mathcal{M})$ is a measurable space, and μ and ν are two measures on that space, show that $\mu + \nu$ is also a measure on that space.

13. Let $\alpha : \mathbb{N} \to [0, \infty)$ be a bounded function. Show that

$$\mu_\alpha(A) = \sum_{k \in A} \alpha(k)$$

is a measure on $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$. Let $f : \mathbb{N} \to \mathbb{C}$ be any function. Express

$$\int f \; d\mu_\alpha$$

as a sum.

14. Let c be the counting measure on the natural numbers, $\mathbb{N}$. Show that $f \in L^1(\mathbb{N}, c)$ if and only if

$$\sum_{k=1}^{\infty} |f(k)| < \infty.$$

ie. if and only if $f(k)$ is absolutely convergent as a series.

15. State the Fubini-Tonelli theorem. Let c be the counting measure on $\mathbb{N}$, and $a_{k,l}$ a double sequence for which

$$\sum_{k=1}^{\infty} \sum_{l=1}^{\infty} |a_{k,l}| < \infty.$$

Use the Tonelli and Fubini theorems on the function $f : \mathbb{N}^2 \to \mathbb{R}$, where $f(k, l) = a_{k,l}$ to show that one can swap the order of summation, ie.

$$\sum_{k=1}^{\infty} \sum_{l=1}^{\infty} a_{k,l} = \sum_{l=1}^{\infty} \sum_{k=1}^{\infty} a_{k,l}.$$

(Note: this is using a sledgehammer to crack a walnut, but I'm really interested in whether you understand how to use Tonelli-Fubini.)

16. Show that if (X, τ_X) and (Y, τ_Y) are both topological spaces, that every continuous function $f : X \to Y$ is Borel measurable.

Appendix A

Notation and Basic Definitions

This section can be thought of as a "cheat-sheet" providing definitions, notation and fundamental results that may come up in the course. By its nature, this must be terse, so you should not be too worried if this seems inordinately complicated.

In particular, you should be comfortable with the subsections on sets, functions, suprema and infima, sequences and series.

This is a rough draft, which I will likely expand upon if need to standardize more concepts and notation.

The exercises are designed for you to confirm your understanding of the subject material.

A.1 Sets

Because we will be discussing "sets of sets" at many points, we will use the terms ***family*** and ***collection*** as synonyms for "set" to clarify exposition.

The empty set is denoted $\emptyset$. We use the symbols $\cap$, $\cup$, $\subset$ and $\supset$ for intersection, union, (proper) subset and (proper) superset respectively. The subset and superset symbols are used strictly: $A \subset B$ does not allow the possibility that $A = B$. We use $\subseteq$ and $\supseteq$, and write "subset" and "superset", when we want to allow the possibility of equality. If A is a subset of some set, we denote the complement of A in that set by A^c. We also define the difference and symmetric difference of two sets by:

$$A \setminus B = \{a \in A : a \notin B\} = A \cap B^c$$

$$A \triangle B = \{x \in A \cup B : a \notin A \cap B\} = (A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A)$$

If $A \cap B = \emptyset$, we say A and B are disjoint. More generally, we say a family of sets is disjoint if any pair of sets from the family are disjoint.

If X is any set, we denote the power set of X by $\mathcal{P}(X)$. The Cartesian product of any two sets is the set of pairs:

$$A \times B = \{(a, b) : a \in A, b \in B\}.$$

Exercises

1. If $A = \{1, 2, 3, 7\}$ and $B = \{3, 4, 6, 7, 8\}$ find $A \setminus B$, $B \setminus A$, $A \triangle B$, and $A \times B$.

2. If A and B are disjoint sets, show $A \setminus B = A$ and $A \triangle B = A \cup B$.

A.1.1 Relations and Functions

A ***relation*** R from A to B is a subset of $A \times B$, and we write aRb if $(a, b) \in R$. A relation for A to A is often said to be a relation on A. A relation on A is:

reflexive if aRa for all $a \in A$;

symmetric if aRb implies bRa for all a, $b \in A$;

antisymmetric if aRb and bRa implies $a = b$ for all a, and $b \in A$;

transitive if aRb and bRc implies aRc for all a, b, $c \in A$.

An ***equivalence relation*** is a relation which is reflexive, symmetric and transitive. A ***partial order*** is a relation which is reflexive, antisymmetric and transitive. A ***preorder*** is a relation which is only reflexive and transitive, so every partial order is a preorder. A ***linear*** or ***total order*** is a partial order for which either aRb or bRa for all a, $b \in A$.

If $\sim$ is an equivalence relation on A, we say that the ***equivalence class*** of $a \in A$ is the set

$$[a] = \{b \in A : a \sim b\}.$$

Clearly $a \sim b$ if and only if $[a] = [b]$. The ***quotient*** of A by $\sim$, $A/\sim$ is the set of all equivalence classes. A set of elements of A is a set of ***equivalence class representatives*** if every equivalence class as precisely one element in the set.

If $\lhd$ is a preorder on A, then so is $\rhd$, where $a \rhd b$ iff $b \lhd a$. If $X \subseteq A$, and $\lhd$ is a preorder on A, an element $a \in A$ is an ***upper bound*** for X if $x \lhd a$ for all $x \in X$; a is a ***lower bound*** if $a \lhd x$ for all $x \in X$. If $\lhd$ is a partial order on A, and $X \subseteq A$, an element $x \in X$ is a ***maximal element*** of X if $y \in X$ with $x \lhd y$ implies $y = x$ and it is a ***minimal element*** if $y \in X$ with $y \lhd x$ implies $x = y$.

A preorder $\lhd$ on A is ***(upwardly) directed*** if every pair of elements $\{a, b\}$ has an upper bound c so that $a \lhd c$ and $b \lhd c$. A set together with a directed preorder is called a ***directed set***. Every total order is automatically directed. A total order on a set A is a ***well ordering*** if every non-empty subset has a minimal element, and we say that A is ***well ordered*** by $\lhd$.

A ***function*** f from A to B is a relation such that for every $a \in A$ there is a unique $b \in B$ such that afb. We write $f : A \to B$ and $f(a) = b$ or $f : a \mapsto b$.

we tend to use "function" alking about functions codomain is $\mathbb{R}$ or $\mathbb{C}$, and when considering ons between more general

We will use the words ***map*** and ***mapping*** interchangeably with function. We call A the ***domain*** of f and B the ***codomain*** of f. If $X \subseteq A$, we say that the ***image*** of X under f is the set

$$f(X) = \{f(x) : x \in X\} \subseteq B.$$

The ***range*** of f is the set $f(A)$. The ***inverse image*** of $X \subset B$ is the set

$$f^{-1}(X) = \{a \in A : f(a) \in X\}.$$

Images and inverse images satisfy the following:

$$f\left(\bigcup_{\alpha\in I} X_\alpha\right) = \bigcup_{\alpha\in I} f(X_\alpha) \tag{A.1}$$

$$f^{-1}\left(\bigcup_{\alpha\in I} X_\alpha\right) = \bigcup_{\alpha\in I} f^{-1}(X_\alpha) \tag{A.2}$$

$$f^{-1}\left(\bigcap_{\alpha\in I} X_\alpha\right) = \bigcap_{\alpha\in I} f^{-1}(X_\alpha) \tag{A.3}$$

$$f^{-1}(X^c) = f^{-1}(X)^c \tag{A.4}$$

A function $f : A \to B$ is ***injective*** or ***one-to-one*** if $f(a) = f(b)$ implies $a = b$ for all $a, b \in A$; it is ***surjective*** or ***onto*** if $f(A) = B$; and it is ***bijective*** if it is both surjective and injective. If f is bijective, then the ***inverse function*** of f is the function $f^{-1} : B \to A$ defined by $f^{-1}(b) = a$ if and only if $f(a) = b$. We then have that $(f^{-1})^{-1} = f$ and $f^{-1}(f(a)) = a$. Additionally $f^{-1}(X)$ is the same set whether you regard it as the inverse image under f or the image under f^{-1}, justifying the notation.

If $\sim$ is an equivalence relation on A, we define the ***quotient map*** $q : A \to A/\sim$ by $q(a) = [a]$.

If X is any set and $Y \subset X$, the ***characteristic function*** of Y, is the function $\chi_Y : X \to \{0, 1\}$ defined by

$$\chi_Y(x) = \begin{cases} 1, & x \in Y \\ 0, & x \notin Y \end{cases}.$$

Exercises

1. If X is any set, show that $A \subseteq B$ is a partial order on $\mathcal{P}(X)$. Show that it is directed. If $\mathcal{F}$ is any collection of subsets of X, show that

   $$\bigcup_{F\in\mathcal{F}} F$$

   is an upper bound for $\mathcal{F}$, and

   $$\bigcap_{F\in\mathcal{F}} F$$

is a lower bound for $\mathcal{F}$. Show that this is not a total ordering except for the trivial cases $|X| = 0$ and $|X| = 1$.

2. Consider the family of intervals $\mathcal{F} = \{[a, \infty) : a \in \mathbb{R}\}$. Show that $A \subseteq B$ is a partial order on $\mathcal{F}$.

3. Let $X \subseteq Y$, and $\triangleleft$ a partial order on Y. Show that the restriction of $\triangleleft$ to X is a partial order on X.

4. Show that $\mathbb{N}$ with the usual order is well ordered. Show that $\mathbb{R}$ with the usual order is totally ordered, but not well ordered.

5. Consider the set S of all convergent sequences of real numbers. Show that the relation $(x_n)_{n=1}^{\infty} \sim (y_n)_{n=1}^{\infty}$ iff
$$\lim_{n\to\infty} x_n = \lim_{n\to\infty} y_n$$
is an equivalence relation.

6. Let S be the set of convergent sequences from the previous Exercise. Show that the function $f : S \to \mathbb{R}$ defined by
$$f((x_n)_{n=1}^{\infty}) = \lim_{n\to\infty} x_n$$
is a surjection.

7. Let S and $\sim$ be as above. Show that the function $g : \mathbb{R} \to S/\sim$ defined by
$$g(x) = [(x_n)_{n=1}^{\infty}],$$
where $x_n = x$ for all $n \in \mathbb{N}$.

8. Verify statements A.1–A.4.

9. If X is any set, and A and $B \subseteq X$, show that
$$\chi_{A\cap B}(x) = \chi_A(x)\chi_B(x).$$

10. If $\triangleleft$ is a preorder on a set X, show that the relation $x \sim y$ when $x \triangleleft y$ and $y \triangleleft x$ is an equivalence relation on X.

    Show that $[x] \triangleleft [y]$ if $x \triangleleft y$ is a well-defined relation on $X/\sim$ and, moreover, it is a partial order.

A.1.2 Cardinality

We denote the cardinality of a set X by $|X|$. Two sets X and Y have the same cardinality if there is a bijection from X to Y. If there is a surjection from X to Y, then $|X| \geq |Y|$; and if there is an injection from X to Y then $|X| \leq |Y|$.

The set $\{1, 2, 3, ..., n\}$ is defined to have cardinality n, and any set with cardinality n is a finite set. All other sets are infinite sets. The set of all natural numbers $\mathbb{N}$, has cardinality $\aleph_0$. Any set with cardinality less than or equal to $\aleph_0$ is called a countable set; and other set is called uncountable. The next biggest cardinality is $\aleph_1$, and so forth. The set of all real numbers, $\mathbb{R}$ has cardinality $\mathbf{c}$, which may or may not be $\aleph_1$, depending on your model of set theory (I don't think we need it in this course, but if we do, we can safely assume that the cardinality of the continuum is $\aleph_1$).

If X is an infinite set, then $|X^n| = |X|$. In particular, this means that $|\mathbb{N}^n| = |\mathbb{Z}^m| = |\mathbb{Q}^k| = \aleph_0$, and $|\mathbb{R}^n| = |\mathbb{C}^m| = \mathbf{c}$, for any natural numbers n, m and k. Explicit bijections can be set up quite easily.

To show that $\aleph_0 \neq \mathbf{c}$ you use Cantor's diagonal argument, and this argument can also be used to show that the set of sequences in a countable set is uncountable.

If X is any set, $|X| < |\mathcal{P}(X)|$. Usually $|\mathcal{P}(X)|$ is denoted $2^{|X|}$.

Exercises

1. If A and B are finite sets, show that $|A \times B| = |A||B|$.

2. If A and B are finite sets, show that $|A \cup B| = |A| + |B| - |A \cap B|$.

3. Show that $|\mathbb{Q}| = \aleph_0$.

4. Let $S_{\mathbb{Q}}$ be the set of all rational sequences. Use Cantor's diagonal argument to show that $|S_{\mathbb{Q}}| > \aleph_0$.

A.1.3 Products of Sets and the Axiom of Choice

Earlier we defined $A \times B = \{(a, b) : a \in A, b \in B\}$. We can define a finite product of sets $A_1, A_2, \ldots, A_n$ in the same way by

$$\prod_{k=1}^{n} A_k = A_1 \times A_2 \times \cdots \times A_n = \{(a_1, a_2, \ldots, a_n) | a_k \in A_k\}.$$

However, for full generality, it is better to consider these n-tuples instead as functions. The n-tuple $\mathbf{a} = (a_1, a_2, \ldots, a_n) \in A_1 \times A_2 \times \cdots \times A_n$ is equivalent to the function $a : \{1, 2, \ldots, n\} \to A_1 \cup A_2 \cup \ldots A_n$ where

$$a(k) = a_k.$$

On the other hand, any function $a : \{1, 2, \ldots, n\} \to A_1 \cup A_2 \cup \ldots A_n$ which satisfies $a(k) \in A_k$ will also give you an n-tuple $\mathbf{a} = (a(1), a(2), \ldots a(n)) \in A_1 \times A_2 \times \cdots \times A_n$. So we can think of

$$\prod_{k=1}^{n} A_k = \left\{ a : \{1, 2, \ldots, n\} \to \bigcup_{k=1}^{n} A_k | a(k) \in A_k \right\}.$$

Notice that the set $\{1, 2, \ldots, n\}$ is indexing the collection of sets. This allows us to generalize to any collection of sets. Let $\{A_\alpha : \alpha \in I\}$ be any collection of sets indexed by an index set I. Then

$$\prod_{\alpha in I} A_\alpha = \left\{ a : I \to \bigcup_{\alpha \in I} A_\alpha | a(\alpha) \in A_\alpha \right\}.$$

In the common case where $A_\alpha = A$ for all $\alpha in I$, we will write

$$A^{|I|} = \prod_{\alpha in I} A.$$

In particular $A \times A \times A \times \ldots \times A = A^n$.

The ***axiom of choice*** states that if $A_\alpha \neq \emptyset$ for all $\alpha in I$, then

$$\prod_{\alpha in I} A_\alpha \neq \emptyset.$$

The reason that this is called the axiom of choice is because it has the immediate corollary that given an arbitrary disjoint collection of non-empty sets A_α for $\alpha \in I$, we can find a set

$$B \subseteq \bigcup_{\alpha \in I} A_\alpha$$

so that $B \cap A_\alpha$ has exactly one element. In other words, we can choose exactly one element from each of an arbitrary collection of sets.

The axiom of choice is equivalent to a number of different statements. The most useful is Zorn's Lemma:

Theorem A.1.1 (Zorn's Lemma)
If X is a partially ordered set, and every totally ordered subset of X has an upper bound, then X has a maximal element.

Some other equivalent statements are:

Theorem A.1.2 (Hausdorff Maximal Principle)
Every partially ordered set has a maximal linearly ordered set.

Theorem A.1.3 (Well Ordering Principle)
For every set X there exists a total order which is a well ordering of X.

Although the Axiom of Choice and its equivalents can be very powerful, it is considered preferable to avoid their use if possible, as any proof which requires their use is neccesarily non-constructive.

Exercises

1. Show that a function $f : \mathbb{R} \to \mathbb{R}$ can be regarded as an element of
$$\prod_{\alpha in \mathbb{R}} \mathbb{R} = \mathbb{R}^{|\mathbb{R}|}.$$
2. Show that the Axiom of Choice, Zorn's Lemma, the Hausdorff Maximal Principle and the Well Ordering Principle are all equivalent.

A.2 Algebraic Systems

We have the usual number systems:

Natural Numbers: $\mathbb{N} = \{1, 2, 3, \ldots\}$.

Integers: $\mathbb{Z} = \{\ldots - 2, -1, 0, 1, 2, \ldots\}$.

Rational Numbers: $\mathbb{Q} = \left\{\frac{p}{q} : p \in \mathbb{Z}, q \in \mathbb{N}, \gcd(p, q) = 1\right\}$.

Real Numbers: $\mathbb{R}$ = the (Cauchy) completion of $\mathbb{Q}$.

Complex Numbers: $\mathbb{C} = \{x + iy : x, y \in \mathbb{R}, i^2 = -1\}$.

There is the obvious chain of inclusions $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$, which we will use without comment.

The sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ are totally ordered by $\leq$. $\mathbb{C}$ is partially ordered by the relation $a \leq b$ iff $a - b \in \mathbb{R}$ and $a - b \leq 0$. These orders are compatible with the standard inclusions of these sets.

It will often be useful to work with infinite quantities. The ***extended real numbers***, $\mathbb{R}^\sharp$ is the set of real numbers with positive and negative points at infinity added.
$$\mathbb{R}^\sharp = \mathbb{R} \cup \{\pm\infty\}.$$
We assume that $-\infty < a < +\infty$, that $a + \infty = \infty$ and that $a - \infty = -\infty$ for each $a \in \mathbb{R}$. Also $a \cdot \infty$ is $+\infty$, 0 and $-\infty$, and $a \cdot -\infty$ is $-\infty$, 0 and $+\infty$, for $a > 0$, $a = 0$ and $a < 0$ respectively. $\infty - \infty$ is undefined. Clearly $\mathbb{R} \subset \mathbb{R}^\sharp$, and we will use this inclusion without comment.

A.2.1 Binary Operations and Groups

An ***binary operation*** $\bullet$ on a set X is a function $\bullet : X \times X \to X$. We will usually write $x \bullet y$ for $\bullet(x, y)$. A binary operation $\bullet$ is:

associative if $x \bullet (y \bullet z) = (x \bullet y) \bullet z$ for all x, y, $z \in X$;

commutative if $x \bullet y = y \bullet x$ for all x, $y \in X$.

An element $e \in X$ is a ***(two-sided) identity*** for $\bullet$ if $x \bullet e = x = e \bullet x$; and an element y is a ***(two-sided) inverse*** for x if $x \bullet y = e = y \bullet x$.

If A and B are subsets of X, and $x \in X$, we define

$$\begin{aligned} A \bullet B &= \{a \bullet b : a \in A, b \in B\} \\ x \bullet A &= \{x\} \bullet A = \{x \bullet a : a \in A\} \\ A \bullet x &= A \bullet \{x\} = \{a \bullet x : a \in A\}. \end{aligned}$$

A set with an associative binary operation is a ***semigroup***, if it also has an identity, it is a ***monoid***.

A ***group*** $\mathbb{G} = (G, \bullet, e)$ is a set G together with an associative binary operation $\bullet$ on G and an identity $e \in G$ for $\bullet$ such that every element $g \in G$ has a unique inverse, usually denoted g^{-1}. If $\bullet$ is also commutative, we say that $\mathbb{G}$ is ***abelian***, and we often suggestively use $+$ for the operation, 0 for the identity and $-g$ for the inverse. In groups where the operation is meant to suggest multiplication, we will often simply write xy for $x \bullet y$, and 1 for the identity.

If A is a subset of G, we define

$$A^{-1} = \{a^{-1} : a \in A\}.$$

If we are using additive notation, we write $-A$ instead, and $A - B$ for $A + (-B)$.

If $\mathbb{G}_1$ and $\mathbb{G}_2$ are groups, a function $\varphi : G_1 \to G_2$ is a ***(group) homomorphism*** if $\varphi(x \bullet y) = \varphi(x) \bullet \varphi(y)$ and $\varphi(x^{-1}) = \varphi(x)^{-1}$ for all x and $y \in G_1$. This implies that $\varphi(e_1) = e_2$. You can combine both conditions into the one check that $\varphi(x \bullet y^{-1}) = \varphi(x) \bullet \varphi(y)^{-1}$.

Exercises

1. Show that $(\mathbb{N}, +)$ is a semigroup, that $(\mathbb{N} \cup \{0\}, +, 0)$ is a monoid, and that $(\mathbb{Z}, +, 0)$ is a group.
2. Let X be a set and F be the set of bijections $f : X \to X$. Show that $(F, \circ, \mathrm{id})$ is a group, where $\circ$ is composition of functions and $\mathrm{id}(x) = x$.

A.2.2 Rings, Fields, Vector Spaces, Algebras

A ***ring*** $\mathbb{K} = (R, +, \cdot, 0)$ is a set R with two binary operations $+$ and $\cdot$ on R, such that $(R, +, 0)$ is an abelian group, and $(R, \cdot)$ is a semigroup; $x \cdot 0 = 0 = 0 \cdot x$; and $\cdot$ is ***left-*** and ***right- distributive*** for $+$:

$$x \cdot (y + z) = x \cdot y + x \cdot z \qquad \text{and} \qquad (x + y) \cdot z = x \cdot z + y \cdot z$$

for all x, y and $z \in R$.

A ring is ***unital*** if there is a multiplicative identity, $1 \in R$, which makes $(R, \cdot, 1)$ a unital semigroup, and a ring is ***abelian*** if the multiplication is abelian. A ***field*** $\mathbb{F} = (F, +, \cdot, 0, 1)$ is a unital abelian ring such that $(F \setminus 0, \cdot, 1)$ is a group.

A ***vector space*** over a field $\mathbb{F} = (F, +, \cdot, 0, 1)$ is a tuple $(V, +, \cdot, 0)$, where $(V, +, 0)$ is an additive group and $\cdot : \mathbb{F} \times V \to V$ is a ***scalar multiplication*** such that

- $\cdot$ is associative: $\alpha(\beta v) = (\alpha\beta)v$.
- 1 is a left identity for $\cdot$: $1v = v$.
- $\cdot$ is right distributive for $+$: $(\alpha + \beta)v = \alpha v + \beta v$.
- $\cdot$ is left distributive for $+$: $\alpha(v + w) = \alpha v + \alpha w$.

Note that we are deliberately using the same notation for the addition operation on the field and the group, even though these are distinct operations; the operation required can be deduced by context. Similarly, we use the same notation for the multiplication operation on the field and the scalar multiplication, even though these are distinct operations.

If V_1 and V_2 are vector spaces over the same field $\mathbb{F}$, a map $\varphi : V_1 \to V_2$ is ***linear*** if

$$\varphi(u + v) = \varphi(u) + \varphi(v) \qquad \text{and} \qquad \varphi(\lambda \cdot v) = \lambda \cdot \varphi(v)$$

for all u and $v \in V_1$ and $\lambda \in \mathbb{F}$. It is often most convenient to combine these two checks into the one check that

$$\varphi(u + \lambda \cdot v) = \varphi(u) + \lambda \cdot \varphi(v).$$

It is automatic that $\varphi(0) = 0$ for a linear map.

If we are considering a vector space over $\mathbb{C}$, we will say that a map is ***antilinear*** if

$$\varphi(u + v) = \varphi(u) + \varphi(v) \qquad \text{and} \qquad \varphi(\lambda \cdot v) = \overline{\lambda} \cdot \varphi(v)$$

for all u and $v \in V_1$ and $\lambda \in \mathbb{C}$. Again, it is not hard to see that $\varphi(0) = 0$ for an antilinear map.

An ***algebra*** over a field $\mathbb{F}$ is a tuple $(A, +, \cdot, \star, 0)$ where $(A, +, \cdot, 0)$ is a vector space over $\mathbb{F}$, $(A, +, \star, 0)$ is a ring, and the scalar product $\cdot$ commutes with the algebra product $\star$:

$$\alpha \cdot (a \star b) = (\alpha \cdot a) \star b = a \star (\alpha \cdot b).$$

As is the case f If there is a multiplicative identity 1 for the ring, we say that the algebra is ***unital***. If the ring is commutative, we say that the algebra is ***commutative***.

Exercises

1. Let $M_n(\mathbb{F})$ be the set of n by n matrices with entries in a field $\mathbb{F}$. Show that $(M_n, +, \cdot, 0)$ is a unital ring, where $\cdot$ is matrix multiplication, 0 is the zero matrix, and the identity matrix is the multiplicative unit.

   Show that $M_n(\mathbb{F})$ is a unital algebra over $\mathbb{F}$.

2. Show that the integers modulo p, where p is a prime number, form a field.

3. Let X be any set, and let $\mathbb{R}(X)$ be the set of all functions $f : X \to \mathbb{R}$. Show that $(\mathbb{R}(X), +, \cdot, 0)$ is a vector space over $\mathbb{R}$, where $+$, $\cdot$ are the usual addition and scalar multiplication of functions, and 0 is the zero function $0(x) = 0$.

4. Let $\mathbb{R}(X)$ be as in the previous exercise. Show that $\mathbb{R}(X)$ is a unital ring.

   Show that $\mathbb{R}(X)$ is a commutative unital algebra over $\mathbb{R}$.

5. Let V, W be vector spaces over a field $\mathbb{F}$. Let $L(V, W)$ be the set of linear maps from V to W. Show that $L(V, W)$ is a vector space over $\mathbb{F}$.

6. Show that if V is a vector space over $\mathbb{F}$, and $L(V) = L(V, V)$, then $L(V)$ is a unital ring where the product is composition of linear operators:

   $$(ST)(v) = S(T(v))$$

   for all $v \in V$, S, $T \in L(V)$, and the multiplicative identity is the identity map.

   Show that $L(V)$ is in fact an algebra over $\mathbb{F}$.

A.3 Real- and Complex-valued Functions

If X is a set, we will often want to consider functions $f : X \to \mathbb{R}$ or $f : X \to \mathbb{C}$. We will denote these sets of functions by $\mathbb{R}(X)$ and $\mathbb{C}(X)$ respectively.

The set of all such functions is a vector space where the functions λf and $f + g$ are defined in the usual way:

$$(\lambda f)(x) = \lambda f(x) \qquad \text{and} \qquad (f + g)(x) = f(x) + g(x)$$

where λ is a scalar and f and g are functions. In fact, this space is also an algebra, where the product fg of two functions f and g, is given by

$$(fg)(x) = f(x)g(x)$$

as usual. We can divide one function by another if the denominator is never zero:

$$\left(\frac{f}{g}\right)(x) = \frac{f(x)}{g(x)},$$

where $g(x) \neq 0$ for all $x \in X$.

We say that $f \leq g$ if $f(x) \leq g(x)$ for all $x \in X$. This is a partial order on these functions. Such a function is ***positive*** if $f \geq 0$. We denote the positive functions by $\mathbb{R}(X)_+$ and $\mathbb{C}(X)_+$. More generally, if A is any collection of functions on X, the positive functions in A are denoted by A_+.

We can also define a function $|f|$ which is the ***absolute value*** function of f by

$$|f|(x) = |f(x)|.$$

Clearly $|f|$ is positive and $f \leq |f|$.

For complex valued functions, we can also define the ***complex conjugate*** function $\overline{f}$ or f^* by

$$\overline{f}(x) = \overline{f(x)},$$

as well as the ***real*** and ***imaginary parts*** of a function

$$(\operatorname{Re} f)(x) = \operatorname{Re}(f(x)) \qquad (\operatorname{Im} f)(x) = \operatorname{Im}(f(x)).$$

Note that $\operatorname{Re} f$ and $\operatorname{Im} f$ are real-valued. As you might expect $f = \operatorname{Re} f + i \operatorname{Im} f$, $\overline{f} = \operatorname{Re} f - i \operatorname{Im} f$ and $|f| = \sqrt{(\operatorname{Re} f)^2 + (\operatorname{Im} f)^2}$.

Exercises

1. Show that for any function $f : X \to \mathbb{C}$, that $f = \operatorname{Re} f + i \operatorname{Im} f$, $\overline{f} = \operatorname{Re} f - i \operatorname{Im} f$ and $|f| = \sqrt{(\operatorname{Re} f)^2 + (\operatorname{Im} f)^2}$.
2. Show that the map $f \mapsto \overline{f}$ is an antilinear map on $\mathbb{C}(X)$.

A.4 Suprema and Infima

If $\lhd$ is a partial order on a set A, and $X \subseteq A$, then the ***supremum*** of X, if it exists, is the ***least upper bound*** of X, ie. a is the supremum of X iff a is an upper bound for X, and if b is any other upper bound, then $a \lhd b$. Similarly, the ***infimum*** of X is the ***greatest lower bound***. The supremum is usually denoted

$$\sup X$$

and the infimum by

$$\inf X.$$

In general, there is no guarantee that such elements exist. A set is ***order complete*** if every bounded set has a supremum and infimum.

We usually work with $\mathbb{R}$ or $\mathbb{R}^\sharp$ ordered by $\leq$, and in this case we say $\sup \emptyset = -\infty$; $\inf \emptyset = +\infty$; if A has no upper bound, then $\sup A = +\infty$; and if A has no lower bound, then $\inf A = -\infty$. Both $\mathbb{R}$ and $\mathbb{R}^\sharp$ are order complete.

For any set $A \subseteq \mathbb{R}$, we automatically have

$$\inf A \leq \sup A$$
$$\sup -A = -\inf A.$$

If $A \subseteq B \subseteq \mathbb{R}$,

$$\sup A \leq \sup B$$
$$\inf A \geq \inf B.$$

Given non-empty subsets A and B of $\mathbb{R}$,

$$\sup(A + B) = \sup A + \sup B$$
$$\inf(A + B) = \inf A + \inf B.$$

and if A and B are subsets of $[0, +\infty)$, we also have that

$$\sup AB = (\sup A)(\sup B)$$
$$\inf AB = (\inf A)(\inf B).$$

Exercises

1. Verify the claims made in this section.

A.4.1 Suprema, Infima and Real-Valued Functions

If X is an arbitrary set, and $f : X \to \mathbb{R}$ is a function, we will sometimes use the alternative notation

$$\sup_{x \in X} f(x) = \sup\{f(x) : x \in X\}$$
$$\inf_{x \in X} f(x) = \inf\{f(x) : x \in X\}.$$

Given functions f and $g : X \to \mathbb{R}$,

$$\sup\{f(x) + g(x) : x \in X\} \leq \sup\{f(x) : x \in X\} + \sup\{g(x) : x \in X\}$$
$$\inf\{f(x) + g(x) : x \in X\} \geq \inf\{f(x) : x \in X\} + \inf\{g(x) : x \in X\}.$$

If $f \leq g$, we have

$$\sup\{f(x) : x \in X\} \leq \sup\{g(x) : x \in X\}$$
$$\inf\{f(x) : x \in X\} \leq \inf\{g(x) : x \in X\}.$$

If $f \geq 0$ and $g \geq 0$, then

$$\sup\{f(x)g(x) : x \in X\} \leq (\sup\{f(x) : x \in X\})(\sup\{g(x) : x \in X\})$$
$$\inf\{f(x)g(x) : x \in X\} \geq (\inf\{f(x) : x \in X\})(\inf\{g(x) : x \in X\}).$$

We define the ***uniform norm*** of a function $f : X \to \mathbb{R}$ or $\mathbb{C}$ by

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

The above facts allow us to conclude that

$$\|\lambda f\|_\infty = |\lambda| \|f\|_\infty,$$
$$\|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty, \text{and}$$
$$\|fg\|_\infty \leq \|f\|_\infty \|g\|_\infty$$

for all scalars λ and functions f and g. Additionally, if $0 \leq f \leq g$,

$$\|f\|_\infty \leq \|g\|_\infty.$$

Given a family of real-valued functions $\mathcal{F}$, we define a function $\bigvee_{f\in\mathcal{F}} f$ by

$$\left(\bigvee_{f\in\mathcal{F}} f\right)(x) = \sup\{f(x) : f \in \mathcal{F}\},$$

and a function $\bigwedge_{f\in\mathcal{F}} f$ by

$$\left(\bigwedge_{f\in\mathcal{F}} f\right)(x) = \inf\{f(x) : f \in \mathcal{F}\}.$$

For pairs of functions we write $f \vee g$ and $f \wedge g$ for these functions respectively.

Exercises

1. Verify that

$$\begin{aligned} \|\lambda f\|_\infty &= |\lambda|\|f\|_\infty, \\ \|f+g\|_\infty &\leq \|f\|_\infty + \|g\|_\infty, \text{and} \\ \|fg\|_\infty &\leq \|f\|_\infty\|g\|_\infty \end{aligned}$$

   for all scalars λ and functions f and g.

2. Show that

$$\left\|\bigvee_{f\in\mathcal{F}} f\right\|_\infty = \sup\{\|f\|_\infty | f \in \mathcal{F}\}$$

   and

$$\left\|\bigwedge_{f\in\mathcal{F}} f\right\|_\infty = \inf\{\|f\|_\infty | f \in \mathcal{F}\}.$$

3. Let $f : X \to \mathbb{R}$, and assume that there is some $\varepsilon > 0$ such that $|f(x)| > \varepsilon$ for all $x \in X$ (ie. f is ***bounded away from 0***). Show that $\|1/f\|_\infty < 1/\varepsilon$.

A.5 Sequences

A ***sequence*** $(x_n)_{n=1}^\infty$ in a set X is a function

$$x : \mathbb{N} \to X : n \mapsto x_n.$$

A property holds ***eventually*** for a sequence if there is some $n_0 \in \mathbb{N}$ such that the property holds for all x_n with $n \geq n_0$. A property holds ***frequently*** if for every $n \in \mathbb{N}$ there is some $m \geq n$ so that the property holds for x_m.

A sequence $(a_n)_{n=1}^\infty$ in $\mathbb{R}$ or $\mathbb{C}$ ***converges*** to a if the sequence is eventually in any neighbourhood of a. More concretely, for every $\varepsilon > 0$, there exists an n_0 such that

$$|a_n - a| < \varepsilon \qquad \text{for all } n \geq n_0.$$

We write "$a_n \to a$ as $n \to \infty$", or

$$\lim_{n\to\infty} a_n = a.$$

For real numbers, we say that $a_n \to +\infty$ if a_n is eventually in any set of the form $(M, +\infty]$, and $a_n \to -\infty$ if a_n is eventually in any set of the form $[-\infty, M)$.

A sequence $(f_n)_{n=1}^\infty$ of real- or complex-valued functions on a set X converges ***pointwise*** to a function f if for each $x \in X$, $f_n(x) \to f(x)$ as $n \to \infty$. The sequence converges ***uniformly*** to f if for all $\varepsilon > 0$,

$$\|f_n(x) - f(x)\|_\infty < \varepsilon$$

eventually. Any uniformly convergent sequence is automatically pointwise convergent.

Exercises

1. Use the definition of the limit of a sequence to *prove* that $1/n^2 \to 0$ as $n \to \infty$.

2. Consider the sequence $\sin n\pi/4$. For which $x \in \mathbb{R}$ is $|\sin n\pi/4 - x| < \varepsilon$ frequently.

3. Let a, $b \in \mathbb{R}$, and let $0 < \varepsilon < |a - b|/2$. Show that a sequence $(x_n)_{n=1}^\infty$ cannot converge if both

$$|x_n - a| < \varepsilon \qquad \text{and} \qquad |x_n - b| < \varepsilon$$

are frequently true.

4. Show that if $f_n \to f$ uniformly, then $f_n \to f$ pointwise. Provide an example that shows the converse doews not hold.

A.5.1 Evaluating Limits

Most of the following should be familiar from calculus or undergraduate real analysis.

Theorem A.5.1 (Sandwich (or Squeeze) Theorem)
If a_n, b_n and c_n are sequences of real numbers such that eventually $a_n \leq b_n \leq c_n$ and

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} c_n = L,$$

then b_n converges to L.

A sequence of real numbers is ***monotone increasing*** on a set if $x_m \geq x_n$ for all $m \geq n$ with m, n in the set; it is ***monotone decreasing*** on a set if $x_m \geq x_n$ for all $m \geq n$, with m, n in the set; it is ***bounded above*** on a set if there is some M with $x_n \leq M$ for all n in the set, and ***bounded below*** on a set if there is some M with $x_n \geq M$ for all n in the set.

Theorem A.5.2 (Monotone Sequence Theorem)
If a_n is a sequence which is eventually monotone increasing, then it is either eventually bounded above, in which case it converges to some $a \leq M$, or it is not, in which case it converges to $+\infty$.

If a_n is a sequence which is eventually monotone decreasing, then it is either eventually bounded below, in which case it converges to some $a \geq M$, or it is not, in which case it converges to $-\infty$.

The Sandwich and Monotone Sequence theorems can be applied to pointwise limits of real-valued functions as well, remembering that $f \leq g$ iff $f(x) \leq g(x)$ for all x.

For sequences of real or complex numbers, or real- or complex- valued functions (for either type of convergence), if $a_n \to a$ and $b_n \to b$ as $n \to \infty$, then $a_n + b_n \to a + b$, $a_n - b_n \to a - b$, and $a_n b_n \to ab$. Also $a_n/b_n \to a/b$, provided $b \neq 0$ (for numbers); or $b(x) \neq 0$ for all x for functions.

Theorem A.5.3 (Cancellation Theorem)
If a_n, b_n and c_n are sequences of real or complex numbers, with b_n and $c_n \neq 0$, and a_n/b_n converges, then

$$\lim_{n\to\infty} \frac{a_n c_n}{b_n c_n} = \lim_{n\to\infty} \frac{a_n}{b_n}$$

The following concrete limits are sometimes useful to know:

$$\begin{aligned} c &\to c \qquad \text{for any number } c \\ n^p &\to +\infty \qquad \text{for } p > 0 \\ \frac{1}{n^p} &\to 0 \qquad \text{for } p > 0 \\ n \sin\frac{1}{n} &\to 1 \\ n(1 - \cos\frac{1}{n}) &\to 0 \end{aligned}$$

For complex numbers, if $z_n = a_n + ib_n$, then $z_n \to z = a + ib$ as $n \to \infty$ if and only if $a_n \to a$ and $b_n \to b$ as $n \to \infty$.

A sequence $(x_n)_{n=1}^{\infty}$ is ***Cauchy*** if for every $\varepsilon > 0$, $|x_n - x_m| < \varepsilon$ eventually. ie. there is some n_0 such that $|x_n - x_m| < \varepsilon$ for all n and m greater than n_0. Every Cauchy sequence in $\mathbb{R}$ or $\mathbb{C}$ converges.

The following theorem is less well-known, but can be invaluable for finding limits:

Theorem A.5.4
If a sequence a_n has the property that every subsequence a_{n_k} has a sub-sub-sequence $a_{n_{k_l}}$ which converges to a, then a_n converges to a.

This theorem in fact holds no matter what sort of sequence we are considering.

Exercises

1. Prove that $\frac{\sin n}{n} \to 0$ as $n \to \infty$.
2. Show that every convergent sequence in $\mathbb{R}$ is a Cauchy sequence.
3. We say that a sequence $(f_n)_{n=1}^{\infty}$ of functions $f_n : X \to \mathbb{R}$ is uniformly Cauchy if for every $\varepsilon > 0$, $\|f_n - f_m\|_\infty < \varepsilon$ everntually. Show that every uniformly convergent sequence of functions is uniformly Cauchy.
4. Show that every uniformly Cauchy sequence of functions is uniformly convergent.

A.5.2 Limit Supremum and Limit Infimum of Sequences

Even if a sequence of real numbers does not converge, we can still extract some information from it. The ***limit supremum*** of a sequence a_n, is the limit of the sequence $x_k = \sup\{a_n : n \geq k\}$, or more concisely,

$$\limsup_{n\to\infty} a_n = \lim_{k\to\infty} \sup\{a_n : n \geq k\}.$$

This sequence either converges to a number or $-\infty$, since it is decreasing.

Similarly, the ***limit infimum*** is the limit of the infimum of the tail of the sequence,

$$\liminf_{n\to\infty} a_n = \lim_{k\to\infty} \inf\{a_n : n \geq k\}.$$

Clearly,

$$\limsup_{n\to\infty} a_n \geq \liminf_{n\to\infty} a_n,$$

and you get equality iff the sequence converges, in which case the limit supremum and the limit infimum are both equal to the limit.

For sequences of real-valued functions, we define the limit supremum function

$$f = \limsup_{n\to\infty} f_n$$

by

$$f(x) = \limsup_{n\to\infty} f_n(x).$$

The limit infimum function is analogously defined. These functions may take values in the extended reals.

Once again

$$\limsup_{n\to\infty} f_n \geq \liminf_{n\to\infty} f_n$$

and we have equality iff f_n converges pointwise, and in this case the limit function is equal to the limit supremum and limit infimum functions.

A.5.3 Topology of $\mathbb{R}$

A subset $U of \mathbb{R}$ is ***open*** if given any $x \in U$ there is a $\varepsilon > 0$ such that if $|x-y| < \varepsilon$ then $y \in U$. Equivalently, U can be written as a union of open intervals. A subset of $\mathbb{R}$ is ***closed*** if its complement is open. The empty set and $\mathbb{R}$ are considered to be both open and closed.

Arbitrary unions of open sets are open, and finite intersections of open sets are open. Arbitrary intersections of closed sets are closed, and finite unions of closed sets are closed.

Closed sets are also characterised by the property that the limit points of any sequence contained in the set are also contained in the set.

A set is ***compact*** if it is closed and bounded. Equivalently, it is compact if every open cover of the set has a finite subcover.

Exercises

1. Find $\limsup \frac{n+1}{n}$.

2. Let $f_n(x) = \frac{n-x^2}{n}$. Find $\limsup f_n$.

3. Let f_n be any sequence of functions $f_n : X \to \mathbb{R}$. Show that

$$\left\|\limsup_{n\to\infty} f_n\right\|_\infty = \limsup_{n\to\infty} \|f_n\|.$$

A.6 Series

A ***series***

$$\sum_{n=1}^{\infty} x_n$$

is a sequence $(x_n)_{n=1}^{\infty}$ in an additive group $(X, +, 0)$, regarded as an infinite sum,

$$\sum_{n=1}^{\infty} x_n = x_1 + x_2 + x_3 + \dots$$

We say that a series of real or complex numbers ***converges*** if the sequence

$$s_n = \sum_{k=1}^{n} x_n$$

of partial sums converges. Similarly, if we have a series of real- or complex-valued functions, we say that the series ***converges pointwise*** (respectively ***uniformly***) if the sequence of partial sums converges pointwise (respectively uniformly).

The limit of the sequence of partial sums is the limit of the series, and if $s_n \to x$ as $n \to \infty$, we write

$$\sum_{n=1}^{\infty} x_n = x.$$

A series of numbers or functions is ***absolutely convergent*** in the appropriate sense if the series

$$\sum_{n=1}^{\infty} |x_n|$$

converges. All absolutely convergent series converge. A series which converges, but does not converge absolutely, is ***conditionally convergent***. Absolutely convergent series have the important property that the limit of the series does not depend on the order of summation: for all bijections $\iota : \mathbb{N} \to \mathbb{N}$,

$$\sum_{n=1}^{\infty} x_n = \sum_{n=1}^{\infty} x_{\iota(n)}$$

iff the series converges absolutely.

The ***geometric series***

$$\sum_{n=1}^{\infty} ar^{n-1} = \frac{a}{1-r}$$

converges absolutely for $|r| < 1$ and diverges otherwise.

The harmonic series

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

diverges, but the alternating harmonic series

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n} = \ln 2$$

converges.

In general, if you need to evaluate a series, you want to either relate it to a power series (see below), hope that it is a telescoping series, or try to explicitly find the partial sums and take a limit.

A.6.1 Series Convergence Tests

Given a sequence of non-negative numbers, we have the following tests for convergence:

Theorem A.6.1 (Divergence Test)
If $\lim a_n \neq 0$, *then* $\sum_{n=1}^{\infty} a_n$ *diverges.*

Theorem A.6.2 (Comparison Test)
If $a_n \geq b_n \geq 0$ *eventually, then if* $\sum_{n=1}^{\infty} a_n$ *converges, so does* $\sum_{n=1}^{\infty} b_n$. *On the other hand, if* $\sum_{n=1}^{\infty} b_n$ *diverges,* $\sum_{n=1}^{\infty} a_n$ *also diverges.*

Theorem A.6.3 (Integral Comparison Test)
If $a_n \geq 0$ *and there is a continuous, positive, decreasing function* $f : [N, \infty)$ *such that* $a_n = f(n)$ *eventually, then if the improper Riemann integral*

$$\int_N^{\infty} f(x)\ dx \text{ converges}$$

so does $\sum_{n=1}^{\infty} a_n$. *On the other hand, if the improper Riemann integral*

$$\int_N^{\infty} f(x)\ dx \text{ diverges}$$

so does $\sum_{n=1}^{\infty} a_n$.

See Section A.7.3 below for the definitions of improper integrals.

Theorem A.6.4 (Limit Comparison Test)
If $a_n \geq 0$ *and* $b_n \geq 0$, *then if*

$$\limsup \frac{a_n}{b_n} < \infty$$

and $\sum_{n=1}^{\infty} b_n$ *converges, then* $\sum_{n=1}^{\infty} a_n$ *converges. On the other hand, if*

$$\liminf \frac{a_n}{b_n} > 0$$

and $\sum_{n=1}^{\infty} b_n$ *diverges, then* $\sum_{n=1}^{\infty} a_n$ *diverges.*

The limit comparison test can be used with limits instead of limit infima and limit suprema, but this way of stating things is more general.

Theorem A.6.5 (Alternating Series Test)
If $a_n = \pm(-1)^n b_n$ *eventually, where* $b_n \geq 0$, $b_{n+1} \leq b_n$, *and* $\lim_{n\to\infty} b_n = 0$, *then* $\sum_{n=1}^{\infty} a_n$ *converges.*

Theorem A.6.6 (Ratio Test)
Given a series $\sum_{n=1}^{\infty} a_n$, *let*

$$L = \limsup \left| \frac{a_n}{a_{n+1}} \right|.$$

If $L < 1$, *then the series converges absolutely. If* $L > 1$, *the series diverges.*

Theorem A.6.7 (Root Test)
Given a series $\sum_{n=1}^{\infty} a_n$*, let*

$$L = \limsup \sqrt[n]{|a_n|}.$$

If $L < 1$*, then the series converges absolutely. If* $L > 1$*, the series diverges.*

In both the root and ratio tests, if $L = 1$, the series may converge absolutely, converge conditionally, or diverge.

A.7 Functions on $\mathbb{R}$

In this section we consider functions $f : D \to \mathbb{R}$ where $D \subseteq \mathbb{R}$. Much of this section is equally valid when considering functions $f : D \to \mathbb{C}$, where $D \subseteq \mathbb{R}$.

We say that the ***limit*** of f as x goes to a is L, and write $f(x) \to L$ as $x \to a$, or

$$L = \lim_{x \to a} f(x),$$

if given any sequence $x_n \to a$ in D, we have

$$L = \lim_{n \to \infty} f(x_n).$$

Equivalently, given any $\varepsilon > 0$, there is a $\delta > 0$ such that $|f(x) - L| < \varepsilon$ for all $x \in D$ with $|x - a| < \delta$.

If there is no such L, we say the limit does not exist.

These limits observe the same rules as limits of sequences.

A function $f : D \to \mathbb{R}$ is ***continuous at*** $a \in D$ if

$$\lim_{x \to a} f(x) = f(a).$$

Equivalently, f is continuous at a if for every $\varepsilon > 0$ there is a $\delta > 0$ such that $|f(x) - f(a)| < \varepsilon$ for all $x \in D$ with $|x - a| < \delta$.

A function is ***continuous on*** $X \subseteq D$ if it is continuous at every $x \in X$.

A function is ***continuous*** if it is continuous on its domain. Equivalently, a function $f : D \to \mathbb{R}$ is continuous if for any open set $U \subseteq \mathbb{R}$ we have that $f^{-1}(U)$ is relatively open in D. In fact it is sufficient that this hold for open intervals.

A function $f : D \to \mathbb{R}$ is ***uniformly continuous*** if for every $a \in D$ and $\varepsilon > 0$ there is a $\delta > 0$ such that $|f(x) - f(a)| < \varepsilon$ for all $x \in D$ with $|x - a| < \delta$. Not that this definition differs from the definition of continuity in that the value of δ does not depend on the point a.

A.7.1 Differentiation

A function is ***differentiable at*** $a \in D$ if the limit

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

exists and is finite. In this case we say

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

A function which is differentiable at a must be continuous at a.

A function is ***differentiable on*** $X \subseteq D$ if it is differentiable at every point $x \in X$. A function is ***differentiable*** if it is differentiable on its domain. The notation Df and $\frac{df}{dx}$ are often used for f'. Higher derivatives can be denoted by f'', f''', $f^{(n)}$; $D^n f$; or $\frac{d^n f}{dx^n}$ as usual.

If $f^{(n)}$ exists and is continuous on D, we say that f is ***n-times continuously differentiable***, and write $f \in C^n(D)$. If $f^{(n)}$ exists for every n, we say that f is ***infinitely differentiable***, and write $f \in C^\infty(D)$.

Amongst other facts, we know that differentiation is linear:

$$D(\lambda f) = \lambda Df \qquad \text{and} \qquad D(f+g) = Df + Dg$$

for all $\lambda \in \mathbb{R}$, and f, g differentiable.

We say that a function $F : D \to \mathbb{R}$ is an ***antiderivative*** of f, if $DF = f$. We denote the set of all antiderivatives of a function f by the ***indefinite integral***

$$\int f(x)\ dx = \{F : D \to \mathbb{R} | DF = f\}.$$

If D is an interval, f is continuous, and F is any one antiderivative, then every antiderivative is of the form $F(x) + C$ for some constant C, and we traditionally write

$$\int f(x)\ dx = F(x) + C$$

where C is the ***constant of integration*** and is assumed to take every possible real value.

A.7.2 Riemann Integration

A ***partition*** $\mathcal{P}$ of an interval $I = [a,b]$ is a collection of points $a = x_0 < x_1 < x_2 < \ldots < x_{n-1} < x_n = b$. The ***norm*** or ***mesh***, $\|\mathcal{P}\|$ of $\mathcal{P}$ is

$$\|\mathcal{P}\| = \max_k \{x_k - x_{k-1}\}.$$

A partition $\mathcal{P}_1$ is a ***refinement*** of $\mathcal{P}_2$, denoted $\mathcal{P}_1 \subseteq \mathcal{P}_2$, if every point in $\mathcal{P}_1$ is also in $\mathcal{P}_2$. If $\mathcal{P}_1 \subseteq \mathcal{P}_2$, then $|\mathcal{P}_1| \geq |\mathcal{P}_2|$. If $\mathcal{P}_1$ and $\mathcal{P}_2$ are two partitions, then there is a partition $\mathcal{P}_1 \cup \mathcal{P}_2$ containing the points in both partitions, which is a refinement of both $\mathcal{P}_1$ and $\mathcal{P}_2$.

If f is a real-valued function on an interval $I = [a,b]$, and $\mathcal{P}$ is a partition of I, then we define

$$U(f, \mathcal{P}) = \sum_{k=1}^{n} (\sup_{x \in [x_{k-1}, x_k]} f(x))(x_k - x_{k-1})$$

to be the ***upper sum*** of f over $\mathcal{P}$, and

$$L(f,\mathcal{P}) = \sum_{k=1}^{n} (\inf_{x\in[x_{k-1},x_k]} f(x))(x_k - x_{k-1})$$

to be the ***lower sum***. It is not hard to see that

$$L(f,\mathcal{P}_1) \le U(f,\mathcal{P}_2)$$

for any two partitions. We let

$$U(f,I) = \inf\{U(f,\mathcal{P}) : \mathcal{P} \text{ a partition of } I\}$$

and

$$L(f,I) = \sup\{L(f,\mathcal{P}) : \mathcal{P} \text{ a partition of } I\}.$$

If $U(f,I) = L(f,I)$, then we say that f is ***Riemann integrable*** and it has ***Riemann integral***

$$\int_a^b f(x)\ dx = U(f,I) = L(f,I).$$

We denote the set of Riemann integrable functions on I by $\mathcal{R}(I)$ or $\mathcal{R}[a,b]$. Every continuous function on I is Riemann integrable.

We know that Riemann integration is linear:

$$\int_a^b \lambda f(x)\ dx = \lambda \int_a^b f(x)\ dx$$

and

$$\int_a^b f(x) + g(x)\ dx = \int_a^b f(x)\ dx + \int_a^b g(x)\ dx$$

for all $\lambda \in \mathbb{R}$ and f, $g \in \mathcal{R}(I)$. It is also order-preserving: if $f \le g$ and f, $g \in \mathcal{R}(I)$, then

$$\int_a^b f(x)\ dx \le \int_a^b g(x)\ dx.$$

Also if $c \in [a,b]$,

$$\int_a^b f(x)\ dx = \int_a^c f(x)\ dx + \int_c^b f(x)\ dx.$$

The fundamental theorem of calculus tells us that if f is continuous on $[a,b]$, the function

$$F(x) = \int_a^x f(x)\ dx$$

is an antiderivative of f. It also says that if F is any antiderivative of a continuous function f on $[a,b]$, then

$$\int_a^b f(x)\ dx = F(b) - F(a).$$

A.7.3 Improper Integrals

If $D = [a, \infty)$, and $f : D \to \mathbb{R}$ is in $\mathcal{R}([a, t])$ for every $t \in D$, we define

$$\int_a^\infty f(x)\ dx = \lim_{t \to \infty} \int_a^t f(x)\ dx,$$

provided the limit exists. We define

$$\int_{-\infty}^b f(x)\ dx$$

analogously.

If $D = [a, b)$, and and $f : D \to \mathbb{R}$ is in $\mathcal{R}([a, t])$ for every $t \in D$, we define

$$\int_a^\infty f(x)\ dx = \lim_{t \to b} \int_a^t f(x)\ dx,$$

provided the limit exists. We define an integral on $(a, b]$ analogously.

For open intervals, we define the integral of a function over that interval by splitting it into half-open subintervals. We can define the integral of a function f with a finite number of discontinuities in an interval I by splitting I into half-open subintervals where f is continuous and using improper integration.

A.7.4 Complex-valued Functions on $\mathbb{R}$

We briefly note here that we can easily define, continuity, derivatives and integration for functions $f : D \to \mathbb{C}$, where $D \subseteq \mathbb{R}$ with the above definitions. In these cases one can show that:

$$\begin{aligned}
\lim_{x \to a} f(x) &= \lim_{x \to a} \operatorname{Re} f(x) + i \lim_{x \to a} \operatorname{Im} f(x), \\
Df &= D(\operatorname{Re} f) + iD(\operatorname{Im} f), \\
\int_a^b f(x)\ dx &= \int_a^b \operatorname{Re} f(x)\ dx + i \int_a^b \operatorname{Im} f(x)\ dx.
\end{aligned}$$

In other words, one can simply use the corresponding operations on the real and imaginary parts of f. Differentiation and integration are then complex-linear maps, and the usual rules of integration and differentiation still apply.

For those of you who have seen some complex function theory, it is worth stressing that we are not dealing with functions defined on general subsets of $\mathbb{C}$, only subsets of $\mathbb{R}$, so we are not concerned with things like the Cauchy condition.

A.8 Series of Functions

Certain series of functions have some significance.

A.8.1 Power Series

A ***power series*** is a series of the form

$$\sum_{n=0}^{\infty} c_n (x-a)^n,$$

where c_n and a are constants, and x is regarded as a variable. Each power series has a ***radius of convergence*** $R \in [0, \infty]$ such that the series converges in the open ball $\{x : |x-a| < R\}$ and diverges on the set $\{x : |x-a| > R\}$.

A power series defines an infinitely differentiable function on its interval of convergence by

$$f(x) = \sum_{n=0}^{\infty} c_n (x-a)^n.$$

Such a function can be differentiated and integrated term-by-term, so that

$$f'(x) = \sum_{n=0}^{\infty} (n+1) c_{n+1} (x-a)^n$$

and

$$\int f(x)\ dx = C + \sum_{n=1}^{\infty} \frac{c_{n-1}}{n} (x-a)^n,$$

where C is the constant of integration. These functions have the same radius of convergence, but may have different convergence properties at the end-points of the interval of convergence.

Some important functions, such as the Gamma, Error, Bessel and Airy functions, are only given their power series.

Given a function f which is infinitely differentiable on some interval I, we define the Taylor series of f about $a \in I$ to be the power series

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (x-a)^n.$$

If $a = 0$, then this is a Maclaurin series. A function is ***analytic*** at a if it is equal to its Taylor series about a inside the region of convergence.

Not every infinitely differentiable function is analytic, the function

$$f(x) = \begin{cases} 0 & x \leq 0, \\ e^{-1/x^2} & x > 0, \end{cases}$$

being the classic counterexample. The Maclaurin series of this function is 0, and has an infinite radius of convergence, but clearly the function is not 0.

The following series are worth while knowing:

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n \qquad R = 1$$
$$e^x = \sum_{n=0}^{\infty} \frac{1}{n!} x^n \qquad R = \infty$$
$$\sin x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1} \qquad R = \infty$$
$$\cos x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n} \qquad R = \infty$$

More examples can be derived by integration, differentiation, and substitution of polynomials. These concrete power series can sometimes be used to evaluate particular series by substituting particular values for x.

A.8.2 Fourier Series

A ***trigonometric polynomial*** is a finite sum of the form

$$f(x) = a_0 + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx),$$

where a_k, b_k are constants. A ***Fourier series*** is a series of the form

$$a_0 + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx)$$

where a_k, b_k are constants. Given a function f defined on all of $\mathbb{R}$ which is periodic of period 2π (or simply a function defined on $(-\pi, \pi]$), we define the ***Fourier coefficients*** of f to be

$$a_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \cos kx \; dx$$

and

$$b_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \sin kx \; dx.$$

By using basic trig integrals, one can show that if f is an infinitely differentiable function given by a convergent Fourier series, then the Fourier coefficients correspond to the coefficients of the Fourier series.

With a little basic knowledge of complex analysis, specifically that $e^{ix} = \cos x + i \sin x$, one can simplify the above by considering complex-valued functions. We use ***(complex) trigonometric polynomials*** of the form

$$f(x) = \sum_{k=-n}^{n} c_k e^{ikx},$$

where c_k are (possibly complex) constants. In this case, if $f : \mathbb{R} \to \mathbb{C}$ is periodic with period 2π, the ***(complex) Fourier coefficients*** are defined by

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ikx} \, dx.$$

Again, if f is an infinitely differentiable function given by a Fourier series, then the Fourier coefficients correspond to the coefficients of the series.

Note that some sources use different periodicities and constants, but the underlying idea is the same.

Appendix B

Strategies

There are some standard "tricks" which you may notice are repeatedly used in real analysis proofs. Of course, you will still find the usual proof tricks, such as arguments by contrapositive and contradiction useful.

Some of the concepts and situations given below won't make sense initially, as we won't have discussed the underlying theory until later in the course, but hopefully by the end of the course this should all make sense.

B.1 Extending from a Subset

Often you want to prove that a given condition holds for everything in a certain set, such as a given class of functions, or all points in a topological space. In many cases, you can prove the result fairly easily for certain simple cases, and you can often extend the result using one or more of these tricks:

1. If you are in a topological space, prove the result for a dense subset, and try to extend it to all cases by taking appropriate limits. This requires you to know or prove that the simple case is in fact dense; and it requires that the condition you are proving behaves "nicely" when you take limits (and you have to be careful checking this, as it is easy to slip up).

   For example, when proving facts about positive measurable functions, we commonly start with simple functions, and take limits. However we need key theorems, such as the Dominated Convergence Theorem and the Monotone Convergence Theorem to be able to get properties to transfer under limits.

   Good candidates for dense sets include: the rational, algebraic, and diadic (or more generally p-adic) numbers; simple functions; continuous functions; polynomials; analytic functions (ie. functions given by power series); differentiable functions; "smooth" functions; rational-valued vectors or functions; etc.

2. If you are in a vector space, prove the result for a spanning subset, and try to extend it to all cases by linearity. This requires you to know or prove that the simple case covers a spanning set; and it requires that the condition you are proving behaves "nicely" when you add and take scalar multiples.

   For example, when proving facts about general integrable functions, we commonly try to prove the result for positive functions, and then use the linearity of integration to extend the result.

   Good candidates for spanning sets include: vector bases, particularly orthogonal or orthonormal bases; positive vectors or functions; real vectors or functions; vectors or functions in the unit ball of some norm; etc.

3. When working with classes of sets (such as σ-algebras and topologies), try to prove the result for a simple generating class (such as a base or sub-base of a topology; or a generating algebra of a σ-algebra) and then try to extend using appropriate unions and intersections.

   For example, when working with Lebesgue measure on $\mathbb{R}$, we often start with intervals, and then extend to measurable sets by taking countable disjoint unions and complements.

   Good candidates for simple classes of sets include: intervals, rectangles and boxes (sometimes we need open or closed as well); balls in a metric space; open sets of a topology; etc.

Sometimes you may need to combine these repeatedly. A tricky proof about the Lebesgue integral could, for example, require you to start with characteristic functions of intervals; extend that to characteristic functions of measurable sets by the third technique; extend it to positive measurable simple functions by linearity; extend it to positive measurable functions by density; and extend it again by linearity to general integrable functions.

B.2 Use All the Hypotheses

Real analysis is a well-developed area of mathematics, so most theorems are "sharp" in the sense that all the hypotheses are in fact required to prove the result. If you find that when you are working a problem you have not used all the assumptions, you may well be missing something.

B.3 Zorn's Lemma

A more advanced technique which may be useful is using Zorn's Lemma. Typically you have some partially ordered set (quite often some collection of subsets of some set ordered by inclusion), and you want to show that if a particular property P holds for every element of a totally ordered subset (often called a chain), then you can find an upper bound for this chain for which P also holds.

Then Zorn's Lemma tells you that there is a maximal element for which P holds. You often finish off the argument by showing that this maximal element is what you want.

Example B.3.1

Let V be a complex vector space with an inner product $\langle \cdot, \cdot \rangle : V \times V \to V$. A set $\{e_\alpha : \alpha \in I\} \subset V$ is orthonormal if

$$\langle e_\alpha, e_\beta \rangle = \begin{cases} 1 & \alpha = \beta \\ 0 & \alpha \neq \beta. \end{cases}$$

Then there is a maximal orthonormal set.

This follows by considering the collection $\mathcal{O}$ of all orthonormal sets in V. This is partially ordered by inclusion, and given any totally ordered subset $\mathcal{C}$ of $\mathcal{O}$, we have that

$$F = \bigcup_{E \in \mathcal{C}} E$$

is also orthonormal, because if we have any two elements $v, w \in F$, with $v \neq w$, then there is some $E \in \mathcal{C}$ which contains *both* of these elements, and so $\langle v, w \rangle = 0$. But F is an upper bound of $\mathcal{C}$.

So we have satisfied the hypotheses of Zorn's Lemma, and so there is at least one maximal orthonormal set (in fact, there are usually many). ◇

Zorn's lemma is equivalent to the axiom of choice, so unnecessary use if frowned upon. However can it make many arguments simpler, even when it may not strictly be required.

B.4 Counterexamples

Sometimes you may thing that some fact from an exercise is not true. In this case, having a good collection of "pathological" cases and common counterexamples can be useful.

When considering topologies, discrete, trivial and cofinite topologies are worth looking at; as are other topologies which don't separate points well. Dense sets, such as the rational numbers in the real numbers, and nowhere dense sets, such as the Cantor set, can also be useful.

For measures and σ-algebras, discrete, trivial and cofinite σ-algebras are useful; counting measure and Lebesgue measure are also useful; the unmeasurable set N discussed in the section on Lebesgue measure is particularly important; as are "thick" Cantor sets which are nowhere dense, but have nonzero measure.

When thinking of functions, keep in mind counterexamples like functions which are continuous everywhere but nowhere differentiable, or functions which are discontinuous everywhere. Simple functions involving some of the sets from above can also be fruitful sources of counter examples.

When considering convergence of functions, you should have examples of sequences of functions which converge in one sense, but fail to converge in another. This might include sequences of functions which converge pointwise, but not uniformly; pointwise or uniformly, but not in L^1. You should also have some examples of sequences of functions of a particular class which don't converge to something of the same class, such as a sequence of continuous functions which converges pointwise to something discontinuous.

B.5 Use a Little Category Theory

Category theory is an abstraction of the many situations where you are considering certain classes of sets (called "objects" in category theory), and the appropriate maps between them (called "arrows" or "morphisms" in category theory). For example there are a lot of similarities when considering:

objects	**arrows**
sets	functions
groups	group homomorphisms
rings	ring homomorphisms
vector spaces	linear maps
topological spaces	continuous functions
measurable spaces	measurable functions
metric spaces	continuous functions
metric spaces	Lipschitz functions
topological vector spaces	continuous linear maps
Banach spaces	bounded linear maps
Hilbert spaces	unitary maps

For example, in each of these cases, composition of arrows gives you another arrow, and the composition is associative; there is an "identity arrow" for each object which leaves the object alone and is an identity for composition. In addition in many of these examples you can take cross products of objects; some arrows are invertible, and two objects linked by an invertible arrow are effectively equivalent as far as category theory is concerned; and in some cases once can build objects out of collections of arrows.

Category theory distills all of these settings (and many more) into an abstract setting where you have a set of objects O, and a set of arrows, or morphisms, A, and for each arrow $f \in A$ there are two elements of O, the domain $\operatorname{dom} f$ and codomain $\operatorname{cod} f$ (this makes (O, A) a directed graph, for those readers keeping track of things). In keeping with the idea that f is modelling a function, we write $f : \operatorname{dom} f \to \operatorname{cod} f$. In addition there is a product on the arrows, where for any arrows f and g with $\operatorname{cod} f = \operatorname{dom} g$, there is an arrow $g \circ f \in A$ with $\operatorname{dom} g \circ f = \operatorname{dom} f$ and $\operatorname{cod} g \circ f = \operatorname{cod} g$. For each object $c \in A$ there is also an identity arrow id_c which has $\operatorname{dom} \mathrm{id}_c = \operatorname{cod} \mathrm{id}_c = c$ and $\mathrm{id}_c \circ f = f$ and $g \circ \mathrm{id}_c = g$ for all arrows f and g with $\operatorname{cod} f = c$ and $\operatorname{dom} g = c$.

We can define many common ideas in terms of these arrows and objects. An arrow $f : a \to b$ is invertible, if there is some other arrow $g : b \to a$ such that $f \circ g = \mathrm{id}_a$ and $g \circ f = \mathrm{id}_b$. If two objects a and b have an invertible map between them, then we say that they are isomorphic (ie. as far as the category can tell, they are identical).

An arrow $f : a \to b$ is monic if given any two arrows $g_1, g_2 : c \to a$, equality $f \circ g_1 = f \circ g_2$ implies $g_1 = g_2$, ie. iff f can be cancelled on the left hand side. This corresponds, in the case of actual sets and functions, to an injective (or 1-to-1) function.

Correspondingly, right cancallation of an arrow gives the analogue of surjectivity.

One can go on with this sort of "abstract nonsense" for quite some time.

The key idea to get from category theory from a practical viewpoint, is that there are natural things to consider depending on your context. If you are working with vector spaces, you probably don't want to condsider group homomorphisms (even though vector spaces are additive groups); if you are working with topological spaces, you really want continuous functions rather than arbitrary functions.

B.6 Pitfalls to Avoid

There are a number of common mis-steps and pitfalls that you need to be careful about.

Perhaps the most common is swapping the order of limits in an expression. This can be done in certain circumstances, but it requires careful justification. Similarly infinite sums, and integrals of functions involve limiting processes by definition, and so you have to be careful in swapping the order of sums, integrals and limits with one-another. Indeed, many of the key theorems of this course specify exactly when you can swap safely.

Another is assuming that something which is true for concrete, visualizable cases represents the "typical" situation. Indeed, such basic examples are usually the exception rather than the rule. For example when considering functions on the real line, most functions you can draw by hand are likely to be smooth almost everywhere; but the vast majority of continuous functions are nowhere smooth, and the vast majority of functions are not even continuous anywhere. Being comfortable with counterexamples can help avoid these problems.

Index