

measure

Aidan Backus

December 2019

Contents

1	Constructing the Lebesgue measure	5
1.1	Algebras of sets	5
1.2	The Borel σ -algebra	7
1.3	The definition of a measure	9
1.4	Premeasures and outer measures	13
1.5	The Carathéodory construction	17
1.6	Lebesgue measure on \mathbb{R}	19
2	More on measures	27
2.1	Measurable maps	27
2.2	Vector-valued measures	28
2.3	Probability	31
2.4	Ergodic systems	33
3	Measurable functions	37
3.1	Simple functions	37
3.2	Measurable functions	40
3.3	Characterizing measurable functions	42
3.4	Convergence of measurable functions	47
3.5	Regularity of measurable functions	52
3.6	Integration of simple functions	54
3.7	The integral in general	57
4	Properties of integration	63
4.1	Integrable functions	63
4.2	Indefinite integrals	66
4.3	Convergence theorems	69
4.4	Product measures	71
4.5	The Lebesgue integral	76
4.6	Changing the order of integration	78
5	Differentiation and regularity	83
5.1	Decomposition of signed measures	83
5.2	Differentiation of measures	83
5.3	Existence of Radon measures	84

5.4	Differentiation of vector-valued functions	84
5.5	Vitali's covering lemma	84
5.6	The maximal inequality	85
5.7	The Lebesgue differentiation theorem	85
5.8	The fundamental theorem of calculus	85
6	Hölder duality and L^p norms	87
6.1	L^p -norms	88
6.2	Convergence in L^p	95
6.3	Duality and representation	96
6.4	Compactness of the unit ball	99
6.5	Interpolation of operators	103
6.6	The mean ergodic theorem	103
7	Harmonic analysis	107
7.1	Haar measures	107
7.2	The Fourier transform	109
7.3	Pontryagin duality	111
7.4	The Fourier transform on \mathbb{R}^d	111
7.5	Applications to quantum mechanics	113
7.6	Applications to PDE	113
7.7	Applications to number theory	114
A	Linear algebra	119
A.1	Normed spaces	119
A.2	Banach spaces	121
A.3	Linear maps	125
A.4	Properties of Banach spaces	126
A.5	Convexity	126
B	Foundations	127
B.1	Axiomatic set theory	127
B.2	Universal properties	132
B.3	Point-set topology	134

Chapter 1

Constructing the Lebesgue measure

A measure is a rule by which we assign “size” to certain sets. The notion of a measure generalizes a handful of familiar notions, which we now review.

Example 1.0.1. If X is a set, we will define a measure known as the *counting measure* μ of X , by declaring that for every finite subset $Y \subseteq X$, $\mu(Y)$ is the cardinality of Y (i.e. the number of elements of Y).

We define the *Lebesgue measure*¹ μ on boxes $[a_1, b_1] \times \cdots \times [a_d, b_d] \subseteq \mathbb{R}^d$ by declaring that

$$\mu([a_1, b_1] \times \cdots \times [a_d, b_d]) = (b_1 - a_1)(b_2 - a_2) \cdots (b_d - a_d).$$

When $d = 1$ this is just the familiar notion of the length of a line segment; when $d = 2$ this is the area of a rectangle; when $d = 3$ this is the volume of a rectangular prism.

In probability theory, one is concerned with sets of “outcomes”; the sets are known as “events”. Given an event E in a sample space X , we define $\mu(E)$ to be the probability that μ occurs. Thus $\mu(X) = 1$.

1.0.2. In all of the above examples, μ has the property that if Y_1, \dots, Y_n is a collection of disjoint sets (thus $Y_i \cap Y_j = \emptyset$ whenever $i \neq j$), then

$$\mu \bigcup_i Y_i = \sum_i \mu(Y_i).$$

In fact this can be made to be true not just for finite, but countable, collections of disjoint sets. So it will be natural to require that the domain of our measure μ be closed under taking countable unions.

1.1 Algebras of sets

Encouraged by the introduction, we study sets R , whose elements are sets, such that R is closed under operations similar to countable union. This will seem obnoxious and abstract at first, but I ask for your patience; it won’t take long to get through this section.

¹Named after Henri Lebesgue (1875-1941), a French mathematician who introduced measure theory in his book *Intégrale, longueur, aire* (*Integral, length, area*).

Definition 1.1.1. By a *ring* R we mean a nonempty set, whose elements are sets, such that if $X, Y \in R$ then the union $X \cup Y$ and set difference $X \setminus Y$ are both in R .

1.1.2. Let R be a ring. One easily checks that if $X_1, \dots, X_n \in R$ then $X_1 \cup \dots \cup X_n \in R$ and $X_1 \cap \dots \cap X_n \in R$ (this follows by induction, because $X \cap Y = X \setminus (X \setminus Y)$). Moreover, since R is nonempty, say $X \in R$, $X \setminus X \in R$, so R contains the empty set.

Definition 1.1.3. Let X be a set. By an *algebra* in X we mean a ring such that $X \in R$ and for every $Y \in R$, $Y \subseteq X$.

Definition 1.1.4. By a σ -ring² R we mean a ring which is closed under countable unions, thus for every sequence $(X_n)_n$ of sets in R , $\bigcup_n X_n \in R$.

We define a σ -algebra in a set X to be a σ -ring which is an algebra in X .

1.1.5. Let R be a σ -ring. In proofs it will frequently be useful to replace sequences of sets in R with sequences of disjoint sets, without leaving R . The below lemma allows us to accomplish this.

Lemma 1.1.6. Let Σ be a σ -ring. Let $(E_n)_n$ be a sequence of sets in Σ . Then there is a sequence of disjoint sets $(F_n)_n$ in Σ such that $F_n \subseteq E_n$ and

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

Proof. Let $F_1 = E_1$ and for $n \geq 2$,

$$F_n = E_n \setminus \bigcup_{i=1}^{n-1} F_i.$$

Then clearly the F_n are disjoint and have the same union as the E_n . □

Exercise 1.1.7. Let X be a set. Show that the power set 2^X – that is, the set of all subsets of X – is a σ -algebra, as is $\{\emptyset, X\}$.

Exercise 1.1.8. Let R be the set of bounded subsets of \mathbb{R} . Show that R is a ring, but neither an algebra in \mathbb{R} nor a σ -ring.

Exercise 1.1.9. Recall the algebraic definition of a ring: a set equipped with an addition and a multiplication satisfying certain axioms. Show that every ring (in our sense) is an algebraic ring, where addition and multiplication are replaced by symmetric difference and intersection, respectively.

Exercise 1.1.10. Let Σ be an infinite σ -ring. Show that Σ has cardinality at least that of $2^{\mathbb{N}}$, the power set of the natural numbers. (Hint: let $\Sigma_0 = \Sigma$. Choose $X_n \in \Sigma_n$ appropriately and let $\Sigma_{n+1} = \{Y \in \Sigma_n : X \cap Y = \emptyset\}$. If done correctly, you should be able to find an injective function $F : 2^{\mathbb{N}} \rightarrow \Sigma$ such that $F(\{n\}) = X_n$).

²Here σ should be thought of as meaning “countable”.

1.2 The Borel σ -algebra

We now consider certain σ -algebras that are built up by a process similar to the notion of “generators and relations” from algebra.

Lemma 1.2.1. The intersection of a nonempty set of rings (resp. σ -rings, algebras, or σ -algebras) is a ring (resp. σ -ring, etc.)

Proof. We prove this for a set of rings (the other cases are similar). Let \mathcal{R} be a set of rings and let R be its intersection. If $X, Y \in R$ then for every $S \in \mathcal{R}$, $X, Y \in S$ so $X \cup Y \in S$ and $X \setminus Y \in S$. Therefore $X \cup Y, X \setminus Y \in R$. \square

1.2.2. Whenever we refer to a “smallest” set X with a property P , we mean that for every set Y with property P , $X \subseteq Y$.

Lemma 1.2.3. If \mathcal{C} is a set of sets, there is a smallest (σ) -ring containing \mathcal{C} . If in fact \mathcal{C} is a collection of subsets of a set X , there is a smallest (σ) -algebra containing \mathcal{C} .

Proof. In the case of a (σ) -ring, let X be the union of all elements of \mathcal{C} . We now prove this claim for the case of a ring; the other cases are similar.

The power set $2^X = \{Y : Y \subseteq X\}$ is clearly a ring which contains every element of \mathcal{C} . Let \mathcal{R} be the set of all rings that contain every element of \mathcal{C} ; since $2^X \in \mathcal{R}$, \mathcal{R} is nonempty. By Lemma 1.2.1, the intersection R of \mathcal{R} is a ring. But for every $C \in \mathcal{C}$, $C \in R$ since C is in every element of \mathcal{R} . Therefore R contains \mathcal{C} . \square

Definition 1.2.4. The smallest σ -algebra containing every element of a set \mathcal{C} is called the σ -algebra generated by \mathcal{C} . We denote it by $\sigma(\mathcal{C})$.

1.2.5. The advantage of defining a σ -algebra by referring to its generators is that it is easy to show that every element of the σ -algebra has a given property. This is quite easy to prove, as the following lemma shows, but so very useful.

Lemma 1.2.6. Let X be a set, and let P be a property that subsets of X can have. Assume that \mathcal{C} is a set whose elements are subsets of X , and every element of \mathcal{C} has property P . If the set of all subsets of X with property P is a σ -algebra, then every element of $\sigma(\mathcal{C})$ has property P .

Proof. Let Σ be the σ -algebra of all subsets of X with property P . Then Σ contains $\sigma(\mathcal{C})$. \square

1.2.7. We now come across the most important example of a σ -algebra. If the reader is unfamiliar with topological spaces, in the below definition it suffices to take $X = \mathbb{R}^d$, and \mathcal{T} the set of all unions of open balls in \mathbb{R}^d , as this will be the most important case.

Definition 1.2.8. Let X be a topological space. Let \mathcal{T} be the topology of X , the set of all open subsets of X . We define the *Borel σ -algebra*³ \mathcal{B} of X by $\mathcal{B} = \sigma(\mathcal{T})$. A *Borel set* in X is an element of \mathcal{B} .

³Named after Émile Borel (1871-1956), a French probabilist who introduced many of the key ideas of measure theory.

1.2.9. Let X be a topological space. It follows from the definition that every open or closed subset of X is Borel, but also that countable unions of complements of countable unions of complements of countable unions of complements of ... of open sets of X is Borel. However, some Borel sets can be even more complicated than that.

Lemma 1.2.10. Let X be a topological space. Assume that P is a property that subsets of X can have, and that every open set has property P . If the set of subsets of X with property P is a σ -algebra, then every Borel set has property P .

Proof. By Lemma 1.2.6. □

1.2.11. When is a set Borel? If $X = \mathbb{N}$, then every set is open and so every set is Borel. But it is not so obvious how to check whether a subset of \mathbb{R}^d is Borel. Certainly any set you will ever “naturally” encounter is Borel, but not every subset of \mathbb{R}^d is Borel.

Lemma 1.2.12. There exists a subset of \mathbb{R}^d which is not Borel. In fact, the set of Borel subsets of \mathbb{R}^d has strictly lower cardinality than the set of all subsets of \mathbb{R}^d .

Proof (omit on first reading). Let $\Sigma_1 = \mathcal{T}$ be the topology of \mathbb{R}^d . Given Σ_α , α a countable ordinal (see Definition B.1.15), let Π_α be the set of all complements of elements of Σ_α . Let $\Sigma_{\alpha+1}$ be the set of countable unions of elements of Π_α . If $\beta < \omega_1$ is not equal to $\alpha + 1$ for any α , let $\Sigma_\beta = \bigcup_{\alpha < \beta} \Sigma_\alpha$. Therefore Σ_α and Π_α are defined for all $\alpha < \omega_1$ by transfinite recursion (see the remarks after Theorem B.1.16). We note that Σ_0 has cardinality \beth_1 by Theorem B.1.24.

Now the mapping $A \mapsto A^c$ is a bijection $\Sigma_\alpha \rightarrow \Pi_\alpha$. Moreover if Π_α has cardinality \beth_1 (see Definition B.1.22) then so does $\Sigma_{\alpha+1}$, since each element of $\Sigma_{\alpha+1}$ can be expressed in terms of a countable number of elements of Π_α , and $\beth_1 \times \beth_1 \times \cdots$ has cardinality \beth_1 by Theorem B.1.23. Finally if β is a countable limit ordinal and for every $\alpha < \beta$, Σ_α is countable, then Σ_β is a countable union of sets of cardinality \beth_1 , hence has cardinality \beth_1 by Theorem B.1.23. It follows by induction that for every countable ordinal α , Σ_α has cardinality \beth_1 .

Now every Borel set is in Σ_α for some α , since it was obtained by applying countable union and complement countably many times to an open set, and thus is in $\Sigma_{\omega_1} = \bigcup_{\alpha < \omega_1} \Sigma_\alpha$. But Σ_{ω_1} is a union of \aleph_1 many sets of cardinality \beth_1 , so has cardinality \beth_1 by Theorem B.1.23, since $\aleph_1 \leq \beth_1$ (which follows by Zermelo’s well-ordering theorem B.1.20). Therefore Σ_{ω_1} is the set of all Borel sets, and has cardinality \beth_1 .

But Σ_{ω_1} is a subset of the power set $2^{\mathbb{R}}$, which has cardinality \beth_2 . Thus there is an element of $2^{\mathbb{R}}$ which is not Borel. □

Exercise 1.2.13. Show that the set of half-open intervals $[a, b)$ generates the Borel σ -algebra of \mathbb{R} .

Exercise 1.2.14. Show that the set of compact subsets of \mathbb{R}^d generates the Borel σ -algebra of \mathbb{R}^d . Show that there exists a metric space X , such that the set of compact subsets of X does not generate the Borel σ -algebra of X . (Hint: Cardinality!)

Exercise 1.2.15. Let us say that a bounded subset A of \mathbb{R} has *Jordan content* if the indicator function 1_A of A is Riemann integrable. Here $1_A(x) = 1$ if $x \in A$ and $1_A(x) = 0$ otherwise. Show that \mathbb{Q} is Borel, but does not have Jordan content. Later we will show that every set with Jordan content is Borel, and that it is reasonable to define the length of any Borel set. So our definitions will generalize those that one learned in a first course in real analysis.

1.3 The definition of a measure

We are almost ready to define a measure. Throughout this section, the reader should refer back to Example 1.0.1 frequently, as the point of the definitions in this section is to generalize those concepts.

1.3.1. Let $(-\infty, \infty]$ denote the set of real numbers, plus another point ∞ which is greater than any real number. We define addition on $[0, \infty]$ by requiring that $\infty + a = \infty$ for any $a \in \mathbb{R}$. We do not define addition on $[-\infty, \infty]$, which would include $-\infty$, because the expression $\infty - \infty$ makes no sense.

Definition 1.3.2. Let \mathcal{C} be a collection of sets. A function $\mu : \mathcal{C} \rightarrow [0, \infty]$ such that for every disjoint sequence of sets $(X_n)_n$ in Σ such that $\bigcup_n X_n \in \mathcal{C}$,

$$\mu \bigcup_n X_n = \sum_{n=1}^{\infty} \mu(X_n),$$

is called a σ -additive function.

Definition 1.3.3. A *measure* is a σ -additive function defined on a σ -algebra Σ on a set X which is not identically ∞ . If we wish to emphasize that the codomain of a measure is $[0, \infty]$, we will call it a *positive measure*.

We call elements of Σ *measurable sets* and call (X, Σ) a *measurable space*. If μ is a measure on Σ , we call (X, Σ, μ) a *measured space*.

If the image of μ is $[0, 1]$, we say that μ is a *probability measure*.

Lemma 1.3.4. For any measure μ , $\mu(\emptyset) = 0$.

Proof. Since μ is a measure, there is a measurable set Y such that $\mu(Y) \neq \infty$. Then $Y = Y \cup \emptyset$ and $Y \cap \emptyset = \emptyset$, so

$$\mu(Y) = \mu(Y) + \mu(\emptyset).$$

Therefore $\mu(\emptyset) = 0$. □

Example 1.3.5. The examples in Example 1.0.1 are σ -additive. However, they are not all defined on σ -algebras; for example, the union of two boxes $\prod_i [a_i, b_i]$ is not a box. Counting measure is defined on the σ -algebra of every subset of X , so counting measure is actually a measure.

In fact, it is not yet clear that there are any interesting measures other than counting measure! We'll have to do a lot of work before we'll be ready to introduce other examples of measures.

Theorem 1.3.6. TODO: Continuity of measure

1.3.7. We now introduce an important class of σ -additive functions defined for certain subsets of \mathbb{R} , known as Stieltjes premeasures⁴. Eventually we will modify their definition so that they are defined for every Borel subset of \mathbb{R} . Since the Borel sets form a σ -algebra, this will define a measure on the Borel sets, known as a Stieltjes measure.

⁴Named after Thomas Stieltjes (1856-1894), who introduced the Riemann-Stieltjes integral, a weighted version of the Riemann integral that inspired the notion of a Stieltjes premeasure.

1.3.8. Recall that a function f on an interval I is said to be “left-continuous” if for every $x \in I$,

$$f(x) = \lim_{\varepsilon \rightarrow 0^+} f(x - \varepsilon).$$

Here the limit is taken over *positive* ε , and so ignores the behavior of f to the right of x . Clearly any continuous function is left-continuous. One can also define right-continuous functions by ignoring the behavior of f to the left of x instead.

Example 1.3.9. The *Heaviside step function*⁵

$$H(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases}$$

is a useful example of a left-continuous function which is not continuous. Clearly the derivative $H'(x)$ exists if $x \neq 0$, and in that case $H'(x) = 0$. We want to say that $H'(0) = \infty$ in some suitable sense, and in fact that for any $\varepsilon > 0$,

$$\int_{-\infty}^{\infty} H'(x) dx = \int_{-\varepsilon}^{\varepsilon} H'(x) dx = 1.$$

Of course we can't do that, because the limit that would define $H'(0)$ does not exist. TODO: Draw a picture.

Definition 1.3.10. Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a nondecreasing, left-continuous function. Define

$$\mu_f([a, b)) = f(b) - f(a).$$

We call μ_f the *Stieltjes premeasure* of f .

1.3.11. A nondecreasing function can only be discontinuous on a countable set (Exercise 1.3.16). Thus we can turn a nondecreasing function f into a left-continuous function g by declaring that if f is continuous at x then $g(x) = f(x)$, and otherwise setting

$$g(x) = \lim_{\varepsilon \rightarrow 0} f(x - \varepsilon).$$

Henceforth we will talk about Stieltjes premeasures of any nondecreasing function, knowing that we may have to redefine them on a countable set in order that the definition make sense.

Example 1.3.12. If $f(x) = x$, then the Stieltjes premeasure of an interval $[a, b)$ under f is just its length $b - a$. This premeasure is known as the *Lebesgue premeasure*. It will be by far the most important premeasure that we study. If the following theory ever seems too abstract, try it out on the Lebesgue premeasure!

⁵Named after Oliver Heaviside (1850-1925), an engineer who introduced an early version of the “distributional calculus” used to solve differential equations. Mathematicians did not consider the distributional calculus rigorous until the 1950s, however.

Example 1.3.13. The Stieltjes premeasure of a differentiable function f can be thought of as “weighted length”; $\mu_f([a, b)) > b - a$ provided that $f' > 1$ on $b - a$, and $\mu_f([a, b)) < b - a$ if $f' < 1$. This is just an expression of the fundamental theorem of calculus: if f is differentiable then

$$\mu_f([a, b)) = f(b) - f(a) = \int_a^b f'(x) \, dx.$$

Since f is nondecreasing, $f' \geq 0$. This is our first clue that there is some connection between integration and measure theory.

Example 1.3.14. The Stieltjes premeasure of the Heaviside function H will allow us to make some sense of our previous waffling about its derivative. If $0 \notin [a, b)$ then $\mu_H([a, b)) = 0$. Otherwise, $a \leq 0 < b$ and $H(b) = 1$, $H(a) = 0$; thus $\mu_H([a, b)) = 1$. If one instead considers a finite set $X = \{x_1, \dots, x_n\}$ and

$$f(x) = \sum_{j=1}^n H(x - x_j),$$

then $\mu_f([a, b))$ is the cardinality of $X \cap [a, b)$.

Theorem 1.3.15. Every Stieltjes premeasure is σ -additive. In particular, Lebesgue premeasure is σ -additive.

Proof. Let μ_f be a Stieltjes premeasure. Let $E_n = [a_n, b_n)$, assume that the E_n are disjoint, and let $E = \bigcup_n E_n$. Suppose that $E = [a, b)$. We must show

$$f(b) - f(a) = \sum_{n=1}^{\infty} f(b_n) - f(a_n). \quad (1.1)$$

To do so, we first note that since f is nondecreasing, the quantities $\mu_f([a_n, b_n))$ are positive, so the sum in (1.1) converges absolutely. Thus we may rearrange the order of the summands without affecting the value of the sum, so we can assume that $a_n \leq a_{n+1}$ for every n , by reordering the intervals E_n . Since the intervals are disjoint it follows that $b_n \leq a_{n+1}$.

We now prove

$$\mu_f(E) \geq \sum_{n=1}^{\infty} \mu_f(E_n).$$

To do this, we fix an N and show that

$$f(b) - f(a) \geq \sum_{n=1}^N f(b_n) - f(a_n). \quad (1.2)$$

Now $b \geq b_N$ and $a \leq a_1$, so $f(b) - f(a) \geq f(b_N) - f(a_1)$, but

$$\sum_{n=1}^N f(b_n) - f(a_n) = f(b_N) - f(a_1) + \sum_{n=1}^{N-1} f(b_n) - f(a_{n+1}).$$

But $b_n \leq a_{n+1}$ so $f(b_n) \leq f(a_{n+1})$, so

$$\sum_{n=1}^{N-1} f(b_n) - f(a_{n+1}) \leq 0.$$

Therefore

$$\sum_{n=1}^N f(b_n) - f(a_n) \leq f(b_N) - f(a_1) \leq f(b) - f(a).$$

This proves (1.2).

Conversely, we must show that

$$\mu_f(E) \leq \sum_{n=1}^{\infty} \mu_f(E_n).$$

It suffices to show that for every $\varepsilon > 0$,

$$f(b) - f(a) \leq \varepsilon + \sum_{n=1}^{\infty} f(b_n) - f(a_n). \quad (1.3)$$

Now choose $b' < b$ such that $f(b') \geq f(b) - \varepsilon/2$ and for each n choose $a'_n < a_n$ such that $f(a'_n) \geq f(a_n) - \varepsilon/2^{n+1}$. Such a'_n and b' exist because f is left-continuous. Now

$$[a'_n, b'] \subseteq [a, b] = \bigcup_{n=1}^{\infty} [a_n, b_n] \subseteq \bigcup_{n=1}^{\infty} (a'_n, b_n).$$

Therefore the (a'_n, b_n) are an open cover of $[a'_n, b]$, so by the Heine-Borel theorem there is an N such that

$$[a'_n, b'] \subseteq \bigcup_{n=1}^N (a'_n, b_n).$$

If any interval is superfluous, we now discard it. Then the way we ordered the intervals, $a'_{n+1} \leq b_n$. TODO: Draw a picture. Moreover, $a'_1 \leq a$ and $b' \leq b_N$. Then

$$\begin{aligned} f(b) - f(a) &\leq f(b') - f(a) + \frac{\varepsilon}{2} \\ &\leq f(b_N) - f(a'_1) + \frac{\varepsilon}{2} \\ &\leq f(b_N) - f(a'_1) + \frac{\varepsilon}{2} + \sum_{n=1}^{N-1} f(b_n) - f(a'_{n+1}) \\ &\leq \frac{\varepsilon}{2} + \sum_{n=1}^N f(b_n) - f(a'_n) \\ &\leq \frac{\varepsilon}{2} + \sum_{n=1}^N f(b_n) - f(a_n) + \frac{\varepsilon}{2^{n+1}} \\ &\leq \varepsilon + \sum_{n=1}^N f(b_n) - f(a_n) \end{aligned}$$

where we used Zeno's paradox

$$\sum_{n=1}^{\infty} \frac{1}{2^n} = 1 \quad (1.4)$$

to sum the geometric series of ε 's. But this estimate is exactly (1.3). \square

Exercise 1.3.16. Show that a nondecreasing function on \mathbb{R} can only be discontinuous on a countable set.

Exercise 1.3.17. Let Σ be a σ -ring, and let $\mu : \Sigma \rightarrow B$ be a function such that whenever $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$,

$$\lim_{n \rightarrow \infty} \mu(A_n) = \mu \left(\bigcup_{n=1}^{\infty} A_n \right).$$

Show that if μ is *additive* (that is, if $A \cap B = \emptyset$, then $\mu(A \cup B) = \mu(A) + \mu(B)$), then μ is σ -additive.

Exercise 1.3.18. Let \mathcal{F} be the vector space of left-continuous, nondecreasing functions and \mathcal{M} the vector space of Stieltjes premeasures. Show that the map $f \mapsto \mu_f$ which assigns a function to its Stieltjes premeasure is a linear map $\mathcal{F} \rightarrow \mathcal{M}$.

1.4 Premeasures and outer measures

Encouraged by the previous section, we now define premeasures in general. Our goal is to define a σ -additive function that can be extended to a measure in a unique way; thus, to define a measure, it will suffice to define a premeasure.

Definition 1.4.1. A set of sets P is said to be a *semiring* if $\emptyset \in P$ and for every $E, F \in P$:

1. $E \cap F \in P$.
2. There are G_1, \dots, G_m disjoint such that

$$E \setminus F = \bigcup_{n=1}^m G_n.$$

A σ -additive function $P \rightarrow [0, \infty]$ which is not identically ∞ is called a *premeasure*.

Example 1.4.2. The set of all half-open intervals $[a, b)$ is a semiring, so a Stieltjes premeasure is a premeasure.

Lemma 1.4.3. Let P be a semiring and μ a premeasure on P . Then:

1. If $E_1, \dots, E_m \in P$ then there are disjoint $F_1, \dots, F_n \in P$ such that

$$(((E_1 \setminus E_2) \setminus E_3) \setminus \cdots) \setminus E_m = \bigcup_{i=1}^n F_i.$$

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

Proof. Exercise 1.4.20. □

Lemma 1.4.4. Let P be a semiring and μ a premeasure on P . Then:

1. If $(E_n)_n$ is a sequence of disjoint sets in P and $E \supseteq \bigcup_n E_n$, $E \in P$, then

$$\sum_{n=1}^{\infty} \mu(E_n) \leq \mu(E).$$

2. If $E \subseteq F$, $E, F \in P$, then $\mu(E) \leq \mu(F)$.

Proof. We first prove the first claim. By Lemma 1.4.3, there are $F_i \in P$ disjoint such that

$$(((E_1 \setminus E_2) \setminus E_3) \setminus \cdots) \setminus E_m = \bigcup_{i=1}^n F_i.$$

In particular, E is the disjoint union of the E_i and F_i . Thus

$$\mu(E) = \sum_i \mu(E_i) + \sum_j \mu(F_j).$$

But μ is nonnegative so $\sum_i \mu(E_i) \geq 0$, thus the claim.

The second claim follows from the first in the case $n = 1$, $E_1 = F$. □

1.4.5. The power of Lemma 1.4.4 is that it does not assume that $\bigcup E_n \in P$.

1.4.6. With these basic properties of premeasures set aside, we now discuss how to extend a premeasure to a measure. First, we note that while premeasures are σ -additive, they also have another useful property, called σ -subadditivity.

Definition 1.4.7. Let P be a set of sets. A function $\mu : P \rightarrow (-\infty, \infty]$ such that for every sequence of $E_i \in P$, if $E \in P$ and $E \subseteq \bigcup_i E_i$, then

$$\mu(E) \leq \sum_{i=1}^{\infty} \mu(E_i),$$

is said to be a σ -subadditive function.

Lemma 1.4.8. Every premeasure is σ -subadditive.

Proof. Suppose that $E = \bigcup_i E_i$. By Lemma 1.4.3, we can write

$$(((E_i \setminus E_{i-1}) \setminus E_{i-2}) \setminus \cdots \setminus E_1) = \bigcup_{j=1}^{k_i} F_i^j$$

where the F_i^j are disjoint, hence

$$\mu(E) = \mu(E_1) + \sum_{i=2}^{\infty} \sum_{j=1}^{k_i} \mu(F_i^j).$$

But $\bigcup_j F_i^j \subseteq E_i$ so by Lemma 1.4.4,

$$\sum_{j=1}^{k_i} \mu(F_i^j) \leq \mu(E_i).$$

Thus

$$\mu(E) \leq \sum_{i=1}^{\infty} \mu(E_i)$$

which was to be shown. \square

1.4.9. Because premeasures are σ -subadditive, it would be natural to extend them to a σ -subadditive function defined on a σ -ring. We do this now.

Definition 1.4.10. Let R be a σ -ring. An *outer measure* is a σ -subadditive function $\mu : R \rightarrow [0, \infty]$ such that $\mu(\emptyset) = 0$.

1.4.11. It follows from the definition that an outer measure μ is monotone; that is, if $E \subseteq F$, $E, F \in P$, then $\mu(E) \leq \mu(F)$.

1.4.12. Let μ be a prmeasure. We will now construct a σ -ring R such that μ extends to an outer measure on R .

Definition 1.4.13. A set \mathcal{H} whose elements are sets is said to be *hereditary* if for every $E \in \mathcal{H}$ and $F \subseteq E$, $F \in \mathcal{H}$.

Lemma 1.4.14. For every semiring P there is a smallest hereditary σ -ring $\mathcal{H}(P)$ such that $P \subseteq \mathcal{H}(P)$.

Proof. Let X be the union of all elements in P . Then the power set 2^X is clearly a hereditary σ -ring such that $P \subseteq 2^X$. The intersection of σ -rings containing P is a σ -ring containing P , so it suffices to show that the intersection of hereditary sets is hereditary.

In fact, let H be a set whose elements are hereditary sets, and let \mathcal{H} be the intersection of H . Let $E \in \mathcal{H}$; then for every $F \subseteq E$, F is contained in every element of H , so $F \in \mathcal{H}$. \square

Definition 1.4.15. Let P be a semiring and μ be a premeasure on P . For every $A \in \mathcal{H}(P)$, define

$$\mu^*(A) = \inf \sum_{n=1}^{\infty} \mu(E_n)$$

where the inf ranges over all sequences of sets $(E_n)_n$, $E_n \in P$, such that $A \subseteq \bigcup_n E_n$. We call μ^* the *outer measure generated by μ* .

Theorem 1.4.16. Let P be a semiring and μ be a premeasure on P . Then the outer measure μ^* generated by μ is an outer measure on $\mathcal{H}(P)$ and for every $E \in P$, $\mu^*(E) = \mu(E)$.

Proof. First, we check $\mu^*(\emptyset) = 0$. In fact, taking $E_n = \emptyset$ for every n , we have $\mu^*(\emptyset) \leq \sum_n 0 = 0$.

We now check that μ^* is σ -subadditive:

Lemma 1.4.17. Suppose that $A \in \mathcal{H}(P)$, $A_n \in \mathcal{H}(P)$, and $A \subseteq \bigcup_n A_n$. Then

$$\mu^*(A) \leq \sum_{n=1}^{\infty} \mu^*(A_n).$$

Proof of lemma. Obviously this is true if some $\mu^*(A_n) = \infty$, so suppose that for every n , $\mu^*(A_n) < \infty$ and let $\varepsilon > 0$. By definition of μ^* , there are $E_i^j \in P$ such that $A_i \subseteq \bigcup_j E_i^j$ and

$$\mu^*(A_i) \geq \sum_{j=1}^{\infty} \mu(E_i^j) - \frac{\varepsilon}{2^i}.$$

Thus $A \subseteq \bigcup_{i,j} E_i^j$ whence

$$\begin{aligned} \mu^*(A) &\leq \sum_{i,j} \mu(E_i^j) \leq \sum_{i,j=1}^{\infty} \mu(E_i^j) \\ &\leq \sum_{i=1}^{\infty} \mu^*(A_i) + \frac{\varepsilon}{2^i} \\ &\leq \varepsilon + \sum_{i=1}^{\infty} \mu^*(A_i). \end{aligned}$$

This was to be shown. □

Finally we check that $\mu^*(A) = \mu(A)$ when $A \in P$. Clearly $\mu^*(A) \leq \mu(A)$. Since μ is σ -subadditive, for any $E_i \in P$ such that $A \subseteq \bigcup_i E_i$, $\mu(A) \leq \sum_i \mu(E_i)$, thus $\mu(A) \leq \mu^*(A)$. □

1.4.18. There is a dual approach to the extension of premeasures, introduced by Lebesgue. He considered not just outer measures but *inner measures* defined by the relation

$$\mu_*(E) = \sup \sum_{n=1}^{\infty} \mu(E_n)$$

where the sup ranges over all sequences of $E_n \in P$ such that $\bigcup_n E_n \subseteq E$ and the E_n are disjoint. Then Lebesgue proposed to study the σ -ring of all sets whose inner and outer measures agree. Note the asymmetry: for inner measure we need to assume that the E_n are disjoint, or else we could “double-count” elements of E . This asymmetry is the origin of several pathologies that make inner measures difficult to work with, and now this approach is considered nothing more than a historical footnote.

Example 1.4.19. An example of an oddity of inner measure comes from trying to compute the inner measure of the set X of irrational numbers in $[0, 1]$. Let P be the semiring of intervals with rational endpoints in $[0, 1]$ and let μ be the Stieltjes premeasure on P defined by $\mu([a, b)) = b - a$. Then $\mu^*(X) = 1$.

To see this, let $x \in [0, 1]$; we will compute $\mu^*(\{x\})$. Let $[x]_n$ be a rational number such that $[x]_n < x < [x]_n + 1/n$; thus $\{x\} \subset [[x]_n, [x]_n + 1/n)$ and so

$$\mu^*(\{x\}) \leq \mu([x]_n, [x]_n + 1/n) = \frac{1}{n}$$

whence $\mu^*(\{x\}) = 0$.

Let $(x_n)_n$ be an enumeration of the countable set $\mathbb{Q} \cap [0, 1]$; by σ -subadditivity,

$$\mu^*(\mathbb{Q} \cap [0, 1]) \leq \sum_{n=1}^{\infty} \mu^*(\{x_n\}) = 0$$

but $X \cup (\mathbb{Q} \cap [0, 1]) = [0, 1]$, so

$$1 = \mu([0, 1]) \leq \mu^*(X) + \mu^*(\mathbb{Q} \cap [0, 1]) = \mu^*(X) + 0.$$

Therefore $\mu^*(X) \geq 1$, but $X \subseteq [0, 1]$ so $\mu^*(X) \leq 1$.

But there are no intervals in P which are contained in X ; thus the only element of P contained in E is \emptyset , so

$$\mu_*(E) = \sup 0 = 0.$$

Thus this seemingly reasonable way of defining an inner measure fails to measure the set of irrational numbers.

Exercise 1.4.20. Prove Lemma 1.4.3.

1.5 The Carathéodory construction

Carathéodory introduced a modern approach that we now consider which avoids the issues with inner measures. The idea is that, while an outer measure may not be a measure, there is a canonically defined σ -ring on which the outer measure will restrict to a measure. The elements of that ring will be called measurable sets.

1.5.1. The Carathéodory construction gives us a method to define measures: first define a premeasure, then check a certain technical hypothesis that we now state; then one has a unique measure which extends the premeasure.

Definition 1.5.2. Let (X, Σ, μ) be a measured space. We say that μ is a *σ -finite measure* if there are countably many $E_n \in \Sigma$ such that $\bigcup_n E_n = X$ and $\mu(E_n) < \infty$. We say that μ is a *complete measure* if for every $E \in \Sigma$ such that $\mu(E) = 0$ and $F \subseteq E$, $F \in \Sigma$.

Theorem 1.5.3 (Carathéodory construction). Let P a semiring of subsets of a set X such that $X \in P$. Let μ be a premeasure on P . Then there is a σ -algebra Σ which contains P and an extension of μ to a complete measure on Σ . Moreover, if μ is σ -finite, then (X, Σ, μ) is the unique measured space with this property.

1.5.4. The rest of this section is devoted to the proof of the Carathéodory construction theorem. We begin by constructing Σ . We will do this by giving a property – namely, “measurability” – that every set in the σ -algebra Σ ought to satisfy. We will then show that the set of all measurable sets forms a σ -algebra; thus, we might as well define Σ to be that σ -algebra.

1.5.5. Let us fix an outer measure μ^* be an outer measure on a hereditary σ -ring \mathcal{H} . In practice, μ^* will be the extension of a premeasure and \mathcal{H} will be the σ -algebra of all subsets of some set.

Definition 1.5.6. A μ^* -measurable set is a set $A \in \mathcal{H}$ such that for every $E \in \mathcal{H}$,

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \setminus A). \quad (1.5)$$

We let $\mathcal{M}(\mu^*)$ denote the set of all μ^* -measurable sets. In the event that (1.5) holds, we say that A *cleanly divides* E , so a set A is μ^* -measurable if for every set $E \in \mathcal{H}$, A cleanly divides E .

1.5.7. Since an outer measure is subadditive, one already has

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

and so one just has to prove the opposite inequality

$$\mu^*(E \cap A) + \mu^*(E \setminus A) \leq \mu^*(E) \quad (1.6)$$

to check that a set is μ^* -measurable.

Definition 1.5.8. We say that a set $A \in \mathcal{H}$ is μ^* -null if $\mu^*(A) = 0$. We let $\mathcal{N}(\mu^*)$ denote the set of all μ^* -null sets.

1.5.9. Clearly \emptyset is always μ^* -null, and any subset of a μ^* -null set is μ^* -null, thus $\mathcal{N}(\mu^*)$ is hereditary. In general \emptyset may be the only μ^* -null set (think of counting measure), but as we will see, Stieltjes measures have lots of null sets. We now show that every μ^* -null set is measurable, and have no effect on measurability. In fact, μ^* -null sets will never matter much.

Lemma 1.5.10. The set $\mathcal{N}(\mu^*)$ of μ^* -null sets is a σ -ring, and $\mathcal{N}(\mu^*) \subseteq \mathcal{M}(\mu^*)$. Moreover, if A is μ^* -measurable and Z is μ^* -null, then $A \cup Z$ and $A \setminus Z$ are μ^* -measurable.

Proof. Let Z be μ^* -null. Then for every E ,

$$\mu^*(E \cap Z) + \mu^*(E \setminus Z) = 0 + \mu^*(E \setminus Z) \leq \mu^*(E)$$

since $\mathcal{N}(\mu^*)$ is hereditary. Thus Z is μ^* -measurable, so $\mathcal{N}(\mu^*) \subseteq \mathcal{M}(\mu^*)$. If A cleanly divides E then clearly so do $A \cup Z$ and $A \setminus Z$, so null sets have no effect on measurability. Countable subadditivity implies that $\mathcal{N}(\mu^*)$ is a σ -ring. \square

Lemma 1.5.11. If A_1, A_2 are μ^* -measurable sets, then so is $A_1 \setminus A_2$.

Proof. Since E is the disjoint union of the sets $E \cap (A_1 \cap A_2)$, $E \cap (A_1 \setminus A_2)$, $E \cap (A_2 \setminus A_1)$, and $E \cap (A_1 \cup A_2)$, this follows from the fact that A_1 and A_2 divide each of the above sets cleanly. We leave the details as Exercise 1.5.13. \square

Lemma 1.5.12. For every outer measure μ^* , the set of measurable sets $\mathcal{M}(\mu^*)$ is a σ -ring and μ^* restricts to a measure on $\mathcal{M}(\mu^*)$.

Proof. \square

TODO: Show that P has measurable sets

TODO: Completions of measures

TODO: Uniqueness of extensions

Leave lots of the above as exercises.

Exercise 1.5.13. Fill in the details in the proof of Lemma 1.5.11.

Exercise 1.5.14. Let (X, Σ, μ) be a measured space. Show that the following are equivalent:

1. μ is σ -finite.
2. There are countably many disjoint $E_i \in \Sigma$, $\mu(E_i) < \infty$, such that $\bigcup_i E_i = X$.
3. There are countably many $E_i \in \Sigma$, $\mu(E_i) < \infty$, such that $E_i \subseteq E_{i+1}$ and $\bigcup_i E_i = X$.

These equivalences are highly useful and will be used throughout the text without explicit mention.

Exercise 1.5.15. Let (X, Σ, μ) be a measured space, and let \mathcal{N} be the σ -algebra of all null subsets of μ . Let $\bar{\Sigma}$ be the σ -algebra generated by Σ and \mathcal{N} . Show that there is a unique extension $\bar{\mu}$ of μ to $(X, \bar{\Sigma})$, and that $(X, \bar{\Sigma}, \bar{\mu})$ is a complete measured space. We call $(X, \bar{\Sigma}, \bar{\mu})$ the *completion* of (X, Σ, μ) .

1.6 Lebesgue measure on \mathbb{R}

In this section we define the most important measure of all: Lebesgue measure on \mathbb{R} .

1.6.1. We begin by recalling that given a nondecreasing left-continuous function f on \mathbb{R} , the Stieltjes premeasure μ_f was given by

$$\mu_f([a, b)) = f(b) - f(a)$$

whenever $a < b$. We can extend the semiring P of half-open intervals to include \mathbb{R} by setting

$$\mu_f(\mathbb{R}) = \lim_{n \rightarrow \infty} f(n) - f(-n),$$

which may be infinite, but always makes sense, since the only way that $\mu_f(\mathbb{R})$ could be $\infty - \infty$ is if

$$\lim_{x \rightarrow -\infty} f(x) = +\infty$$

which is impossible since f is increasing. Since the Stieltjes premeasure is a premeasure, it follows from the Carathéodory construction that there is a σ -algebra Σ containing P (and therefore containing the Borel σ -algebra $\mathcal{B}(\mathbb{R})$) and a measure, which we also call μ_f , on Σ which extends μ_f .

Lemma 1.6.2. Every Stieltjes measure is σ -finite.

Proof. \mathbb{R} is covered by sets of the form $[n, n+1)$, $n \in \mathbb{Z}$, of which there are countably many. Now $\mu_f([n, n+1)) = f(n+1) - f(n)$ is finite, so \mathbb{R} is σ -finite. \square

1.6.3. By the Carathéodory construction theorem, if a premeasure extends to a σ -finite complete measure μ , then μ is the unique such extension. It follows from the previous lemma that μ_f is well-defined. More precisely, μ_f is the unique complete Borel measure obtained by extending the Stieltjes premeasure associated to f .

Definition 1.6.4. The *Stieltjes measure* arising from a nondecreasing left-continuous function f is the complete Borel measure μ_f obtained from the Stieltjes premeasure of f .

Lemma 1.6.5. If μ_f is a Stieltjes measure then for every open interval (α, β) one has

$$\mu_f((\alpha, \beta)) = f(\beta) - \lim_{n \rightarrow \infty} f(\alpha + 1/n).$$

In particular, if f is continuous then $\mu_f((\alpha, \beta)) = f(\beta) - f(\alpha)$.

Proof. By continuity of measure,

$$\mu_f((\alpha, \beta)) = \lim_{n \rightarrow \infty} \mu_f([\alpha + 1/n, \beta)) = \lim_{n \rightarrow \infty} f(\beta) - f(\alpha + 1/n)$$

which is what we wanted. \square

1.6.6. In particular, if $f(x) = x$, then $\mu_f((\alpha, \beta)) = \beta - \alpha$. This fact will be used so often that we will frequently use it without reference.

Definition 1.6.7. The *Lebesgue measure* on \mathbb{R} is the Stieltjes measure arising from the Lebesgue premeasure; that is, the Lebesgue measure is μ_f where $f(x) = x$.

1.6.8. Let μ denote Lebesgue measure; then $\mu([\alpha, \beta)) = \beta - \alpha$ is the length of the interval $[\alpha, \beta)$. Therefore μ is the natural generalization of “length” to as many subsets of \mathbb{R} as one reasonably can generalize it to. Indeed, if A is a Lebesgue measurable subset of \mathbb{R} (e.g. if A is Borel) then

$$\mu(A) = \inf \left\{ \sum_{j=1}^{\infty} \beta_j - \alpha_j : A \subseteq \bigcup_{i=1}^{\infty} [\alpha_i, \beta_i) \right\}. \quad (1.7)$$

Theorem 1.6.9. Lebesgue measure μ is *translation-invariant* in the sense that if A is a Lebesgue measurable set, and $A + x = \{a + x : a \in A\}$, then $\mu(A) = \mu(A + x)$. Conversely, if ν is a translation-invariant complete Borel measure such that $\nu([0, 1)) = 1$, then $\mu = \nu$.

Proof. If $A = [\alpha, \beta)$ is an interval, then $\mu(A + x) = \beta + x - \alpha - x = \beta - \alpha = \mu(A)$. If A is

an arbitrary measurable set, then by (1.7),

$$\begin{aligned}\mu(A+x) &= \inf \left\{ \sum_{j=1}^{\infty} \beta_j - \alpha_j : A+x \subseteq \bigcup_{i=1}^{\infty} [\alpha_i, \beta_i) \right\} \\ &= \inf \left\{ \sum_{j=1}^{\infty} \beta_j + x - \alpha_j - x : A \subseteq \bigcup_{i=1}^{\infty} [\alpha_i, \beta_i) \right\} \\ &= \inf \left\{ \sum_{j=1}^{\infty} \beta_j - \alpha_j : A \subseteq \bigcup_{i=1}^{\infty} [\alpha_i, \beta_i) \right\} \\ &= \mu(A).\end{aligned}$$

Conversely, ν is σ -finite, since by translation invariance, the sets $[n, n+1)$ all have ν -measure 1, there are countably many of them, and they cover \mathbb{R} . Now if $[\alpha, \beta)$ is an arbitrary interval, we can write

$$[\alpha, \beta) = \bigcup_{n \in \mathbb{Z}} [\alpha, \beta) \cap [n, n+1)$$

and use countable additivity to see that

$$\nu([\alpha, \beta)) = \sum_{n \in \mathbb{Z}} \nu([\alpha, \beta) \cap [n, n+1)).$$

So it suffices to show that $\nu([\alpha, \beta)) = \mu([\alpha, \beta))$ whenever $[\alpha, \beta) \subseteq [n, n+1)$ for some n , in order that this also be true for any interval $[\alpha, \beta)$. Then by translation invariance, it suffices to check this when $\alpha = 0$, in which case $\beta \in (0, 1)$.

Suppose first that β is rational, say $\beta = p/q$. Then by translation invariance and countable additivity,

$$\nu([\alpha, \beta)) = \nu([0, p/q)) = \sum_{j=0}^{p-1} \nu([j, (j+1)/q)) = p\nu([0, 1/q)).$$

But

$$1 = \nu([0, 1)) = \sum_{j=0}^{q-1} \nu([j, (j+1)/q)) = q\nu([0, 1/q))$$

which implies that $\nu([0, 1/q)) = 1/q$ and hence $\nu([0, p/q)) = p/q$; therefore $\nu([\alpha, \beta)) = \beta - \alpha$ whenever $\beta - \alpha$ is rational. If $\beta - \alpha$ is irrational, choose $\beta_n > \beta_{n+1} > \dots > \beta$ so that $\beta_n \rightarrow \beta$ and $\beta_n - \alpha$ is rational for all n ; then $\nu([\alpha, \beta_n)) = \beta_n - \alpha$, so continuity of measure implies that $\nu([\alpha, \beta)) = \beta - \alpha$.

Since $\mu([\alpha, \beta)) = \nu([\alpha, \beta))$ for all $\alpha < \beta$, uniqueness of the Carathéodory construction for σ -finite measures implies that $\mu = \nu$. \square

1.6.10. The above argument is an example of a “bootstrapping” strategy that is common in measure theory. Once one has established a simple case (here $\nu([0, 1)) = 1$), it is often possible to propagate that case to a slightly more complicated case ($\nu([0, 1/q)) = 1/q$, say), and then repeat this process until all sets of interest have been considered. If you find yourself stuck on an exercise, it is worth trying to prove the claim in the simplest possible case and then iteratively improving the cases your proof works for until all cases are hit.

1.6.11. Theorem 1.6.9 shows that we could have defined Lebesgue measure axiomatically. We would have had to demanded that:

1. Countable additivity: Lebesgue measure be a measure.
2. Borel: Every open interval be Lebesgue measurable.
3. Completeness: Every subset of a Lebesgue null set be null.
4. Translation-invariance: A translate of a Lebesgue measurable set have the same Lebesgue measurable.
5. Calibration: The interval $[0, 1)$ have Lebesgue measure 1.

Certainly all of these conditions seem quite tame; then the Carathéodory construction would have proven existence and uniqueness. However, it is convenient to pass through the notion of a Stieltjes measure along the way, as more general Stieltjes measures are quite useful in their own right.

1.6.12. One may wonder if *every* subset of \mathbb{R} is Lebesgue measurable. Under certain reasonable set-theoretic hypotheses, this is false; we discuss a counterexample, Vitali's set, in Example 3.3.20. But the next-best thing is true, as any set that an analyst, algebraist, topologist, or applied mathematician will ever have to work with will turn out to be measurable. We discuss this in Example 3.3.18. This is closely related to a theorem of Solovay (TODO:Cite) which shows that in a certain sense, those set-theoretic hypotheses *cannot* be avoided in the construction of a nonmeasurable set: a slightly different logical setup would imply that every subset of \mathbb{R} is measurable.

1.6.13. Lebesgue measure is remarkably well-behaved, along with satisfying the five axioms above. To see an example of another good property of Lebesgue measure, we need a new definition. If the reader is uncomfortable with the abstract definition of a locally compact Hausdorff space, they may take $X = \mathbb{R}$ in the following definition without losing any insight.

Definition 1.6.14. Let X be a locally compact Hausdorff space. Suppose that μ is a Borel measure on X such that:

1. *Outer regularity*: For every Borel set W ,

$$\mu(W) = \inf_U \mu(U)$$

where the infimum is taken over all open sets $U \supseteq W$.

2. *Inner regularity for open sets*: For every open set U ,

$$\mu(U) = \sup_K \mu(K)$$

where the supremum is taken over all compact sets $K \subseteq U$.

3. *Local finiteness*: For every compact set K , $\mu(K) < \infty$.

Then we say that μ is a *Radon measure*.

Theorem 1.6.15. Every Stieltjes measure on \mathbb{R} is Radon. In particular, Lebesgue measure is Radon.

Proof. Let f be a nondecreasing left-continuous function. If K is compact then K is bounded, so K is contained in an interval $[\alpha, \beta]$ so $\mu_f(K) \leq f(\beta) - f(\alpha)$. Therefore μ_f is locally finite.

For inner regularity, note that every open set U can be written as a disjoint union of countably many open intervals, so it suffices to check when U is the interval (α, β) . Now if $K \subset (\alpha, \beta)$ is compact then there is an n such that $K \subseteq [\alpha + 1/n, \beta]$, so $\mu_f(K) \leq f(\beta) - f(\alpha + 1/n)$. On the other hand,

$$\lim_{n \rightarrow \infty} \mu_f([\alpha + 1/n, \beta]) = f(\beta) - \lim_{n \rightarrow \infty} f(\alpha + 1/n) = \mu_f((\alpha, \beta)).$$

Therefore μ_f is inner regular.

For outer regularity, let W be a Borel set (actually, any Stieltjes measurable set); then

$$\mu(W) = \inf_{(\alpha_n), (\beta_n)} \sum_n f(\beta_n) - f(\alpha_n)$$

where the inf ranges over all sequences of α_n and β_n such that $W \subseteq \bigcup_n [\alpha_n, \beta_n]$. Fix any such sequences. Now $E_m^n = (\alpha_n - m^{-1}2^{-n}, \beta_n)$ is an open cover of W for any m , so

$$\mu(W) \leq \sum_n f(\beta_n) - f(\alpha_n) < \frac{1}{m} + \sum_n f(\beta_n) - f(\alpha_n) = \sum_n f(\beta_n) - f(\alpha_n) + m^{-1}2^{-n}.$$

Taking $m \rightarrow \infty$ and minimizing $\sum_n f(\beta_n) - f(\alpha_n)$ by varying the α_n and β_n we collapse the above inequalities into infs. \square

1.6.16. The proof of Theorem 1.6.15 shows that Stieltjes measures are not just outer regular for Borel sets, but outer regular for any Stieltjes measurable set whatsoever. However, this fact is rarely useful; as discussed below, we are mainly interested in equivalence classes of sets modulo null sets, and every Lebesgue measurable set is equivalent under that equivalence relation to a Borel set.

1.6.17. After this chapter, we will mainly be interested in not measurable sets, but rather equivalence classes of measurable sets under the equivalence relation “symmetric difference is a null set”. That is, we think of two measurable sets A, B as the same if the symmetric difference $A \Delta B = (A \setminus B) \cup (B \setminus A)$ is null. For example, every countable set is the same as the empty set. One can give a different construction of the Lebesgue measure if one is *only* interested in equivalence classes of measurable sets, rather than measurable sets themselves; see Exercises 1.6.27 through 1.6.29.

1.6.18. Let us now construct a useful example of an uncountable, null compact set, known as the *standard Cantor set*. This set will frequently be useful as an example, and will be treated at length in the exercises.

Let $C_0 = [0, 1]$, and given C_n a finite union of closed intervals, let C_{n+1} consist of C_n with the open middle-thirds of each interval in C_n removed. TODO: Draw a picture. Let C , the standard Cantor set, be defined by $C = \bigcap_n C_n$.

1.6.19. It is often convenient to view the Cantor set in the following way. Let 2 be a shorthand for the set $\{0, 1\}$. Let 2^ω denote the space of sequences in 2 , thus 2^ω is the set of all functions $\mathbb{N} \rightarrow \{0, 1\}$. For each $x \in 2^\omega$, we obtain a point $f(x) \in C$ in the following way. Assume that D_{n-1} is one of the closed intervals obtained in stage $n - 1$ of the Cantor set construction, thus D_{n-1} is an interval in C_{n-1} . If $x_n = 0$, let D_n be the left closed interval in D_{n-1} , and otherwise let D_n be the right closed interval. Then $D_n \subseteq D_{n-1} \cap C_n$. The intersection of closed nonempty intervals is a closed nonempty interval, so we obtain such an interval $D \subseteq C$. But a consequence of Exercise 1.6.23 is that the Cantor set contains no nontrivial intervals, so D consists of a single point $f(x)$. The map f is injective since if $f(x) = f(y)$ then, at the n th stage of the Cantor set construction, both x and y agreed on which subinterval to pass to, thus $x = y$. It is surjective since every point $p \in C$ must be in some interval D_n at stage n , and if D_n is the left interval of D_{n-1} then we can set $x = 0$, and otherwise set $x = 1$; then $f(x) = p$. One can similarly interpret the Cantor set as the set of all paths through the full infinite binary tree. This perspective can be highly useful in some exercises and in applications. For more on Cantor sets, see TODO:Cite Pugh.

Exercise 1.6.20. Let $x \in \mathbb{R}$ and let f be a nondecreasing, left-continuous function. Show that

$$\mu_f(\{x\}) = \lim_{\varepsilon \rightarrow 0} f(x + \varepsilon) - f(x).$$

Conclude that if f is continuous, then μ_f makes all countable sets null.

Exercise 1.6.21. Let f be a nondecreasing, left-continuous function. Show that there is a nonempty open set U such that $\mu_f(U) = 0$ iff f is not strictly increasing (i.e. there are $x < y$ with $f(x) = f(y)$).

Exercise 1.6.22. Let $x \in \mathbb{R}$. The *Dirac measure* at x , denoted δ_x , is the Stieltjes measure arising from the function

$$y \mapsto \begin{cases} 0, & y \leq x \\ 1, & y > x. \end{cases}$$

A set is *Dirac measurable* at x if it is measurable with respect to δ_x . Show that every subset of \mathbb{R} is Dirac measurable at x , and compute its Dirac measure.

Exercise 1.6.23. Show that the standard Cantor set is compact and Lebesgue null, has the same cardinality as \mathbb{R} , and contains no interval except for points and the empty set.

Exercise 1.6.24. Let μ denote Lebesgue measure on \mathbb{R} . Show that there is a Borel probability measure ν on \mathbb{R} , such that for every countable set A , $\nu(A) = 0$, and if $\mu(A) > 0$, then $\nu(A) = 1$. (Hint: Define ν in such a way that, if C denotes the standard Cantor set, then $\nu(C) = 1$, and use Exercise 1.6.23.)

Exercise 1.6.25. Show that the set of Lebesgue measurable subsets of \mathbb{R} has cardinality equal to that of the power set $2^\mathbb{R}$ of \mathbb{R} . Conclude that there exists a Lebesgue measurable set which is not Borel. (Hint: Exercise 1.6.23.)

Exercise 1.6.26. Show that for every $\alpha \in [0, 1]$ there exists a Cantor-like subset C_α of $[0, 1]$ whose Lebesgue measure is α . In particular, C_α should be compact and contain no open subset of \mathbb{R} . This is known as a *fat Cantor set*. Show that there does not exist a fat Cantor set of measure 1 in $[0, 1]$.

Exercise 1.6.27. In this exercise and the following we give a different construction of Lebesgue measure, other than the axiomatic definition or the construction using Stieltjes measures.

Let Σ_0 be the set of all intervals (open, closed, or half-open) with rational endpoints. Given $I \in \Sigma_0$, define $\nu(I)$ to be the length of I . If $I, J \in \Sigma_0$, then $I \setminus J \in \Sigma_0$, so we can define

$$d_0(I, J) = \nu(I \setminus J) + \nu(J \setminus I).$$

Show that d_0 is a semimetric, so the completion (Σ, d) of (Σ_0, d_0) is a complete metric space, and ν is a continuous function on Σ .

Show that the “union” of countably many elements of Σ is well-defined, as is the “intersection” of countably many elements and “complement” of one element. That is, show that these operations drop to equivalence classes of elements of Σ_0 under the equivalence relation “distance is 0”, and then extend uniquely to Σ . So Σ can be thought of as an “abstract σ -ring” in some sense, even though its elements are not sets, but something more abstract.

Exercise 1.6.28. Let all definitions be as Exercise 1.6.27. Given $J \in \Sigma_0$, define $[J]$ to be the equivalence class of J in Σ . Then, given $I \in \Sigma$, let

$$\nu(I) = d(I, [\emptyset]).$$

Show that ν is “ σ -additive” in the sense that whenever $(E_n)_n$ is a sequence in Σ with $E_n \cap E_m = [\emptyset]$ whenever $n \neq m$, then

$$\nu\left(\bigcup_n E_n\right) = \sum_n \nu(E_n).$$

So we can think of ν as an “abstract measure” on the abstract σ -ring Σ . We call ν the *physical Lebesgue measure*.

Exercise 1.6.29. Let μ be Lebesgue measure and ν physical Lebesgue measure as in Exercise 1.6.28, defined on the abstract σ -ring Σ . Let Γ be the σ -algebra of Lebesgue measurable subsets of \mathbb{R} . Show that there is a surjective map $\pi : \Gamma \rightarrow \Sigma$ such that:

1. For every rational interval I , $\pi(I)$ is the equivalence class of I in Σ .
2. For every sequence $(E_n)_n$ in Γ , $\pi(\bigcup_n E_n) = \bigcup_n \pi(E_n)$, and similarly for intersection and complement.
3. For every $E \in \Gamma$, $\mu(E) = \nu(\pi(E))$.

So we can think of $\pi(E)$ as the equivalence class of E under the equivalence relation “symmetric difference is null”, and ν correctly computes the Lebesgue measurable of a set.

Chapter 2

More on measures

In this chapter we discuss vector-valued measures and the relationships between measured spaces. These turn out to be especially useful in applications, including probability theory, so we introduce the languages of probability theory and ergodic theory in this chapter as well.

2.1 Measurable maps

In topology, a continuous map is one that pulls back open sets to open sets. Here we have no open sets to work with – only measurable spaces – but this turns out to be exactly what we want.

Definition 2.1.1. Let (X, Σ) and (Y, Γ) be measurable spaces. A *measurable map* $f : X \rightarrow Y$ is a map such that for every measurable set $A \in \Gamma$, $f^{-1}(A)$ is measurable.

A *measurable isomorphism* is a measurable map $f : X \rightarrow Y$ such that f is a bijection and f^{-1} is measurable. If a measurable isomorphism exists, we say that X, Y are *isomorphic* as measurable spaces, or that Σ and Γ are *isomorphic* as σ -algebras.

Example 2.1.2. Isomorphism of measurable spaces is a very weak notion. If X is an uncountable *Polish space* – a complete separable metric space – then we call its Borel σ -algebra Σ the *standard Borel algebra*. A theorem of a branch of logic known as descriptive set theory implies that there is only one standard Borel algebra, up to isomorphism. So any two Polish spaces, when viewed as measurable spaces, are the same!

Definition 2.1.3. Let (X, Σ, μ) be a measured space, (Y, Γ) a measurable space, and $f : X \rightarrow Y$ a measurable map. We define $f_*(\mu)$ on (Y, Γ) by

$$f_*(\mu)(E) = \mu(f^{-1}(E))$$

and call $f_*(\mu)$ the *pushforward measure* of μ by f .

Example 2.1.4. Let \mathbb{T} be the circle, viewed for now as the set $\{z \in \mathbb{C} : |z| = 1\}$. Then \mathbb{T} has its Borel σ -algebra Γ . Let Σ be the Borel σ -algebra of \mathbb{R} and μ the Lebesgue measure restricted to Σ . Then we have a continuous map $f : [0, 1] \rightarrow \mathbb{T}$ defined by

$$f(x) = e^{2\pi i x}.$$

Every continuous map between Borel measurable spaces is measurable, so f is. The push-forward measure $\nu = f_*(\mu)$ is known as the *Lebesgue measure* on \mathbb{T} . The diligent reader will check that ν agrees with the notion of arc length defined in multivariable calculus (Exercise 2.1.9).

Definition 2.1.5. Let (X, μ) and (Y, ν) be measured spaces. A *measure-preserving map* is a measurable map $f : X \rightarrow Y$ such that

$$\nu = f_*(\mu).$$

If f is a measurable isomorphism, we call f a *measure-preserving isomorphism*.

Definition 2.1.6. Let X be a set, (Y, Γ) a measurable space, and $F : X \rightarrow Y$ a map. The *pullback σ -algebra* $F^*\Gamma$ is the σ -algebra of sets of the form $F^{-1}(A)$ where $A \in \Gamma$.

2.1.7. We need to check that the definition of a pullback σ -algebra makes sense. Namely, we must show that if $F : X \rightarrow Y$ is a map, and Γ a σ -algebra on Y , then $F^*\Gamma$ is a σ -algebra. This is part of the content of Exercise 2.1.11, which also shows that $F^*\Gamma$ is the smallest σ -algebra for which F is measurable.

Exercise 2.1.8. Show that measurable spaces with measurable maps form a category.

Exercise 2.1.9. By an *arc* we mean a compact connected subset of \mathbb{T} . Recall the definition of the length of an arc from calculus. Show that the length of an arc equals its Lebesgue measure.

Exercise 2.1.10. Let S^d denote the unit sphere in \mathbb{R}^{d+1} , and let N be its open northern hemisphere. Let U be the open unit ball in \mathbb{R}^d , so $U \times \{0\}$ is a subset of the unit ball of \mathbb{R}^{d+1} . Show that the map $p : N \rightarrow \mathbb{R}^{d+1}$, $p(x_1, \dots, x_{d+1}) = (x_1, \dots, x_d, 0)$, is a homeomorphism between N and $U \times \{0\}$. So the pushforward of Lebesgue measure on \mathbb{R}^d by p^{-1} gives a Borel measure on N , which we call *Lebesgue measure* on N . Use this fact to define Lebesgue measure of any Borel subset of S^d . Show that when $d = 2$, Lebesgue measure on S^d agrees with the usual notion of surface area, up to a constant factor.

Exercise 2.1.11. Let $F : X \rightarrow Y$ be a map, Γ a σ -algebra on Y . Show that $F^*\Gamma$ is a σ -algebra on X and $F : (X, F^*\Gamma) \rightarrow (Y, \Gamma)$ is measurable. Conversely, show that $F^*\Gamma$ is the intersection of all σ -algebras Σ on X such that $F : (X, \Sigma) \rightarrow (Y, \Gamma)$ is measurable.

2.2 Vector-valued measures

Recall that by definition, a measure μ is a σ -additive function $\mu : \Sigma \rightarrow (-\infty, \infty]$, where Σ is a σ -algebra. However, the definition of σ -additive function makes sense even if $(-\infty, \infty]$ is replaced by a more general codomain, and we will have use for this when we integrate functions in more general codomains than the real numbers.

2.2.1. Let B be a Banach space, as discussed in Appendix A.2. If the reader is unfamiliar with Banach spaces, they can take $B = \mathbb{C}^d$ and not lose any insight. If (x_n) is a sequence in B , then (A.3) is the definition of the infinite sum $\sum_n x_n$.

Definition 2.2.2. Let Σ be a σ -ring and B a Banach space. A *vector-valued measure* on Σ , or simply a *measure*, is a function $\mu : \Sigma \rightarrow B$ such that whenever (E_n) are countably many disjoint sets in Σ ,

$$\mu\left(\bigcup_n E_n\right) = \sum_n \mu(E_n).$$

Example 2.2.3. The most important examples of vector-valued measures will be of the following form. Let ν be a positive measure on a measurable space X , and let $f : X \rightarrow B$ be an “integrable function”. We will get back to what this means later – but, taking the definition of integration as a black box, set

$$\mu(E) = \int_E f(x) \, d\nu(x).$$

Then μ is a vector-valued measure, and its “derivative” is f . This will be the setting in which we generalize the fundamental theorem of calculus.

2.2.4. Frequently, positive measures are easier to work with than vector-valued measures, so we will mainly spend this section developing tools to convert problems about vector-valued measures into problems about positive measures.

Definition 2.2.5. Let μ be a measure. The *total variation measure* $|\mu|$ of μ is defined by

$$|\mu|(E) = \sup_{E_1, \dots, E_n} \sum_{i=1}^n \|\mu(E_i)\|_B$$

where the supremum ranges over finite sequences E_1, \dots, E_n of disjoint measurable sets such that $E = \bigcup_i E_i$.

Theorem 2.2.6. Every total variation measure is a positive measure.

Proof. Let μ be a measure. Then $|\mu|$ is nonnegative. If $E \subseteq F$ are measurable sets, write $F = E \cup (F \setminus E)$ to see that $|\mu|(E) \leq |\mu|(F)$.

To see that $|\mu|$ is σ -additive, suppose that F, G are disjoint measurable sets, $E = F \cup G$. Write $F = \bigcup_{i \leq m} F_i$ and $G = \bigcup_{j \leq n} G_j$ where the F_i, G_j are all disjoint; then

$$|\mu|(E) \geq \sum_{i=1}^m \|\mu(F_i)\|_B + \sum_{j=1}^n \|\mu(G_j)\|_B.$$

Thus $|\mu|(E) \geq |\mu|(F) + |\mu|(G)$, so by induction if $E = \bigcup_{i \leq m} E_i$ where the E_i are disjoint, then $|\mu|(E) \geq \sum_{i \leq m} |\mu|(E_i)$. Using the monotonicity, it follows that if (E_i) is a disjoint countable sequence of sets and $E = \bigcup_i E_i$, then

$$|\mu|(E) \geq \sum_{i=1}^{\infty} |\mu|(E_i).$$

Conversely, suppose that $E = \bigcup_j E_j = \bigcup_{i \leq n} F_i$ where (E_j) is a disjoint countable sequence and (F_i) is a disjoint finite sequence. Then

$$\begin{aligned} \sum_{i=1}^n \|\mu(F_i)\|_B &= \sum_{i=1}^n \left\| \mu \left(\bigcup_{j=1}^{\infty} (F_i \cap E_j) \right) \right\|_B \leq \sum_{i=1}^n \sum_{j=1}^{\infty} \|\mu(F_i \cap E_j)\|_B \\ &= \sum_{j=1}^{\infty} \sum_{i=1}^n \|\mu(F_i \cap E_j)\|_B \leq \sum_{j=1}^{\infty} |\mu|(E_j). \end{aligned}$$

Here we used the fact that $E_j = \bigcup_{i \leq n} F_i \cap E_j$. It follows that $|\mu|(E) \leq \sum_{i=1}^{\infty} |\mu|(E_i)$. \square

Theorem 2.2.7 (triangle inequality and reverse triangle inequality). Let μ, ν be measures on Σ with values in B . Then $|\mu + \nu| \leq |\mu| + |\nu|$ and

$$||\mu|(E) - |\nu|(E)| \leq |\mu - \nu|(E).$$

Proof. If $E = \bigcup_{i \leq n} E_i$, the E_i disjoint, then

$$\sum_{i=1}^n \|(\mu + \nu)(E_i)\|_B \leq \sum_{i=1}^n \|\mu(E_i)\|_B + \|\nu(E_i)\|_B \leq |\mu|(E) + |\nu|(E)$$

which implies the two desired inequalities. \square

2.2.8. Let μ be a vector-valued measure. The total variation measure $|\mu|$ can be viewed as the positive measure that “best dominates” $\|\mu\|_B$ in the sense that for every measurable set E , $|\mu|(E) \geq \|\mu(E)\|_B$. On the other hand, if μ is already a positive measure, then the definition of the total variation measure $|\mu|$ still makes sense, and additivity of μ implies that $\mu = |\mu|$. Thus we can use total variation to extend definitions of measures to this more general setting.

Definition 2.2.9. A vector-valued measure μ is σ -finite if $|\mu|$ is σ -finite. A set E is μ -null if E is $|\mu|(E)$ -null. The measure μ is a *complete measure* if every μ -null set is complete. The *completion* of μ is the vector-valued measure obtained by extending the domain of μ to contain all μ -null sets.

Exercise 2.2.10. Show that the completion of a vector-valued measure is well-defined. That is, if (X, Γ) is a measurable space and μ is a vector-valued measure on Γ , then show that there is a σ -algebra Σ on X which contains every measurable set and every μ -null set, such that (X, Σ, μ) is a complete measured space.

Exercise 2.2.11. Give an example of a vector-valued measured space (X, μ) such that $\mu(X) = 0$, but $|\mu|$ is not identically zero. Thus, explain why we cannot define a μ -null set to simply be a set E such that $\mu(E) = 0$.

2.3 Probability

In 1933, Andrey Kolmogorov put probability theory – which was long viewed as not part of rigorous mathematics – on a sound footing by establishing three axioms that probability theory ought to satisfy. These axioms, however, are exactly the definition of a probability measure! Thus measure theory is an invaluable tool in applied mathematics, such as statistics. However, even the myopic reader who only cares about pure, abstract mathematics will find probability useful to their work.

In this section we do little more than develop a language, but in later sections we will find a great use for it.

Definition 2.3.1. A *probability space* (Ω, Σ, P) is a measure space with $P(\Omega) = 1$. Elements of Σ are called *events* and elements of X are called *outcomes*. The quantity $P(A)$, where A is an event, is called the *probability* that A is true.

2.3.2. The intuition here is that the events form an algebra Σ of statements that could be true or false about the world, with varying probabilities. By an algebra of statements, more precisely, one means that it is meaningful to conjoin them using grammar: $A \cup B$ is the event that A happens or B happens, while $A \cap B$ is the probability that both A happens and B happens. The event $\Omega \setminus A$ is the probability that A does not happen, and so on. The event Ω is vacuously true; it represents the statement “ $2 + 2 = 4$ ”. The event \emptyset , on the other hand, is vacuously false; it represents the statement “ $2 + 2 = 5$ ”. Thus probability theory sits somewhere between analysis and logic in the world of mathematics, and the language used in the following definition is motivated.

Definition 2.3.3. Let (Ω, Σ, P) be a probability space.

The event $A \cup B$ is called *A or B*, or the *disjunction* of A, B , and the event $A \cap B$ is called *A and B*, or the *conjunction* of A, B . The event $\Omega \setminus A$ is called *not A*, or the *negation* of A .

The event Ω is said to be *true*, and the negation \emptyset of true is said to be *false*. An event A is said to be *almost surely true* if $P(A) = 1$, and *almost surely false* if $P(A) = 0$.

A set $\Sigma_0 \subseteq \Sigma$ of events is said to be *mutually exclusive* if for every $A, B \in \Sigma_0$, $A \cap B$ is almost surely false.

2.3.4. The σ -additivity of the measure P says that if a countable set Σ_0 of events is mutually exclusive, its disjunction A satisfies

$$P(A) = \sum_{B \in \Sigma_0} P(B),$$

thus for example we have the *inclusion-exclusion formula*

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

valid for any two events E, F familiar from elementary probability theory.

2.3.5. So far we have only discussed events. But what role do outcomes play? Outcomes are “microstates” and each outcome carries all possible information about a particular state

that the universe could be – the position and momentum of every last molecule, the value of every card in every player’s hand. But we are just puny mortal observers of the universe; we cannot possibly have access to all that information! So outcomes are not terribly useful, and one generally avoids ever mentioning outcomes directly. What we mortal observers can observe are random variables.

Definition 2.3.6. Let (Ω, P) be a probability space and E be a measurable space. A *random variable* of type E , also known as an *observable* of type E , is a measurable map $\Omega \rightarrow E$.

2.3.7. The points of the space E represent possible values that an experiment can return. The random variable $X : \Omega \rightarrow E$ represents an experiment. If ω is an outcome, $X(\omega)$ is the result of the experiment X if the universe is in state ω . One generally takes $E = \mathbb{R}$ with its Borel σ -algebra, since most experiments return numerical data, but there are other examples of useful types E ; for example, points of E are often sequences, vectors, or graphs. If the type of X is clear – especially if it is \mathbb{R} – we will suppress it.

2.3.8. Let X be a random variable (of type \mathbb{R}). The sets $[x, \infty)$ are Borel, so the sets $\{X \geq x\} = \{\omega \in \Omega : X(\omega) \geq x\}$ are actually events. The probability of the event $\{X \geq x\}$ is frequently of interest, so we make the following definition.

Definition 2.3.9. Let X be a random variable of type E . The *distribution* μ_X of X is the pushforward X_*P of P on E . If Y is another random variable and $\mu_X = \mu_Y$, we say that X, Y are *identically distributed*.

2.3.10. Unraveling the definitions, if $A \subseteq E$ is a measurable set,

$$\mu_X(A) = P(X \in A) = P(\{\omega \in \Omega : X(\omega) \in A\}).$$

For example, if $E = \mathbb{R}$ then $\mu_X([x, \infty))$ is the probability that $X \geq x$. We think of two identically distributed random variables as being isomorphic, but not the same.

2.3.11. Recall that in elementary probability theory two events A, B are said to be independent if $P(A) = P(A|B)$, where $P(A|B)$ (“the probability of A given B ”) is by definition $P(A \cap B)/P(B)$. The intuition is that if we know that B is true, then we want to restrict to subsets of B and rescale the probability measure P so that B is almost surely true. However, if A and B really have nothing to do with each other, the probability of A given B is just the probability of A . We can write this in a more symmetrical form, which also avoids the issue of division by zero, as in the following definition.

Definition 2.3.12. Let (Ω, Σ, P) be a probability space and (E, Γ) a measurable space.

Countably many events $A_n \in \Sigma$ are said to be *independent* if

$$P\left(\bigcap_n A_n\right) = \prod_n P(A_n).$$

Countably many σ -algebras $\mathcal{A}_n \subseteq \Sigma$ are said to be *independent* if for every $A_n \in \mathcal{A}_n$, the events A_n are independent. Countably many random variables X_n of type E are said to be *independent* if the pullback σ -algebras $X_n^*\Gamma$ are independent.

Random variables X, Y are said to be *iid* if they are independent and identically distributed.

2.3.13. The idea in the definition of independent random variables is that the pullback σ -algebra $X^*\Gamma$ is the set of events that can be checked as true or false by measuring X .

Example 2.3.14. Suppose that I am playing a game of cards. Every time I draw a card, I put it back in the deck and then shuffle again. Let X_n be 1 if I draw the queen of hearts on turn n , and 0 otherwise. Then the distribution of X_n is $(51/52)\delta_0 + (1/52)\delta_1$, where δ_x is the Dirac measure at x . Since I shuffle between turns, the X_n are independent. So the X_n are iid.

Exercise 2.3.15. Verify the inclusion-exclusion formula

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n (-1)^{k+1} \sum 1 \leq i_1 < \cdots < i_k \leq n P\left(\bigcap_{j=1}^k A_{i_j}\right)$$

valid for any events A_1, \dots, A_n .

Exercise 2.3.16 (Skohorod representation). Let μ be a Borel probability measure on \mathbb{R} . Show that there is a probability space Ω and a random variable $X : \Omega \rightarrow \mathbb{R}$ of distribution μ .

Exercise 2.3.17. Let X_n be random variables and suppose that for every $x \in \mathbb{R}$, the events $X_n \geq x$ are independent. Show that the X_n are independent random variables.

Exercise 2.3.18. Let X_n be independent random variables of type E . Suppose that $f_n : E \rightarrow E$ are measurable maps. Show that the $f_n(X_n)$ are still independent.

Exercise 2.3.19. Let $X, Y : \Omega \rightarrow E$ be random variables. A *morphism of random variables* $X \rightarrow Y$ is a measure-preserving map $F : \Omega \rightarrow \Omega$ such that $Y \circ F = X$. Two morphisms F, G are *equal almost surely* if $P(F = G) = 1$. Two random variables X, Y are *isomorphic* if there are morphisms $F : X \rightarrow Y$ and $G : Y \rightarrow X$ such that $F \circ G$ and $G \circ F$ are the identity almost surely (so G is almost surely the inverse of F).

Show that being equal almost surely is an equivalence relation on the set of morphisms of random variables. Show that random variables $\Omega \rightarrow E$ and morphisms between them, modulo the equivalence relation of being equal almost surely, form a category. Show that two random variables are isomorphic iff they are identically distributed.

2.4 Ergodic systems

This section is used nowhere else in this book, except in some exercises; it is meant to showcase an application.

Measure-preserving maps from a measured space to itself are important in probability and thermodynamics; their study is known as *ergodic theory*.

Definition 2.4.1. A *measure-preserving system* (X, Σ, P, f) consists of a probability space (X, Σ, P) and a measure-preserving map $f : X \rightarrow X$.

2.4.2. We may suppress X or Σ from the notation when they are clear or unhelpful to specify.

2.4.3. The intuition for measure-preserving systems is as follows. We have a probability space X whose events A are properties that the world could have, and the probability $P(A)$ is the probability that the world is in that state. The map f represents the passing of time; each application of f represents the passing of a unit of time. However, the probability that the universe has property A is the same today as it is tomorrow.

Definition 2.4.4. An *ergodic system* is a measure-preserving system (X, Σ, P, f) for which, whenever $f(A) \subseteq A$ and $A \in \Sigma$, then $P(A) = 0$ or $P(A) = 1$.

2.4.5. Ergodic systems are the most important examples of measure-preserving systems. The reason is that ergodic systems behave “randomly”. In an ergodic system, the universe might have property A today, but this is no evidence that the universe will have property A tomorrow, because every day all the properties that the universe could have get completely mixed up every time f is applied.

Let (X, P, f) be an ergodic system. Imagine that X is an egg; then P represents picking a random point inside the egg, and f represents the action of scrambling the egg, say with chopsticks. Though X is initially in an orderly state, the ergodicity of the system means that all this order is lost and all the points of X get mixed up. This is illustrated mathematically by the following example.

Example 2.4.6. One ergodic system is known as the *irrational rotation*. Let $\theta \in [0, 2\pi]$ and suppose that θ/π is irrational. Then the map $f(z) = e^{i\theta}z$ on the circle \mathbb{T} , which rotates the circle by θ , defines an ergodic system (\mathbb{T}, μ, f) , where μ is Lebesgue measure. The reader who is familiar with some Fourier analysis may show this in Exercise 2.4.12.

2.4.7. As a sample of the power of ergodic theory, let us prove the following theorem that “breaks the laws of thermodynamics”. If (X, f) is a measure-preserving system and $x \in X$, we let $f^n(x)$ denote $f \circ \cdots \circ f$, where there are n copies of f . The theorem is due to Carathéodory, which is naturally why it is named after Poincaré.

Theorem 2.4.8 (Poincaré recurrence). Let (X, Σ, P, f) be a measure-preserving system. For every $A \in \Sigma$, define $A^b = \{x \in A : \forall n \ f^n(x) \notin A\}$. Then $P(A^b) = 0$.

Proof. We first note that

$$A^b = A \setminus \bigcup_{n \in \mathbb{N}} f^n(A)$$

which is measurable since each of the $f^n(A)$ are.

If $m > n$, then $f^n(A^b) \cap f^m(A^b)$ is empty: if $x \in f^m(A^b)$, so that $f^{-m}(x) \in A^b$, then by definition of A^b ,

$$f^{-n}(x) = f^{m-n}(f^{-m}(x)) \notin A.$$

Therefore, for any $N \in \mathbb{N}$,

$$P(A^b) = \frac{1}{N} \sum_{n=1}^N P(f^n(A^b)) = \frac{1}{N} P\left(\bigcup_{n=1}^N f^n(A^b)\right) \leq \frac{1}{N} P(X) = \frac{1}{N}.$$

So $P(A^b) = 0$. □

2.4.9. Let us explain to the reader who is familiar with thermodynamics how to interpret this theorem.

Let X be the set of microstates that the universe can be in and let us assume for simplicity that X is finite. Let P be the normalized counting measure, so that

$$P(A) = \frac{\text{card } A}{\text{card } X}$$

and $\Sigma = 2^X$. Thus $P(A)$ represents the probability of drawing a particular microstate uniformly at random.

Let $f : X \rightarrow X$ be the map that sends the state of the universe right now to the state of the universe one second into the future. We can think of sets A as macrostates, where $x \in A$ iff x is consistent with A . By Poincaré recurrence, if the universe is currently in macrostate A , then there is an N such that N seconds into the future, the universe will again be in macrostate A . In particular, the entropy of the universe will return to what it currently is, apparently contradicting the second law of thermodynamics. This is no contradiction, however, because N is much larger than the lifespan of the universe.

2.4.10. We will not discuss ergodic systems much, but ergodic theory is a crucial application of measure theory. For example, it can be used to prove the strong law of large numbers from probability, and actually very powerful generalizations thereof. We refer the reader to TODO:Cite Couden's book to learn about ergodic theory.

Exercise 2.4.11. Show that Poincaré recurrence fails for infinite measure spaces by giving a counterexample.

Exercise 2.4.12. Show that the irrational rotation with Lebesgue measure forms an ergodic system. (Hint: Let A be a subset of the circle which the irrational rotation sends into itself. Consider the Fourier series of 1_A . You are allowed to assume basic facts about Fourier series.)

Chapter 3

Measurable functions

Throughout this chapter, let (X, Σ) be a measurable space and let B be a Banach space. We would like to consider functions $f : X \rightarrow B$ which “respect” Σ . Then, given a measure μ defined on Σ , we will be able to define the integral $\int_X f \, d\mu$ of f with respect to μ .

3.1 Simple functions

When we prove a theorem T about functions in measure theory, we want to follow the following template:

1. Let \mathcal{F} be the set of functions for which T is true. Show that \mathcal{F} contains all functions which are “sufficiently simple.”
2. Show that \mathcal{F} is closed under linear combination, so is a vector space.
3. Show that \mathcal{F} is closed under taking limits of appropriate type.
4. Conclude that every appropriate function lies in \mathcal{F} .

In this section we treat the “sufficiently simple” functions.

Definition 3.1.1. A *simple function* is a function $f : X \rightarrow B$ such that the image of f is finite, and for every b in the image of f , $f^{-1}(b)$ is a measurable set. The set of simple functions is denoted $\mathbf{Simp}(X \rightarrow B)$.

3.1.2. The simple functions have a particularly convenient canonical form. To define them, we first characterize the \mathbb{R} -valued simple functions.

Definition 3.1.3. Let $Y \subseteq X$ be a measurable set. The *indicator function* of Y , denoted 1_Y , is the function $X \rightarrow \{0, 1\}$, defined by $1_Y(y) = 1$ if $y \in Y$ and $1_Y(y) = 0$ if $y \notin Y$.

3.1.4. Note that every indicator function is simple, since its image is $\{0, 1\}$, the preimage of 0 is Y^c , the preimage of 1 is Y , and Y, Y^c are both measurable. Conversely, if $f \in \mathbf{Simp}(X \rightarrow$

\mathbb{C}), f is a linear combination of indicator functions. In fact, if $\{y_1, \dots, y_n\}$ is the image of f , and the preimage of y_i is Y_i , then

$$f(x) = \sum_{i=1}^n y_i 1_{Y_i}(x).$$

Indeed, the Y_i are disjoint since they are preimages of distinct real numbers, so $y_i 1_{Y_i}(x) = y_i$ iff $f(x) = y_i$, and $f(x) = 0$ otherwise. This characterization also works for B -valued simple functions: if $\{b_1, \dots, b_n\}$ is the image of f , and the preimage of b_i is Y_i , then

$$f(x) = \sum_{i=1}^n b_i 1_{Y_i}(x).$$

3.1.5. We now show that $\mathbf{Simp}(X \rightarrow B)$ is closed under various operations.

Definition 3.1.6. A vector space A over \mathbb{C} is called an *algebra* if it is equipped with a multiplication $A \times A \rightarrow A$ which is associative, distributes over addition, and satisfies, for every $c, d \in \mathbb{C}$ and $x, y \in A$,

$$(cx)(dy) = (cd)(xy).$$

If the multiplication of A has an identity, we call the identity 1 and call A a *unital algebra*.

3.1.7. In particular, a collection of functions is an algebra iff it is closed under multiplication. See Exercise 3.1.10.

Lemma 3.1.8. $\mathbf{Simp}(X \rightarrow B)$ is a vector space. In particular, $\mathbf{Simp}(X \rightarrow \mathbb{C})$ is a unital algebra.

Proof. Let $f, g \in \mathbf{Simp}(X \rightarrow B)$, say

$$f(x) = \sum_{i=1}^n y_i 1_{Y_i}(x)$$

and

$$g(x) = \sum_{j=1}^m z_j 1_{Z_j}(x).$$

We first claim that $f + g$ is simple. In fact, the image of $f + g$ is contained in $\{y_i + z_j : 1 \leq i \leq n, 1 \leq j \leq m\}$, and the preimage of $y_i + z_j$ under $f + g$ is $Y_i \cap Z_j$.

The proof that $\mathbf{Simp}(X \rightarrow B)$ is closed under scaling is similar. Clearly $\mathbf{Simp}(X \rightarrow B)$ is nonempty, so this implies that $\mathbf{Simp}(X \rightarrow B)$ is a vector space.

Now if $B = \mathbb{C}$, function multiplication is defined, and

$$fg(x) = \sum_{i=1}^n y_i 1_{Y_i}(x) \sum_{j=1}^m z_j 1_{Z_j}(x) = \sum_{i,j} y_i z_j 1_{Y_i \cap Z_j}(x).$$

Therefore fg is simple. Moreover, the function $f(x) = 1$ is an identity for multiplication. \square

Lemma 3.1.9. Let $f, g \in \mathbf{Simp}(X \rightarrow \mathbb{C})$. Then $|f| \in \mathbf{Simp}(X \rightarrow \mathbb{C})$. In fact, if $f, g \in \mathbf{Simp}(X \rightarrow \mathbb{R})$, then $\max(f, g) \in \mathbf{Simp}(X \rightarrow \mathbb{R})$ and $\min(f, g) \in \mathbf{Simp}(X \rightarrow \mathbb{R})$.

Moreover, if $f \in \mathbf{Simp}(X \rightarrow B)$, the function $x \mapsto \|f(x)\|$ is in $\mathbf{Simp}(X \rightarrow \mathbb{C})$.

Proof. Left as Exercise 3.1.11. □

Exercise 3.1.10. Look up or prove the following algebraic theorems:

1. Let $\varphi : \mathbb{C} \rightarrow A$ be a morphism of rings. Then A is an algebra, where scalar multiplication is defined by $za = \varphi(z)a$, for any $z \in \mathbb{C}$ and $a \in A$. We call φ the *canonical inclusion* of \mathbb{C} into A .
2. If A is a unital algebra with multiplicative identity 1_A , then the map $\varphi(z) = 1_A z$, $\varphi : \mathbb{C} \rightarrow A$, is a morphism of rings, and is in fact the canonical inclusion of \mathbb{C} into A .
3. If V is a vector space of functions $X \rightarrow \mathbb{C}$, then V is an algebra over \mathbb{C} iff V is closed under function multiplication.

None of these theorems actually require that we are working over \mathbb{C} ; they are valid if \mathbb{C} is replaced with any field. However, you are not expected to check that.

Exercise 3.1.11. Prove Lemma 3.1.9.

Exercise 3.1.12. If $(A_n)_n$ is a sequence of measurable sets, we define $\limsup_{n \rightarrow \infty} A_n$ by the relation

$$1_{\limsup_{n \rightarrow \infty} A_n} = \limsup_{n \rightarrow \infty} 1_{A_n}.$$

Then:

1. Show that $\limsup_{n \rightarrow \infty} A_n$ is a well-defined measurable set.
2. Show that $x \in \limsup_{n \rightarrow \infty} A_n$ iff there are infinitely many $n \in \mathbb{N}$ such that $x \in A_n$.
3. Prove the *Borel-Cantelli lemma*: If μ is a nonnegative measure on X such that

$$\sum_{n=1}^{\infty} \mu(A_n) < \infty,$$

then $\limsup_{n \rightarrow \infty} A_n$ is μ -null.

The Borel-Cantelli lemma is a primitive example of a “zero-one law”, a theorem that implies that certain sets must either be null or have null complement.

Exercise 3.1.13. Given $x \in [0, 1)$, we may assign x a canonical decimal expansion by not letting x have a decimal expansion that ends with an infinitely repeating string of nines. We let $x_j \in \{0, \dots, 9\}$ denote the j th entry in the canonical decimal expansion of x (so for example $(\pi - 3)_3 = 1$). Let us say that x has *uniform expansion* at $n \in \mathbb{N}$ if for every $d \in \{0, \dots, 9\}$, the set of $j \in \{1, \dots, 10n\}$ such that $x_j = d$ has cardinality n . In other words, x has uniform expansion at n iff for every $d \in \{0, \dots, 9\}$ the probability that a $j \in \{1, \dots, 10n\}$ drawn uniformly at random satisfies $x_j = d$ is $1/10$.

Let A be the set of $x \in [0, 1)$ such that there are infinitely many $n \in \mathbb{N}$ such that x has uniform expansion at n . Show that A is a Borel set, and that A is Lebesgue null. (Hint: Use the Borel-Cantelli lemma, Exercise 3.1.12.)

3.2 Measurable functions

We want to know which functions $X \rightarrow B$ are “good” from the perspective of measure theory. To accomplish this, recall the notion of pointwise convergence.

Definition 3.2.1. A sequence of functions $f_n : X \rightarrow B$ is said to *converge pointwise* to a function f if for every $x \in X$,

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

3.2.2. Let μ be a complete measure on Σ . If μ is not already complete, we can always expand Σ by adjoining the μ -null sets to Σ , so this assumption is no loss in generality.

3.2.3. In measure theory, a property is said to hold *almost everywhere* if it holds everywhere except a null set, and hold *almost nowhere* if it only holds on a null set. We will refer to measurable functions whose definition only makes sense almost everywhere as if they were defined on all of X .

Example 3.2.4. Let μ be Lebesgue measure on \mathbb{R} and $f(x) = 1/x$. Then $f(0)$ is not defined, but $\{0\}$ is a μ -null set. So we can view f as a function $\mathbb{R} \rightarrow \mathbb{C}$, even though it is only defined almost everywhere.

3.2.5. Because μ -null sets are not very important, we want to view two functions that are equal almost everywhere as actually being the same function, modulo “measurement error”. For example, if $u(x)$ denotes the temperature in the air at a point $x \in \mathbb{R}^3$ measured by some thermometer, and we measure that $u(x) = 0$ at every x close to 0, but $u(0) = 1$, then we must have made an error in the measurement of $u(0)$, and might as well view this measurement u as “the same as” the measurement of temperature which is identically 0 in a neighborhood of 0.

Definition 3.2.6. Let μ be a measure on Σ . A sequence of functions $f_n : X \rightarrow B$ is said to *converge pointwise almost everywhere* with respect to μ , or simply *almost converge*, to a function $f : X \rightarrow B$ if there is a null set Z such that on $X \setminus Z$, $f_n \rightarrow f$ pointwise. In this case, we write

$$f = \lim_{n \rightarrow \infty} f_n,$$

noting that the limit is meant almost everywhere if unclear from context.

3.2.7. Note that in Definition 3.2.6, we allow the functions f_n , or their limit f , to be undefined on a null set, which is then viewed as a subset of the bad set Z where the f_n may not converge to f .

Example 3.2.8. Any sequence of functions which converges pointwise almost converges. As an example of a sequence of functions which almost converges but doesn’t converge pointwise, let $f_n(x) = x^n$, $X = [0, 1]$. Then $f_n \rightarrow 0$ everywhere except 1, so $f_n \rightarrow 0$ almost everywhere with respect to Lebesgue measure.

Definition 3.2.9. A *measurable function* $f : X \rightarrow B$ is a function such that there is a sequence $f_n \in \mathbf{Simp}(X \rightarrow B)$ such that $f_n \rightarrow f$ almost everywhere. Let $\mathcal{M}(X \rightarrow B)$ denote the set of measurable functions.

3.2.10. The notation $\mathcal{M}(X \rightarrow B)$ will make more sense when we define the L^p -spaces later on. We will abuse notation and write \mathcal{M} if X, B are clear from context, and may also write something like $\mathcal{M}(X \rightarrow B, \mu)$ if μ is not clear from context.

Lemma 3.2.11. $\mathcal{M}(X \rightarrow B)$ is a vector space. In particular, $\mathcal{M}(X \rightarrow \mathbb{C})$ is an algebra such that if $f, g \in \mathcal{M}(X \rightarrow \mathbb{C})$, then so are $\max(f, g)$, $\min(f, g)$, and $|f|$. Moreover, if $f \in \mathcal{M}(X \rightarrow B)$, then $x \mapsto \|f(x)\|$ is in $\mathcal{M}(X \rightarrow \mathbb{C})$.

Proof. Let $f, g \in \mathcal{M}(X \rightarrow B)$, and suppose that $f_n \in \mathbf{Simp}(X \rightarrow B)$, $f_n \rightarrow f$. Similarly let $g_n \rightarrow g$, $g_n \in \mathbf{Simp}(X \rightarrow B)$. Then $f_n + g_n \in \mathbf{Simp}(X \rightarrow B)$ by Lemma 3.1.8 and $f_n + g_n \rightarrow f + g$.

We leave the other claims as an exercise for the reader. \square

3.2.12. Let N be the set of all functions f which are zero almost everywhere, thus there is a null set Z such that on $X \setminus Z$, $f = 0$.

Lemma 3.2.13. The set N of functions that are zero almost everywhere is a vector subspace of $\mathcal{M}(X \rightarrow B)$.

Proof. Let $f \in N$, and suppose that Z is the set where f is nonzero. We first claim that f is measurable; in fact, the sequence $f_n = 0$ converges to f pointwise except on Z , hence almost everywhere.

Now if $g \in N$, and g is nonzero on a set W , then $f + g$ is nonzero on a subset of the null set $Z \cup W$; since μ is complete, any subset of $Z \cup W$ is null. The argument for scalars is similar. Clearly $0 \in N$ so N is nonempty. \square

3.2.14. Whenever W is a vector subspace of a vector space V , we may form its quotient space V/W of equivalence classes, where two elements $f, g \in V$ are viewed as equivalent if $f - g \in W$. In particular, if we take the quotient $\mathcal{M}(X \rightarrow B)/N$, two functions f, g are equivalent iff $f - g$ is zero almost everywhere.

Definition 3.2.15. Let N be the space of all functions that are zero almost everywhere is a vector subspace. We denote its quotient space

$$M(X \rightarrow B) = \frac{\mathcal{M}(X \rightarrow B)}{N}.$$

We will abuse terminology and refer to equivalence classes $f \in M(X \rightarrow B)$ as “functions”, and a representative of an equivalence class f as a *version* of f .

3.2.16. Again, we may write $M(X \rightarrow B, \mu)$ and similar notations to mean $M(X \rightarrow B)$. In general, we will want to work with M whenever possible rather than \mathcal{M} .

Exercise 3.2.17. Let μ be a probability measure on a measurable space X . Let $(f_n)_n$ be a sequence of measurable functions on (X, μ) , and f a measurable function on (X, μ) . Assume that for every $\varepsilon > 0$, one has

$$\sum_{n=1}^{\infty} \mu(\{x \in X : \|f_n(x) - f(x)\| > \varepsilon\}) < \infty.$$

Show that $f_n \rightarrow f$ almost everywhere.

This characterization is highly useful in probability, where one is typically not able to refer directly to elements of X , but only measurable sets, but still wants to be able to discuss convergence almost everywhere. (Hint: Use the Borel-Cantelli lemma, Exercise 3.1.12.)

3.3 Characterizing measurable functions

The current definition of M is unwieldy. Its elements are equivalence classes of functions, themselves defined to be the limits of simple functions, whose definition was natural but already little long. Here we give another characterization of measurability that is somewhat easier to work with, and will readily imply that every function that is relevant to analysis is measurable.

3.3.1. Throughout, we as usual fix a complete measured space (X, Σ, μ) and a Banach space B .

Definition 3.3.2. A function $f : X \rightarrow B$ is *almost separably valued* if there is a null set Z such that $f(X \setminus Z)$ is separable in the topology of B .

3.3.3. Being almost separably valued is a good condition. It means that, modulo a harmless null set, the image of f consists of points which can be approximated by points that lie in a countable set C ; and elements of C then are likely to admit finitary descriptions. Think of how difficult \mathbb{R} would be to work with if we did not have \mathbb{Q} , whose elements are described as pairs of natural numbers!

3.3.4. Thankfully, most Banach spaces that arise naturally in analysis turn out to be separable; certainly any finite-dimensional vector space has this property, and all but one Banach space that we will consider in this text will be separable. Certainly any function into a separable Banach space is almost separably valued. So Definition 3.3.2 will turn out to be a slightly annoying technical condition, and not at all of import, in practice. The reader who is only interested in the case $B = \mathbb{C}$, which is reasonable to do on one's first reading, can forget about this hypothesis altogether.

Definition 3.3.5. The *carrier*¹ of a function $f : X \rightarrow B$ is the set $\{x \in X : f(x) \neq 0\}$.

3.3.6. Now B has a norm, so it has open balls $B(x, r) = \{y \in B : \|x - y\| < r\}$, and in particular B has a topology: its open sets are the unions of the balls $B(x, r)$. So we can define its Borel σ -algebra in the usual way: it is the smallest σ -algebra containing the topology of B .

3.3.7. Our goal in this section is to prove the following theorem:

Theorem 3.3.8. Let $f : X \rightarrow B$ be a function with carrier C , possibly only defined almost everywhere. Then the following are equivalent:

1. f is measurable.
2. f is almost separably valued and for every open set $U \subseteq B$, $f^{-1}(U) \cap C$ is measurable.
3. f is almost separably valued and for every closed set $K \subseteq B$, $f^{-1}(K) \cap C$ is measurable.
4. f is almost separably valued and for every Borel set $W \subseteq B$, $f^{-1}(W) \cap C$ is measurable.

¹Some books prefer the term “support”, but we use “support” to mean the closure of the carrier, whenever (X, μ) has a topology.

Corollary 3.3.9. Let Γ be the Borel σ -algebra of B , so $B = (B, \Gamma)$ is a measurable space whose measurable sets are exactly the Borel sets. Then a function $f : X \rightarrow B$ is measurable iff f is almost separably valued and for every measurable $Y \subseteq B$, $f^{-1}(Y)$ is measurable.

3.3.10. The reader should compare Corollary 3.3.9 to the result which says that a function f is continuous iff the preimage of an open set is open. It says that a measurable function $X \rightarrow B$ is, modulo sets of measure zero, the same thing as a measurable map $X \rightarrow B$, where B is equipped with its Borel σ -algebra.

3.3.11. Before we prove Theorem 3.3.8, we need several lemmata which are useful in their own right.

Lemma 3.3.12. Let $f_n : X \rightarrow B$ be a sequence of functions converging pointwise to a function f . For every open $U \subseteq B$, we define $U_n = \{y \in U : \inf_{x \notin U} \|x - y\| > 1/n\}$. Then

$$f^{-1}(U) = \bigcup_{n=1}^{\infty} \bigcup_{K=1}^{\infty} \bigcap_{k=K}^{\infty} f_k^{-1}(U_n).$$

TODO: Draw a picture of U_n

Proof. The following are equivalent:

1. $x \in f^{-1}(U)$.
2. $f(x) \in U$.
3. There are n, K such that for every $k \geq K$, $f_k(x) \in U_n$.
4. There are n, K such that for every $k \geq K$, $x \in f_k^{-1}(U_n)$.
5. There are n, K such that $x \in \bigcap_{k \geq K} f_k^{-1}(U_n)$.
6. $x \in \bigcup_n \bigcup_K \bigcap_{k \geq K} f_k^{-1}(U_n)$.

Indeed, for every $i \in \{1, \dots, 6\}$, the i th entry in the above list is clearly equivalent to the $i + 1$ th entry. \square

Definition 3.3.13. A function $f : X \rightarrow B$ is *separably valued* if the image of f is separable in B .

Lemma 3.3.14. Let $f : X \rightarrow B$ be a function with carrier C . Then the following are equivalent:

1. f is the pointwise limit of simple functions.
2. f is separably valued and for every open set $U \subseteq B$, $f^{-1}(U) \cap C$ is measurable.
3. f is separably valued and for every open ball $U \subseteq B$, $f^{-1}(U) \cap C$ is measurable.

Proof. We first show that 1 implies 2. Let $f_n \rightarrow f$ pointwise, $f_n \in \mathbf{Simp}(X \rightarrow B)$ and for every n , let $\{b_1^n, \dots, b_{k(n)}^n\}$ be the image of f_n . Let K be the closure of $K_0 = \{b_i^n : n \in \mathbb{N}, i \in \{1, \dots, k(n)\}\}$. Then K_0 is countable and dense in K , so K is separable; moreover, $f(X) \subseteq K$, so f is separably valued. Now let $U \subseteq B$ be an open set; then $f^{-1}(U) \cap C = f^{-1}(U \setminus \{0\})$, and $U \setminus \{0\}$ is open.

Thus we must show that if V is an open set which does not contain 0, $f^{-1}(V)$ is measurable. Let $V_k = \{y \in V : \inf_{x \notin V} \|x - y\| > 1/n\}$. Clearly $f_n^{-1}(V_k)$ is measurable since there are only finitely many points of $f_n(X)$ in V_k , each with a measurable preimage, and Lemma 3.3.12 implies

$$f^{-1}(V) = \bigcup_{n=1}^{\infty} \bigcup_{K=1}^{\infty} \bigcap_{k=K}^{\infty} f_k^{-1}(V_n)$$

which is measurable since the measurable sets form a σ -algebra.

Clearly 2 implies 3 so it suffices to show that 3 implies 1. Let $\{b_i : i \in \mathbb{N}\}$ be dense in $f(X)$. Let

$$C_{ij} = \{x \in C : \|f(x) - b_i\| < 1/j\}.$$

Then C_{ij} is a preimage of a union of open balls, so C_{ij} is measurable. Now it would be reasonable to define for every $x \in C_{ij}$, $f_n(x) = b_i$, except that the C_{ij} are not disjoint.

To rectify this problem, let

$$E_{ijn} = C_{ij} \setminus \bigcup_{\substack{(i,j) < (k,\ell) \leq (n,n) \\ 1 \leq i,j \leq n}} C_{k\ell}$$

where $(i,j) \leq (k,\ell)$ iff $j < \ell$ or $j = \ell$ and $i \leq k$. Then if n is fixed, the E_{ijn} are disjoint, $E_{ijn} \subseteq C_{ij}$. Now let

$$f_n = \sum_{i,j=1}^n b_i 1_{E_{ijn}}.$$

Sublemma 3.3.15. $f_n \rightarrow f$ pointwise.

Proof. Let $x \in X$. If $f(x) = 0$, then for every n , $f_n(x) = 0$, so $f_n(x) \rightarrow f(x)$.

Otherwise, $x \in C$. Let $\varepsilon > 0$. Let $N_1 > 1/\varepsilon$ and choose N_2 so that

$$\|f(x) - b_{N_2}\| < \frac{1}{N_1}.$$

Now let $N = \max(N_1, N_2)$. Then $x \in C_{N_2 N_1}$, so if $n > N$,

$$(k, \ell) = \max_{(N_1, N_2) \leq (i,j) \leq (n,n)} (i, j),$$

then $x \in E_{k\ell n}$. Therefore $f_n(x) = b_k$ and

$$\|f(x) - b_k\| < \frac{1}{\ell} \leq \frac{1}{N_1} < \varepsilon.$$

But $f_n(x) = b_k$, so $\|f_n(x) - f(x)\| < \varepsilon$. □

This implies 1. □

Lemma 3.3.16. Let $f : X \rightarrow B$ be a function with carrier C . The following are equivalent:

1. For every closed set $K \subseteq B$, $f^{-1}(K) \cap C$ is measurable.
2. For every open set $U \subseteq B$, $f^{-1}(U) \cap C$ is measurable.
3. For every Borel set $W \subseteq B$, $f^{-1}(W) \cap C$ is measurable.

Proof. Obviously 3 implies 1.

Now assume 1. Then let $U \subseteq B$ be open,

$$K_n = \{y \in B : \inf_{x \notin U} \|x - y\| \geq \frac{1}{n}\}.$$

Then K_n is closed and $U = \bigcup_n K_n$. But $f^{-1}(U) \cap C = \bigcup_n f^{-1}(K_n) \cap C$, and the $f^{-1}(K_n) \cap C$ are measurable, so 2 follows.

To see that 2 implies 3, note that the set Γ of all sets Y such that $f^{-1}(Y) \cap C$ is measurable is a σ -algebra. Indeed Γ contains \emptyset so is nonempty, is closed under complement since $f^{-1}(B \setminus Y) = X \setminus f^{-1}(Y)$, and is closed under countable union since $f^{-1}(\bigcup_n Y_n) = \bigcup_n f^{-1}(Y_n)$. Since Γ contains the open sets, Γ contains the Borel sets. □

Proof of Theorem 3.3.8. We can show that f is measurable iff f is almost separably valued and for every open set U , $f^{-1}(U) \cap C$ is measurable. In fact, Lemma then immediately shows that for every open set U , $f^{-1}(U) \cap C$ is measurable iff the same is true when “open set” is replaced with “closed set” or “Borel set”.

First assume that f is measurable. Then there are $f_n \in \mathbf{Simp}(X \rightarrow B)$ and a null set Z such that $f_n \rightarrow f$ pointwise on $X \setminus Z$ and f is defined on $X \setminus Z$. So $f_n 1_{X \setminus Z} \rightarrow f 1_{X \setminus Z}$ (where $f 1_{X \setminus Z} = 0$ if f is undefined), so $f 1_{X \setminus Z}$ meets the criteria of Lemma 3.3.14; since null sets are measurable, so does f .

Conversely, if Z is a null set such that $f(X \setminus Z)$ is separable and for every open set U , $f^{-1}(U) \cap C$ is measurable, then $f 1_{X \setminus Z}$ is separably valued and for every open set U , $(f 1_{X \setminus Z})^{-1}(U) \cap C$ is measurable, so Lemma 3.3.14 implies the claim. □

We now prove a very important theorem about measurable functions. Note that the analogous result for continuous functions (or even Riemann integrable functions) does not hold.

Theorem 3.3.17. Suppose that f_n are measurable functions which almost converge to a function f . Then f is measurable.

Proof. We apply Newberger’s theorem. Since the f_n are almost separably valued, there are null sets Z_n such that $K = \overline{\bigcup_n f_n(X \setminus Z_n)}$ is closed. Therefore f is almost separably valued. Then Lemma 3.3.12 implies that for every open set $U \subseteq B$,

$$f^{-1}(U) \cap C = \bigcup_{n=1}^{\infty} \bigcup_{K=1}^{\infty} \bigcup_{k=K}^{\infty} f_n^{-1}(U_n) \cap C$$

which implies that $f^{-1}(U)$ is a countable union of countable unions of countable intersections of measurable sets, hence is measurable. □

3.3.18. Let \mathbb{R} be equipped with its usual Lebesgue measure (or really any Borel measure μ on any space with a countable dense subset, such that for every countable set Z , $\mu(Z) = 0$). Then any continuous function $\mathbb{R} \rightarrow B$ pulls back open sets to open (hence Borel) sets and has separable image (since the image of \mathbb{Q} is dense in the image of \mathbb{R}), hence is measurable by Theorem 3.3.8. Any pointwise limit of continuous functions is also measurable; for example any function with only a discrete set of discontinuities, or with only jump discontinuities. A monotone function has only jump discontinuities, so is measurable. A similar argument applies for any left-continuous or right-continuous function. And of course, we can modify any of the above functions on a countable set (or any null set!) to get another measurable function. See Exercise 3.3.22.

TODO: Draw a picture of continuous functions converging to a finitely disct function or a function with jump disc.

Example 3.3.19. An example of a function which is not almost separably valued is given by Example A.4.3. Let δ be as in that example and let $f(x) = \delta_x$. Then for any uncountable $Y \subseteq \mathbb{R}$, $f(Y)$ is an uncountable discrete set, so removing a null set cannot possibly help us here.

Example 3.3.20 (Vitali's set). Recall that we defined Lebesgue measure on \mathbb{T} , μ , in Example 2.1.4. If we rotate \mathbb{T} by an angle θ , then the rotation preserves μ ; that is, if $A \subseteq \mathbb{T}$ is measurable, and $A + \theta = \{e^{i(\varphi+\theta)} : e^{i\varphi} \in A\}$, then $\mu(A) = \mu(A + \theta)$. The group \mathbb{Q} of rational numbers under addition acts on A by rotation: if $q \in \mathbb{Q}$, define a map $e^{i\theta} \mapsto e^{i(\theta+2\pi q)}$ from the circle to itself.

Let \mathcal{O} be the set of all orbits of the action of \mathbb{Q} ; that is, elements of \mathcal{O} are sets

$$\{e^{i(\varphi+2\pi q)} : q \in \mathbb{Q}\}.$$

Each orbit is countable since it is indexed by \mathbb{Q} , yet \mathbb{T} is uncountable, so \mathcal{O} must be uncountable. Now the axiom of choice B.1.19 implies that there is a set $X \subset \mathbb{T}$ which contains exactly one element from each orbit in \mathcal{O} . Suppose that X is measurable. Then $\bigcup_{q \in \mathbb{Q}} X + q = \mathbb{T}$ and the union is disjoint, so

$$1 = \mu(\mathbb{T}) = \sum_{q \in \mathbb{Q}} \mu(X + q) = \sum_{q \in \mathbb{Q}} \mu(X),$$

so $\mu(X) > 0$ (so that it sums to 1) yet $\mu(X) = 0$ (since the sum of infinitely many copies of any nonzero number is infinite and hence > 1), a contradiction.

Therefore X is nonmeasurable, and so is its indicator function 1_X .

3.3.21. But should you, dear reader, worry about nonmeasurable functions? Example 3.3.18 shows that functions that appear in mainstream analysis, algebra, or number theory are measurable. Example 3.3.19 does suggest that in deeply infinitary parts of functional analysis, one may have to worry about functions which are not separably valued, and Vitali's set shows that in logic, where the axiom of choice (and its evil friends such as the axiom of power set and the axiom schema of replacement) is of import, nonmeasurable sets may emerge. But (TODO cite me) there exist models of set theory with certain axioms weakened where Vitali's construction fails and every almost separably valued function is measurable, and someone

with a more constructivist bent may conclude from that theorem that indeed every function is measurable in “reality” (whatever that means). Strichartz (TODO cite me) summarized these observations by remarking that “wise-guys who like using the axiom of choice will have to worry about ... wolves under the bed, etc.”

Exercise 3.3.22. Prove the assertions in Example 3.3.18.

Definition 3.3.23. Define the *Baire space* B_n inductively: let B_0 be the space of continuous functions $[0, 1] \rightarrow \mathbb{C}$, and let B_{n+1} be the space of pointwise limits of functions in B_n . If $f \in B_n$, we say that f is *n-Baire*.

Unfortunately, there is another notion of Baire space, having to do with the Baire category theorem, which is not the same as the Baire space that we just defined.

Exercise 3.3.24. Show that every Baire function is Borel. (The converse is true if one allows for α -Baire functions, α any countable ordinal, but you are not being asked to prove that.)

Exercise 3.3.25. Show that the derivative of a differentiable function is 1-Baire. Show that the function $1_{\mathbb{Q} \cap [0,1]}$ is 2-Baire but not 1-Baire.

Exercise 3.3.26. Find a function which is 3-Baire but not 2-Baire. More generally, find a function which is $(n + 1)$ -Baire but not n -Baire.

3.4 Convergence of measurable functions

We have already discussed two means by which measurable functions may converge to other measurable functions: pointwise and almost pointwise. From this point onwards, pointwise convergence will be largely irrelevant; following our philosophy that null sets are important, almost pointwise will usually be the desired property.

3.4.1. Throughout, we fix a complete measured space (X, Σ, μ) and a Banach space B .

Definition 3.4.2. A sequence of functions f_n converge to a function f *uniformly* if for every $\varepsilon > 0$ there is an N such that for every $n > N$, $\sup \|f_n - f\|_B < \varepsilon$.

3.4.3. Now, by analogy with the notion of a property holding “almost everywhere” (everywhere except a null set), we introduce the notion of a property holding *nearly everywhere*; that is, everywhere except a set of measure $\varepsilon > 0$. The property holds for arbitrarily small ε , but the parameters in the property (the δ s, N s, and so on) may become arbitrarily “bad” as $\varepsilon \rightarrow 0$.

Definition 3.4.4. A sequence of functions $f_n : X \rightarrow B$ *converges nearly uniformly* to a function $f : X \rightarrow B$ if for every $\varepsilon > 0$ there is a set E_ε such that $|\mu|(X \setminus E_\varepsilon) < \varepsilon$ and $f_n \rightarrow f$ uniformly on E_ε .

3.4.5. Thus $f_n \rightarrow f$ nearly uniformly iff for every $\varepsilon > 0$ there is a E such that for every $\delta > 0$ there is an N such that $|\mu|(E) < \varepsilon$ and for every $n > N$, $\sup_E \|f_n - f\|_B < \delta$. Clearly a sequence which converges nearly uniformly converges almost pointwise; indeed,

the sets E_ε where the sequence fails to converge uniformly have a null intersection E , and if the sequence fails to converge pointwise at a point x , then $x \in E$. In particular, the nearly uniform limit of a sequence of measurable functions is measurable.

Lemma 3.4.6. Suppose that $f_n : X \rightarrow B$ and $g_n : X \rightarrow B$, $f_n \rightarrow f$ and $g_n \rightarrow g$ nearly uniformly. Then the following limits also hold nearly uniformly:

1. $f_n + g_n \rightarrow f + g$.
2. For every $c \in \mathbb{C}$, $cf_n \rightarrow cf$.
3. If $B = \mathbb{C}$, $f_n g_n \rightarrow fg$.
4. $(x \mapsto \|f_n(x)\|_B) \rightarrow (x \mapsto \|f(x)\|_B)$.
5. $\max(f_n, g_n) \rightarrow \max(f, g)$.
6. $\min(f_n, g_n) \rightarrow \min(f, g)$.

Proof. Routine and omitted. □

Theorem 3.4.7 (Egorov). Let X be a finite measure space (thus $\mu(X) < \infty$) and suppose that $f_n \in M(X \rightarrow B)$ almost converge to f , and the f_n are measurable. Then $f_n \rightarrow f$ nearly uniformly.

Proof. After throwing away a harmless null set, we may assume that $f_n \rightarrow f$ pointwise. Now define

$$E_m^n = \{x \in X : \exists k \geq n (\|f(x) - f_k(x)\|_B \geq \frac{1}{m})\}.$$

Since $x \mapsto \|f(x) - f_k(x)\|_B$ is measurable, by Theorem 3.3.8, E_m^n is measurable. Moreover, if m is fixed, the E_m^n shrink as n increases and $\bigcap_n E_m^n = \emptyset$, since $f_n \rightarrow f$. By Lemma TODO, since $|\mu|(X) < \infty$, $\lim_n \mu(E_m^n) = 0$. So for every $\varepsilon > 0$ and every m we may find $n(m)$ such that

$$\mu(E_m^{n(m)}) < \frac{\varepsilon}{2^m}.$$

Now let $F = E \setminus \bigcup_m E_m^{n(m)}$, so

$$\mu(E \setminus F) \leq \sum_{m=1}^{\infty} \mu(E_m^{n(m)}) < \varepsilon.$$

So it suffices to show that for every $\delta > 0$ there is a N such that for every $n > N$, $\sup_F \|f_n - f\|_B < \delta$. Indeed, if $1/m < \delta$ and $N = n(m)$, then for every $x \in F$, $x \notin E_m^N$, so if $n > N$ then $\|f(x) - f_n(x)\|_B < \delta$. □

Definition 3.4.8. A sequence of functions $f_n : X \rightarrow B$ is said to be a *nearly uniform Cauchy sequence* if for every $\varepsilon > 0$ there is a measurable set E_ε such that $\mu(X \setminus E_\varepsilon) < \varepsilon$ and for every $\delta > 0$ there is an N such that for every $n_1, n_2 > N$, $\sup \|f_{n_1} - f_{n_2}\|_B < \delta$.

3.4.9. Since the hypothesis of being nearly uniform Cauchy is not altered if we change the functions f_n on a null set, we can work with equivalence classes of functions (equivalent iff equal almost everywhere) rather than functions themselves. This will be important when we demand that the limit of a Cauchy sequence be unique; it will not be a unique function everywhere, but only almost everywhere.

Lemma 3.4.10. Let $f_n \in M(X \rightarrow B)$ be a nearly uniform Cauchy sequence. Then there is a unique $f \in M(X \rightarrow B)$ such that $f_n \rightarrow f$ nearly uniformly.

Proof. For every m we can find E_m such that $|\mu|(X \setminus E_m) < 1/m$ and f_n is uniformly Cauchy on E_m . Let $E = \bigcup_m E_m$. Then $|\mu|(X \setminus E) < 1/m$ for every m , so $|\mu|(X \setminus E) = 0$, and it is okay if we leave f undefined on $X \setminus E$. As for if $x \in E$, we can choose an m such that $x \in E_m$. Since the f_n are a nearly uniform Cauchy sequence, the $f_n(x)$ are a Cauchy sequence, which converge to some $y \in B$ since B is a Banach space. Therefore we may let $f(x) = y$.

Now the $f_n \rightarrow f$ nearly uniformly. In fact, for every $\varepsilon > 0$ we may take $m > 1/\varepsilon$ and let $E_\varepsilon = E_m$. Then $f_n \rightarrow f$ uniformly on E_ε .

As for uniqueness, the $f_n \rightarrow f$ almost pointwise, and pointwise limits are unique (so that almost pointwise limits are unique almost everywhere). But f was only defined up to measure zero, so this is no loss. \square

Definition 3.4.11. Let $f_n : X \rightarrow B$ be a sequence of measurable functions, $f : X \rightarrow B$. We say that f_n converges in measure to f if for every $\varepsilon > 0$ the set

$$\lim_{n \rightarrow \infty} |\mu|(\{x \in X : \|f_n(x) - f(x)\|_B > \varepsilon\}) = 0.$$

If μ is a probability measure, we instead say that μ converges in probability.

3.4.12. We sometimes abbreviate the set in the above definition as $\{\|f_n - f\|_B > \varepsilon\}$. In other words, $f_n \rightarrow f$ in measure iff for every $\varepsilon > 0$ and every $\delta > 0$ there is an N such that for every $n > N$,

$$|\mu|(\{\|f_n - f\|_B > \varepsilon\}) < \delta,$$

and collapsing quantifiers this happens iff for every $\varepsilon > 0$ there is an N such that for every $n > N$,

$$|\mu|(\{\|f_n - f\|_B > \varepsilon\}) < \varepsilon.$$

3.4.13. What is the intuition for convergence in measure? Suppose that we are scientists running experiments in an attempt to compute the value of a function f . As the number of test subjects n goes to infinity, the experimental data f_n should converge to f , but in what sense? Let X be the set of possible outcomes and $P(E)$ the probability that one of the outcomes in E occurs, thus P is a probability measure. Then $P(\{\|f_n(x) - f(x)\|_B > \varepsilon\})$ is the probability that we got an experimental error of size at least ε ; as $n \rightarrow \infty$, this probability becomes vanishingly small. However, on the off-chance that an error of size at least ε occurs, we have no control over how bad the error may be! Thus $f_n \rightarrow f$ in probability and a priori we can prove no stronger.

Lemma 3.4.14. If $f_n \rightarrow f$ in measure, then f is unique almost everywhere, hence as an element of $M(X \rightarrow B)$.

Proof. Suppose that $f_n \rightarrow g$ in measure as well. Then for every $\varepsilon > 0$ and n ,

$$\{\|f - g\|_B > \varepsilon\} \subseteq \{\|f - f_n\|_B > \varepsilon/2\} \cup \{\|g - f_n\|_B > \varepsilon/2\}.$$

If n is large enough, the right hand side has measure at most ε . □

Definition 3.4.15. Let $f_n : X \rightarrow B$ be a sequence of functions. We say that the f_n are *Cauchy in measure* if for every $\varepsilon > 0$ there is a N such that for every $n_1, n_2 > N$,

$$|\mu|(\{\|f_{n_1} - f_{n_2}\|_B > \varepsilon\}) < \varepsilon.$$

Lemma 3.4.16. Suppose that $f_n : X \rightarrow B$ and $g_n : X \rightarrow B$, $f_n \rightarrow f$ and $g_n \rightarrow g$ in measure. Then the following limits also hold in measure:

1. $f_n + g_n \rightarrow f + g$.
2. For every $c \in \mathbb{C}$, $cf_n \rightarrow cf$.
3. If $B = \mathbb{C}$, $f_n g_n \rightarrow fg$.
4. $(x \mapsto \|f_n(x)\|_B) \rightarrow (x \mapsto \|f(x)\|_B)$.
5. $\max(f_n, g_n) \rightarrow \max(f, g)$.
6. $\min(f_n, g_n) \rightarrow \min(f, g)$.

Proof. Routine and omitted. □

3.4.17. How does convergence in measure relate to other modes of convergence? It does not imply almost pointwise convergence – imagine a sequence of functions racing back and forth along $[0, 1]$, their supports getting smaller with every time they turn around. TODO: Draw a picture. Nor does it follow from almost pointwise convergence on infinite measure sets – just take $f_n = 1_{[n, n+1]}$ as a counterexample. But convergence in measure is weaker than nearly uniform convergence, hence from almost pointwise convergence on finite measure sets. TODO: Draw a diagram.

Lemma 3.4.18. Suppose that $f_n \rightarrow f$ nearly uniformly; then $f_n \rightarrow f$ in measure.

Proof. Let $\varepsilon > 0$. Then there is a set E_ε on which $f_n \rightarrow f$ uniformly such that $\mu(X \setminus E_\varepsilon) < \varepsilon$, thus if n is large enough $\sup_{E_\varepsilon} |f_n - f| < \varepsilon$, so $\{\|f_n - f\|_B > \varepsilon\} \subseteq X \setminus E_\varepsilon$ and hence $\mu(\{\|f_n - f\|_B > \varepsilon\}) < \varepsilon$. So $f_n \rightarrow f$ in measure. □

Corollary 3.4.19. If $\mu(X) < \infty$ and $f_n \rightarrow f$ almost pointwise, then $f_n \rightarrow f$ in measure.

Proof. By Egorov's theorem and Lemma 3.4.18. □

3.4.20. We now come to a critical result which implies that Cauchy in measure not only implies convergence in measure, but other modes of convergence as well. This result has been called the *Riesz-Weyl theorem* or the *fundamental theorem of integration*.

Theorem 3.4.21 (fundamental theorem of integration). Suppose that f_n is a Cauchy sequence in measure. Then there is a subsequence of f_{n_k} and a unique $f \in M(X \rightarrow B)$ such that:

1. The f_{n_k} are a nearly uniform Cauchy sequence.
2. $f_{n_k} \rightarrow f$ nearly uniformly.
3. $f_n \rightarrow f$ in measure.

Proof. Let $n_1 = 1$, and choose $n_{k+1} > n_k$ such that if $m_1, m_2 \geq n_{k+1}$, then

$$|\mu|(\{\|f_{m_1} - f_{m_2}\|_B \geq 2^{-k}\}) < 2^{-k}.$$

Let $g_k = f_{n_k}$.

Sublemma 3.4.22. g_k is nearly uniformly Cauchy.

Proof. Let $\varepsilon > 0$ and choose K so that $\sum_{k \geq K} 2^{-k} < \varepsilon$. Let

$$F = X \setminus \bigcup_{k \geq K} \{\|g_k - g_{k+1}\|_B > 2^{-k}\}.$$

Then $\mu(X \setminus F) < \varepsilon$.

Now let $\delta > 0$ and choose N so large that $N \geq K$, $2^{1-N} < \delta$. Then if $j > \ell > N$, $x \in F$,

$$\begin{aligned} \|g_j(x) - g_\ell(x)\|_B &= \|g_j(x) - g_{j-1}(x) + g_{j-1}(x) - g_{j-2}(x) + \cdots + g_{\ell+1}(x) - g_\ell(x)\|_B \\ &\leq \sum_{i=\ell}^{j-1} \|g_{i+1}(x) - g_i(x)\|_B \\ &\leq \sum_{i=\ell}^{j-1} 2^{-i} < 2^{1-N} < \delta. \end{aligned}$$

Therefore the g_k are a uniform Cauchy sequence on F and hence nearly uniform on X . \square

So by Lemma 3.4.10, there is an f such that $\lim_k g_k = f$ nearly uniformly. But then Lemma 3.4.18 implies that $\lim_k g_k = f$ in measure. But

$$\{\|f - f_n(x)\|_B > \varepsilon\} \subseteq \{\|f - g_k\|_B > \varepsilon/2\} \cup \{\|f_n - g_k\|_B > \varepsilon/2\}, \quad (3.1)$$

and the g_k are a subsequence of the Cauchy-in-measure sequence f_n , hence the right hand side of (3.1) is $< \varepsilon$ if n, k are large enough. So $f_n \rightarrow f$ in measure. Uniqueness follows by Lemma 3.4.14. \square

Corollary 3.4.23. If $f_n \rightarrow f$ in measure, then there is a subsequence of f_{n_k} such that $f_{n_k} \rightarrow f$ nearly uniformly.

Proof. The f_n are Cauchy in measure, so the fundamental theorem of integration implies that there is a subsequence that is nearly uniformly Cauchy. Uniqueness of a nearly uniform limit implies that the subsequence must converge to f . \square

Exercise 3.4.24. Find an example which shows that the hypothesis of finite measure in Egorov's theorem cannot be omitted.

3.5 Regularity of measurable functions

Before we define the integral, we pause to use Egorov's theorem to prove a partial converse to the theorem which said that every continuous function was measurable. Of course not every measurable function is continuous (most measure spaces don't even come with a topology, but also, any familiar discontinuous function on \mathbb{R} will be measurable), but we will do the best that we can this section.

3.5.1. Throughout this section, fix a locally compact Hausdorff space X . If the reader is not familiar with such notions, they can take $X = \mathbb{R}^c$.

3.5.2. Recall that a function $f : \mathbb{R}^c \rightarrow \mathbb{C}^d$ is *smooth* if for every point x and every vector of natural numbers (k_1, \dots, k_c) , the partial derivative

$$\partial_1^{k_1} \partial_2^{k_2} \dots \partial_c^{k_c} f(x)$$

exists; here $\partial_i^{k_i}$ is the operator that takes a function to its k_i th partial derivative along its i th basis vector. Clearly every smooth function is differentiable and hence continuous.

Lemma 3.5.3 (Urysohn). Let $K_0, K_1 \subseteq X$ be disjoint closed sets. Then there is a continuous function $f : X \rightarrow [0, 1]$ such that $f|_{K_0} = 0$ and $f|_{K_1} = 1$. Moreover, if $X = \mathbb{R}^c$, we can even assume that f is smooth.

3.5.4. We refer the reader to a book on topology such as (TODO cite me) for the proof of Urysohn's lemma. (Or in an appendix?) We will also need the fact that in every Hausdorff space, a compact set is closed. Again we refer to (TODO cite me) for the proof, though when $X = \mathbb{R}^c$, this follows from the Heine-Borel theorem.

3.5.5. Now the open sets of X generate a topology, namely the Borel σ -algebra of X . Therefore X is a measurable space in a natural way. By a *Borel measure* on X we mean a measure defined on the Borel σ -algebra of X .

Lemma 3.5.6. Every continuous function is measurable with respect to the Borel σ -algebra.

Proof. Let $f : X \rightarrow B$ be continuous; then the f -preimage of an open set is open, hence Borel, hence measurable. We now appeal to Theorem 3.3.8. \square

Definition 3.5.7. View X as a complete measured space (X, Σ, μ) , where μ is a Borel measure and Σ is the σ -algebra of μ -measurable sets (so that Σ is generated by the Borel sets and μ -null sets). Let $f : X \rightarrow B$ be a function. We say that f is a *nearly continuous function* if for every $\varepsilon > 0$ there is a set E such that $\mu(X \setminus E) < \varepsilon$ and $f|_E$ is continuous.

3.5.8. You should check that your favorite discontinuous function (that isn't the indicator function of the Vitali set) is nearly continuous. For example the function $x \mapsto 1/x$ is nearly continuous because if we discard a small neighborhood of 0, then it is continuous.

3.5.9. Throughout the rest of the section, we fix a positive Radon measure μ on X , and take its completion, so if Σ denotes the σ -algebra generated by Borel sets and μ -null sets, then (X, Σ, μ) is a complete measured space, which we also denote by X . Thus X is equipped with a topology, a σ -algebra, and a measure; so X is a lot like the real line \mathbb{R} , and when we refer to measurable functions, nearly continuous functions, and so on, we do so with respect to Σ .

Lemma 3.5.10. If $f : X \rightarrow \mathbb{C}^d$ is a measurable function, then there is a sequence of continuous functions $f_n : X \rightarrow \mathbb{C}^d$ such that $f_n \rightarrow f$ almost pointwise. If $X = \mathbb{R}^c$ then we can even take the f_n to be smooth.

Proof. We first check this when f is the indicator function of a compact set K . By outer regularity, for every n there is an open set $U_n \supseteq K$ such that $\mu(U_n \setminus K) < 1/n$, and after taking intersections we can assume that $U_n \supseteq U_{n+1}$. The complement $X \setminus U_n$ is closed, so by Urysohn's lemma there is a continuous function $f_n : X \rightarrow [0, 1]$, smooth if $X = \mathbb{R}^c$, such that $f_n|_K = 1$ and $f_n|_{(X \setminus U_n)} = 0$. Now $f_n \rightarrow f$ almost pointwise. Indeed, if $x \in K$, then for every n , $f_n(x) = 1$; otherwise, unless x is in $\bigcap_n U_n \setminus K$, which is a null set, there is an N such that $x \notin U_N$, and hence for every $n > N$, $x \notin U_n$, so $f_n(x) = 0$. TODO: Draw a picture.

We now check when f is the indicator function of an open set U . By inner regularity, there is a sequence of compact sets K_m such that $K_m \subseteq K_{m+1}$ and $\mu(U \setminus K_m) < 1/m$. In particular, $\lim_m 1_{K_m} = 1_U$ almost pointwise. Now there are sequences of continuous functions (smooth if $X = \mathbb{R}^c$) $f_n^m : X \rightarrow [0, 1]$ such that $\lim_n f_n^m = 1_{K_m}$ almost pointwise, thus

$$|f_n^n(x) - 1_U(x)| \leq |f_n^n(x) - 1_{K_n}(x)| + |1_{K_n}(x) - 1_U(x)|.$$

The right-hand side vanishes as $n \rightarrow \infty$ so $f_n^n \rightarrow 1_U$ almost pointwise. TODO: Draw a picture.

A similar argument applies when f is the indicator function of a Borel set W . Indeed, by outer regularity, there is a sequence of open sets U_m such that $U_m \supseteq W$, $\mu(U_m \setminus W) < 1/m$. We can find a sequence of continuous (smooth?) functions $f_n^m : X \rightarrow [0, 1]$ such that $\lim_n f_n^m = 1_{U_m}$ almost pointwise, and $\lim_m 1_{U_m} = 1_W$ almost pointwise, so $f_n^n \rightarrow 1_W$ almost pointwise.

If f is the indicator function of a measurable set, then we can modify f on a null set and replace it with the indicator function of a Borel set.

If $f \in \mathbf{Simp}(X \rightarrow \mathbb{C})$, then f is a linear combination of indicator functions of measurable sets f_1, \dots, f_n , so we can find sequences approximating each of the summands f_i and use linearity.

If f is an arbitrary measurable function $X \rightarrow \mathbb{C}$, we can approximate f by simple functions.

If f is a vector of measurable functions $X \rightarrow \mathbb{C}^d$, we can approximate each of the components of f by continuous (smooth?) functions. \square

Theorem 3.5.11 (Luzin). If $\mu(X) < \infty$, then a function $f : X \rightarrow \mathbb{C}^d$ is measurable iff f is nearly continuous.

Proof. If f is nearly continuous, then for every $\varepsilon > 0$ we can find a set E_ε on which f is continuous, hence measurable, and $\mu(X \setminus E_\varepsilon) < \varepsilon$. Taking their union, we conclude that f is measurable at almost every point of X , and hence everywhere on X . Note that this direction was already valid for any Borel measure and any Banach space codomain.

Conversely, suppose that f is measurable. By Lemma 3.5.10, we can find a sequence of continuous (or even smooth) functions f_n with $f_n \rightarrow f$ almost pointwise, and hence nearly uniformly by Egorov's theorem. The uniform limit of a sequence of functions is continuous, so the nearly uniform limit is nearly continuous. \square

Exercise 3.5.12. Show that there exists a measurable function on \mathbb{R}^d which is not nearly smooth.

Exercise 3.5.13. Let μ be a σ -finite positive Radon measure on \mathbb{R}^d . Show that for every Borel set A such that $\mu(A) > 0$, there is an $R > 0$ such that $0 < \mu(B(0, R) \cap A) < \infty$. Since this is in particular true for Lebesgue measure, it is often no loss of generality to suppose that the sets one is working with have finite Lebesgue measure.

Exercise 3.5.14. Let μ be Lebesgue measure on \mathbb{R} . Show that there is a Borel set $A \subseteq [0, 1]$ such that for every open interval $U \subseteq [0, 1]$, $0 < \mu(A \cap U) < \mu(U)$. (Hint: Start with a fat Cantor set, c.f. Exercise 1.6.26, and then put more fat Cantor sets in the leftover intervals, but be careful.)

Exercise 3.5.15. Let μ be Lebesgue measure on \mathbb{R} . Let $A \subseteq \mathbb{R}$ be a Borel set such that $\mu(A) > 0$. Show that for every $\varepsilon > 0$ there is an open interval U such that $\mu(U \cap A) > (1 - \varepsilon)\mu(U)$. (Hint: Use the pigeonhole principle.)

Exercise 3.5.16. Prove *Steinhaus' theorem*: Let A be a Borel subset of \mathbb{R} with positive Lebesgue measure. Then the set $\{x - y : (x, y) \in A^2\}$ contains an open set U such that $0 \in U$.

Exercise 3.5.17. Let $G \subset \mathbb{R}$ be a proper subgroup of \mathbb{R} under addition. Show that if G is Borel then G is null. (Hint: use Exercise 3.5.16, Steinhaus' theorem).

3.6 Integration of simple functions

We are ready to define the integral, at least for simple functions.

3.6.1. A priori, our intuitive definition of integration as “the net signed area under the graph” is problematic. Consider the function $f = 1_{[0, \infty)} - 1_{(-\infty, 0]}$. What is the net signed area under the graph of f ? Well, to the left of 0, it is $-\infty$, and to the right it is $+\infty$, so we run into our usual pesky foe, $\infty - \infty$. To put off this problem for now, we dodge the issue by declaring that we will for now only try to integrate functions whose integrals will be finite.

3.6.2. For the rest of the chapter, fix a measure μ , either valued in \mathbb{C} (which is a one-dimensional Banach space) or $(-\infty, \infty]$. The reason that we do not allow μ to be more generally vector-valued is that we need to be able to multiply elements of the Banach space B by $\mu(E)$.

Definition 3.6.3. An *integrable simple function* is a function $f \in \mathbf{Simp}(X \rightarrow B)$ such that for every nonzero b in the image of f , $f^{-1}(b)$ has finite measure. We denote the set of integrable simple functions by $\mathbf{ISF}(X \rightarrow B)$.

3.6.4. We want the integral to be, at first, a linear map $\mathbf{ISF}(X \rightarrow B) \rightarrow B$. To motivate it, let's suppose that $B = \mathbb{C}$, $X = \mathbb{R}$, μ is Lebesgue measure, and E is an interval. Then $\int_{-\infty}^{\infty} 1_E$ had better be the area of the rectangle $E \times [0, 1]$ (TODO draw a picture), hence $\int_{-\infty}^{\infty} 1_E = \mu(E)$. In order for linearity to hold, if $c \in \mathbb{C}$, we must then have $\int_{-\infty}^{\infty} c 1_E = c\mu(E)$.

3.6.5. For every integrable simple function f , f can be written in terms of the indicator functions in a unique way; namely, if $\{b_1, \dots, b_n\}$ is the image of f ,

$$f(x) = \sum_{i=1}^n b_i 1_{f^{-1}(b_i)}(x). \quad (3.2)$$

In particular, the indicator functions span $\mathbf{ISF}(X \rightarrow \mathbb{C})$.

Definition 3.6.6. We call (3.2) the *canonical representation* of the integrable simple function f with image $\{b_1, \dots, b_n\}$.

Definition 3.6.7. Let $f \in \mathbf{ISF}(X \rightarrow B)$ and suppose that (3.2) is the canonical representation of f . Let E be a measurable set. We define the *integral* of f to be

$$\int_E f \, d\mu = \sum_{i=1}^n b_i \mu(f^{-1}(b_i) \cap E).$$

3.6.8. We will occasionally write $\int f$ or similar to mean $\int_X f \, d\mu$, but only when X and μ are understood. If we need a dummy variable, we may write $\int_E f(x) \, d\mu(x)$, and if μ is understood we may even write $\int_E f(x) \, dx$. If E is an interval $[a, b]$, we may write \int_a^b to mean \int_E . For example, once we will have adequately defined integration,

$$\int_0^{2\pi} \sin x \, dx = 0$$

as one would expect.

Lemma 3.6.9. Let $f, g \in \mathbf{ISF}(X \rightarrow B)$. Then:

1. *Linearity:* For every $c \in \mathbb{C}$,

$$\int_X cf + g \, d\mu = c \int_X f \, d\mu + \int_X g \, d\mu.$$

2. The *triangle inequality*:

$$\left\| \int f(x) \, d\mu(x) \right\|_B \leq \int \|f(x)\|_B \, d|\mu|(x).$$

3. The *change-of-variables formula*: Let (Y, ν) be a measured space. Suppose that $h : Y \rightarrow X$ is a measurable map and μ is the pushforward measure $\mu = h_*\nu$. Then

$$\int_X f \, d\mu = \int_Y f \circ h \, d\nu.$$

4. If $B = \mathbb{R}$ and $f \leq g$, then $\int f \leq \int g$.

5. If E, F are disjoint measurable sets, then

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

6. If $E \subseteq F$ are measurable sets, $B = \mathbb{R}$, and μ is a nonnegative measure, then

$$\int_E f \, d\mu \leq \int_F f \, d\mu.$$

7. $\int \|f(x)\|_B \, d|\mu|(x) = 0$ if and only if $f = 0$ almost everywhere.

Proof. See Exercise 3.6.19. □

3.6.10. We want to extend the integral from $\mathbf{ISF}(X \rightarrow B)$ to the space $M(X \rightarrow B)$ of all measurable functions. But we cannot merely define $\int f = \lim_n \int f_n$, where the f_n are simple functions that converge to f . Indeed, if $f_n = 1_{[n, n+1]}$ then $f_n \rightarrow 0$ pointwise, but $\lim_n \int f_n = 1$, and we certainly do not want $\int 0 = 1$! We need another notion of convergence, which will turn out to be closely related to convergence in measure.

Definition 3.6.11. Given $f \in \mathbf{ISF}(X \rightarrow B)$, let

$$\|f\|_1 = \int_X \|f(x)\|_B \, d|\mu|(x). \quad (3.3)$$

One calls $\|f\|_1$ the L^1 -norm of f .

3.6.12. As the reader should check, the L^1 norm is a seminorm. See Exercise 3.6.20. Now when we say that $f_n \rightarrow f$ in L^1 , we mean that $\|f_n - f\|_1 \rightarrow 0$. Similarly if we say that the f_n are Cauchy in L^1 , we mean that $\|f_n - f_m\|_1 \rightarrow 0$.

3.6.13. It is natural to want to extend the L^1 norm to the completion of the space $\mathbf{ISF}(X \rightarrow B)$, on which it will actually be a norm. (See Theorem A.2.5 for more on that.)

Definition 3.6.14. The completion of $\mathbf{ISF}(X \rightarrow B)$ is known as $L^1(X \rightarrow B)$.

3.6.15. Formally, $L^1(X \rightarrow B)$ consists of Cauchy sequences of simple functions modulo Cauchy equivalence. We will define the integral as a linear map $L^1(X \rightarrow B) \rightarrow B$:

Definition 3.6.16. Let $f \in L^1(X \rightarrow B)$ and suppose that E is a measurable set. Choose a Cauchy sequence of $f_n \in \mathbf{ISF}(X \rightarrow B)$ such that $f_n \rightarrow f$ in L^1 . We define the *integral* of f to be

$$\int_E f \, d\mu = \lim_{n \rightarrow \infty} \int_E f_n \, d\mu.$$

3.6.17. Let's check that Definition 3.6.16 makes sense. When we say that $f_n \rightarrow f$ in L^1 , we mean that f is the equivalence class of the Cauchy sequence $(f_n)_n$. So if there was another Cauchy sequence $(g_n)_n$ with $g_n \rightarrow f$ in L^1 ,

$$\begin{aligned} \left\| \lim_{n \rightarrow \infty} \int_E g_n - f_n \, d\mu \right\| &\leq \lim_{n \rightarrow \infty} \int_E \|g_n(x) - f_n(x)\|_B \, d|\mu|(x) \\ &\leq \lim_{n \rightarrow \infty} \int_X \|g_n(x) - f_n(x)\|_B \, d|\mu|(x) \\ &= \lim_{n \rightarrow \infty} \|g_n - f_n\|_1 = 0, \end{aligned}$$

the last equality following because the g_n and the f_n are Cauchy equivalent. Therefore the choice of Cauchy sequence in Lemma 3.6.16 is immaterial, and any choice will return the same integral. As a consequence, the L^1 norm extends to all of $L^1(X \rightarrow B)$ by the equation (3.3).

Lemma 3.6.18. The conclusion of Lemma 3.6.9 holds for any $f, g \in L^1(X \rightarrow B)$, not just $f, g \in \mathbf{ISF}(X \rightarrow B)$.

Proof. Left as Exercise 3.6.21. □

Exercise 3.6.19. Prove Lemma 3.6.9.

Exercise 3.6.20. Show that the L^1 -norm is a seminorm on \mathbf{ISF} , and that $\|f\|_1 = 0$ iff $f = 0$ almost everywhere.

Exercise 3.6.21. Prove Lemma 3.6.18.

Exercise 3.6.22. Let $A = \{1, \dots, d\}$ with counting measure. Therefore *every* function $A \rightarrow \mathbb{R}$ is an integrable simple function, and we can identify $f \in \mathbf{ISF}(A \rightarrow \mathbb{R})$ with the vector $(f(1), \dots, f(d))$. So we can identify $\mathbf{ISF}(A \rightarrow \mathbb{R})$ with \mathbb{R}^d . Show that the topology induced by the L^1 -norm on \mathbb{R}^d is the usual euclidean topology on \mathbb{R}^d . What is the unit ball in the L^1 -norm shaped like?

3.7 The integral in general

The conclusion of the previous section, wherein we treated the integral on $L^1(X \rightarrow B)$, was really quite silly. We want to integrate functions, not equivalence classes of Cauchy sequences of simple functions. This is analogous to how we want to study real numbers, not equivalence classes of Cauchy sequences of rational numbers. We need some sort of isomorphism which sends functions to members of L^1 .

Theorem 3.7.1 (fundamental theorem of integration, part II). Suppose that $f_n \rightarrow f$ in L^1 . Then there is a function f' such that $f_n \rightarrow f'$ in measure and there is a subsequence of f_{n_k} such that $f_{n_k} \rightarrow f'$ nearly uniformly, and hence almost pointwise.

Proof. We have to show that the f_n are Cauchy in measure; the other claims follow from the first part of the fundamental theorem of integration.

To see that the f_n are Cauchy in measure, we reason by contraposition, so suppose that we are given a subsequence f_{k_n} and a $\varepsilon > 0$ such that for every n, m ,

$$\{\|f_{k_m} - f_{k_n}\|_B > \varepsilon\} > \varepsilon.$$

Then

$$\begin{aligned}
\|f_{k_m} - f_{k_n}\|_1 &= \int_X \|f_{k_m}(x)f_{k_n}(x)\|_B d\mu(x) \\
&\geq \int_{\{\|f_{k_m} - f_{k_n}\|_B > \varepsilon\}} \|f_{k_m}(x)f_{k_n}(x)\|_B d\mu(x) \\
&\geq \int_{\{\|f_{k_m} - f_{k_n}\|_B > \varepsilon\}} \varepsilon d\mu(x) \\
&= \varepsilon \int_{\{\|f_{k_m} - f_{k_n}\|_B > \varepsilon\}} d\mu(x) \\
&= \varepsilon |\mu|(\{\|f_{k_m} - f_{k_n}\|_B > \varepsilon\}) > \varepsilon^2 > 0.
\end{aligned}$$

Therefore the f_n are not Cauchy in L^1 , so they do not converge in L^1 . \square

3.7.2. Now if $f_n \rightarrow f$ in L^1 and $f_n \rightarrow f'$ in measure, then it is tempting to identify f' with f , but we need to check that the choice of L^1 Cauchy sequence does not matter.

Theorem 3.7.3 (fundamental theorem of integration, part III). Suppose that $f_n \rightarrow f$ in L^1 and $f_n \rightarrow f'$ in measure. If $g_n \rightarrow f$ in L^1 , then $g_n \rightarrow f'$ in measure. Conversely, if the g_n are Cauchy in L^1 and $g_n \rightarrow f'$ in measure, then $g_n \rightarrow f$ in L^1 .

Proof. Suppose that $g_n \rightarrow f$ in L^1 . Then the sequence $(f_1, g_1, f_2, g_2, \dots)$ is Cauchy in L^1 , hence Cauchy in measure; but it has a subsequence which converges in measure to f' in measure, so the mother sequence must also converge in measure to f' , and hence every subsequence, including the g_n , must converge in measure to f' .

As for the converse, we use part I of the fundamental theorem of integration to show that there are subsequences f_{n_k}, g_{n_k} which converge nearly uniformly to f' , and are Cauchy in L^1 . Let $h_k = f_{n_k} - g_{n_k}$. Then the h_k is Cauchy in L^1 , converge nearly uniformly to 0, and if $h_k \rightarrow 0$ in L^1 then the f_{n_k} and g_{n_k} are Cauchy equivalent.

Lemma 3.7.4. $h_k \rightarrow 0$ in L^1 .

Proof. Let $\varepsilon > 0$, and choose N so that if $n_1, n_2 \geq N$ then $\|h_{n_1} - h_{n_2}\|_1 < \varepsilon$. We claim that $\|h_N\|_1 \lesssim \varepsilon$ where the implied constant only depends on the sequence and not on the index N , so that if $n > N$ then

$$\|h_n\|_1 \leq \|h_n - h_N\|_1 + \|h_N\|_1 \lesssim \varepsilon.$$

But $h_N \in \mathbf{ISF}(X \rightarrow B)$, so the carrier E of h_N satisfies $\mu(E) < \infty$, and h_N is bounded, thus $\|h_N(x)\| \lesssim 1$, where the implied constant does not depend on x or the index N , but only on the sequence.

Since $h_n \rightarrow 0$ nearly uniformly, there is a measurable set $F \subseteq E$ such that $|\mu|(E \setminus F) < \varepsilon$ and $h_n \rightarrow 0$ uniformly on F . Therefore

$$\int_{E \setminus F} \|h_n(x)\|_B d|\mu|(x) \lesssim \int_{E \setminus F} |\mu|(x) = |\mu|(E \setminus F) < \varepsilon.$$

Meanwhile, if N is large enough, then $\|h_N(x)\|_B < \varepsilon$ for every $x \in F$,

$$\int_F \|h_N(x)\|_B d|\mu|(x) < \varepsilon|\mu|(F) \leq \varepsilon|\mu|(E).$$

So

$$\|h_N\|_1 = \int_X \|h_N(x)\|_B d|\mu|(x) \lesssim \varepsilon$$

which was to be shown. \square

Hence the f_{n_k} and g_{n_k} are Cauchy equivalent. Since the mother sequences f_n and g_n are Cauchy, if they have subsequences that are Cauchy equivalent, then the f_n and g_n are Cauchy equivalent, so since $f_n \rightarrow f$ in L^1 , $g_n \rightarrow f$ in L^1 . \square

3.7.5. Summarizing, if $f \in L^1$, then there is an $f' \in M$ with the following property: for every L^1 Cauchy sequence $f_n \in \mathbf{ISF}$ such that $f_n \rightarrow f$ in L^1 , such that $f_n \rightarrow f'$ in measure, nearly uniformly along a subsequence, and almost pointwise along a subsequence. Moreover, L^1 is the completion of \mathbf{ISF} , so such a L^1 Cauchy sequence must exist.

Corollary 3.7.6. Let $f \in M$, and suppose the $f_n \in \mathbf{ISF}$ are an L^1 Cauchy sequence. Then the following are equivalent:

1. $f_n \rightarrow f$ in measure.
2. $f_n \rightarrow f$ nearly uniformly.
3. $f_n \rightarrow f$ almost everywhere.

Proof. That 1 implies 2 implies 3 is the content of the fundamental theorem of integration. Now if $f_n \rightarrow f$ almost everywhere, Egorov's theorem furnishes a subsequence f_{n_k} which converges to f nearly uniformly, hence in measure. But the f_n are Cauchy in L^1 , hence in measure; if the mother sequence is Cauchy and a subsequence converges, the mother sequence converges, so $f_n \rightarrow f$ in measure. \square

3.7.7. Elements of M are functions up to the equivalence relation of being equal almost everywhere. That is,

$$M = \frac{\mathcal{M}}{\mathcal{N}}$$

where \mathcal{M} is the space of all measurable functions and \mathcal{N} is the space of measurable functions which are zero almost everywhere. Therefore if $f_n \rightarrow f$ in L^1 and $f_n \rightarrow f'$ in measure (equivalently, nearly uniformly, or almost everywhere), we can actually identify f with f' and think of f' as a “function” by choosing a version of f' . Henceforth we will not make a distinction between elements of L^1 , elements f of M which admit an L^1 Cauchy sequence of simple functions which converge to f in measure, and versions of f . Therefore we are (finally!) ready to define the integral in general.

Definition 3.7.8. Let $f \in M(X \rightarrow B)$, and suppose that there is a L^1 Cauchy sequence $f_n \in \mathbf{ISF}(X \rightarrow B)$ such that $f_n \rightarrow f$ in measure. Then we say that $f \in L^1(X \rightarrow B)$, define for any measurable set E the *integral*

$$\int_E f \, d\mu = \lim_{n \rightarrow \infty} \int_E f_n \, d\mu,$$

and the L^1 norm

$$\|f\|_1 = \int_X \|f(x)\|_B \, d|\mu|(x).$$

If the integral $\int_X f \, d\mu$ exists, we say that f is *integrable* or *summable*.

Theorem 3.7.9. Let $f, g \in L^1(X \rightarrow B)$. Then:

1. *Linearity:* For every $c \in \mathbb{C}$,

$$\int_X cf + g \, d\mu = c \int_X f \, d\mu + \int_X g \, d\mu.$$

2. The *triangle inequality*:

$$\left\| \int_X f(x) \, d\mu(x) \right\|_B \leq \int_X \|f(x)\|_B \, d|\mu|(x).$$

3. The *change-of-variables formula*: Let (Y, ν) be a measured space. Suppose that $h : Y \rightarrow X$ is a measurable map and μ is the pushforward measure $\mu = h_*\nu$. Then

$$\int_X f \, d\mu = \int_Y f \circ h \, d\nu.$$

4. If $B = \mathbb{R}$ and $f \leq g$, then $\int f \leq \int g$.
5. If E, F are disjoint measurable sets, then

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

6. If $E \subseteq F$ are measurable sets, $B = \mathbb{R}$, and μ is a nonnegative measure, then

$$\int_E f \, d\mu \leq \int_F f \, d\mu.$$

7. $\int \|f(x)\|_B \, d\mu(x) = 0$ if and only if $f = 0$ almost everywhere.

Proof. This follows from the previous version of the theorem by density in L^1 . □

3.7.10. We can extend the definition of the integral even further. If $f \in M(X \rightarrow B)$ has a nonnegative version, we write $f \geq 0$.

Definition 3.7.11. Let $f \in M(X \rightarrow B)$, and assume that $f \geq 0$. Let E be a measurable set. If for every L^1 Cauchy sequence $f_n \in \mathbf{ISF}(X \rightarrow B)$, f_n does not converge to $1_E f$ in measure, then we write

$$\int_X f \, d\mu = \infty$$

provided that μ is a nonnegative measure.

3.7.12. Now if $f \in M(X \rightarrow \mathbb{C})$, we can write $f = f_a + if_b$ where f_a, f_b are real-valued, and $\int f = \int f_a + i \int f_b$. So integration of complex-valued functions reduces to integration of real-valued functions. Moreover, if f is real-valued then we can write $f = f_+ - f_-$ where $f_{\pm} \geq 0$. So $\int f = \int f_+ - \int f_-$. This makes sense even if one (but not both!) of the $\int f_{\pm}$ is infinite. Thus the only measurable functions which cannot be integrated are those whose integrals would be $\infty - \infty$. That's a far cry from the Riemann integral, whose definition was quite restricted!

3.7.13. While in practice we identify functions which are equal almost everywhere, sometimes it is convenient to work with functions, rather than their equivalence classes. Recall that \mathcal{M} is the space of measurable functions; analogously we let \mathcal{L}^1 denote the space of functions whose equivalence classes are in L^1 . If $f \in \mathcal{L}^1$, we let $[f]$ denote the equivalence class of f and define

$$\int_E f \, d\mu = \int_E [f] \, d\mu.$$

This definition then extends to those elements of \mathcal{M} whose integral is not $\infty - \infty$. Notice that $\|\cdot\|_1$ is a seminorm on \mathcal{L}^1 , and a norm on L^1 .

3.7.14. Let us now recast the above in the language of probability theory.

Definition 3.7.15. Let (Ω, P) be a probability space. Let X be a random variable of type B on Ω . The *expected value* of X is

$$EX = \int_{\Omega} X \, dP$$

provided that $X \in L^1$. If in addition $X^2 \in L^1$ and $B = \mathbb{R}$, then the *variance* of X is

$$\text{Var } X = E((X - EX)^2).$$

The *standard deviation* of X is $\sqrt{\text{Var } X}$.

Exercise 3.7.16. Let μ be a positive measure. Prove *Chebyshev's inequality*, which says that if $f \in M$, then for any $t > 0$ and $p > 0$,

$$\mu(\|f(x)\|_B \geq t) \leq t^{-p} \int_{\|f\|_B \geq t} \|f\|_B^p \, d\mu.$$

Here $\|f\|_B \geq t$ is the set $\{x \in X : \|f(x)\|_B \geq t\}$.

Exercise 3.7.17. Let δ_x be the Dirac measure at $x \in \mathbb{R}$, as in Exercise 1.6.22. Show that every function $f : \mathbb{R} \rightarrow B$ is Dirac measurable and compute the integral $\int_{\mathbb{R}} f \, d\delta_x$.

Exercise 3.7.18. Let X be a random variable of type \mathbb{R} . Show that

$$\text{Var } X = E(X^2) - (EX)^2 \geq 0$$

with equality iff X is almost surely constant, as long as the definition of $\text{Var } X$ makes sense.

Exercise 3.7.19. Let X be a random variable of type \mathbb{R} and distribution μ . Show that

$$EX = \int_{-\infty}^{\infty} x \, d\mu(x)$$

as long as the definition of EX makes sense.

Chapter 4

Properties of integration

In the previous chapter we defined integration. Now we show that integration is quite robust; most importantly, it commutes with most types of limits. Integration with Lebesgue measure also agrees with the Riemann integral whenever the Riemann integral is defined.

4.1 Integrable functions

Let us treat the properties of integrable functions. Throughout this section, let $X = (X, \Sigma, \mu)$ be a complete measured space, B a Banach space, and $L^1 = L^1(X \rightarrow B)$. We view L^1 as the space of integrable functions \mathcal{L}^1 modulo the space of functions f with $\|f\|_1 = 0$, but by the fundamental theorem of integration, L^1 is naturally isomorphic to the completion of the space **ISF** of integrable simple functions.

Lemma 4.1.1. Let f_n be an L^1 Cauchy sequence and $f \in M$. Then $f_n \rightarrow f$ nearly uniformly iff $f_n \rightarrow f$ in measure iff $f_n \rightarrow f$ in L^1 .

Proof. This was already true for **ISF**, and L^1 is the completion of **ISF**. □

Lemma 4.1.2. If $f_n \rightarrow f$ in L^1 , then $\int f_n \rightarrow \int f$.

Proof. We have

$$\left\| \int_E f_n d\mu - \int_E f d\mu \right\|_B \leq \int_E \|f_n(x) - f(x)\|_B d|\mu|(x) = \|f_n - f\|_1$$

which converges to 0. □

4.1.3. If f, f' are versions of the same element of M , then their carriers C, C' are equal up to a set of measure zero; indeed, the set of points x such that $f(x) = 0$ but $f'(x) \neq 0$ is null, but that set is $C \setminus C'$. Thus we can speak of the carrier of an equivalence class of functions, which is a set modulo sets of measure zero.

Lemma 4.1.4. If $f \in L^1$, then the carrier of f is σ -finite.

Proof. Let $f_n \in \mathbf{ISF}$, $f_n \rightarrow f$. Then the f_n have finite-support carriers C_n , and the carrier of f is contained in the union $\bigcup_n C_n$. □

Lemma 4.1.5. Let $f \in L^1$. Then for every $\varepsilon > 0$ there is a measurable set E such that $|\mu|(E) < \infty$ and

$$\int_{X \setminus E} \|f(x)\|_B d|\mu|(x) < \varepsilon.$$

Proof. Let $g \in \mathbf{ISF}$, $\|f - g\|_1 < \varepsilon$. Let E be the carrier of g . Then $g = 0$ on $X \setminus E$, so

$$\int_{X \setminus E} \|f(x)\|_B d|\mu|(x) = \int_{X \setminus E} \|f(x) - g(x)\|_B d|\mu|(x) \leq \|f - g\|_1 < \varepsilon.$$

□

Definition 4.1.6. Let $f \in M$. We say that f is *almost bounded*, or $f \in L^\infty$, if there is a version of f which is bounded. In that case, we define the L^∞ norm of f by

$$\|f\|_\infty = \inf_{f'} \sup_{x \in X} \|f(x)\|_B$$

where the inf is taken over all versions f' of f such that f' is bounded.

4.1.7. We leave it to the reader to check that L^∞ is a vector space and $\|\cdot\|_\infty$ is a norm.

Lemma 4.1.8. L^∞ is a Banach space.

Proof. We are given an L^∞ Cauchy sequence of $f_n \in L^\infty$, then we can choose bounded versions f'_n , and for almost every $x \in X$,

$$\|f'_n(x) - f'_m(x)\|_B \leq \|f_n - f_m\|_\infty \rightarrow 0$$

so the $f'_n(x)$ form a Cauchy sequence in the Banach space B , and hence converge to a vector that we call $f(x)$. We claim that $f'_n \rightarrow f$ almost uniformly; that is, there is a null set away from which $f'_n \rightarrow f$ uniformly. Indeed, given $\varepsilon > 0$ we can find N such that for almost every x and every $n_1, n_2 > N$, $\|f'_{n_1}(x) - f'_{n_2}(x)\|_B < \varepsilon$; this implies that $\|f'_n(x) - f(x)\|_B < \varepsilon$ if $n > N$. Therefore

$$\|f_n - f\|_\infty \leq \sup_x \|f'_n(x) - f(x)\|_B < \varepsilon$$

where the sup is taken over all x on the set on which $f'_n \rightarrow f$ uniformly, hence over almost every $x \in X$. □

4.1.9. The space \mathcal{L}^∞ is defined to be the space of all versions of all elements of L^∞ , and on \mathcal{L}^∞ , $\|\cdot\|_\infty$ is a seminorm. Its normalization is L^∞ .

4.1.10. It is worth contrasting the norms L^1 and L^∞ . L^1 only cares about the “area under the graph”; a function f whose graph is a narrow but very tall spike is tiny in L^1 . Meanwhile L^∞ only cares about “height of the graph”; that same function f would have an enormous L^∞ norm. On the other hand, a function which is very wide but shallow would be tiny in L^∞ but huge in L^1 . TODO: Draw a picture. Later we will define L^p norms which serve as a “weighted average” between the two extremes.

Lemma 4.1.11. The space $L^1 \cap L^\infty$ of almost bounded integrable (equivalence classes of) functions is dense in L^1 .

Proof. One has $\mathbf{ISF} \subseteq L^1 \cap L^\infty \subseteq L^1$, but \mathbf{ISF} is dense in L^1 , so $L^1 \cap L^\infty$ is as well. \square

4.1.12. How are the L^∞ and L^1 norms related? Well, if the measure of X is finite, then the graph of the function cannot be “too wide”, as our next lemma shows.

Lemma 4.1.13. If $\mu(X) < \infty$ then

$$\|f\|_1 \leq \mu(X) \|f\|_\infty.$$

Proof. We check

$$\|f\|_1 \leq \int_X \|f(x)\| d\mu(x) \leq \int_X \|f\|_\infty d\mu = \|f\|_\infty \int_X d\mu = \|f\|_\infty \mu(X).$$

Easy as that! \square

4.1.14. Conversely, if the graph of the function cannot be “too skinny”, then we have the opposite bound. To make this precise, we need the notion of a “granular measure”.

Definition 4.1.15. A *granular measure* is a measure μ such that there is a $\delta > 0$ such that for every measurable set E , either $E = \emptyset$ or $|\mu|(E) \geq \delta$.

4.1.16. For example, counting measure is δ -granular, with $\delta = 1$. Another example is Lebesgue measure restricted to the σ -algebra generated by the intervals $[n, n+1]$, which is also 1-granular.

Lemma 4.1.17. If μ is δ -granular, then

$$\|f\|_\infty \leq \frac{\|f\|_1}{\delta}.$$

Proof. If $f = 0$ almost everywhere, then both sides of the claimed equation are 0. Otherwise, let $E_n = \{\|f\| > \|f\|_\infty - 1/n\}$; then $|\mu|(E_n) \geq \delta$ and $E_n \supseteq E_{n+1}$, so let $E = \bigcap_n E_n$; either $|\mu|(E_n) < \infty$ for some n or $|\mu|(E_n) = \infty$ for all n . In the latter case, there is an n large enough that $\|f\|_\infty - 1/n > 0$ and

$$\|f\|_1 \geq \int_{E_n} \|f(x)\| d|\mu|(x) \geq (\|f\|_\infty - \frac{1}{n}) |\mu|(E) = \infty$$

and there is nothing to prove. Otherwise, $|\mu|(E) = \lim_n |\mu|(E_n) \geq \delta$ and

$$\|f\|_1 \geq \int_E \|f(x)\| d|\mu|(x) \geq \|f\|_\infty \delta,$$

proving the claim. \square

4.1.18. Let us now show that for Radon measures, and in particular Lebesgue measure, continuous functions are integrable on compact sets. If the reader is unfamiliar with locally compact Hausdorff spaces, they may as usual take $X = \mathbb{R}^d$ and μ Lebesgue measure.

Definition 4.1.19. Let X be a locally compact Hausdorff space, and suppose that μ is a Radon measure on X . A *locally integrable function* is a function f such that for every compact set K , $f|_K$ is an integrable function on K . The space of locally integrable functions modulo null sets is denoted L^1_l . A *almost locally bounded function* is a function f such that for every compact set K , $f|_K$ is almost bounded on K . The space of almost locally bounded functions modulo null sets is denoted L^∞_l .

Lemma 4.1.20. Let X be a locally compact Hausdorff space, and suppose that μ is a Radon measure on X . Let f be a continuous function; then $f \in L^1_l \cap L^\infty_l$.

Proof. First note that every continuous function is bounded on a compact set K . Therefore $f \in L^\infty_l$, and since μ is Radon, $|\mu|(K) < \infty$; therefore

$$\int_K \|f(x)\| d|\mu|(x) \leq \sup_{x \in K} \|f(x)\| \cdot |\mu|(K) < \infty.$$

Therefore $f \in L^1_l$. □

Exercise 4.1.21. Show that every set with Jordan content (see Exercise 1.2.15) is Borel.

4.2 Indefinite integrals

In calculus, one defined the indefinite integral g of a continuous function $f : [a, b] \rightarrow \mathbb{R}$ by the relation

$$g(x) = \int_a^x f(t) dt.$$

By the fundamental theorem of calculus, g is differentiable and $g' = f$. So we would like to define the indefinite integral of an arbitrary measurable function, and we would like it to have good regularity properties, but this is problematic; we used the order structure of \mathbb{R} to choose the interval $[a, x]$, but there is no such thing as an order on an arbitrary measure space, and there is no such thing as a differentiable function on an arbitrary measure space either.

But what if we thought of the indefinite integral as a function of the *set* $[a, x]$ rather than the number x ? As it turns out, this will fix both problems, and also the pesky issue of needing to choose a (the somewhat arbitrary choice of a being the reason for the constant of integration that has caused calculus students so much grief).

4.2.1. Throughout, let $X = (X, \Sigma, \mu)$ be a complete measured space, with μ valued in $(-\infty, \infty]$ or \mathbb{C} , and B a Banach space.

Definition 4.2.2. Let $f \in M$ and suppose that $\int f$ is defined. For every measurable set E , define

$$\nu(E) = \int_E f d\mu.$$

We call ν the *indefinite integral* of f . We also write $f = d\nu/d\mu$ and call f the *Radon-Nikodym derivative* of ν . If ν is a measure which is an indefinite integral, we say that ν is *Radon-Nikodym differentiable*.

4.2.3. The hypothesis that $\int f$ is defined rules out the possibility that $\nu(E) = \infty - \infty$. It is satisfied, for example, if $f \in L^1$ or $f \geq 0$.

Lemma 4.2.4. Let $f, g \in M$. Suppose that the indefinite integrals ν_f, ν_g of f, g are defined. Then for every measurable set E ,

$$\|\nu_f(E) - \nu_g(E)\|_B \leq \|f - g\|_1.$$

Proof. One checks

$$\|\nu_f(E) - \nu_g(E)\|_B = \left\| \int_E f - g \, d\mu \right\| \leq \|f - g\|_1$$

which proves the claim. \square

Theorem 4.2.5. Suppose that $f \in L^1$, and ν is the indefinite integral of f . Then ν is a finite measure.

Proof. Finiteness follows from $f \in L^1$, so we just need to check that ν is countably additive. We first check this when $f \in \mathbf{ISF}$. Indeed, if $f = b1_E$ is the canonical representation of f and the F_j are a sequence of disjoint measurable sets with union F , then

$$\nu(F) = b\mu\left(E \cap \bigcup_j F_j\right) = \sum_j b\mu(E \cap F_j) = \sum_j \nu(F_j).$$

Otherwise, f is a linear combination of functions with canonical representation of the form $b1_E$ and the claim still follows.

Now if $f \in L^1$, then for every $\varepsilon > 0$ there is a $g \in \mathbf{ISF}$ such that $\|f - g\|_1 < \varepsilon$; let ρ be the indefinite integral of g . Then for every measurable set E , $\|\nu(E) - \rho(E)\|_B \leq \|f - g\|_1 < \varepsilon$.

If all but finitely many of the F_j are empty, then there is an N such that

$$\nu(F) = \int_F f \, d\mu = \sum_{j < N} \int_{F_j} f \, d\mu = \sum_{j < N} \nu(F_j) = \sum_j \nu(F_j).$$

So it suffices to show that as $N \rightarrow \infty$, the partial sum $\sum_{j < N} \nu(F_j)$ converges to $\nu(F)$. Let $F^N = \bigcup_{j < N} F_j$, so $\nu(F^N) = \sum_{j < N} \nu(F_j)$.

We already showed that ρ is a measure, so if N is large enough then for every $n > N$,

$$\|\rho(F) - \rho(F^N)\|_B < \varepsilon.$$

In particular,

$$\|\nu(F) - \nu(F^N)\|_B \leq \|\nu(F) - \rho(F)\|_B + \|\rho(F) - \rho(F^N)\|_B + \|\rho(F^N) - \nu(F^N)\|_B < 3\varepsilon.$$

This implies $\nu(F^N) \rightarrow \nu(F)$. \square

Radon-Nikodym differentiable measures have a particularly easy-to-understand total variation.

Theorem 4.2.6. Let ν be a Radon-Nikodym differentiable measure. Then for any measurable set E ,

$$|\nu|(E) = \int_E \left\| \frac{d\nu}{d\mu}(x) \right\| d|\mu|(x).$$

Proof. Let $f = d\nu/d\mu$.

Suppose that E is a measurable set and $E = \bigcup_i E_i$, a finite disjoint union. Then

$$\sum_i \|\nu(E_i)\| = \sum_i \left\| \int_{E_i} f d\mu \right\| \leq \int_E \|f(x)\| d\mu(x).$$

Taking the supremum over all such finite disjoint unions we see that $|\nu|(E) \leq \int_E \|f\|$.

We first check the converse when $f \in \mathbf{ISF}$. Let $f = \sum_{i \leq k} b_i 1_{F_i}$ where the F_i are disjoint. Let $F = \bigcup_i F_i$. Let us write $E \cap F_i = \bigcup_{j \leq n_i} G_{ij}$ where the G_{ij} are disjoint. Then

$$|\nu|(E) \geq \sum_{i,j} \|\nu(G_{ij})\|_B = \sum_{i,j} \|b_i\|_B |\mu(G_{ij})| = \sum_i \|b_i\|_B \sum_j |\mu(G_{ij})|.$$

That is,

$$|\nu|(E) \geq \int_E \|f(x)\| d|\mu|(x).$$

In general, if $f, g \in L^1$ are the derivatives of ν, λ respectively, then

$$\| |\nu|(E) - |\lambda|(E) \| \leq \|\nu - \lambda\|(E) \leq \int_E \|(f - g)(x)\|_B d|\mu|(x) \leq \|f - g\|_{L^1},$$

by the reverse triangle inequality, Theorem 2.2.7. Now a straightforward approximation argument shows that the ISF case extends to all of L^1 . \square

Definition 4.2.7. A measure ν is *absolutely continuous* with respect to μ if for every $\varepsilon > 0$ there is $\delta > 0$ such that for every measurable set E , if $|\mu|(E) < \delta$, then $|\nu|(E) < \varepsilon$.

Theorem 4.2.8. For every $f \in L^1$, the indefinite integral of f is absolutely continuous.

Proof. Let $\varepsilon > 0$ be given. Since $\mathbf{ISF} \subseteq L^1 \cap L^\infty$, there is $g \in L^\infty$ such that $\|f - g\|_1 \leq \varepsilon/2$. Suppose that E is a measurable set such that $|\mu|(E) < \varepsilon/2\|g\|_\infty$. Then

$$\begin{aligned} \int_E \|f(x)\| d|\mu|(x) &\leq \|f - g\|_1 + \int_E \|g(x)\| d|\mu|(x) \\ &\leq \|f - g\|_1 + |\mu|(E)\|g\|_\infty \leq \varepsilon \end{aligned}$$

which completes the proof. \square

4.2.9. Later we will show the Radon-Nikodym theorem, which says that a measure ν is absolutely continuous iff every μ -null set is automatically ν -null, which happens iff ν is differentiable.

4.3 Convergence theorems

One of the cornerstone of calculus and its applications is the ability to interchange an integral with another sort of limit. Unfortunately, this sort of manuever is not valid in general.

Example 4.3.1. Let $f_n = 1_{[n, n+1]}$. Then $f_n \rightarrow 0$ pointwise but

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} f_n(x) dx = 1.$$

4.3.2. Roughly speaking, there are two settings in which it is acceptable to commute an integral with a limit. For one, if all functions involved are “dominated” by some function in L^1 , then it is usually safe to do so; one way to make this precise is the dominated convergence theorem.

Theorem 4.3.3 (dominated convergence). Let $f_n \in L^1(X \rightarrow B)$, and suppose that $f_n \rightarrow f$ almost pointwise. If there is a $g \in L^1(X \rightarrow [0, \infty))$ such that for every n and almost every $x \in X$,

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

then $f \in L^1(X \rightarrow B)$ and $f_n \rightarrow f$ in $L^1(X \rightarrow B)$.

Proof. By Lemma 4.1.5, there is a measurable set E with $|\mu|(E) < \infty$ such that

$$\int_{X \setminus E} g(x) d|\mu|(x) < \varepsilon.$$

Thus

$$\int_{X \setminus E} ||f_n(x) - f_m(x)||_B \leq 2 \int_{X \setminus E} g(x) d|\mu|(x) < 2\varepsilon.$$

Let ν be the indefinite integral of g . By Theorem 4.2.8, there is $\delta > 0$ so small that if $|\mu|(G) < \delta$ then $|\nu|(G) < \varepsilon$. Since E has finite measure, Egorov’s theorem implies that there is a measurable set $F \subseteq E$ such that $|\mu|(E \setminus F) < \delta$ and $f_n \rightarrow f$ in $L^\infty(F)$. But then $|\nu|(E \setminus F) < \varepsilon$, and so

$$\int_{E \setminus F} ||f_n - f_m(x)||_B \leq 2 \int_{E \setminus F} g(x) d|\mu|(x) = 2|\nu|(E \setminus F) < 2\varepsilon.$$

In particular, $\int_{X \setminus F} ||f_n - f_m||_B d|\mu| < 4\varepsilon$. Since F has finite measure and $f_n \rightarrow f$ in $L^\infty(F)$, $f_n \rightarrow f$ in $L^1(F)$ and so (f_n) is Cauchy in L^1 ; so choose N so large that if $n, m \geq N$ then

$$||f_n - f_m||_{L^1(F)} < \varepsilon.$$

Then

$$||f_n - f_m||_1 \leq ||f_n - f_m||_{L^1(F)} + \int_{X \setminus F} ||f_n - f_m||_B d|\mu| < 5\varepsilon$$

so (f_n) is Cauchy in L^1 . Since $f_n \rightarrow f$ almost pointwise, it follows that $f_n \rightarrow f$ in L^1 . \square

Example 4.3.4. We will compute

$$\lim_{n \rightarrow \infty} \int_0^n \left(1 + \frac{x}{n}\right)^n e^{-2x} dx.$$

First we compute

$$\lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n e^{-2x} = e^x e^{-2x} = e^{-x}.$$

Now, we have a pesky changing bound of integration, so we write

$$\lim_{n \rightarrow \infty} \int_0^n \left(1 + \frac{x}{n}\right)^n e^{-2x} dx = \lim_{n \rightarrow \infty} \int_0^\infty 1_{[0,n]}(x) \left(1 + \frac{x}{n}\right)^n e^{-2x} dx.$$

Since the integrand converges to e^{-x} , after finitely many n (which are allowed to discard, since we only care about the limiting behavior), we get

$$1_{[0,n]}(x) \left(1 + \frac{x}{n}\right)^n e^{-2x} \leq 2e^{-x}.$$

On the other hand,

$$\int_0^\infty e^{-x} dx = 1,$$

so $x \mapsto 2e^{-x}$ is in L^1 and by dominated convergence,

$$\lim_{n \rightarrow \infty} \int_0^n \left(1 + \frac{x}{n}\right)^n e^{-2x} dx = \int_0^\infty \lim_{n \rightarrow \infty} 1_{[0,n]}(x) \left(1 + \frac{x}{n}\right)^n e^{-2x} dx = \int_0^\infty e^{-x} dx = 1$$

as desired. These sorts of problems are very common on preliminary exams for graduate students, so the reader should probably master them.

4.3.5. The other situation in which it is acceptable to interchange an integral with a limit is when we can strongly use the order structure of \mathbb{R} in the codomain. So, we will need to restrict to the case when $B = [0, \infty)$ for the rest of the section, and thus formulate the so-called monotone convergence theorem.

TODO: Monotone convergence

Corollary 4.3.6. Let $f_n \geq 0$ be nonnegative integrable functions and μ a nonnegative measure on X . Then

$$\int_X \sum_n f_n d\mu = \sum_n \int_X f_n d\mu.$$

Proof. The sequence of partial sums is increasing, so we can apply monotone convergence. \square

Corollary 4.3.7. Let $f \geq 0$ be a nonnegative measurable function and μ a nonnegative measure. Then

$$\int_X f d\mu = \sup_s \int_X s d\mu$$

where the supremum is taken over all nonnegative $s \in \mathbf{ISF}(X \rightarrow [0, \infty))$.

Proof. We first show that there is a sequence of simple functions $\leq f$ converging to f monotonically. Fix n , and let $A_n = \{0, 1/n, 2/n, \dots, n - 2/n, n - 1/n, n\}$. For each $y \in A_n$, let $E_y^n = f^{-1}([y, y + 1/n))$. Let

$$f_n = \sum_{y \in A_n} y 1_{E_y^n}.$$

Now let $s_n = \max_{m \leq n} f_m$. Then s_n is simple and $s_n \leq s_{n+1}$, and $s_n \rightarrow f$ almost everywhere. So

$$\lim_{n \rightarrow \infty} \int_X s_n d\mu = \int_X f d\mu$$

by monotone convergence. Therefore

$$\int_X f d\mu \leq \sup_s \int_X s d\mu.$$

Monotonicity of the integral implies the other direction. □

Exercise 4.3.8. Let f be a measurable B -valued function. Show that if there $g \in L^1(X \rightarrow [0, \infty))$ with $\|f\|_B \leq g$ almost everywhere, then $f \in L^1$.

Exercise 4.3.9. Suppose that $|\mu|(X) < \infty$ (for example, μ is a probability measure). Let (f_n) be a sequence of functions in L^∞ such that there is $C > 0$ such that for every n , $\|f_n\|_{L^\infty} \leq C$. Show that if $f_n \rightarrow f$ almost pointwise, then $f_n \rightarrow f$ in L^1 .

Exercise 4.3.10. Consider the *gamma function*

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx.$$

Show that Γ is well-defined when $z > 0$ and is infinitely differentiable there. Then show that $\Gamma(n+1) = n!$ if $n \in \mathbb{N}$.

Exercise 4.3.11. Show that

$$\lim_{n \rightarrow \infty} \int_0^1 \frac{n^{3/2} x}{1 + n^2 x^2} dx = 0.$$

4.4 Product measures

Previous our development of the Lebesgue measure has been totally one-dimensional: we have defined the measure of a measurable subset of the line \mathbb{R} . We would like to do the same for higher-dimensional spaces.

We first review the notion of a product set. Suppose that we are given sets X_α , where α ranges over a set A . The *Cartesian product* $\prod_{\alpha \in A} X_\alpha$ is by definition the set of maps $x : A \rightarrow \bigcup_{\alpha \in A} X_\alpha$ such that for every $a \in A$, $x(a) \in X_a$. We usually write x_α or $\pi_\alpha(x)$ to mean x_α . The maps

$$\pi_\beta : \prod_{\alpha \in A} X_\alpha \rightarrow X_\beta$$

are known as *canonical projections* and the sets X_α are known as *factors*.

We mainly will be interested in the case when $A = \{1, \dots, n\}$ is a finite set, in which case we write $X_1 \times \dots \times X_n$ to mean the product of sets X_i , $i \in A$. An element of $X_1 \times \dots \times X_n$ can be written as an n -tuple (x_1, \dots, x_n) , where $x_i \in X_i$. For example, \mathbb{R}^n is a product of n copies of \mathbb{R} , and its elements are n -tuples of real numbers.

Lemma 4.4.1. Suppose that X_α are nonempty sets. Then $\prod_\alpha X_\alpha$ is nonempty.

Proof. We first note that we can assume that the X_α are disjoint. Indeed, if they are not, we can replace them with

$$X'_\alpha = X_\alpha \times \{\alpha\}.$$

Then elements of X'_α are pairs (x, α) where $x \in X_\alpha$. There is an obvious bijection $X_\alpha \rightarrow X'_\alpha$, $x \mapsto (x, \alpha)$, so we identify the two sets X_α and X'_α . Henceforth we replace X_α with X'_α and hence assume the X_α are disjoint.

Define a map $f : \bigcup_{\alpha \in A} X_\alpha \rightarrow A$ by declaring that if $x \in X_\alpha$ then $f(x) = \alpha$. Since the X_α are all nonempty, f is surjective. By the axiom of choice, Axiom B.1.19, there is an injective map $g : A \rightarrow \bigcup_{\alpha \in A} X_\alpha$ such that $f \circ g$ is the identity, and so $g(\alpha) \in X_\alpha$. Define an element x of X_α by letting $x_\alpha = g(\alpha)$. \square

If A is *finite* – the case that is the most interesting to us – then the use of the axiom of choice in the above argument is unnecessary (but otherwise it cannot be avoided, because if every product of nonempty sets is nonempty, then the axiom of choice is necessarily true). The use of the axiom of choice in the above argument is a hint that infinite products may be rather ill-behaved in measure theory.

Having discussed Cartesian products of sets, we now move on to products of measurable spaces.

Definition 4.4.2. Let $(X_\alpha, \Sigma_\alpha)$ be measurable spaces. A *measurable rectangle* in $\prod_\alpha X_\alpha$ is a Cartesian product $\prod_\alpha Y_\alpha$, where $Y_\alpha \in \Sigma_\alpha$ and all but finitely many of the Y_α are equal to X_α . The set of measurable rectangles is denoted $\bigoplus_\alpha \Sigma_\alpha$.

The measurable rectangles do not form a σ -algebra in general. For example, in \mathbb{R}^2 , the diagonal $\{(x, x) : x \in \mathbb{R}\}$ is not a rectangle, but will be in the σ -algebra generated by the rectangles, as we will later show.

The rather awkward requirement that finitely many of the Y_α are equal to the X_α can be explained by the following lemma.

Lemma 4.4.3. Let $(X_\alpha, \Sigma_\alpha)$ be measurable spaces. Then $\bigoplus_m \Sigma_m$ is a semiring in $\prod_m X_m$.

Proof. Let E, F be measurable rectangles. Since all but finitely many of the E_α are X_α , we can rename those α such that $E_\alpha \neq X_\alpha$ to be natural numbers. That is, the only E_α which are not X_α will be called E_1, \dots, E_m . Then if any unrenamed α has $F_\alpha \neq X_\alpha$, we can rename those α to $m+1, \dots, n$. Thus, we can assume that

$$A = \{1, \dots, n\} \cup B$$

where for every $\beta \in B$, $E_\beta = F_\beta = X_\beta$. One can then ignore the E_β and F_β entirely, and so assume that $B = \emptyset$. Then, arguing by induction, one can assume that $n = 2$. So assume that $E = E_1 \times E_2$ and $F = F_1 \times F_2$.

Now products commute with intersections, so $E \cap F$ is also a product of measurable sets, hence a measurable rectangle. One similarly checks that

$$(E_1 \times E_2) \setminus (F_1 \times F_2) = (E_1 \times (E_2 \setminus F_2)) \cup (E_1 \setminus F_1) \times (E_2 \cap F_2).$$

The above union is disjoint. \square

Therefore it is reasonable to want to define a premeasure on $\bigoplus_m E_m$, which we do shortly.

Definition 4.4.4. Let $(X_\alpha, \Sigma_\alpha)$ be measurable spaces, and let $X = \prod_\alpha X_\alpha$. The *product σ -algebra* $\bigotimes_\alpha \Sigma_\alpha$ is the σ -algebra on X generated by measurable rectangles in X . We call $(X, \bigotimes_\alpha \Sigma_\alpha)$ the *product measurable space* of the $(X_\alpha, \Sigma_\alpha)$.

Let $(\prod_\alpha X_\alpha, \bigotimes_\alpha \Sigma_\alpha)$ be a product measurable space. We will usually just denote this space by $\prod_\alpha X_\alpha$, leaving $\bigotimes_\alpha \Sigma_\alpha$ understood, since usually $\bigotimes_\alpha \Sigma_\alpha$ is the only interesting σ -algebra on $\prod_\alpha X_\alpha$.

We leave it to the categorically-minded reader to check that the product measurable space satisfies the universal property of products, and leave everyone else to quizzically wonder what such a sentence means. This is another sign that our definition of measurable space, with its bizarre clause that all but finitely many of the factors are trivial, is “correct”.

We recall that a measure μ is complex-valued if for every measurable E , $\mu(E)$ is a complex number or ∞ . We will restrict to complex-valued measures because we need to be able to multiply the measures of sets. Actually, if μ is complex-valued, then we can define its *complex conjugate* $\bar{\mu}$ by $\bar{\mu}(E) = \overline{\mu(E)}$. Then we can define the *real part* $\operatorname{Re} \mu = (\mu + \bar{\mu})/2$ and *imaginary part* $\operatorname{Im} \mu = (\mu - \bar{\mu})/2i$. Then $\mu = \operatorname{Re} \mu + i \operatorname{Im} \mu$. Thus whenever we work with complex-valued measures, we can replace them with real-valued measures whenever necessary. For a real-valued measure, we define $\mu_+ = (\mu + \mu)/2$ and $\mu_- = (\mu - \mu)/2$, thus μ_\pm are nonnegative measures and $\mu_+ - \mu_- = \mu$. So, when working with products of measured spaces, we will state theorems that are for complex-valued measures, but then prove them for nonnegative measures, since every complex-valued measure can be written as a sum of nonnegative measures.

Definition 4.4.5. Let $(X_1, \Sigma_1, \mu_1), \dots, (X_n, \Sigma_n, \mu_n)$ be measured spaces, where all the μ_m are complex-valued measures. We define a function $\bigoplus_m \mu_m = \mu_1 \oplus \dots \oplus \mu_n$ on $\bigoplus_m \Sigma_m$ by

$$\left(\bigoplus_m \mu_m \right) (E) = \mu_1(E_1) \mu_2(E_2) \cdots \mu_n(E_n).$$

We take the convention $0 \times \infty = 0$ whenever necessary.

We note that if we have an *infinite* collection of measured spaces $(X_\alpha, \Sigma_\alpha, \mu_\alpha)$, $\alpha \in A$ it is reasonable to define $\bigoplus_\alpha \mu_\alpha$ whenever we can guarantee that the infinite product $\prod_\alpha \mu_\alpha(E_\alpha)$ converges. For example this happens if, for every α , μ_α is a probability measure. However, this case can be rather tricky, due to the technicalities in the definition of a product of infinitely many measurable spaces. We discuss this in more detail in Example 4.4.10.

Lemma 4.4.6. Let $(X_1, \Sigma_1, \mu_1), \dots, (X_n, \Sigma_n, \mu_n)$ be measured spaces, where all the μ_m are complex-valued measures. Then $\bigoplus_m \mu_m$ is a premeasure on $\bigoplus_m \Sigma_m$.

Proof. We must show that $\bigoplus_m \mu_m$ is σ -additive, and it suffices to check when $n = 2$, by induction. By the usual reduction we can assume that the μ_m are nonnegative. In that case we change notation and write $\eta = \mu \oplus \nu$, where (X, μ) and (Y, ν) are measured spaces.

Suppose that $E \times F$ is a rectangle which is a disjoint union of rectangles $E_n \times F_n$. Then

$$1_E(x)1_F(y) = 1_{E \times F}(x, y) = \sum_{n=1}^{\infty} 1_{E_n \times F_n}(x, y) = \sum_{n=1}^{\infty} 1_{E_n}(x)1_{F_n}(y).$$

Therefore for any x ,

$$1_E(x)\nu(F) = 1_E(x) \int_Y 1_F(y) d\nu(y) = \int_Y \sum_{n=1}^{\infty} 1_{E_n}(x)1_{F_n}(y) d\nu(y).$$

Thus by Corollary 4.3.6,

$$1_E(x)\nu(F) = \sum_{n=1}^{\infty} 1_{E_n}(x) \int_Y 1_{F_n}(y) d\nu(y) = \sum_{n=1}^{\infty} 1_{E_n}(x)\nu(F_n).$$

Applying Corollary 4.3.6 again we see that

$$\eta(E \times F) = \mu(E)\nu(F) = \sum_{n=1}^{\infty} \mu(E_n)\nu(F_n) = \sum_{n=1}^{\infty} \eta(E_n \times F_n).$$

This is what we needed to prove. \square

Corollary 4.4.7. Let $(X_1, \Sigma_1, \mu_1), \dots, (X_n, \Sigma_n, \mu_n)$ be measured spaces, where all the μ_m are complex-valued measures. Then $\bigoplus_m \mu_m$ extends to a measure on $\bigotimes_m \Sigma_m$, which is unique and σ -finite if the μ_m are all σ -finite.

Proof. Existence is obvious by Lemma 4.4.6 and the Carathéodory construction. As for uniqueness, we use σ -finiteness of μ_m to find measurable sets $E_m^k \subseteq X_m$ such that $E_m^k \subseteq E_m^{k+1}$, $\bigcup_k E_m^k = X_m$, and $\mu_m(E_m^k) < \infty$. Then $\prod_m E_m^k \subseteq \prod_m E_m^{k+1}$, $\bigoplus_m \mu_m(\prod_m E_m^k) = \prod_m \mu_m(E_m^k) < \infty$, and $\bigcup_k \prod_m E_m^k = \prod_m X_m$. This implies that the extension of $\bigoplus_m \mu_m$ to a measure on $\bigotimes_m \Sigma_m$ is σ -finite and therefore unique. \square

Definition 4.4.8. Let $(X_1, \Sigma_1, \mu_1), \dots, (X_n, \Sigma_n, \mu_n)$ be measured spaces, where all the μ_m are complex-valued measures. Let $\bigotimes_m \mu_m = \mu_1 \otimes \dots \otimes \mu_n$ be the extension of the premeasure $\bigoplus_m \mu_m$ to $\bigotimes_m \Sigma_m$. We call $\bigotimes_m \mu_m$ the *product measure* of the μ_m and $(\prod_m X_m, \bigotimes_m \Sigma_m, \bigotimes_m \mu_m)$ the *product measured space*.

Let us give some examples of product measures. We first consider the simplest example, which any reader who has played a children's card game is familiar with.

Example 4.4.9. Let $A = \{1, \dots, n\}$, equipped with the σ -algebra consisting of *every* subset of A , and consider a function $\beta : A \rightarrow [0, 1]$ such that $\sum_{m=1}^n \beta(m) = 1$. Then β defines a probability measure μ by

$$\mu(\{a_1, \dots, a_k\}) = \sum_{j=1}^k \beta(a_j).$$

For example if $\beta = 1/n$, then μ is the *uniform probability measure* which sends every set E to its cardinality divided by n . If one has a set of n cards, and the probability of drawing card m is $\beta(m)$, then $\mu(E)$ is the probability of drawing a card in the set E .

Now we consider the Cartesian power $A^\ell = A \times \cdots \times A$ (ℓ factors). Elements of A^ℓ are vectors of ℓ elements of A , and if $\mu^\ell = \bigotimes_\ell \mu$ is the product measure on A^ℓ , then $\mu^\ell(E_1 \times E_2 \times \cdots \times E_\ell)$ can be interpreted as the probability of first drawing a card in $E_1 \subseteq A$, then in $E_2 \subseteq A$, and so on, and then in $E_\ell \subseteq A$, with replacement. In particular, if $E^\ell = E \times \cdots \times E$ is the Cartesian power of a set $E \subseteq A$, then $\mu^\ell(E^\ell) = \mu(E)^\ell$ is the probability of drawing a card in E ℓ times in a row, with replacement.

Example 4.4.10. Let A, β be as in Example 4.4.9. Now let us consider an infinitely long game, where one draws an infinite sequence of cards (the logicians would say that the player draws ω cards, because of TODO:Appendix) with replacement. Let A^ω be the set of sequences with values in A , viewed as a measurable space by endowing it with the σ -algebra generated by the measurable rectangles. By a cardinality argument similar to the one given in Lemma 1.2.12, one can show that not every subset of A^ω is measurable. On the other hand, the reader who is familiar with Cantor spaces (TODO:Appendix), and with the notion of the product of topological spaces (TODO:Appendix), will check that if we endow each copy of A with the discrete topology (which is its unique Hausdorff topology) and then endow A^ω with the product topology, A^ω is Cantor, and every Borel set is measurable (and conversely, that every measurable set is Borel).

Since μ is probability, an infinite product of numbers $\mu(E_n)$ will converge. Therefore if E is a rectangle in A^ω , and π_n is the canonical projection onto the n th factor,

$$\mu^\omega(E) = \prod_n \mu(\pi_n(E))$$

is well-defined, and the reader can check (using the fact that for all n large enough, $\mu(\pi_n(E))$ must be either 0 or 1 – why?) that μ^ω is a premeasure on the measurable rectangles, and hence a Borel probability measure. In the case that $n = 2$ and μ is uniform (so $\beta = 1/2$), then we say that μ^ω is the *standard Cantor measure*. It is of essential importance in probability theory and logic, among other fields.

Example 4.4.10 motivates the idea that the product of Borel σ -algebras should be the Borel σ -algebra of the product spaces. Unfortunately, this is not true in general. We include the following example for the reader's amusement, but it is not terribly important and can be omitted.

Example 4.4.11. Let κ be an uncountable cardinal as in TODO:Appendix, let $A = \{1, 2\}$ with its discrete topology, and let $X = A^\kappa$ be the Cartesian power, consisting of one factor of A for each ordinal of cardinality less than κ . Let π_α be projection onto the α th factor. Let Δ be the diagonal, so $x \in \Delta$ iff there is a $y \in A$ for every $\alpha < \kappa$, $\pi_\alpha(x) = y$.

Since X is a product of discrete (hence Hausdorff) spaces, X is Hausdorff, so Δ is closed TODO:Appendix and hence Borel. On the other hand, if Δ was measurable, then (as the set-theoretically minded reader can check) for all but countably many α , $\pi_\alpha(\Delta) = A$, contradicting the fact that there are uncountably many α and $\pi_\alpha(\Delta) = \{y\}$.

The above example is highly pathological. The below lemma covers most interesting cases. We remind the reader that if X_m are metric spaces with metrics d_m then $\prod_m X_m$ can be given the metric

$$d(x, y) = \max_m d_m(\pi_m(x), \pi_m(y)). \quad (4.1)$$

Lemma 4.4.12. Let X_1, \dots, X_n be separable metric spaces. Then the Borel σ -algebra on $\prod_m X_m$ is the product of the Borel σ -algebras on X_m .

Proof. By induction we can assume $n = 2$, and then change notation to write $X = X_1$, $Y = X_2$. We let $\mathcal{B}(Z)$ denote the Borel σ -algebra of the metric space Z .

By a *Borel cylinder* in $X \times Y$ we mean a set of the form $\pi_X^{-1}(E)$ or $\pi_Y^{-1}(F)$ where E is Borel in X and F is Borel in Y . We mean similarly for a *Borel cylinder*. We leave it to the reader to check that $\mathcal{B}(X) \otimes \mathcal{B}(Y)$ is generated by the Borel cylinders. Clearly every Borel cylinder is Borel, so this implies that every element of $\mathcal{B}(X) \otimes \mathcal{B}(Y)$ is Borel. TODO: Draw a picture of a cylinder.

Conversely, since X, Y are separable there are countable dense subsets $E, F \subseteq X, Y$. Then $E \times F$ is countable and dense in $X \times Y$. Let $B(x, y, r)$ denote the ball of radius r centered at (x, y) ; then $B(x, y, r) = B_X(x, r) \times B_Y(y, r)$ if we are using the metric 4.1. Here $B_X(x, r)$ is a ball in X and similarly for B_Y . Let \mathcal{S} be the set of $B(x, y, r)$ with $(x, y) \in E \times F$ and $r \in \mathbb{Q}$; then any open set in $X \times Y$ is a countable union of sets in \mathcal{S} and so \mathcal{S} generates $\mathcal{B}(X \times Y)$. Therefore $\mathcal{B}(X \times Y) \subseteq \mathcal{B}(X) \otimes \mathcal{B}(Y)$. \square

In the following section we use Lemma 4.4.12 to define the Lebesgue integral in general.

4.5 The Lebesgue integral

Let μ be a Stieltjes measure. Then μ is a Borel measure on \mathbb{R} , and by Lemma 4.4.12, a product of d copies of μ gives rise to a Borel measure on \mathbb{R}^d . We will mainly be interested in the case when μ is the Lebesgue measure.

Definition 4.5.1. Let μ be the Lebesgue measure on \mathbb{R} . The Borel measure $\mu^d = \bigotimes_{i=1}^d \mu$ on \mathbb{R}^d is called the *Lebesgue measure* on \mathbb{R}^d . If $f \in L^1(\mathbb{R}^d, \mu^d)$, we say that f is *Lebesgue integrable* and call $\int f d\mu^d$ the *Lebesgue integral* of f .

If $d = 2$ then the Lebesgue measure of a rectangle, or indeed any of the classical shapes, is just its area. Similarly if $d = 3$ then the Lebesgue measure of a rectangular prism, or any other classical shape, is just its volume. Thus Lebesgue measure generalizes the basic notions of Euclidean geometry to arbitrary (Borel) subsets of \mathbb{R}^d .

In general Lebesgue measure is so important that we usually refer to it implicitly. For example, we will usually just write $\int f$ or $\int f(x) dx$ for the Lebesgue integral of f .

In this section we record the basic properties of the Lebesgue measure.

Theorem 4.5.2. The Lebesgue measure is a σ -finite Radon measure.

Proof. By induction on d . When $d = 1$ this is the content of Theorem 1.6.15. Now $\mu^d = \mu^{d-1} \otimes \mu$, and we know that both μ^{d-1} and μ are Radon.

First we check local finiteness. By the Heine-Borel theorem every compact set is bounded and hence is contained in a compact rectangle in \mathbb{R}^d , which is a product of a compact rectangle in \mathbb{R}^{d-1} and a compact interval, both of which have finite measure. TODO: Draw a picture. This makes σ -finiteness easy to prove, because $[-n, n]^d$ is a compact (hence finite measure) rectangle that grows to be all of \mathbb{R}^d .

Now we check inner regularity on open rectangles. An open rectangle U in \mathbb{R}^d is a product of an open rectangle $U^* = \prod_{i < d} \pi_i(U)$ in \mathbb{R}^{d-1} and an open interval $\pi_d(U)$. Now if K is compact in U , then obviously $\mu(K) \leq \mu(U)$. Conversely, for every $\varepsilon > 0$, we can find a compact interval $K_d \subset \pi_d(U)$ with $\mu^d(\pi_d(U) \setminus K_d) < \varepsilon$ and a compact rectangle $K^* \subset U^*$ with $\mu^d(U^* \setminus K^*) < \varepsilon$. So $K = K^* \times K_d$ also has $\mu^d(U \setminus K)$ arbitrarily small.

Every open set U is a countable union of almost disjoint¹ open rectangles U_n , which can be approximated from within by compact sets $K_n^m \subset U_n$ with $\mu(U_n \setminus K_n^m) < \varepsilon 2^{-m} 2^{-n}$. Then the K_n^m are disjoint and $L_n = \bigcup_{m \leq n} K_n^m$ is compact. Moreover if $\mu^d(U) = \infty$ then $\mu^d(L_n) \rightarrow \infty$; otherwise

$$\mu^d(U) = \sum_n \mu^d(U_n) \leq \sum_n \mu^d(K_n^n) + \frac{1}{2^n 2^n} < \varepsilon + \mu^d(L_n) + \sum_{m > n} \mu^d(K_n^m)$$

and

$$\sum_{m > n} \mu^d(K_n^m) \leq \sum_{m > n} \mu^d(U_m) < \varepsilon$$

if n is large enough, since the sequence of $\mu^d(U_m)$ is absolutely summable. Therefore the L_n approximate U from within.

The proof of outer regularity is similar to the proof of inner regularity, and we leave it as an exercise. The reader may wish to use “half-open rectangles” in the proof of outer regularity. \square

Theorem 4.5.3. If A is Borel in \mathbb{R}^d and $x \in \mathbb{R}^d$, then the translation $A + x$ has the same Lebesgue measure as A .

Proof. See Exercise 4.5.6. \square

Theorem 4.5.4. Let $f : [\alpha, \beta] \rightarrow \mathbb{R}$ be a bounded Riemann integrable function and let R be its Riemann integral. Then

$$\int_{\alpha}^{\beta} f(x) dx = R.$$

Proof. The Riemann integral approximates f from below by step functions f_n on $[\alpha, \beta]$ which converge to f pointwise, and R is the limit of the f_n as $n \rightarrow \infty$. Since f is a bounded function on a set of finite measure it is in L^1 , and then dominated convergence implies that the Lebesgue integral is the limit of the integrals of the f_n . \square

Thus, we really have generalized the familiar notion of integration that many students learn about in high school or their first year of undergraduate education. We now prove that the integral of a function is the area under its graph.

¹in the sense that their intersection is Lebesgue null

Theorem 4.5.5. Let $f \in L^1(\mathbb{R}^d \rightarrow [0, \infty))$. Let $U = \{(x, y) : x \in \mathbb{R}^d, 0 \leq y \leq f(x)\} \subset \mathbb{R}^{d+1}$. Then

$$\mu^{d+1}(U) = \int_{\mathbb{R}^d} f(x) \, dx.$$

Proof. All functions here are nonnegative (so that we do not have to talk about “net signed area”), so by monotone convergence and continuity of measure, it suffices to check this for simple functions, and then by linearity it suffices to check this for an indicator function $f = 1_A$. But this is obvious:

$$\begin{aligned} U &= \{(x, y) : x \in \mathbb{R}^d, 0 \leq y \leq f(x)\} \\ &= \{(x, y) : x \in A, 0 \leq y \leq 1\} \cup \{(x, y) : x \notin A, 0 \leq y \leq 0\} \\ &= (A \times [0, 1]) \cup (A^c \times \{0\}). \end{aligned}$$

Since $\mu^{d+1}(A^c \times \{0\}) = 0$,

$$\mu^{d+1}(U) = \mu^{d+1}(A \times [0, 1]) = \mu^d(A) \mu^1([0, 1]) = \mu^d(A) = \int_{\mathbb{R}^d} 1_A(x) \, dx.$$

That proves the claim. □

Be aware, however, that there are Riemann integrable functions which are not Lebesgue integrable. The reason is that, to avoid integrals of the form $\infty - \infty$, we demanded that Lebesgue integrals converge absolutely, while the Riemann integral is allowed to converge conditionally. See Exercise 4.5.9 for an example of this phenomenon.

Exercise 4.5.6. Prove Theorem 4.5.3. (Hint: Use Theorem 1.6.9.)

Exercise 4.5.7. The *euclidean group* is the group of invertible functions $\mathbb{R}^d \rightarrow \mathbb{R}^d$ generated by translations, rotations around the origin, and reflections. Equivalently, it is the group generated by translations and orthogonal linear maps. An element of the euclidean group is called a *rigid motion*. Show that if A is a rigid motion and μ denotes Lebesgue measure, then $A_*\mu = \mu$. This generalizes Theorem 4.5.3.

Exercise 4.5.8. Let $A : \mathbb{R}^d \rightarrow \mathbb{R}^d$ be an invertible linear map, μ Lebesgue measurable, and E a Lebesgue measurable set. Show that $A_*\mu(E) = |\det A| \mu(E)$. (Hint: Use Exercise 4.5.7 to show that without loss of generality, we may assume that A is positive. Now use the spectral theorem.)

Exercise 4.5.9. Define the *sampling function* $\text{sinc} : \mathbb{R} \rightarrow \mathbb{R}$ by $\text{sinc } x = \sin x / x$ for $x \neq 0$ and $\text{sinc } 0 = 1$. Let $F(r)$ be the Riemann integral of sinc on $[0, r]$. Show that $\lim_{r \rightarrow \infty} F(r)$ exists and is finite. However, show that sinc is not Lebesgue integrable on $[0, \infty)$.

4.6 Changing the order of integration

In this section we prove the following extremely useful theorem. We let $f(x, \cdot)$ denote the function $y \mapsto f(x, y)$ whenever f is a function of two variables; similarly for $f(\cdot, y)$.

Theorem 4.6.1 (Fubini). Let (X, Σ, μ) and (Y, Γ, ν) be σ -finite complex-valued measured spaces. Let $f : X \times Y \rightarrow \mathbb{C}$ be a nonnegative $(\Sigma \otimes \Gamma)$ -measurable function. Then the following are equivalent:

1. $f \in L^1(X \times Y \rightarrow \mathbb{C}, \mu \otimes \nu)$.
2. For almost every $x \in X$, $f(x, \cdot) \in L^1(Y \rightarrow \mathbb{C}, \nu)$ and the function

$$x \mapsto \int_Y f(x, y) d\nu(y)$$

is in $L^1(X \rightarrow \mathbb{C}, \mu)$.

3. For almost every $y \in Y$, $f(\cdot, y) \in L^1(X \rightarrow \mathbb{C}, \mu)$ and the function

$$y \mapsto \int_X f(x, y) d\mu(x)$$

is in $L^1(Y \rightarrow \mathbb{C}, \nu)$.

Moreover,

$$\int_{X \times Y} f d(\mu \otimes \nu) = \int_X \int_Y f(x, y) d\mu(x) d\nu(y) = \int_Y \int_X f(x, y) d\nu(y) d\mu(x). \quad (4.2)$$

We will weaken the hypotheses on this theorem somewhat before the end of the section. Because of (4.2) we can *define* the *double integral* by

$$\iint_{X \times Y} f d\mu d\nu = \int_{X \times Y} f d(\mu \otimes \nu).$$

One similarly defines triple integrals, quadruple integrals, et cetera, by induction. Indeed, the hypothesis that only two measure spaces are in play in the statement of Fubini's theorem can be completely done away with, by induction, and one can consider arbitrary finite products of measure spaces. We will also do away with the hypothesis that f is nonnegative, at the price of requiring that $f \in L^1$. This is necessary to protect ourselves from our old enemy, $\infty - \infty$. We *cannot* do away with the σ -finite hypothesis, however TODO:show this.

TODO: State the Banach space valued version.

Before we prove Fubini's theorem, let us record three of its many applications.

Corollary 4.6.2. Let $(x_{i,j})_{i,j=1}^\infty$ be absolutely summable, thus

$$\sum_{i=1}^\infty \sum_{j=1}^\infty |x_{i,j}| < \infty,$$

or nonnegative, thus for every i, j , $x_{i,j} \geq 0$. Then

$$\sum_{i=1}^\infty \sum_{j=1}^\infty x_{i,j} = \sum_{j=1}^\infty \sum_{i=1}^\infty x_{i,j}.$$

Proof. This is an immediate consequence of Fubini's theorem applied to counting measure. \square

Corollary 4.6.3. One has the *Gaussian integral formula*

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}.$$

Proof. We first note that

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \int_{-\infty}^{\infty} e^{-y^2} dy$$

since all we have done is replace a dummy variable. Therefore it suffices to show that

$$\int_{-\infty}^{\infty} e^{-x^2} dx \int_{-\infty}^{\infty} e^{-y^2} dy = \pi. \quad (4.3)$$

Clearly $x \mapsto e^{-x^2}$ is nonnegative, so by Fubini's theorem we can replace the product integral by a double integral:

$$\int_{-\infty}^{\infty} e^{-x^2} dx \int_{-\infty}^{\infty} e^{-y^2} dy = \iint_{\mathbb{R}^2} e^{-(x^2+y^2)} dA(x, y).$$

Here A (short for area) is Lebesgue measure on \mathbb{R}^2 . Now $\{0\}$ is a null set so we can discard it, and the reader who recalls their calculus class will diligently check that $\mathbb{R}^2 \setminus \{0\} \cong \mathbb{R}_+ \times [0, 2\pi)$ according to the map $(r \cos \theta, r \sin \theta) \mapsto (r, \theta)$, where the open half-line \mathbb{R}_+ is given the Borel measure μ with

$$\mu(E) = \int_E r dr$$

and $[0, 2\pi)$ is given Lebesgue measure. That is,

$$dA(r \cos \theta, r \sin \theta) = d\mu(r) d\theta = r dr d\theta.$$

To check this claim, one just needs to show that the Borel sets in $\mathbb{R}^2 \setminus \{0\}$ are generated by “rectangles” (which here are sectors $\{(r \cos \theta, r \sin \theta) : r \in [r_1, r_2], \theta \in [\theta_1, \theta_2]\}$), and that for every such rectangle R , which we identify with the (plain old) rectangle $[r_1, r_2] \times [\theta_1, \theta_2]$, its area satisfies

$$A(R) = (\theta_2 - \theta_1) \int_{r_1}^{r_2} r dr.$$

TODO: Draw a picture of a sector. Once the reader verifies this, they are entitled to apply Fubini's theorem again, and

$$\iint_{\mathbb{R}^2} e^{-(x^2+y^2)} dA(x, y) = \int_0^{2\pi} \int_0^{\infty} r e^{-r^2} dr d\theta = 2\pi \int_0^{\infty} r e^{-r^2} dr.$$

Now this is really just a calculus problem: if $s = r^2$ then $ds = 2r dr$, and so one easily checks that

$$\int_0^{\infty} r e^{-r^2} dr = \frac{1}{2}.$$

Then (4.3) immediately follows. \square

Corollary 4.6.4. Let X, Y be independent random variables of type \mathbb{R} . If X, Y are in L^1 , then $E(XY) = (EX)(EY)$, and if $X^2, Y^2 \in L^1$, then $\text{Var}(X + Y) = \text{Var } X + \text{Var } Y$.

Proof. Let μ_X, μ_Y be the distributions of X, Y . Let $A = A_1 \times A_2 \subseteq \mathbb{R}^2$ be a Borel rectangle. Then

$$\mu_X \otimes \mu_Y(A) = \mu_X(A_1)\mu_Y(A_2) = P(X \in A_1)P(Y \in A_2) = P(X \in A_1 \cap Y \in A_2)$$

by independence. On the other hand, the random variable (X, Y) of type \mathbb{R}^2 has distribution

$$\nu(A) = P(X \in A_1 \cap Y \in A_2)$$

so $\nu = \mu_X \otimes \mu_Y$. It follows that

$$E(|X||Y|) = \int_{\mathbb{R}^2} \prod_{j=1}^n |x||y| \, d\nu(x, y) = \iint_{\mathbb{R}^2} |x||y| \, d\mu_X(x) \, d\mu_Y(y)$$

by Exercise 3.7.19 and Fubini's theorem. This integral simplifies to

$$\iint_{\mathbb{R}^2} |x||y| \, d\mu_X(x) \, d\mu_Y(y) = \int_{-\infty}^{\infty} |x| \, d\mu_X(x) \int_{-\infty}^{\infty} |y| \, d\mu_Y(y) = (E|X|)(E|Y|)$$

by Exercise 3.7.19. So by Fubini's theorem we can repeat the above argument with absolute values dropped and conclude $E(XY) = (EX)(EY)$.

Now set $X' = X - EX$, $Y' = Y - EY$. Then X', Y' are independent and $EX' = EY' = 0$, so $E(X'Y') = 0$. Thus

$$\begin{aligned} \text{Var}(X + Y) &= E((X' + Y')^2) = E((X')^2) + E((Y')^2) + 2E(X'Y') \\ &= \text{Var } X' + \text{Var } Y' = \text{Var } X + \text{Var } Y \end{aligned}$$

as desired. □

Monotone Classes

TODO: Prove Fubini

Exercise 4.6.5 (Cavalieri's principle). Let $E \subseteq \mathbb{R}^2$ be a measurable set, μ^d Lebesgue on \mathbb{R}^d , and $E_x = \{y \in \mathbb{R} : (x, y) \in E\}$. Show that

$$\mu^2(E) = \int_{-\infty}^{\infty} \mu^1(E_x) \, dx$$

(so that, in particular, $x \mapsto \mu^1(E_x)$ exists almost everywhere and is measurable).

Exercise 4.6.6. Let f be the function $f(x, y) = y^{-2}$ for $x < y$ and $f(x, y) = x^{-2}$ for $y < x$. Show that f is defined almost everywhere and measurable, and the iterated integrals $\int_0^1 \int_0^1 f(x, y) \, dx \, dy$ and $\int_0^1 \int_0^1 f(x, y) \, dy \, dx$ are finite, but $f \notin L^1([0, 1]^2)$. What went wrong?

Exercise 4.6.7 (weak law of large numbers). Let X_n be a countable sequence of iid random variables of type \mathbb{R} . Assume that $\text{Var } X_n$ makes sense and $EX_n = 0$. Show that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n X_j = 0$$

in probability using Chebyshev's inequality, Exercise 3.7.16.

Exercise 4.6.8 (Stone-Weierstrass theorem for polynomials). One form of the Stone-Weierstrass theorem is the statement that for every continuous function $f : [0, 1] \rightarrow \mathbb{R}$, there exists a sequence of polynomials that converge to f in L^∞ . Surprisingly, this deterministic statement has a simple probabilistic proof, which we now outline.

Fix $x \in [0, 1]$. Show that there is a probability space (Ω, P) which admits iid random variables X_j with $P(X_j = 1) = x$ and $P(X_j = 0) = 1 - x$. (Hint: One constructive way to do this is to say that Ω is the Cantor set $\prod_j \{0, 1\}$.)

Let $\bar{X}_n = \sum_{j \leq n} X_j/n$ be the average of the first n random variables X_j and show that $B_n(x) = E(f(\bar{X}_n))$ is a polynomial in x . For every $\varepsilon > 0$ there is a $\delta > 0$, which does not depend on x , such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \varepsilon$. Conclude that

$$|B_n(x) - f(x)| \lesssim \varepsilon + P(|\bar{X}_n - x| > \delta)$$

where the constant is allowed to depend on f but not x . Use Chebyshev's inequality, Exercise 3.7.16, to conclude the proof.

Chapter 5

Differentiation and regularity

5.1 Decomposition of signed measures

5.2 Differentiation of measures

As an application, we define the conditional expectation of a random variable. Recall from elementary probability theory that if A is an event (which is not almost surely false) and X is a random variable, then the conditional expectation of X given A is

$$E(X|A) = \frac{E(1_A X)}{P(A)}, \quad (5.1)$$

the expected value of X if we rescale A to be the entire probability space. However, in probability theory, one often needs to discuss the conditional expectation of X with respect to not just an event, but an entire σ -algebra of events. More precisely, let \mathcal{F} be a σ -algebra. If $A \in \mathcal{F}$ and we define $E(X|A)$ by (5.1), then we forget everything X except its mean on A . The conditional expectation of X on \mathcal{F} , by definition, will be a random variable that only remembers the conditional expectations of X with respect to every event in \mathcal{F} .

Definition 5.2.1. Let (Ω, Σ, P) be a probability space, $\mathcal{F} \subseteq \Sigma$ a σ -algebra, and $X \in L^1(\Omega \rightarrow \mathbb{R})$ a random variable. The *conditional expectation* of X given \mathcal{F} is a measurable function

$$E(X|\mathcal{F}) : (\Omega, \mathcal{F}) \rightarrow \mathbb{R}$$

such that

$$E(1_A E(X|\mathcal{F})) = E(1_A X)$$

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

Example 5.2.2. A *measurable partition* of Ω is a set of mutually exclusive events A_i such that $\bigcup_i A_i = \Omega$. If (A_i) is a measurable partition and \mathcal{F} is the smallest σ -algebra containing every A_i , then $E(X|\mathcal{F})$ is constant on each A_i , namely

$$E(X|\mathcal{F})|_{A_i} = E(1_{A_i} X)$$

is the mean of X on A_i .

Corollary 5.2.3. Let (Ω, Σ, P) be a probability space, $\mathcal{F} \subseteq \Sigma$ a σ -algebra, and $X \in L^1(\Omega \rightarrow \mathbb{R})$ a random variable. The conditional expectation $E(X|\mathcal{F})$ is well-defined in the sense that it exists, and if Y is also a conditional expectation, then $E(X|\mathcal{F}) = Y$ almost surely.

Proof. Let μ be the measure

$$\mu(A) = E(1_A X)$$

defined for $A \in \mathcal{F}$. Then μ is absolutely continuous with respect to P , so it has a Radon-Nikodym derivative; we set

$$E(X|\mathcal{F}) = \frac{d\mu}{dP}.$$

Then

$$E(1_A E(X|\mathcal{F})) = \int_A \frac{d\mu}{dP} dP = \int_A d\mu = \mu(A) = E(1_A X)$$

as desired.

For the uniqueness, assume Y is also a conditional expectation and let A be the event that $Y \neq E(X|\mathcal{F})$. Then $A \in \mathcal{F}$ so

$$E(1_A |Y - E(X|\mathcal{F})|) = E(1_A |X - X|) = 0.$$

But $|Y - E(X|\mathcal{F})| > 0$ on A so this implies $E(1_A) = 0$, thus A is almost surely false. \square

Exercise 5.2.4. Let $X \in L^1$ be a random variable and \mathcal{F} a σ -algebra. Show that if the pullback σ -algebra induced by X is independent of \mathcal{F} then

$$E(X|\mathcal{F}) = E(X).$$

Show that

$$E(E(X|\mathcal{F})) = E(X).$$

Show that if X is \mathcal{F} -measurable then $E(X|\mathcal{F}) = X$.

Exercise 5.2.5. Verify that the monotone and dominated convergence theorems, and Fatou lemma, are valid when EX_n is replaced with $E(X_n|\mathcal{F})$, \mathcal{F} a σ -algebra.

5.3 Existence of Radon measures

5.4 Differentiation of vector-valued functions

5.5 Vitali's covering lemma

Definition 5.5.1. If B is a ball, say $B = B(x, r)$, we let kB denote the *dilated ball* $kB = B(x, kr)$.

Lemma 5.5.2. Let X be a metric space and let \mathcal{B} be a finite set of balls in X . Then there is a subset \mathcal{B}_0 of \mathcal{B} consisting of disjoint balls such that

$$\bigcup_{B \in \mathcal{B}} B \subseteq \bigcup_{B_0 \in \mathcal{B}_0} 3B_0.$$

TODO prove me

Theorem 5.5.3 (Vitali's covering lemma). Let X be a separable metric space and let \mathcal{B} be a set of balls in X . Let

$$R(\mathcal{B}) = \sup_{B(x,r) \in \mathcal{B}} r$$

be the supremum of radii of the balls in \mathcal{B} . If $R(\mathcal{B}) < \infty$ then for every $\varepsilon > 0$ there is a countable subset $\mathcal{B}_0(\varepsilon)$ of \mathcal{B} consisting of disjoint balls such that

$$\bigcup_{B \in \mathcal{B}} B \subseteq \bigcup_{B_0 \in \mathcal{B}_0(\varepsilon)} (3 + \varepsilon)B_0. \quad (5.2)$$

TODO Prove me

Exercise 5.5.4. Show that there is a counterexample to Vitali's covering lemma when $R(\mathcal{B})$ is infinite.

Exercise 5.5.5. Show that a version of Vitali's covering lemma holds for any metric space. That is, show that if X is a metric space and \mathcal{B} is a set of balls in X such that $R(\mathcal{B}) < \infty$, then for every $\varepsilon > 0$ there is a subset (not necessarily countable) $\mathcal{B}_0(\varepsilon)$ of \mathcal{B} consisting of disjoint balls such that (5.2) holds. You will need to use some kind of set-theoretic machinery to pull this off.

5.6 The maximal inequality

Theorem 5.6.1 (weaktype Hardy-Littlewood maximal inequality). Let $d \geq 1$ and $f \in L^1(\mathbb{R}^d)$. Let μ denote Lebesgue measure. Then for every $\lambda > 0$,

$$\mu(\{Mf > \lambda\}) \leq 3^d \frac{\|f\|_{L^1(\mathbb{R}^d)}}{\lambda}.$$

TODO prove me

Exercise 5.6.2. Let $f \in L^1(\mathbb{R}^d)$. Show that if $Mf \in L^1(\mathbb{R}^d)$ then $f = 0$. It may help to first try this when f has compact support.

5.7 The Lebesgue differentiation theorem

5.8 The fundamental theorem of calculus

Chapter 6

Hölder duality and L^p norms

Previously we have considered a menagerie of different modes of convergence: pointwise convergence, almost pointwise convergence, uniform convergence, nearly uniform convergence, convergence in measure, convergence in L^1 , convergence in L^∞ ... In this chapter we introduce many more.

It is worth asking why we would bother to do this. Our previously considered modes of convergence served the purpose of allowing us to “approximate”, in various senses, arbitrary measurable functions by simple functions, or other functions that are similarly easy to understand. However, these notions of convergence were rather badly behaved. For example, we say that a mode of convergence M is *topologizable* if there is a topology T such that a sequence of functions converges in T iff it converges in M . Almost pointwise convergence, for one, is not topologizable. In particular, one cannot find a Banach space B such that almost pointwise convergence is equivalent to convergence in B . The only modes of convergence that we have considered so far which are induced by the norm of a Banach space are convergence in L^1 , uniform convergence, and convergence in L^∞ .

But why do we like Banach spaces? Well, Banach spaces have an algebraic structure, since they are vector spaces. More critically, if B is a Banach space with norm $\|\cdot\|$, and $f \in B$, then $\|f\|$ can reasonably be viewed as the “size” of f . In applications, being able to size functions is crucial. Many theorems in PDE amount to proving that if a function f , that we can think of as “initial data” for some dynamical system, is “small enough” (in the sense that $\|f\| < \delta$ for some δ that does not depend on f), then the dynamical system stabilizes to 0. As a concrete example, models of the spread of an infection inside a human body often have the property that if the initial data f is small enough (thus very few virions are present in the body), then the infection will die out.

Another useful feature of Banach spaces is the notion of duality. Given a Banach space B of functions, one can introduce the dual space B^* of bounded linear maps $B \rightarrow \mathbb{C}$; then B^* is a Banach space, whose elements can often be canonically identified with functions that we want to think of as “dual to” the functions in B . One then has an “inner product”

$$\langle \cdot, \cdot \rangle : B^* \times B \rightarrow \mathbb{C}$$

where, if f is the function which is canonically identified with a linear map $\lambda \in B^*$, and $g \in B$, then

$$\langle f, g \rangle = \lambda(g).$$

Under favorable circumstances, this inner product will behave very similarly to the inner products that one learns about in linear algebra. For example, two functions f, g are said to be *orthogonal* provided that $\langle f, g \rangle = 0$. One can then generalize such tools as the spectral theorem and the Cauchy-Schwarz inequality, so that we will be able to bring the power of linear algebra to bear on measurable functions.

In this chapter we will introduce p -norms for $p \in (1, \infty)$, which are the norms defined on Banach spaces L^p , and generalize both L^1 and L^∞ . The duality theory for L^p is known as *Hölder duality*. Hölder duality is especially effective when $p = 2$, since there is a canonical isomorphism $L^2 \rightarrow (L^2)^*$, and it is the case $p = 2$ that will be especially important when we consider applications in the following chapter.

Throughout, we fix a measured space (X, μ) and a Banach space B to serve as a codomain. Later we will need to assume that μ is σ -finite, but for now the general setting will be quite sufficient. Recall that we identify two functions which are equal almost everywhere, and that if we really need to refer to a particular representative of an equivalence class f , we call it a “version” of the function f . This identification will be critical for our Banach spaces to be well-defined!

6.1 L^p -norms

Recall that we defined the L^1 -norm

$$\|f\|_1 = \int_X \|f\| \, d\mu.$$

Thus $\|f\|_1$ could be small even if $\|f\|$ was quite large or even infinite, as long as the set where $\|f\|$ was large had a small measure. We defined the L^∞ -norm $\|f\|_\infty$ to be the infimum of all M such that $\|f\| \leq M$ almost everywhere; thus $\|f\|_\infty$ could be small even if $\|f\|$ did not tend to 0 anywhere, as long as $\|f\|$ did not ever go to infinity. TODO: Draw a picture. We seek to consider a weighted average of the two extremes.

Definition 6.1.1. Let $p \in (1, \infty)$. The L^p -norm of a measurable function f is

$$\|f\|_p = \left(\int_X \|f\|^p \, d\mu \right)^{1/p}. \quad (6.1)$$

We let $L^p(X \rightarrow B, \mu)$, often abbreviated to L^p or similar, be the vector space of all measurable functions f such that $\|f\|_p < \infty$.

Here L stands for “Lebesgue”; thus L^p is sometimes known as a *Lebesgue space*. However, the term “Lebesgue space” also refers in the literature to the measured space $[0, 1]$ with Lebesgue measure. Worse, while the key fact about L^p spaces is Hölder duality, a “Hölder space” is a space of continuous functions satisfying a certain inequality. So, to avoid confusion, we will simply call the spaces that we defined above *L^p spaces*.

We will shortly prove that the above definition makes sense, but, assuming that $\|\cdot\|_p$ is a norm, what does it measure? If f is a function that we can think of as a “wave” of amplitude A , supported on a set of measure V , then one approximately has

$$\|f\|_p \approx AV^{1/p}. \quad (6.2)$$

Here and throughout we take the very important convention that $1/\infty = 0$. (To see a precise version of (6.2), if f is the wave $f(x) = Ae^{ix}1_E$ where E is a set of measure V , then, for Lebesgue measure,

$$\|f\|_p = \|Ae^{ix}1_E\|_p = A\|1_E\|_p = A\mu(E)^{1/p} = AV^{1/p},$$

which partially justifies the intuition (6.2). In more advanced courses one introduces the *Sobolev norms* $W^{s,p}$ which not only takes into account amplitude and support but also frequency; if f has frequency N then $\|f\|_{s,p} \approx AN^s V^{1/p}$. In fact, $W^{0,p} = L^p$. We will discuss the case $p = 2$ in the exercises.

Since f is measurable and $\|f\|$ is nonnegative, $\|f\|_p$ is well-defined; either $\|f\|^p \in L^1(X \rightarrow \mathbb{C})$, in which case $f \in L^p(X \rightarrow B)$, or $\|f\|_p = \infty$. We need to check that $\|\cdot\|_p$ satisfies the definition of a norm, and that L^p is a vector space, in order for the above definition to make sense.

We first check the first two properties of a norm.

Lemma 6.1.2. One has $\|f\|_p = 0$ iff $f = 0$, and $\|cf\|_p = |c| \cdot \|f\|_p$.

Proof. It is clear that $|c| \cdot \|f\|_p = \|cf\|_p$. If $f = 0$, then $\|f\|_p = 0$. Conversely, if $f \neq 0$ on a positive measure set, say E , then

$$\|f\|_p^p \geq \int_E \|f\|^p d\mu > 0.$$

(Here we are using the fact that we identify functions which are equal almost everywhere!) \square

Thus we just need to check the triangle inequality. Before we do that, however, we will need to prove some inequalities and equations which will turn out to be highly useful in their own right. We first assert that

$$\| \|f\|^q \|_p = \|f\|_{pq}^q \quad (6.3)$$

whenever $p, q \in [1, \infty]$ and $f \in L^{pq}$ or $\|f\|^q \in L^p$; this follows straight from the definitions.

Lemma 6.1.3. One has

$$\exp(\theta x + (1 - \theta)y) \leq \theta e^x + (1 - \theta)e^y \quad (6.4)$$

whenever $\theta \in [0, 1]$ and $x, y \in \mathbb{R}$.

Proof. See Exercise 6.1.14. \square

Recall that $[1, \infty]$ is the set of all real numbers ≥ 1 , along with ∞ . Here and after we take the convention $1/\infty = 0$ and $1/0 = \infty$.

Definition 6.1.4. Let $p, q \in [1, \infty]$. If

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

we say that p, q are *Hölder duals*, and write $q = p^*$.

For example, $1^* = \infty$, and $2^* = 2$. One checks at once that p^* is unique, that $p^{**} = p$, and that if $p < q$ then $q^* < p^*$.

Theorem 6.1.5 (Young's product inequality). If $p \in (1, \infty)$ and $q = p^*$ then for every $a, b \geq 0$,

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

Proof. By (6.4) one has

$$ab = \exp(\log a + \log b) = \exp\left(\frac{\log a^p}{p} + \frac{\log b^q}{q}\right) \leq \frac{\exp(\log a^p)}{p} + \frac{\exp(\log b^q)}{q} = \frac{a^p}{p} + \frac{b^q}{q};$$

here we are taking $\theta = 1/p$ and $1 - \theta = 1/q$, and letting \log denote the natural logarithm. \square

Theorem 6.1.6 (Hölder's inequality). Let $p, q \in [1, \infty]$. Let $f \in L^p(X \rightarrow \mathbb{C})$, $g \in L^q(X \rightarrow \mathbb{C})$, and $q = p^*$. Then $fg \in L^1(X \rightarrow \mathbb{C})$ and

$$\|fg\|_1 \leq \|f\|_p \cdot \|g\|_q. \quad (6.5)$$

Proof. Without loss of generality, we can assume that $p \leq q$. If $p = 1$ then $q = \infty$, and this is just the triangle inequality for L^1 and L^∞ . Otherwise, $1 < p < q < \infty$, so Young's inequality for products applies.

Let $F = f/\|f\|_p$ and $G = g/\|g\|_q$. Then $\|F\|_p = \|G\|_q = 1$. By Young's inequality for products,

$$\begin{aligned} \|FG\|_1 &= \int_X \|FG\| \leq \int_X \frac{\|F\|^p}{p} + \frac{\|G\|^q}{q} \\ &= \frac{\|F\|_p^p}{p} + \frac{\|G\|_q^q}{q} = \frac{1}{p} + \frac{1}{q} = 1. \end{aligned}$$

Multiplying both sides by $\|f\|_p \cdot \|g\|_q$ we see that

$$\|fg\|_1 = \|F\|_p \cdot \|G\|_q \|fg\|_1 \leq \|f\|_p \|g\|_q.$$

This is what we wanted to prove. \square

To motivate Hölder's inequality, suppose that f is a wave of amplitude A_1 and supported on a set of measure V_1 . Suppose that g has amplitude A_2 and supported on a set of measure V_2 . Then the left-hand side of (6.5) is maximized when one has "constructive interference", which is only possible when f and g are supported on the same set, thus $V_1 = V_2$. In that case fg has amplitude $A_1 A_2$ and support on a set of measure $V_1^{1/p} V_2^{1/p^*} = V_1^{1/p+1/p^*} = V_1$. This is exactly what Hölder's inequality would predict.

Theorem 6.1.7 (Minkowski's triangle inequality). Let $p \in [1, \infty]$ and $f, g \in L^p$. Then $f + g \in L^p$ and

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p. \quad (6.6)$$

Thus L^p is a normed space with norm $\|\cdot\|_p$.

Proof. If $f + g = 0$ then there is nothing to prove. Otherwise, $\|f + g\|_p > 0$ and

$$\begin{aligned} \|f + g\|_p^p &= \int_X \|f + g\|^p = \int_X \|f + g\| \cdot \|f + g\|^{p-1} \\ &= \int_X \|f\| \cdot \|f + g\|^{p-1} + \int_X \|g\| \cdot \|f + g\|^{p-1} \\ &= \left\| \|f\| \cdot \|f + g\|^{p-1} \right\|_1 + \left\| \|g\| \cdot \|f + g\|^{p-1} \right\|_1. \end{aligned}$$

Now we apply Hölder's inequality. In fact,

$$\left\| \|f\| \cdot \|f + g\|^{p-1} \right\|_1 \leq \|f\|_p \cdot \left\| \|f + g\|^{p-1} \right\|_{p^*}.$$

But (6.3) says that

$$\left\| \|f + g\|^{p-1} \right\|_{p^*} = \|f + g\|_q^{p-1}$$

where $q = p^*(p - 1)$. Thus

$$\|f + g\|_p^p \leq (\|f\|_p + \|g\|_p) \|f + g\|_q^{p-1}.$$

But $1/p^* = 1 - 1/p$, so $p = q$. Therefore, dividing both sides by $\|f + g\|_p^{p-1}$, one concludes (6.6).

We have verified the triangle inequality, so $\|\cdot\|_p$ is a norm. If $f, g \in L^p$ and $c \in \mathbb{C}$ then $f + g \in L^p$ by the triangle inequality, and $cf \in L^p$ as we have already shown. So L^p is a normed vector space. \square

Note that the proof of Minkowski's inequality makes sense even for functions which are not valued in \mathbb{C} , since we never actually multiply the functions themselves, only their norms. So $L^p(X \rightarrow B)$ is a Banach space even if $B \neq \mathbb{C}$.

We stop to consider the case that μ is counting measure.

Definition 6.1.8. Let A be a set. Let ν be counting measure on A . We define $\ell^p(A) = L^p(A \rightarrow B, \nu)$. If $A = \mathbb{N}$, we leave A as understood and simply let $\ell^p = L^p(\mathbb{N} \rightarrow B, \nu)$. Here ℓ^p is pronounced "little LP."

Thus ℓ^p is the space of sequences x in B with

$$\sum_{n=1}^{\infty} |x_n|^p < \infty.$$

This space comes up in many useful examples, and has many nice properties.

Let us state some useful corollaries of Hölder's inequality.

Corollary 6.1.9 (Hölder interpolation). Let $1 \leq s \leq r \leq t \leq \infty$ and

$$\frac{1}{r} = \frac{\theta}{s} + \frac{1-\theta}{t}$$

where $\theta \in [0, 1]$. Then, for every $f \in L^s(X \rightarrow \mathbb{C}) \cap L^t(X \rightarrow \mathbb{C})$, $f \in L^r(X \rightarrow \mathbb{C})$ and

$$\|f\|_r \leq \|f\|_s^\theta \cdot \|f\|_t^{1-\theta}. \quad (6.7)$$

Proof. Since

$$\frac{\theta r}{s} + \frac{(1-\theta)r}{t} = 1,$$

we can use Hölder's inequality to check

$$\|f\|_r^r = \int_X \|f\|^{\theta r} \cdot \|f\|^{(1-\theta)r} \leq \left\| \|f\|^{\theta r} \right\|_{\frac{s}{\theta r}} \cdot \left\| \|f\|^{(1-\theta)r} \right\|_{\frac{t}{(1-\theta)r}};$$

taking r th roots of both sides and using (6.3), we prove (6.7). \square

Corollary 6.1.10 (Hölder's inequality with induction). Let $p_1, \dots, p_m, r \in [1, \infty]$ with

$$\frac{1}{r} = \sum_{j=1}^m \frac{1}{p_j}.$$

Then

$$\left\| \prod_{j=1}^m f_j \right\|_r \leq \prod_{j=1}^m \|f_j\|_{p_j}$$

for any f_1, \dots, f_m .

Proof. See Exercise 6.1.15. \square

Corollary 6.1.11 (Cauchy-Schwarz inequality). Whenever $f, g \in L^2(X \rightarrow \mathbb{C})$ one has

$$\|fg\|_1 \leq \|f\|_2 \cdot \|g\|_2. \quad (6.8)$$

In particular, if v, w are vectors in \mathbb{C}^d , then

$$|\langle v, w \rangle|^2 \leq \langle v, v \rangle \cdot \langle w, w \rangle. \quad (6.9)$$

Proof. The first claim (6.8) is just Hölder's inequality when $p = 2$. The second claim (6.9) is (6.8) specialized to $\mathbb{C}^d = \ell^p(\{1, \dots, d\} \rightarrow \mathbb{C})$. \square

Corollary 6.1.12 (Hölder comparison). Suppose that $p \leq q$ and f is measurable $X \rightarrow \mathbb{C}$. If μ is a finite measure, then

$$\|f\|_p \leq \mu(X)^{1/p-1/q} \|f\|_q \quad (6.10)$$

and so $L^q \subseteq L^p$. On the other hand, if μ is a granular measure, then $L^p \subseteq L^q$ and $\|f\|_q \lesssim_{p,q} \|f\|_p$.

Proof. The inequality (6.10) is equivalent to

$$\|1\|_1 \|f\|^p \leq \|1\|_{q/(q-p)} \cdot \|f\|^p \|1\|_{q/p}$$

by (6.3) and the fact that $\|1\|_{q/(q-p)}^p = \mu(X)^{1/p-1/q}$. Thus (6.10) is a consequence of Hölder's inequality. We leave the proof of the case when μ is granular as Exercise 6.1.16. \square

Hölder comparison is especially useful when μ is a probability measure, in which case it can be interpreted as the statement that every bounded random variable has a finite standard deviation and finite expected value, or when μ is counting measure, in which case it can be interpreted as saying that every absolutely summable sequence is bounded.

Let us finish the section by giving an example of Banach-valued L^p spaces which frequently come up in applications. Let $T > 0$. An *evolution equation* (say, on an open set $U \subseteq \mathbb{R}^d$) is a partial differential equation satisfied by functions $u : [0, T] \times U \rightarrow \mathbb{C}$ relating the “time derivatives” of u (i.e. those in the domain $[0, T]$) to the “space derivatives” (those in U). For instance, the Schrödinger equation

$$-i \frac{\partial u}{\partial t} + \Delta u = f,$$

where f is given, t is the time variable $\in [0, T]$, Δ is the Laplace operator on U , so

$$\Delta v(t, x_1, \dots, x_d) = \sum_{i=1}^d \frac{\partial^2 v}{\partial x_i^2}(t, x_1, \dots, x_d),$$

and $i^2 = -1$, is an evolution equation. In the theory of evolution equations it is frequently useful to define a function $u(t) : U \rightarrow \mathbb{C}$ by $u(t)(x) = u(t, x)$, so we think of u not as a function of two variables $[0, T] \times U \rightarrow \mathbb{C}$ but a function on $[0, T]$ which returns functions on U . In this case we often assume that there is a $q \in [1, \infty]$ such that for every $t \in [0, T]$, $u(t) \in L^q(U)$. So we view u as a function $[0, T] \rightarrow L^q(U)$, and we will often assume there is a $p \in [1, \infty]$ such that $u \in L^p([0, T] \rightarrow L^q(U))$. This motivates the following definition.

Definition 6.1.13. A *mixed norm* is a norm of the form

$$\|u\|_{L^p([0, T] \rightarrow L^q(U))} = \left(\int_0^T \|u(t)\|_{L^q(U)}^p dt \right)^{1/p}$$

defined on functions $u : [0, T] \rightarrow L^q(U \rightarrow \mathbb{C})$.

Later in the text we will have exercises extending the results in the text to spaces with mixed norms.

Exercise 6.1.14. Prove (6.4) using the fact that $\exp' = \exp$.

Exercise 6.1.15. Prove Corollary 6.1.10.

Exercise 6.1.16. Complete the proof of Corollary 6.1.12.

Exercise 6.1.17. Suppose that $f \in \bigcap_{p \in [1, \infty]} L^p$. Show that

$$\|f\|_\infty = \lim_{p \rightarrow \infty} \|f\|_p.$$

To do this, try using Corollary 6.1.12.

Definition 6.1.18. Let V be a vector space. A map $\|\cdot\| : V \rightarrow [0, \infty)$ is called a *quasinorm* if for every $f \in V$ and $c \in \mathbb{C}$, $\|f\| = 0$ iff $f = 0$, $\|cf\| = |c| \cdot \|f\|$, and one has a *quasitriangle inequality* $\|f + g\| \lesssim \|f\| + \|g\|$.

Exercise 6.1.19. One can also define $\|f\|_p$ when $p \in (0, 1)$, but then $\|\cdot\|_p$ is not well-behaved. In fact, define the L^p -quasinorm by (6.1) for $p \in (0, 1)$. Show that $\|\cdot\|_p$ is in fact a quasinorm, but the triangle inequality does not hold.

Exercise 6.1.20. Let $f \in \bigcap_{p \in (0, \delta)} L^p$, where L^p is meant in the sense of Exercise 6.1.19 and $\delta > 0$ is a small constant. Let C be the carrier of f . Show that

$$\mu(C) = \lim_{p \rightarrow 0} \|f\|_p.$$

Thus we may define the L^0 -quasinorm $\|f\|_0 = \mu(C)$. Show that the L^0 -quasinorm is a quasinorm, but the triangle inequality does not hold.

Exercise 6.1.21. This exercise is an example of a common phenomenon in PDE and physics: if p is chosen correctly, then $\|u(t)\|_{L^p}$ can often be viewed as the “energy” of the scalar field $u(t)$. Here $u : [0, T] \rightarrow L^p(U)$, and $U \subseteq \mathbb{R}^d$ is an open set. Consider the heat equation

$$\partial_t u - \Delta u = f,$$

where f is given and u is unknown. The *heat energy* of u at time t is $\|u(t)\|_{L^2(U)}$.

Show that if u is a smooth solution to the homogeneous heat equation $\partial_t u - \Delta u = 0$ such that $u = 0$ on the boundary ∂U , then whenever $t \leq t'$,

$$\|u(t')\|_{L^2(U)} \leq \|u(t)\|_{L^2(U)}.$$

(Hint: What is the derivative of $t \mapsto \|u(t)\|_{L^2(U)}$?) Conclude that if \bar{U} is compact,

$$\partial_t u - \Delta u = \partial_t v - \Delta v$$

on U , and $u = v = 0$ on ∂U , then $u = v$ almost everywhere (and hence everywhere, since u, v are smooth).

Exercise 6.1.22 (Dirichlet’s principle). Let U be a bounded open subset of \mathbb{R}^d , and let $f : \partial U \rightarrow \mathbb{R}$ be a continuous function defined on the boundary of U . *Dirichlet’s problem* for the Laplace equation

$$\Delta u = 0 \tag{6.11}$$

is to find a smooth function u such that (6.11) holds on U and $u = f$ on ∂U . Define the *Dirichlet energy* of u to be

$$E(u) = \|\nabla u\|_{L^2(U)}.$$

Let \mathcal{A} be given by

$$\mathcal{A} = \{u \in C^2(\bar{U}) \cap L^2(U) : u|_{\partial U} = f\}$$

where $C^2(\bar{U})$ is the space of continuous functions $\bar{U} \rightarrow \mathbb{R}$ such that $u|_U$ is twice-differentiable.

1. Show that u is a solution of Dirichlet’s problem iff $u \in \mathcal{A}$ and u minimizes the Dirichlet energy $E(u)$ among elements of \mathcal{A} . (Hint: For one direction, you should use the Cauchy-Schwarz inequality; for the other, let $g(t) = E(u + tv)$ where $v \in \mathcal{A}$, and show that $0 = g'(0)$.)
2. Show that if $f = 0$ and u is a solution of Dirichlet’s problem then $E(u) = 0$.
3. Show that there is at most one solution to Dirichlet’s problem.

Later, when we have more advanced techniques at our disposal, we will actually solve Dirichlet’s problem (Exercise 6.4.17).

6.2 Convergence in L^p

We now show that L^p is a Banach space. First we need to relate convergence in L^p to other modes of convergence, especially convergence in measure.

Theorem 6.2.1 (Markov's inequality). For any measurable function f and $\varepsilon > 0$,

$$\mu(\{\|f\| \geq \varepsilon\}) \leq \left(\frac{\|f\|_p^p}{\varepsilon^p} \right). \quad (6.12)$$

Proof. Let

$$E = \{x \in X : \|f\| \geq \varepsilon\}.$$

Then

$$\varepsilon^p 1_E \leq \|f\|^p. \quad (6.13)$$

Indeed, if $x \in E$ then $\varepsilon^p 1_E(x) = 0$; otherwise $\varepsilon^p 1_E(x) = \varepsilon^p \leq \|f(x)\|^p$. Integrating both sides of (6.13) $d\mu$ we get (6.12). \square

We now remind the reader what convergence in a normed space means: one has $f_n \rightarrow f$ in a normed space V provided that for every $\varepsilon > 0$ and every n large enough, $\|f_n - f\|_V < \varepsilon$. Similarly, $(f_n)_n$ is Cauchy in V provided that for every $\varepsilon > 0$ and every n, m large enough, $\|f_n - f_m\|_V < \varepsilon$.

Corollary 6.2.2. If $f_n \rightarrow f$ in L^p then $f_n \rightarrow f$ in measure, and there is a subsequence which converges to f pointwise. Moreover, if $(f_n)_n$ is Cauchy in L^p , then $(f_n)_n$ is Cauchy in measure, and there is a subsequence which is Cauchy almost everywhere.

Proof. By hypothesis and Markov's inequality, one has

$$\mu(\{\|f_n - f\|_p \geq \varepsilon\}) \leq \frac{\|f_n - f\|_p^p}{\varepsilon^p}$$

which implies convergence in measure. The proof of Cauchyness in measure is similar, replacing f with f_m in the above inequality. Convergence almost everywhere follows from the fundamental theorem of integration. \square

Theorem 6.2.3. L^p is a Banach space.

Proof. Suppose that $(f_n)_n$ is Cauchy in L^p ; then there is a subsequence $(f_{n_k})_k$ and a f such that

$$\lim_{k \rightarrow \infty} f_{n_k} = f$$

almost everywhere. Therefore for every n and x ,

$$\|f(x) - f_m(x)\|^p = \lim_{k \rightarrow \infty} \|f_{n_k}(x) - f_m(x)\|^p.$$

Therefore by Fatou's lemma,

$$\|f(x) - f_m(x)\|_p^p \leq \lim_{k \rightarrow \infty} \int_X \|f_{n_k}(x) - f_m(x)\|^p d\mu = \lim_{k \rightarrow \infty} \|f_{n_k} - f_m\|_p^p.$$

Taking the limit as $m \rightarrow \infty$ of both sides and using the fact that $(f_n)_n$ was Cauchy in L^p ,

$$\lim_{m \rightarrow \infty} \|f(x) - f_m(x)\|_p^p \leq \lim_{k \rightarrow \infty} \lim_{m \rightarrow \infty} \|f_{n_k} - f_m\|_p^p = 0.$$

This implies that $f_m \rightarrow f$ in L^p ; in particular, $f \in L^p$. □

TODO: Monotone and dominated convergence

TODO: ISF is dense

TODO: Separability

Exercise 6.2.4 (strongtype Hardy-Littlewood maximal inequality). Let $p \in (1, \infty]$, $f \in L^p(\mathbb{R}^d)$, and let Mf denote the Hardy-Littlewood maximal function of f . Show that

$$\|Mf\|_{L^p(\mathbb{R}^d)} \lesssim_{p,d} \|f\|_{L^p(\mathbb{R}^d)}$$

using the weaktype Hardy-Littlewood maximal inequality.

6.3 Duality and representation

We now come to the important notion of Hölder duality, which we should motivate before stating.

“Duality” is a funny word that many mathematicians use, but only a few would dare try to give a precise definition of. For our purposes, we can think of an “algebraic” object X (a ring, an abelian group, et cetra) as being “dual” to a “geometric” object X^* (a topological space, a manifold, et cetra) if there is a “good bijection” between “good” maps $X \rightarrow Y$ (say homomorphisms) and “good” maps $Y^* \rightarrow X^*$ (continuous maps, smooth maps, et cetra), and another good bijection between good maps $Y \rightarrow X$ and good maps $X^* \rightarrow Y^*$. The data consisting of the good bijections, for every such pair (X, X^*) , is then known as a “duality theory”. We leave all the words in scare quotes undefined; the reader who is familiar with modern algebra may try to replace them with precisely defined terms, perhaps using the language of category theory or universal algebra. However, there is a good chance that the resulting notion will not quite be satisfied by Hölder duality, because in the edge cases $p = 1$ and $p = \infty$, Hölder duality is rather ill-behaved.

One duality theory that every mathematician is familiar with is duality on finite-dimensional vector spaces. Finite-dimensional vector spaces are both “algebraic” and “geometric”, and indeed if V is a finite-dimensional vector space over a field K , then there is an isomorphism $V \rightarrow V^*$, where V^* is the finite-dimensional vector space of all linear maps $V \rightarrow K$. If V consists of column vectors over K , then V^* can be canonically identified with a space of row vectors over K . For any linear map $V \rightarrow W$ one obtains its adjoint map $W^* \rightarrow V^*$. In particular, if we select an isomorphism $V \rightarrow V^*$, then we can talk about the self-adjoint maps $V \rightarrow V$, which have especially nice properties (for example, being orthogonally diagonalizable). Moreover, there is a “canonical” isomorphism $V \rightarrow V^{**}$. Here we again use scare quotes for a word that we will refuse to define precisely.

We try to recover the same ideas in higher generality, replacing finite-dimensional vector spaces with L^p spaces. Since every finite-dimensional vector space over \mathbb{C} is indeed isomorphic to $\ell^p(\{1, \dots, n\} \rightarrow \mathbb{C})$ for some n and any p , this is indeed a generalization.

Throughout, we will assume that μ is a σ -finite measure valued in \mathbb{C} and that B is a Hilbert space with inner product denoted

$$(f, g) \mapsto f\bar{g}.$$

(TODO: Explain this? Is this allowed by mixed norm duality??) Actually, since μ is valued in \mathbb{C} , it is no loss of generality to assume that μ is nonnegative; so when convenient we will take $\mu = \mu$.

Definition 6.3.1. Let X be a normed space. The *dual space* X^* of X is the Banach space of all linear maps $X \rightarrow \mathbb{C}$ such that

$$|Tf| \lesssim \|f\|_X, \quad (6.14)$$

with the norm

$$\|T\|_{X^*} = \sup_{\|f\|_X \leq 1} \|Tf\|.$$

An element of X^* is known as a *linear functional* or a *covector* on X .

To check that X^* is indeed a Banach space, one needs to use the fact that \mathbb{C} is itself a Banach space; see Exercise 6.3.8.

Recall that p^* is the Hölder dual of p ; we want a canonical isomorphism

$$L^{p^*} \cong (L^p)^*$$

whenever possible.

Lemma 6.3.2. Let $p \in [1, \infty]$. One has an injective linear map $\lambda : L^{p^*} \rightarrow (L^p)^*$ defined by

$$\lambda(g)(f) = \int_X f\bar{g} \, d\mu. \quad (6.15)$$

Proof. By Hölder's inequality,

$$|\lambda(g)(f)| = \|fg\|_1 \leq \|f\|_p \cdot \|g\|_q,$$

which implies (6.14). □

We will henceforth omit the mention of λ whenever possible, and write

$$\langle f, g \rangle = \lambda(g)(f)$$

whenever $f \in L^{p^*}$ and $g \in L^p$.

Theorem 6.3.3 (Hölder duality, part I). Let $p \in [1, \infty]$. Suppose that λ is as in (6.15). Then $\|\lambda g\|_{(L^p)^*} = \|g\|_{p^*}$. Moreover, if $p < \infty$, λ is an isomorphism.

TODO

Corollary 6.3.4. If $p \in [1, \infty]$, there is an injective linear map $\varphi : L^p \rightarrow (L^p)^{**}$ defined by

$$\langle \varphi(f), g \rangle = \langle g, f \rangle$$

which preserves norms. Moreover, if $p \in (1, \infty)$ then φ is an isomorphism.

TODO

Henceforth, we simply write $L^p = (L^p)^{**}$ and $(L^p)^* = L^{p^*}$ (whenever these spaces are isomorphic!) to emphasize that these isomorphisms are in some sense “natural”; they do not depend on a choice of basis. In particular, we will always identify a function $f \in L^{p^*}$ with the corresponding linear map $f : L^p \rightarrow \mathbb{C}$; to avoid writing awkward things like $f(g)$, we will continue to use $\langle f, g \rangle$ to mean the evaluation of f at $g \in L^p$ when thinking of f as a linear map.

Corollary 6.3.5. For every bounded linear map $\psi : L^2 \rightarrow \mathbb{C}$ there is a unique $f \in L^2$ such that for every $g \in L^2$, $\langle f, g \rangle = \psi(g)$. In particular, $\langle \cdot, \cdot \rangle$ is an inner product, making L^2 into a Hilbert space (TODO).

Theorem 6.3.6 (Hölder duality, part II). Let $p, q \in (1, \infty)$. Then for every bounded linear map $T : L^p \rightarrow L^q$ there is a unique bounded linear map $T^* : L^{q^*} \rightarrow L^{p^*}$ satisfying

$$\langle Tf, g \rangle = \langle f, T^*g \rangle.$$

Definition 6.3.7. The map T^* is called the *adjoint* of T .

TODO proof

Exercise 6.3.8. Prove that if X is a normed space, then X^* is Banach.

Exercise 6.3.9. Show that $(\ell^1)^{**} \neq \ell^1$. To do this, first show that if $(L^p)^*$ is separable, then L^p is separable.

Exercise 6.3.10. Show that $(\ell^1)^{**} \neq \ell^1$ directly, by filling in the following outline. Let $\mu : 2^{\mathbb{N}} \rightarrow \{0, 1\}$ be an additive (but not σ -additive) function such that for every finite set $A \subset \mathbb{N}$, $\mu(A) = 0$. Here μ can be built using Zorn’s lemma¹ or a similar substitute. If $A \subseteq \mathbb{N}$, one can define

$$\langle \mu, 1_A \rangle = \mu(A)$$

and then extend this definition to $\langle \mu, x \rangle$, whenever $x \in \ell^\infty$. If μ is induced by an element of ℓ^1 , however, one can conclude a contradiction.

¹Any proof of that $(\ell^1)^{**} \neq \ell^1$ must use the axiom of choice somehow, because in Solovay’s model of set theory, wherein there are no nonmeasurable functions, there are also no elements of $(\ell^1)^{**}$ which are not elements of ℓ^1 .

6.4 Compactness of the unit ball

An annoying technical difficulty with L^p spaces is that, unlike \mathbb{C}^d , the unit ball of L^p is not compact (Exercise 6.4.9). This is rather problematic: in many situations in analysis, one is given a sequence and needs to extract a limit point from the sequence. In this section we use Hölder duality to introduce a new topology on L^p in which the unit ball is compact, but which does not arise from a norm.

Definition 6.4.1. Let $p \in (1, \infty)$. We say that a sequence $(f_n)_n \in L^p$ *converges weakly* in L^p to $f \in L^p$ provided that

$$\lim_{n \rightarrow \infty} \langle f_n, g \rangle = \langle f, g \rangle$$

for every $g \in L^{p^*}$.

To avoid confusion, we will say that a sequence *converges strongly* in L^p if it converges in norm. The words “strong” and “weak” are motivated by the following lemma:

Lemma 6.4.2. Let $p \in (1, \infty)$. If a sequence converges strongly in L^p , then it converges weakly in L^p .

Proof. One has

$$\langle f_n - f, g \rangle \leq \|f_n - f\|_p \|g\|_{p^*}$$

for every $g \in L^{p^*}$. □

The idea of weak convergence is the following. Say that $f(x)$ is the temperature of a bucket of water at point x . Then one measures $f(x)$ by taking a thermometer, which we model by g , chosen so that the thermometer picks up the temperature in a small ball near x (so g has compact support centered near x). Then the thermometer returns $\langle f, g \rangle$. So $f_n \rightarrow f$ weakly in L^p if for every thermometer $g \in L^{p^*}$, the thermometer g thinks that the error $f_n - f$ is negligible.

We now show that weak convergence is induced by a topology, so it is in some sense “better behaved” than, say, convergence almost pointwise.

Definition 6.4.3. Let $p \in (1, \infty)$ and $g \in L^{p^*}$. We define the *weakstar seminorm*

$$p_g(f) = |\langle f, g \rangle|.$$

The *weakstar topology* on L^p is the topology induced by the weakstar seminorms $\{p_g : g \in L^{p^*}\}$.

The term “weakstar” is motivated by the fact that the weakstar topology will induce weak convergence, and that the weakstar topology is induced by duality (hence “star”).

There are several suspicious things about this definition. First, we check that p_g is a seminorm. Indeed, one has a triangle inequality

$$p_g(f_1 + f_2) = |\langle f_1 + f_2, g \rangle| = |\langle f_1, g \rangle + \langle f_2, g \rangle| \leq p_g(f_1) + p_g(f_2)$$

and also $p_g(cf) = cp_g(f)$. We recall that a set of seminorms induces a topology: the open neighborhoods of $f_0 \in L^p$ are generated by sets of the form

$$B_g(f_0, \varepsilon) = \{p_g(f_0 - f) < \varepsilon\}$$

which are *open balls* for the seminorm g .

Lemma 6.4.4. A sequence (or net) converges weakly in L^p iff it converges in every weakstar seminorm of L^p , which happens iff it converges in the weakstar topology of L^p .

Proof. The first equivalence is due to the definition of the weakstar seminorm; the latter is a general fact about topologies induced by seminorms. \square

Now we come to the key fact about the weakstar topology.

Theorem 6.4.5 (Banach-Alaoglu compactness). The closed unit ball of L^p is compact in the weakstar topology.

The general proof uses a black box in the form of Tychonoff's theorem² so we start by proving a special case, which is the particularly useful one in practice.

Theorem 6.4.6 (Alaoglu's sequential compactness theorem). Suppose that L^{p^*} is separable. Then for every bounded sequence $(f_n)_n$ in L^p , there is a $f \in L^p$ such that a subsequence $(f_{n_k})_k$ converges to f weakly in L^p .

Proof. Let D be a countable dense subset of the unit sphere $\{g \in L^{p^*} : \|g\|_{p^*} = 1\}$ of L^{p^*} . Let

$$d(f_1, f_2) = \sum_{g \in D} 2^{-n} p_g(f_1 - f_2).$$

Since $p_g(f_1 - f_2) \leq \|g\|_{p^*} \|f_1 - f_2\|_p$ and $\|g\|_{p^*} \leq 1$, we can sum the geometric series to conclude that $d(f_1, f_2) \leq 1$. We claim that d is a semimetric.³ We just need to check the triangle inequality, and indeed,

$$d(f_1, f_3) = \sum_{g \in D} 2^{-n} p_g(f_1 - f_3) \leq \sum_{g \in D} \frac{p_g(f_1 - f_2) + p_g(f_2 - f_3)}{2^n} = d(f_1, f_2) + d(f_2, f_3)$$

since p_g is a seminorm.

Lemma 6.4.7. A sequence converges in d iff it converges weakly in L^p .

Proof. If $f_n \rightarrow f$ weakly in L^p then for every $g \in L^{p^*}$, $p_g(f_n - f) \rightarrow 0$, so $d(f_n, f) \rightarrow 0$.

Conversely, if $(f_n)_n$ is a sequence in L^p and $f \in L^p$, and $d(f_n, f) \rightarrow 0$, then $p_g(f_n - f) \rightarrow 0$ for every $g \in D$. But if $\|g\|_{p^*} = 1$ is arbitrary, and $g_m \in D$ have $g_m \rightarrow g$, then

$$\lim_{n \rightarrow \infty} p_g(f_n - f) = \lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} p_{g_m}(f_n - f) = 0$$

so that $p_g(f_n - f) \rightarrow 0$ for every g in the unit sphere, and hence in all of L^{p^*} by rescaling. \square

We now use a variant of Cantor's diagonal argument to show that the unit ball of L^p under the semimetric d is compact, which, along with Lemma 6.4.7, implies Alaoglu's theorem. Fix $(f_n)_n$, say $\|f_n\|_p \lesssim 1$, and an enumeration $(g_m)_m$ of D . Then

$$|\langle f_n, g_1 \rangle| \lesssim \|g_1\|_{p^*} = 1$$

²The general proof must use a particularly strong form of the axiom of choice, as most results about compactness of very large spaces must, so this is unavoidable.

³One can use Exercise 6.4.10 to show that d is actually a metric, but we won't need this.

so by compactness of the closed unit ball of \mathbb{C} , there is a subsequence $(f_{n_k}^1)_k$ such that $(\langle f_{n_k}^1, g_1 \rangle)_k$ is a Cauchy sequence.

Suppose we have defined $(f_{n_k}^j)_k$; then $|\langle f_{n_k}^j, g_1 \rangle| \lesssim 1$ so there is a subsequence $(f_{n_k}^{j+1})_k$ of $(f_{n_k}^j)_k$ such that $(\langle f_{n_k}^{j+1}, g_{j+1} \rangle)_k$ is a Cauchy sequence. Moreover, since $(f_{n_k}^{j+1})_k$ is a subsequence of $(f_{n_k}^\ell)_k$ whenever $\ell \leq j$, $(\langle f_{n_k}^{j+1}, g_\ell \rangle)_k$ is also a Cauchy sequence.

Let $f_{n_k} = f_{n_k}^j$. Then for any j , after finitely many entries $(f_{n_k})_k$ is a subsequence of $(f_{n_k}^j)_k$, so $(\langle f_{n_k}, g_j \rangle)_k$ is a Cauchy sequence, and hence we can set

$$\langle f, g_j \rangle = \lim_{k \rightarrow \infty} \langle f_{n_k}, g_j \rangle.$$

This defines f as a map $D \rightarrow \mathbb{C}$; then f extends to a bounded linear map $L^{p*} \rightarrow \mathbb{C}$ uniquely, since D was dense in the unit sphere of L^{p*} . By Hölder duality, this implies that $f \in (L^p)^{**} = L^p$. By construction, $d(f_{n_k}, f) \rightarrow 0$, so the unit ball of L^p is compact under the semimetric d . \square

Proof of the Banach-Alaoglu theorem. Let B be the closed unit ball of L^p and B^* the closed unit ball of L^{p*} . Then any $f \in B$ maps B^* into a closed subset of $D = \{z \in \mathbb{C} : |z| \leq 1\}$. Thus B is a closed subset of the space D^{B^*} of functions $B^* \rightarrow D$. Moreover, the definition of the weakstar topology is exactly the definition of the product topology when restricted to D^{B^*} . So D^{B^*} is compact by Tychonoff's theorem and hence B , a closed subset of D^{B^*} , is as well. \square

One application of compactness theorems, which we discuss in the exercises, is to show that certain PDE have solutions. The idea is to consider a set of “admissible functions” \mathcal{A} , show that \mathcal{A} is compact in some function space X , and then show that an admissible function $u \in \mathcal{A}$ is a solution iff u minimizes the norm $\|u\|_X$. A famous example of this is the Riemann mapping theorem, which says that for every simply connected open set $U \subseteq \mathbb{C}$ there is a holomorphic bijection with holomorphic inverse $U \rightarrow \{z \in \mathbb{C} : z < 1\}$. We sketch the relevant part of the proof of the Riemann mapping theorem in Exercises 6.4.14 and 6.4.15, and the solution of Dirichlet's problem in Exercise 6.4.17.

For the latter, we need a definition that we will use in some more exercises later on, and is of fundamental importance in applications. Recall that $\nabla^i u$ is the vector of all partial derivatives of order i of u , so $\nabla^i u : U \rightarrow \mathbb{C}^{d^i}$ if $U \subseteq \mathbb{R}^d$ is open and $u : U \rightarrow \mathbb{C}$. In particular, $\nabla^0 u = 0$. Since \mathbb{C}^{d^i} is a Banach space (of dimension d^i), one can consider the norms $\|\nabla^i u\|_{L^p}$.

Definition 6.4.8. Let $U \subseteq \mathbb{R}^d$ be an open set and $s \in \mathbb{N}$. The *Sobolev norm* is the norm

$$\|u\|_{H^s(U)} = \sum_{i=0}^s \|\nabla^i u\|_{L^2}.$$

The space $H^s(U)$ is the completion of the space of smooth functions $u : U \rightarrow \mathbb{C}$ such that $\|u\|_{H^s(U)} < \infty$.

In particular, a sequence $(u_n)_n$ converges (strongly) in H^s (say, to u) if for every $i \in \{0, \dots, s\}$, $\nabla^i u_n \rightarrow \nabla^i u$ strongly in L^2 . One can also define weak convergence in H^s by simply requiring that the derivatives converge weakly in L^2 .

Exercise 6.4.9. Let X be a Banach space. Show that the unit ball of X is compact (in the norm topology, not the weakstar topology) iff X is finite-dimensional.

Exercise 6.4.10. Show that the weakstar topology is Hausdorff; that is, if $f_1, f_2 \in L^p$, then there are $g_1, g_2 \in L^{p^*}$ and a $\varepsilon > 0$ such that $B_{g_1}(f_1, \varepsilon) \cap B_{g_2}(f_2, \varepsilon)$ is empty.

Exercise 6.4.11 (Riemann-Lebesgue). Show that if $f_n(x) = \sin nx$, then $f_n \rightarrow 0$ weakly in $L^2([0, 2\pi])$ but not strongly.

Exercise 6.4.12. Show that if $f_n = 1_{\{n\}}$, then $f_n \rightarrow 0$ weakly in ℓ^2 but not strongly.

Exercise 6.4.13 (Banach-Saks). Let $(x_n)_n$ be a bounded sequence in L^2 . Show that there is a subsequence $(x_{n_k})_k$ and a $x \in L^2$ such that

$$\lim_{n \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N x_{n_k} = x$$

strongly in L^2 .

Exercise 6.4.14 (Arzelà-Ascoli). Modify the proof of Alaoglu's sequential compactness theorem to show that, if X is a compact metric space, then $C(X)$ (with L^∞ norm) satisfies the following result: a subset $A \subseteq C(X)$ is compact iff A is closed, bounded, and *equicontinuous* in the sense that for every $\varepsilon > 0$ there is a $\delta > 0$ such that for every $f \in A$ and every $x_1, x_2 \in X$ if $d(x_1, x_2) < \delta$ then $|f(x_1) - f(x_2)| < \varepsilon$. (Here δ does not depend on f or x_1, x_2 .)

Exercise 6.4.15 (Montel's little theorem – for those who know complex analysis). Let U be an open subset of \mathbb{C} , and let $\mathcal{H}(U)$ be the space of holomorphic functions $U \rightarrow \mathbb{C}$. Suppose that $\mathcal{F} \subset \mathcal{H}(U)$ is locally uniformly L^∞ ; that is, for every compact set $K \subset U$ there is a $M_K > 0$ such that for every $f \in \mathcal{F}$, $\|f|_K\|_\infty \leq M_K$. (Hint: You may need to use Cauchy's integral formula and Exercise 6.4.14.) Show that every sequence in \mathcal{F} has a subsequence which converges in L^∞ .

Exercise 6.4.16 (Banach-Alaoglu in Hilbert spaces). Let H be any Hilbert space. Show that the unit ball of H is weakstar compact.

Exercise 6.4.17 (solving Dirichlet's problem). Let U be an open subset of \mathbb{R}^d such that the boundary ∂U is a smooth manifold, and let $f : \partial U \rightarrow \mathbb{R}$ be a continuous function. If you're not comfortable with smooth manifolds, you may assume that U is a ball, in which case ∂U is a sphere. We are in the position to show that Dirichlet's problem, as defined in Exercise 6.1.22, has a solution, if we are willing to take a result known as *Weyl's lemma* as a black box.

Let \mathcal{A} be the admissible set, and let E be the Dirichlet energy, defined in 6.1.22. Proceed as follows:

1. Let $(u_n)_n$ be a sequence of functions in \mathcal{A} such that $E(u_n) \rightarrow \inf_{v \in \mathcal{A}} E(v)$. Show that $(u_n)_n$ is a bounded sequence in H^1 .
2. Show that if $(v_n)_n$ is a sequence of functions which converge in H^1 to a function $v \in H^1$, then $E(v_n) \rightarrow E(v)$.

3. Invoke Exercise 6.4.16.
4. Show that if u is smooth, then u solves Dirichlet's problem iff for every smooth function v which is zero in an open set containing ∂U ,

$$\langle u, \Delta v \rangle_{L^2} = 0. \quad (6.16)$$

The problem of finding a $u \in L^2(U)$ (not necessarily smooth) such that for every smooth v which is zero near ∂U , (6.16) holds, is called the *weak Dirichlet problem*.

5. Invoke Weyl's lemma, which says that if u solves the weak Dirichlet problem and ∂U is a smooth manifold, then u is smooth.

The above proof outline is quite common in PDE and its applications.

6.5 Interpolation of operators

TODO

Exercise 6.5.1 (Riesz-Thorin interpolation in mixed norms). Let $p_1, p_2, q_1, q_2, r \in [1, \infty]$. Let $T : L^{r_1} + L^{r_2} \rightarrow L^{p_0}(X \rightarrow L^{q_0}) + L^{p_1}(X \rightarrow L^{q_1})$ be a linear operator. Establish the Riesz-Thorin inequality

$$\|T\|_{L^{r_\theta} \rightarrow L^{p_\theta}(X \rightarrow L^{q_\theta})} \leq \|T\|_{L^{r_0} \rightarrow L^{p_0}(X \rightarrow L^{q_0})}^{1-\theta} \|T\|_{L^{r_1} \rightarrow L^{p_1}(X \rightarrow L^{q_1})}^\theta$$

whenever $\theta \in [0, 1]$.

6.6 The mean ergodic theorem

Let us give an application of the above theory to ergodic theory.

Throughout we fix a measure-preserving system (Ω, P, T) . Recall that by definition this consists of a probability space (Ω, P) and a measure-preserving map $T : \Omega \rightarrow \Omega$. The measure-preserving system is ergodic if, for every event A , if $T^{-1}A \subseteq A$, then A is either almost surely true, or almost surely false.

Definition 6.6.1. The *Koopman operator* U acts on L^2 by $Uf = f \circ T$.

Recall that a linear operator L on an inner-product space is said to be an isometry if $\langle Lf, Lg \rangle = \langle f, g \rangle$.

Lemma 6.6.2. If $f \in L^2$ then

$$E(Uf) = Ef.$$

In particular, U is unitary.

Proof. This is obvious if $f = 1_A$. We can extend to simple functions by taking linear combinations, then use monotone convergence. \square

Definition 6.6.3. The *time average* of f is

$$A_n f = \frac{1}{n} \sum_{j=0}^{n-1} U^j f.$$

The *space average* of f is Ef .

The idea of time and space averages is as follows. Fix x , the state of the universe today. Then T represents the passing of time from today to tomorrow. Every day, we measure the random variable f , and their average after n days is $A_n f(x)$. Our goal will be to show that $A_n f \rightarrow Ef$ as long as (Ω, P, f) is ergodic, so in order to approximate the expected value of f , we can just keep testing f every day over a long period of time and then take their mean.

To do this, we need two general facts about contractions on Hilbert spaces.

Lemma 6.6.4. Let H be a Hilbert space, $U : H \rightarrow H$ a bounded linear operator such that $\|U\| \leq 1$. If $Ug = g$ then $U^*g = g$.

Proof. If $Ug = g$ then

$$0 = \|U^*g - g\|^2 = \|U^*g\|^2 - \|g\|^2 \leq \|Ug\|^2 - \|g\|^2 = 0$$

so the inequalities collapse and $U^*g = g$. □

Lemma 6.6.5. Let H be a Hilbert space, $U : H \rightarrow H$ a bounded linear operator such that $\|U\| \leq 1$. Let $V = \{f \in H : Uf = f\}$ and let $\Pi : H \rightarrow V$ be the orthogonal projection. Then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} U^j f = \Pi(f). \quad (6.17)$$

Proof. Consider the orthogonal decomposition

$$U = V \oplus V^\perp.$$

We can break up into the cases $f \in V$ and $f \in V^\perp$. If $f \in V$ then obviously both sides of (6.17) are f . If $f \in V^\perp$, $\Pi(f) = 0$. Also

$$\left\| \frac{1}{n} \sum_{j=0}^{n-1} U^j f \right\| \leq \|f\|$$

since U is unitary, so the map $S_n f = \sum_{j=0}^{n-1} U^j f$ is bounded. Therefore

$$g_n = \frac{S_n^* S_n f}{n^2}$$

is a bounded sequence. Since H has a weakstar compact unit ball (Exercise 6.4.16, though in the applications we need this is just the Banach-Alaoglu theorem in L^2), g_n has a weakstar limit g . Now

$$(1 - U^*)(S_n^*/n) = \frac{1}{n} \sum_{j=0}^{n-1} (1 - U^*)(U^*)^{j-1} = \frac{1 - (U^*)^n}{n}$$

satisfies

$$\|(1 - U^*)(S_n^*/n)\| \lesssim n^{-1}$$

so $(1 - U^*)g_n \rightarrow 0$ and hence $(1 - U^*)g = 0$. By the previous lemma, $Ug = g$. Therefore $f \perp g$, so $\Pi(f) = 0$. \square

Theorem 6.6.6 (mean ergodic theorem). Let $f \in L^2$ and let $V = \{g \in L^2 : Ug = g\}$. Let $\Pi : L^2 \rightarrow V$ be the orthogonal projection. Then

$$\lim_{n \rightarrow \infty} A_n f = \Pi(f)$$

in L^2 .

Proof. Expanding out the definitions of A_n and Π , we get (6.17). \square

This theorem is mainly interesting in the case that (Ω, P, T) is an ergodic system. In that case, the only random variables in V are constant almost surely, so $\Pi(f)$ is a constant. TODO: Conditional expectation argument - $\Pi = E$.

Chapter 7

Harmonic analysis

We now sketch some of the applications of measure theory to harmonic analysis as an example. It is reasonable to take the results that we are using, most notably Fubini's theorem and the Riesz-Markov representation theorem, as a black box and peruse this section before we prove them, to get motivation for the utility of such results.

7.1 Haar measures

We begin by introducing measure theory on abelian groups. Recall that a group is called abelian if its group operation is commutative. In that case we will usually write its operations additively; that is, if G is an abelian group, we write $g_1 + g_2$ for the composition of g_1 and g_2 , we let $-g$ denote the inverse of an element g , and let 0 denote the identity of G , so that $g - g = 0$.

Definition 7.1.1. A *locally compact abelian group* G is an abelian group $(G, +)$ equipped with a topology, under which it is locally compact and Hausdorff, $+$ is a continuous function $G \times G \rightarrow G$, and $g \mapsto -g$ is continuous.

We note that the finite product of locally compact abelian groups is a locally compact abelian group. Thus \mathbb{R}^d is a locally compact abelian group, as is the d -dimensional lattice \mathbb{Z}^d , and the d -dimensional torus \mathbb{T}^d . In addition, any finite abelian group is a locally compact abelian group for the discrete topology (wherein every set is open).

We may define, for any set $A \subseteq G$, its translate

$$A + y = \{x + y \in G : x \in A\}.$$

Definition 7.1.2. A *Haar measure* on G is a nonnegative Radon measure μ on G with the following properties:

1. *Translation invariance:* For every Borel set B and every $x \in G$, $\mu(B + x) = \mu(B)$.
2. *Normalization:* If G is compact, then $\mu(G) = 1$.

If G is compact, we call μ a *Haar probability measure*.

Let us give examples of Haar measure. First, Lebesgue measure is a Haar measure on \mathbb{R} and on \mathbb{T} , and counting measure is a Haar measure on \mathbb{Z} . On finite abelian groups G , we define the *uniform probability measure* $\mu(A) = |A|/|G|$ where $|\cdot|$ denotes cardinality. The uniform probability measure of a finite abelian group is Haar.

Recall that if G_1, G_2 are abelian groups, then one can form a group $G_1 \oplus G_2$, called the direct sum of G_1, G_2 , on the set $G_1 \times G_2$ by pointwise operations.

Lemma 7.1.3. The finite product of Haar measures is a Haar measure on the direct sum.

Proof. See Exercise 7.1.6. □

Thus we have Haar measures on, for example, $\mathbb{T}^{d_1} \times \mathbb{Z}^{d_2}$. One can extend this to infinite products, but the statement and proof become much trickier, for algebraic reasons (direct sum and direct product no longer coincide), set-theoretic reasons (the Borel and Baire σ -algebras no longer coincide, and one needs to use Tychonoff's theorem to show that the infinite product of compact spaces is compact), and topological reasons (the infinite product of locally compact spaces is not locally compact). Since we will only ever need Haar measure in finite dimensions, we dodge these issues by restricting to finite products.

We now show that every locally compact abelian group has an essentially unique Haar measure.

Theorem 7.1.4. Let G be a locally compact abelian group. Then there is a Haar measure μ on G . If G is compact, then μ is unique. Otherwise, if ν is another Haar measure, then there is a $c > 0$ such that $\nu = c\mu$.

TODO: Prove Haar's theorem using the Riesz-Markov theorem.

Henceforth if G is a group, we will omit reference to its Haar measure and Borel σ -algebra whenever possible, writing dx instead of $d\mu(x)$ in integrals for example, and $L^1(G)$ to mean $L^1(G, \mu)$. Of course if G is not compact then one could always rescale its Haar measure, but this is rarely important, so we might as well choose whatever normalization is convenient.

If G is compact, then since Haar measure is a probability measure, we can refer to the expected value of a function $f \in L^1(G)$.

Definition 7.1.5. Let μ be a probability measure on a measurable space X , and $f \in L^1(X, \mu)$. Define the *expected value* of f ,

$$Ef = \int_X f \, d\mu.$$

In this case we may refer to f as a *random variable* drawn from X , and refer to $\mu(E)$ as the *probability* of E whenever $E \subseteq X$ is measurable.

Though we will not treat probability theory in detail, it will be at times convenient to use the language of probability theory when thinking about compact abelian groups.

Exercise 7.1.6. Prove Lemma 7.1.3.

Exercise 7.1.7. Let μ be Lebesgue measure, and let ν be a Haar measure for the positive real numbers under multiplication. Compute the Radon-Nikodym derivative of ν with respect to μ .

Exercise 7.1.8. Let G be a compact abelian group with Haar measure μ , $E \subseteq G$ a Borel set such that $\mu(E) > 0$. Show that for every $N \in \mathbb{N}$ there are $g_1, \dots, g_N \in G$ such that

$$\mu\left(\bigcup_{n=1}^N g_n + E\right) \geq 1 - (1 - \mu(E))^N.$$

(Hint: what is the expected value of $\mu\left(\bigcup_{n=1}^N g_n + E\right)$?)

7.2 The Fourier transform

To motivate what follows, we first consider \mathbb{T} and consider a smooth function $f : \mathbb{T} \rightarrow \mathbb{C}$. Such a function is in L^∞ since \mathbb{T} is compact, thus in L^2 since \mathbb{T} has finite measure. To ease notation, we identify \mathbb{T} with $[0, 1)$ and f with a periodic function on $[0, 1]$.

In applications, it is frequently useful to decompose f into a linear combination of waves, namely

$$f(x) = \sum_{\xi=-\infty}^{\infty} \hat{f}(\xi) e^{2\pi i x \xi}. \quad (7.1)$$

For example, if the function $\hat{f} : \mathbb{Z} \rightarrow \mathbb{C}$ has finite support, then f really is just a linear combination of plane waves $e_\xi(x) = e^{2\pi i x \xi}$, which are extremely easy to work with in calculus since they are eigenfunctions of the derivative operator $u \mapsto u'$. On the other hand, the waves e_ξ are an orthonormal basis of $L^2(\mathbb{T})$; that is,

$$\langle e_\xi, e_\eta \rangle = \int_{\mathbb{T}} e^{2\pi i x (\xi - \eta)} dx = \delta_\xi^\eta,$$

so the representation (7.1) converges for the norm of $L^2(\mathbb{T})$, and \hat{f} exists and is unique for any $f \in L^2(\mathbb{T})$. We can recover \hat{f} from f by taking the orthogonal projections

$$\hat{f}(\xi) = \langle f, e_\xi \rangle = \int_{\mathbb{T}} f(x) e^{-2\pi i x \xi} dx. \quad (7.2)$$

Putting together (7.1) and (7.2), we obtain the *Fourier inversion formula*

$$f(x) = \sum_{\xi=-\infty}^{\infty} \int_{\mathbb{T}} f(y) e^{2\pi i (x-y) \xi} dy \quad (7.3)$$

which is valid for any $f \in L^2(\mathbb{T})$. Though at first this just seems like a trick using the structure of $L^2(\mathbb{T})$, it suggests some sort of “duality” between \mathbb{T} and \mathbb{Z} . Moreover, these ideas work in much higher generality.

Recall that if G, H are abelian groups, a *morphism* $\psi : G \rightarrow H$ is a map ψ such that for any $g_1, g_2 \in G$,

$$\psi(g_1 + g_2) = \psi(g_1) + \psi(g_2). \quad (7.4)$$

By \mathbb{C}^* we mean the group of nonzero complex numbers under multiplication. Thus \mathbb{C}^* is a locally compact abelian group. We will write the group operation of \mathbb{C}^* multiplicatively, even though \mathbb{C}^* is abelian. Thus morphisms into \mathbb{C}^* satisfy the relation

$$\psi(g_1 + g_2) = \psi(g_1)\psi(g_2)$$

rather than (7.4). We let $U(1)$ denote the subgroup of \mathbb{C}^* consisting of numbers of absolute value 1; so $U(1)$ is isomorphic to \mathbb{T} , but is written multiplicatively rather than additively.

Definition 7.2.1. Let G be a locally compact abelian group. By a *character* of G we mean a continuous morphism $G \rightarrow U(1)$.

If ξ_1, ξ_2 are characters on a locally compact abelian group G , we define their sum

$$(\xi_1 + \xi_2)(g) = \xi_1(g)\xi_2(g).$$

It is straightforward to check that $\xi_1 + \xi_2$ is a character on G , and that $+$ defines the operation of an abelian group \hat{G} .

We want to be able to do measure theory on \hat{G} , so we need to turn it into a locally compact abelian group. We recall that a sequence of continuous functions f_n is said to converge uniformly provided that f_n converges in L^∞ .

Definition 7.2.2. Let ψ_n be a sequence of characters of a locally compact abelian group G . We say that ψ_n *converges locally uniformly* to a character ψ if for every compact set $K \subseteq G$, the restrictions $\psi_n|K$ converge to $\psi|K$ uniformly on K .

We now give \hat{G} a suitable topology, by declaring that $\psi_n \rightarrow \psi$ in the topology of \hat{G} iff $\psi_n \rightarrow \psi$ locally uniformly.

Definition 7.2.3. Let G be a locally compact abelian group. The *Pontryagin dual* \hat{G} of G is the locally compact abelian group of characters of G under addition and locally uniform convergence.

Later we will prove that $\widehat{\hat{G}} = G$. First, let us compute the Pontryagin duals of some familiar groups.

Example 7.2.4. Let G be a finite abelian group. We claim that G is isomorphic to \hat{G} . To see this, we first appeal to the classification of finite abelian groups to write

$$G \cong \frac{\mathbb{Z}}{n_1} \oplus \cdots \oplus \frac{\mathbb{Z}}{n_k}.$$

Here \mathbb{Z}/n denotes the cyclic group of order n . It is a straightforward exercise in group theory to see that this implies that

$$\hat{G} \cong \frac{\widehat{\mathbb{Z}}}{n_1} \oplus \cdots \oplus \frac{\widehat{\mathbb{Z}}}{n_k}$$

(indeed, it follows from the universal property of \oplus), so it suffices to check the claim when $G = \mathbb{Z}/n$.

Let g be the generator of \mathbb{Z}/n and let $kg = g + \cdots + g$ (k copies of g). A character $\xi : \mathbb{Z}/n \rightarrow \mathbb{C}^*$ is determined by $\xi(g)$, and $1 = \xi(ng) = \xi(g)^n$ since \mathbb{Z}/n is cyclic. Therefore $\xi(g)$ is an n th root of unity $e^{2\pi i \ell/n}$. If $\xi_j(g) = e^{2\pi i \ell_j/n}$, then $\xi_1(g) + \xi_2(g) = e^{2\pi i(\ell_1 + \ell_2)/n}$. Since $\xi \mapsto \xi(g)$ is an injective function between two sets of the same cardinality, it is a bijection. The map $\xi \mapsto \xi(g)$ therefore gives an isomorphism between $\widehat{\mathbb{Z}/n}$ and the group of n th roots of unity of \mathbb{C}^* under multiplication, the latter of which is isomorphic to \mathbb{Z}/n since it is generated by $e^{2\pi i/n}$ and has n elements.

Example 7.2.5. TODO: Duality of \mathbb{T} and \mathbb{Z} .

We leave the following important duality result to the reader.

Theorem 7.2.6. The map $\eta : \mathbb{R} \rightarrow \hat{\mathbb{R}}$ defined by $\eta(\xi)(x) = e^{2\pi i x \xi}$ is an isomorphism.

Motivated by the duality of \mathbb{T} and \mathbb{Z} , we now define the Fourier transform, a generalization of the Fourier series to arbitrary locally compact abelian groups.

Definition 7.2.7. Let G be a locally compact abelian group, and $f \in \mathbf{IF}(G)$. The *Fourier transform* \hat{f} of f is defined by

$$\hat{f}(\xi) = \int_G f(x) \xi(-x) dx.$$

We extend the Fourier transform to all of $L^2(G)$ by density.

It is not obvious that the above definition makes sense. Certainly if $f \in \mathbf{IF}(G)$ then the definition of \hat{f} is uncontroversial, but why does it extend to $L^2(G)$? This happens because of the following theorem:

Theorem 7.2.8 (Plancherel). TODO: <https://www.math.ucla.edu/~tao/247b.1.07w/notes9.pdf> need to prove Fourier inversion first

7.3 Pontryagin duality

7.4 The Fourier transform on \mathbb{R}^d

blah blah

Exercise 7.4.1. If μ is a Borel probability measure on \mathbb{R} , we can define its *Fourier transform* by

$$\hat{\mu}(\xi) = \int_{-\infty}^{\infty} e^{-ix\xi} d\mu(x).$$

Assume that μ is absolutely continuous, and its Radon-Nikodym derivative with respect to Lebesgue measure is denoted μ' . Show that

$$\mu'(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ix\xi} \hat{\mu}(\xi) d\xi.$$

(Hint: Let φ be a suitably chosen Schwartz function and let $F(x) = \mu((-\infty, x])$ be the cumulative distribution function of μ . Show that

$$F(x) = \lim_{\varepsilon \rightarrow 0} \frac{1}{2\pi} \int_{-\infty}^x \int_{-\infty}^{\infty} e^{iy\xi} \varphi(\varepsilon^2 \xi^2) \hat{\mu}(\xi) d\xi dy$$

and commute the limit with the iterated integral.) Conclude that you have given an alternative proof of the Radon-Nikodym theorem for Lebesgue measure.

Exercise 7.4.2 (Hausdorff-Young inequality). Show that if p^* is the Hölder dual to p and $p \in [1, 2]$ then for every Schwartz function f ,

$$\|\hat{f}\|_{L^{p^*}(\mathbb{R}^d)} \leq \|f\|_{L^p(\mathbb{R}^d)},$$

so the Fourier transform can be extended to a linear map $L^p(\mathbb{R}^d) \rightarrow L^{p^*}(\mathbb{R}^d)$.

Exercise 7.4.3 (restriction estimates). Suppose that $p \in [1, 2]$ and $q \in [1, \infty]$. Let σ denote Lebesgue measure on the unit sphere S^{d-1} (defined in Exercise 2.1.10). Suppose that, for every Schwartz function f , the *restriction estimate*

$$\|\hat{f}\|_{L^q(S^{d-1}, \sigma)} \lesssim \|f\|_{L^p(\mathbb{R}^d)}$$

holds. Show that for every $f \in L^p(\mathbb{R}^d)$, the restriction $\hat{f}|_{S^{d-1}}$ is defined for σ -almost every point of S^{d-1} . Restriction estimates are an active area of research in harmonic analysis; the above restriction estimate is valid whenever $p = 2$ and $q \in [1, 4d/(3d+1)]$.

Exercise 7.4.4. Let $s \in \mathbb{R}$. Define the *Sobolev measure* μ_s so that its Radon-Nikodym derivative with respect to Lebesgue measure is $(1 + |\xi|^s)^2$. Show that for any $s \in \mathbb{N}$ and $u \in H^s(\mathbb{R}^d)$ one has

$$\|u\|_{H^s} \lesssim \|u\|_{L^2(\mu_s)} \lesssim \|u\|_{H^s}.$$

So we define the *Sobolev norm* H^s , s any real number, by setting $\|u\|_{H^s} = \|u\|_{L^2(\mu_s)}$. This allows us to define the *Sobolev space* H^s in general.

Exercise 7.4.5. This Exercise should be done after Exercise 7.4.4. Let $s \in \mathbb{R}$. Show that there is an isomorphism $(H^s)^* \rightarrow H^{-s}$ – so while we can think of H^s as its own dual (since H^s is a Hilbert space), we can also think of H^{-s} and H^s as dual spaces.

Exercise 7.4.6. Let $a : \mathbb{R}^2 \rightarrow \mathbb{C}$ be a smooth function such that $a \in L^\infty$ and all partial derivatives of a are in L^∞ . Define the *pseudodifferential operator*

$$Af(x) = \int_{-\infty}^{\infty} a(x, \xi) \hat{f}(\xi) e^{ix\xi} d\xi.$$

We call A the *Kohn-Nirenberg quantization* of a , and a the *symbol* of A . Show that Af exists whenever $f \in \mathcal{S}(\mathbb{R})$ and A maps $\mathcal{S}(\mathbb{R})$ to itself. Show that the Kohn-Nirenberg quantization $a(x, \xi) = x$ satisfies $Af(x) = xf(x)$, and the Kohn-Nirenberg quantization of $a(x, \xi) = \xi$ satisfies $Af = f'$, up to a multiplicative constant.

7.5 Applications to quantum mechanics

The following three sections are independent of each other. We include this section first because it is arguably the easiest of the three. Here we prove the uncertainty principle of Heisenberg, which says that one cannot know both the position and the momentum of a subatomic particle.

In quantum mechanics, one typically models a particle P with one degree of freedom by a Schwartz function f , its *wavefunction*, such that $\|f\|_{L^2} = 1$. If $A \subseteq \mathbb{R}$ is Borel, the probability of measuring that P is inside A is $\|f\|_{L^2(A)}^2$. Moreover, the momentum of P also has a wavefunction, given by \hat{f} ; the probability of measure that the momentum of P is inside A is $\|\hat{f}\|_{L^2(A)}^2$.

We now introduce the *position* operator $Xf(x) = xfx$ and *momentum* operator $Df(x) = f'(x)/2\pi i$. By Lemma TODO, D is the conjugation of X by the Fourier transform, as one might expect. Notice that X, D both preserve Schwartz space.

Now if we have two operators A, B on Schwartz space, we can introduce their *commutator* $[A, B] = AB - BA$. The commutator measures how badly A, B fail to commute. In the case of position and momentum, the answer is “quite badly”.

Theorem 7.5.1 (von Neumann’s canonical commutation relations). One has

$$[X, D] = -\frac{1}{2\pi i}.$$

This can be easily verified by the reader using the product rule for differentiation.

We now recall that the expected value of a random variable is its mean value. We leave it as an exercise to check that the expected value of the position of a particle with wavefunction f is

$$\int_{-\infty}^{\infty} x |f(x)|^2 dx.$$

Similarly,

$$\int_{-\infty}^{\infty} \xi |\hat{f}(\xi)|^2 d\xi$$

is the expected value of the momentum. In statistics, one defines the variance of a random variable A to be the difference between the expected value of A^2 and the expected value of A , squared. This is a measure of how far a “typical” measurement will deviate from the expected value. TODO: Move to probability ch?

TODO: Heisenberg

7.6 Applications to PDE

TODO: Elliptic equation have solutions

7.7 Applications to number theory

In this section we use the theory that we have developed thus far to prove the famous functional equation for the Riemann zeta function. The proof technique is of great importance to number theory, where it leads one into the theory of modular forms, L -functions, the Langlands program, and Tate's thesis. However, the author confesses complete ignorance of such matters, and refers the brave reader to `TODO:Cite`.

We start by recalling the basic definitions from number theory.

Definition 7.7.1. The *Riemann zeta function* is the function

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s},$$

defined when $\operatorname{Re} s > 1$.

Then ζ is holomorphic in the half-plane $\Omega = \{s \in \mathbb{C} : \operatorname{Re} s > 1\}$. Indeed, the summands n^{-s} are holomorphic in s and the partial sums converge locally uniformly in Ω (that is, converge in $L^\infty(K)$ for any compact $K \subset \Omega$), and the locally uniform limit of holomorphic functions is holomorphic (`TODO:Cite a complex analysis book`). This function is interesting to number theorists because of the following theorem.

Theorem 7.7.2 (Euler's product formula). For every $s \in \Omega$,

$$\zeta(s) = \prod_p \frac{1}{1 - p^{-s}},$$

where the product ranges over all prime numbers p .

Proof. We induct on the primes. First,

$$\frac{1}{2^s} \zeta(s) = \sum_{n \not\equiv 0 \pmod{2}} n^{-s}$$

so

$$\zeta(s) \left(1 - \frac{1}{2^s}\right) = \sum_{n \not\equiv 0 \pmod{2}} n^{-s}.$$

Now suppose that q is a prime and

$$\zeta(s) \prod_p \left(1 - \frac{1}{p^s}\right) = \sum_n n^{-s} \tag{7.5}$$

where the product is taken over all primes $p < q$, and the sum is taken over all n such that for each prime $p < q$, $n \not\equiv 0 \pmod{p}$. Then

$$\frac{1}{q^s} \zeta(s) \prod_p \left(1 - \frac{1}{p^s}\right) = \sum_n n^{-s}$$

where the product is again taken over all primes $p < q$ and the sum is taken over all n such that $n \equiv 0 \pmod q$ and for each prime $p < q$, $n \not\equiv 0 \pmod p$. Therefore (7.5) holds when the product is taken over all primes $p \leq q$ and the sum is taken over all n such that for every $p \leq q$, $n \not\equiv 0 \pmod p$. This completes the induction. Taking $q \rightarrow \infty$ in (7.5) we see that

$$\zeta(s) \prod_p \left(1 - \frac{1}{p^s}\right) = 1$$

where the product is taken over all primes p and the infinite product converges by a monotone convergence argument (since $\sum_n n^{-s}$, the sum taken over all natural numbers n , converges). Indeed, $1 \not\equiv 0 \pmod p$ for any prime p , but any $n \geq 2$ eventually divides some prime. The claim now follows. \square

Now recall that the *gamma function* is the function

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$$

which converges if $\operatorname{Re} z > 0$.

Definition 7.7.3. Define the *Riemann xi function*

$$\xi(s) = \pi^{-s/2} \Gamma(s/2) \zeta(s).$$

Theorem 7.7.4 (Riemann's functional equation). For any $s \in \Omega$,

$$\xi(s) = \xi(1-s). \quad (7.6)$$

Therefore $\xi(s)$, and hence $\zeta(s)$, can be defined for any $s \neq 1$, and ζ extends to a holomorphic function on $\mathbb{C} \setminus \{1\}$.

Riemann's functional equation was first conjectured by Euler, and is one of the cornerstones of number theory. One of the most fundamental open problems in mathematics is the *Riemann hypothesis*, which asserts that if s is a zero of ζ with $\operatorname{Re} s \in [0, 1]$, then $\operatorname{Re} s = 1/2$. If the Riemann hypothesis were true, then it would give sharp bounds on the growth of the prime-counting function.

We prove (7.6) later in this section; the holomorphy claim is easy to check for any student of complex analysis once (7.6) is established. First we relate (7.6) to harmonic analysis.

We shall recall some algebraic machinery. The trivial group is denoted 0 ; for any group G , there are unique morphisms $0 \rightarrow G$ and $G \rightarrow 0$, namely the trivial morphism, which we leave unlabeled when clear from context. We let 0 denote the trivial morphism between any two groups when we do need to specify it.

A *short exact sequence* of abelian groups is a commutative diagram

$$0 \longrightarrow A \xrightarrow{\alpha} B \xrightarrow{\beta} C \longrightarrow 0$$

where A, B, C are abelian groups, all the arrows are morphisms, and the kernel of any arrow is the image of the arrow before it. This is equivalent to requiring that α is injective, β is

surjective, and the image of α is the kernel of β , thus for any $b \in B$, $\beta(b) = 0$ iff there is an $a \in A$ with $b = \alpha(a)$. Whenever B is a subgroup of A , we get a short exact sequence

$$0 \longrightarrow B \xrightarrow{\iota} A \xrightarrow{q} \frac{A}{B} \longrightarrow 0 \quad (7.7)$$

where ι is the inclusion map $\iota(b) = b$ and q is the quotient map.

If $\alpha : A \rightarrow B$ is a continuous morphism between locally compact abelian groups, then we can define its Pontryagin dual $\hat{\alpha} : \hat{B} \rightarrow \hat{A}$ by

$$\hat{\alpha}(\xi) = \xi \circ \alpha.$$

Note that this is not the same as taking the Fourier transform of α ; we will always use Roman letters for functions and Greek letters for morphisms, to avoid confusion. If $B \subseteq A$, and $\iota : B \rightarrow A$ is the inclusion morphism $\iota(b) = b$, then $\hat{\iota}(\xi) = \xi \circ \iota = \xi|_B$, thus $\hat{\iota}$ is the restriction map $\hat{A} \rightarrow \hat{B}$.

Theorem 7.7.5. Pontryagin duality is *exact* in the sense that whenever one has a diagram of locally compact abelian groups and continuous morphisms between them

$$A \xrightarrow{\alpha} B \xrightarrow{\beta} C$$

such that the image of α is the kernel of β , then the image of $\hat{\beta}$ is the kernel of $\hat{\alpha}$.

Proof. Let $H = \hat{\beta}(\hat{C})$. Then $\hat{\beta} \circ \hat{\alpha} = \widehat{\beta \circ \alpha} = \hat{0} = 0$. Therefore $H \subseteq \ker \hat{\alpha}$. Applying the Pontryagin duality theorem, we see that if $H \subset \ker \hat{\alpha}$ then $\widehat{\ker \hat{\alpha}} = \alpha(A)$ is a proper subgroup of $\hat{H} = \ker \beta$, which is impossible. \square

Corollary 7.7.6. Let A be a locally compact abelian group, B a closed subgroup of A . Then the Pontryagin dual of (7.7),

$$0 \longrightarrow \frac{\hat{A}}{\hat{B}} \xrightarrow{\hat{q}} \hat{A} \xrightarrow{\hat{\iota}} \hat{B} \longrightarrow 0,$$

is a short exact sequence of locally compact abelian groups.

Proof. This immediately follows from Theorem (7.7.5). The reader can check that since B is closed, all the morphisms are continuous and A/B is a locally compact abelian group. \square

Theorem 7.7.7 (Poisson summation). Let B be a closed subgroup of a locally compact abelian group A . If $f \in L^1(A)$ let

$$f^B(x + B) = \int_B f(x + y) dy.$$

Then for any $\xi \in \widehat{A/B}$, $\widehat{f^B}(\xi) = \hat{f}([\xi])$ where $[\xi](x) = \xi(x + B)$ for any $x \in A$.

If $f^B \in L^1(\widehat{A/B})$, then for almost all $x \in G$

$$\int_B f(x + y) dy = \int_{\widehat{A/B}} \hat{f}([\xi])[\xi](x) d\xi. \quad (7.8)$$

If f^B is continuous and everywhere defined, then (7.8) is valid for every $x \in G$.

Proof. If $\xi \in \widehat{A/B}$ then $\xi|B = 0$. Indeed, $\xi|B = \hat{\iota}(\xi) = \hat{\iota}(\hat{q}(B)) = 0$, by Corollary 7.7.6. Therefore $\xi(x+y) = \xi(x)$ for any $(x, y) \in A \times B$. So

$$\begin{aligned} \widehat{f^B}(\xi) &= \int_{A/B} f^B(x+B) \overline{\xi(x)} d(x+B) \\ &= \iint_{A/B \times B} f(x+y) \overline{\xi(x+y)} dy d(x+B) \\ &= \int_A f(x) \overline{[\xi](x)+B} dx = \hat{f}([\xi]). \end{aligned}$$

Now if $f^B \in L^1(\widehat{A/B})$, then by the Fourier inversion formula,

$$f^B(x+B) = \widehat{\widehat{f^B}}(-x+B) = \int_{\widehat{A/B}} \hat{f}([\xi]) \overline{[\xi](x)} d\xi$$

almost everywhere, and everywhere under the necessary hypotheses. Substituting the definition of f^B in the above equation we obtain (7.8) \square

Corollary 7.7.8 (Poisson summation on \mathbb{R}). Let f be a Schwartz function on \mathbb{R} . Then for every $x \in \mathbb{R}$,

$$\sum_{k=-\infty}^{\infty} f(x+k) = \sum_{\xi=-\infty}^{\infty} \hat{f}(\xi) e^{-2\pi i x \xi}. \quad (7.9)$$

Proof. Plugging \mathbb{R} and \mathbb{Z} into (7.7) we obtain a short exact sequence

$$0 \longrightarrow \mathbb{Z} \longrightarrow \mathbb{R} \longrightarrow \mathbb{T} \longrightarrow 0.$$

Dualizing, we obtain a bijection $\mathbb{Z} \rightarrow \widehat{\mathbb{T}} = \mathbb{Z}$. Since f is Schwartz, $f \in L^1(\mathbb{R})$ and \hat{f} is Schwartz, so that $\hat{f}|_{\mathbb{Z}} \in L^1(\mathbb{Z})$. Then plugging in f into (7.8) we obtain (7.9). \square

Corollary 7.7.9. Let f be a Schwartz function on \mathbb{R} . Then

$$\sum_{k=-\infty}^{\infty} f(k) = \sum_{\xi=-\infty}^{\infty} \hat{f}(\xi). \quad (7.10)$$

Definition 7.7.10. Define, for $\operatorname{Re} z > 0$,

$$\Theta(z) = \sum_{k=-\infty}^{\infty} e^{-z\pi k^2},$$

the *Jacobi theta function* of one variable.

Lemma 7.7.11. For every $\operatorname{Re} z > 0$,

$$\sqrt{z}\Theta(z) = \Theta(z^{-1}).$$

Proof. By uniqueness of analytic continuation, a theorem of complex analysis, it suffices to check when $\text{Im } z = 0$. Let $f_z(x) = e^{-z\pi x^2}$. Then $\hat{f}_1 = f_1$ by Lemma TODO, so by Lemma TODO,

$$\sqrt{z}\hat{f}_z = f_{1/z}.$$

Moreover, f_z is Schwartz, so the claim now follows by applying (7.10). \square

Proof of (7.6). Note that dt/t is the Haar measure of \mathbb{R}^+ , the positive real numbers under multiplication. One has

$$\xi(s) = \sum_{n=1}^{\infty} \int_0^{\infty} n^{-s} t^{s/2} \pi^{-s/2} e^{-t} \frac{dt}{t} = \sum_{n=1}^{\infty} \int_0^{\infty} \left(\frac{t}{n^2 \pi} \right)^{s/2} e^{-t} \frac{dt}{t}.$$

Simplifying, we see that

$$\xi(s) = \int_0^{\infty} t^{s/2} \frac{\Theta(t) - 1}{2} \frac{dt}{t} = \left[\int_0^1 + \int_1^{\infty} \right] t^{s/2} \frac{\Theta(t) - 1}{2} \frac{dt}{t}.$$

Now $\Theta - 1$ is Schwartz, so the integral over $[1, \infty]$ converges locally uniformly in s , hence to an entire function in s . Applying the transform $t \mapsto 1/t$, which preserves the Haar measure of \mathbb{R}^+ , and using Lemma 7.7.11,

$$\int_0^1 t^{s/2} \frac{\Theta(t) - 1}{2} \frac{dt}{t} = \int_1^{\infty} t^{-s/2} \frac{\sqrt{t}\Theta(t) - 1}{2} \frac{dt}{t}$$

thus

$$\xi(s) = \int_1^{\infty} t^{s/2} \frac{\Theta(t) - 1}{2} \frac{dt}{t} + \int_1^{\infty} t^{(1-s)/2} \frac{\Theta(t) - 1}{2} \frac{dt}{t} + \frac{1}{2} \left[\int_1^{\infty} t^{(1-s)/2} \frac{dt}{t} - \int_1^{\infty} t^{-s/2} \frac{dt}{t} \right].$$

But

$$-\frac{1}{2} \left[\int_1^{\infty} t^{(1-s)/2} \frac{dt}{t} - \int_1^{\infty} t^{-s/2} \frac{dt}{t} \right] = \frac{1}{s} + \frac{1}{1-s}.$$

Therefore

$$\xi(s) = \int_1^{\infty} t^{s/2} + t^{(1-s)/2} \frac{\Theta(t) - 1}{2} \frac{dt}{t} - \left(\frac{1}{s} + \frac{1}{1-s} \right). \quad (7.11)$$

The right-hand side of (7.11) does not change when one replaces s with $1-s$, thus $\xi(s) = \xi(1-s)$. \square

Appendix A

Linear algebra

A.1 Normed spaces

The reader should be familiar with this material before reading Chapter 1.

When one first learns what a “vector” is, they are told that a vector is comprised of a length and a direction. However, the algebraic definition of a vector space does not satisfy this property; nothing in the definition of a vector space allows one to canonically assign lengths to vectors. In this section we correct this matter by introducing a notion of length.

We take all vector spaces to be over the real numbers \mathbb{R} or the complex numbers \mathbb{C} (preferably the latter). We let $[0, \infty)$ denote the nonnegative real numbers.

Definition A.1.1. A *seminormed space* is a vector space V equipped with a function

$$\begin{aligned} V &\rightarrow [0, \infty) \\ v &\mapsto \|v\|, \end{aligned}$$

known as a *seminorm*, such that for any $v, w \in V$ and c a scalar,

$$\|v + w\| \leq \|v\| + \|w\|,$$

the *triangle inequality*, and

$$\|cv\| = |c| \cdot \|v\|.$$

The quantity $\|v\|$ is called the *length* of v .

A *normed space* is a seminormed space V such that the only $v \in V$ such that $\|v\| = 0$ is $v = 0$. A seminorm with this property is called a *norm*.

Every vector space can be turned into a seminormed space in a trivial way, namely by setting $\|v\| = 0$ for every v . However, we will have no use for this.

Example A.1.2. The most important example of a normed space has as its base space \mathbb{R}^d . We define

$$\|(x_1, \dots, x_d)\|_2^2 = \sum_{i=1}^d |x_i|^2,$$

thus $\|\cdot\|_2$ is the *Euclidean norm* on \mathbb{R}^d . Then the triangle inequality is just the usual triangle inequality for Euclidean geometry.

Example A.1.3. We can define several other norms on \mathbb{R}^d , closely related to the Euclidean norm $\|\cdot\|_2$. First we let

$$\|(x_1, \dots, x_d)\|_\infty = \max_{i=1}^d |x_i|.$$

We then define, for any $p \in [1, \infty)$,

$$\|(x_1, \dots, x_d)\|_p^p = \sum_{i=1}^d |x_i|^p.$$

Thus when $p = 2$ we recover the Euclidean norm, when $p = 1$ we recover the “sum norm” $\|(x_1, \dots, x_d)\|_1 = \sum_i |x_i|$, and in the limit $p \rightarrow \infty$ we obtain the $\|\cdot\|_\infty$ norm.

A normed space is a metric space in a natural way, namely the distance between two vectors v, w is defined to be $\|v - w\|$. Thus we have access to the usual notion of sequences, convergence, etc. for normed spaces; so the equation

$$\lim_{n \rightarrow \infty} v_n = w$$

means that for every $\varepsilon > 0$ there is an N such that for every $n \geq N$, $\|v_n - w\| < \varepsilon$. Similarly a sequence of v_n is Cauchy if for every $\varepsilon > 0$ there is an N such that for every $n, n' \geq N$, $\|v_n - v_{n'}\| < \varepsilon$.

The notion of convergence makes sense in seminormed spaces, but it is no longer true that the limit of a sequence need be unique. In fact, consider the seminorm on \mathbb{R}^2 defined by

$$\|(x, y)\| = |x|.$$

Let $(x_n, y_n)_n$ be a sequence in \mathbb{R}^2 and $x \in \mathbb{R}$. Then for every $\lim_n x_n = x$, then for every $y \in \mathbb{R}$, $\lim_n (x_n, y_n) = (x, y)$. In the language of point-set topology, seminormed spaces are not Hausdorff (nor do they even satisfy Axiom T_0).

We fix the problem with seminormed spaces by observing that it is always possible to turn a seminormed space into a normed space.

Theorem A.1.4. For every seminormed space V there is a normed space V' , called the *normalization* of V , such that:

1. There is a surjective linear map $\pi : V \rightarrow V'$.
2. For every normed space W and every linear map $T : V \rightarrow W$ such that for every v , $\|Tv\| = \|v\|$, there is a linear map $T' : V/\ker V \rightarrow W$ such that $\|T'v\| = \|v\|$ and the diagram

$$\begin{array}{ccc} V & \xrightarrow{T} & W \\ & \searrow \pi & \nearrow T' \\ & \frac{V}{\ker V} & \end{array} \quad (\text{A.1})$$

commutes.

In the language of category theory, the diagram (A.1) is called the *universal property of the normalization*. In category theory one proves that whenever an algebraic object satisfies a universal property, the object is unique up to a unique choice of isomorphism. Thus the normalization (and the closely related completion that we discuss in Theorem A.2.5) is unique.

The reader should omit the proof of Theorem A.1.4 on first reading.

Proof of Theorem A.1.4. In fact, we define the *kernel* of a seminormed space V by $\ker V = \{v \in V : \|v\| = 0\}$. The kernel is a subspace of V , so we can take the quotient space $V/\ker V$ and form a short exact sequence

$$0 \longrightarrow \ker V \longrightarrow V \xrightarrow{\pi} \frac{V}{\ker V} \longrightarrow 0 \quad (\text{A.2})$$

(so the composite of any two arrows in the diagram (A.2) is the zero map, π is the natural projection of V onto $V/\ker V$, and $\ker V$ is the kernel of π). We define a norm on $V/\ker V$ by

$$\|\pi(v)\| = \|v\|.$$

We define the normalization of V to be $V/\ker V$.

To check the universal property, let $T : V \rightarrow W$ be a linear map into a normed space W such that $\|Tv\| = \|v\|$. In particular, if $v \in \ker V$, $\|Tv\| = 0$ so $Tv = 0$ and hence $v \in \ker T$. Thus we may define $T'(\pi(v)) = Tv$; this is well-defined because if $\pi(v) = \pi(v')$, then $v - v' \in \ker V$ and hence in $\ker T$. \square

We conclude this section with a useful consequence of the triangle inequality.

Lemma A.1.5 (reverse triangle inequality). For every v, w in a normed space,

$$|||v|| - ||w||| \leq \|v - w\|.$$

Proof. One has

$$\|v\| = \|v + w - w\| \leq \|v - w\| + \|w\|$$

so

$$\|v\| - \|w\| \leq \|v - w\|.$$

Similarly,

$$\|w\| - \|v\| \leq \|v - w\|.$$

But $\max(\|v\| - \|w\|, \|w\| - \|v\|) = |||v|| - ||w|||$. \square

A.2 Banach spaces

The reader should be familiar with this material before reading Chapter 1.

The defining property of \mathbb{R} is that every Cauchy sequence in \mathbb{R} converges. However, this property is not true for normed spaces, as the following example will show.

Example A.2.1. Let $C[0, 1]$ denote the vector space of continuous functions $[0, 1] \rightarrow \mathbb{C}$. We turn $C[0, 1]$ into a normed space by introducing a Euclidean-type norm,

$$\|f\|_2^2 = \int_0^1 |f(x)|^2 dx.$$

It is easy to check that $\|f\|_2$ is a seminorm, and to see that it is a norm, note that if $\int_0^1 |g|^2 = 0$, then for every $\varepsilon > 0$ there is a partition of $[0, 1]$ into intervals I_1, \dots, I_n such that for every i , the length of I_i is $< \varepsilon$ and the midpoint I_i^* of I_i satisfies $|g(I_i^*)| < \varepsilon$. Thus the set of points x such that $|g(x)| < \varepsilon$ is ε -dense (i.e. for every y there is an x such that $|x - y| < \varepsilon$ and $|g(x)| < \varepsilon$). Taking $\varepsilon \rightarrow 0$ we see that the set of points x such that $g(x) = 0$ is dense (i.e. for every $\delta > 0$ and every y we can find an x with $|x - y| < \delta$ and $g(x) = 0$). But g is continuous so $g = 0$.

Now define

$$f_n(x) = \begin{cases} 0, & x \leq 1/2 - 1/n \\ n(x - 1/2 - 1/n), & 1/2 - 1/n \leq x \leq 1/2 + 1/n \\ 2, & x \geq 1/2 + 1/n \end{cases}.$$

TODO: Draw a picture To see that $(f_n)_n$ is Cauchy, note that if $m > n$ then

$$\|f_n - f_m\|_2^2 = \int_{1/2-1/n}^{1/2+1/n} (m(x - 1/2 - 1/m) - n(x - 1/2 - 1/n))^2 dx$$

and the integrand satisfies

$$(m(x - 1/2 - 1/m) - n(x - 1/2 - 1/n))^2 \leq 8.$$

Thus

$$\|f_n - f_m\|_2^2 \leq \int_{1/2-1/n}^{1/2+1/n} 8 dx = \frac{16}{n}.$$

Therefore

$$\|f_n - f_m\|_2 \leq \frac{4}{\sqrt{n}} \rightarrow 0$$

as $n \rightarrow \infty$. However, it is not too hard to check that if

$$f(x) = \begin{cases} 0, & x \leq 1/2 \\ 1, & x > 1/2 \end{cases}$$

then

$$\lim_{n \rightarrow \infty} \|f_n - f\|_2 = 0;$$

yet f is not continuous, and $\|\cdot\|_2$ is a norm, so f_n cannot converge to any continuous function.

Definition A.2.2. A *Banach space* is a normed space for which every Cauchy sequence converges.

Example A.2.3. The p -norms $\|\cdot\|_p$ that we defined on \mathbb{R}^d turn \mathbb{R}^d into a Banach space. This is an exercise in the fact that every Cauchy sequence on \mathbb{R} converges.

The advantage of working in Banach spaces rather than general normed spaces is that it is meaningful to talk about infinite sums in Banach spaces. Indeed, if $(x_n)_n$ is a sequence of elements in a Banach space X , we define

$$\sum_{n=1}^{\infty} x_n = \lim_{N \rightarrow \infty} \sum_{n=1}^N x_n \quad (\text{A.3})$$

whenever the limit on the right-hand side of (A.3) makes sense; thus, $y = \sum_n x_n$ if for every $\varepsilon > 0$, there is an N such that

$$\left\| y - \sum_{n=1}^N x_n \right\| < \varepsilon.$$

In fact, the right-hand side of (A.3) makes sense as long as the partial sums are Cauchy; thus, for every $\varepsilon > 0$, there is an N such that

$$\left\| \sum_{n=N}^{\infty} x_n \right\| < \varepsilon.$$

By the triangle inequality, it in fact suffices to show that for every $\varepsilon > 0$ there is an N such that

$$\sum_{n=N}^{\infty} \|x_n\| < \varepsilon \quad (\text{A.4})$$

to show that the $(x_n)_n$ are summable.

Definition A.2.4. Let $(x_n)_n$ be a sequence in a Banach space. If for every $\varepsilon > 0$ there is an N such that (A.4) holds, we say that $(x_n)_n$ is *absolutely convergent* or *absolutely summable*.

Just as we had a universal way to turn any seminormed space into a normed space, its normalization, we have a universal way to turn any normed space (hence any seminormed space) into a Banach space by adding limits to each of its Cauchy sequences.

Theorem A.2.5. For every normed space V there is a Banach space W , called the *completion* of V , such that:

1. There is an injective linear map $\iota : V \rightarrow W$ such that for every v , $\|\iota(v)\| = \|v\|$.
2. The image of ι is dense in W .
3. The completion satisfies the *universal property of the completion*: for any Banach space X and any linear map $T : V \rightarrow X$ such that for every v , $\|Tv\| = \|v\|$, then there is a linear map $T' : W \rightarrow X$ such that the diagram

$$\begin{array}{ccc} V & \xrightarrow{T} & X \\ & \searrow \iota & \nearrow T' \\ & W & \end{array}$$

commutes.

As with the normalization, the completion is unique for category-theoretic reasons, and the reader should omit the proof of this theorem on first reading.

Proof. To do this, let V be a normed space and let $\mathbf{Cau}(V)$ be the vector space of all Cauchy sequences in V . (We let x denote a Cauchy sequence $(x_n)_n$.) Then $\mathbf{Cau}(V)$ is a seminormed space, where

$$\|x\| = \lim_{n \rightarrow \infty} \|x_n\|.$$

(To see that the limit exists, note that if x is a Cauchy sequence then $n \mapsto \|x_n\|$ is a Cauchy sequence in \mathbb{R} , so it converges.) Let W be the normalization of $\mathbf{Cau}(V)$ and $\pi : \mathbf{Cau}(V) \rightarrow W$ the natural projection.

We claim that W is the completion of V . First, W is a Banach space, because if $x \in \mathbf{Cau}(W)$, then we can choose $\tilde{x} \in \mathbf{Cau}(\mathbf{Cau}(V))$ such that for every n ,

$$\pi(\tilde{x}_n) = x_n.$$

Now \tilde{x}_n is a Cauchy sequence, say $(\tilde{x}_{n,m})_m$. If we replace \tilde{x}_n with a subsequence (in m) of \tilde{x}_n , $\pi(\tilde{x}_n)$ will not change. Thus we might as well assume that for any two m, m' ,

$$\|\tilde{x}_{n,m} - \tilde{x}_{n,m'}\| < \frac{1}{n}.$$

Now let

$$\tilde{y}_n = \tilde{x}_{n,n}.$$

We want $x_n \rightarrow \pi(\tilde{y})$ but it's not obvious that \tilde{y} is a Cauchy sequence.

Let $\varepsilon > 0$, so there is an $N > \varepsilon^{-1}$ such that for every $n, n' \geq N$,

$$\|x_n - x_{n'}\| < \varepsilon$$

and for any j ,

$$\begin{aligned} \|\tilde{y}_n - \tilde{y}_{n'}\| &= \|\tilde{x}_{n,n} - \tilde{x}_{n',n'}\| \\ &\leq \|\tilde{x}_{n,n} - \tilde{x}_{n,j}\| + \|\tilde{x}_{n,j} - \tilde{x}_{n',j}\| + \|\tilde{x}_{n',j} - \tilde{x}_{n',n'}\| \\ &\leq \frac{1}{n} + \|\tilde{x}_{n,j} - \tilde{x}_{n',j}\| + \frac{1}{n'} \\ &< 2\varepsilon + \|\tilde{x}_{n,j} - \tilde{x}_{n',j}\|. \end{aligned}$$

But

$$\lim_{j \rightarrow \infty} \|\tilde{x}_{n,j} - \tilde{x}_{n',j}\| = 0$$

and since j was large enough we can take $\|\tilde{x}_{n,j} - \tilde{x}_{n',j}\| < \varepsilon$. Then

$$\|\tilde{y}_n - \tilde{y}_{n'}\| \lesssim \varepsilon.$$

Therefore $\tilde{y} \in \mathbf{Cau}(V)$, and we can define

$$y = \pi(\tilde{y}).$$

We similarly choose $N > \varepsilon^{-1}$ such that for every $n, n' \geq N$,

$$\|\tilde{y}_n - \tilde{y}_{n'}\| < \varepsilon.$$

Then

$$\|\tilde{x}_{n',n} - \tilde{y}_n\| \leq \|\tilde{x}_{n',n} - \tilde{x}_{n',n'}\| + \|\tilde{y}_{n'} - \tilde{y}_n\| \lesssim \varepsilon.$$

But

$$\lim_{m \rightarrow \infty} \|\tilde{x}_{n,m} - \tilde{y}_m\| = \|x_n - y\|$$

so $\lim_n x_n = y$.

We can define

$$\iota(v) \mapsto \pi(v, v, v, \dots). \quad (\text{A.5})$$

Clearly the limit of the Cauchy sequence on the right-hand side of (A.5) is v , so $\|\iota(v)\| = \|v\|$.

To see that $\iota(V)$ is dense in W , let $x \in W$ and choose $\tilde{x} \in \mathbf{Cau}(V)$ with $\pi(\tilde{x}) = x$. Now \tilde{x} is a Cauchy sequence in V , so let

$$y_n = \iota(\tilde{x}_n);$$

then $y \in \mathbf{Cau}(W)$, and the reader can check that $\lim_n y_n = x$.

To check the universal property, let $T : V \rightarrow X$ be a linear map of V into a Banach space X such that $\|Tv\| = \|v\|$. Thus T is a bounded linear map (c.f. Definition A.3.1). First define $T'(\iota(v)) = v$ for every v , thus T' is defined on the dense subspace V' of W . But then T' is continuous, so it extends uniquely to a linear map by Lemma A.3.3. We leave it to the reader to check $\|T'w\| = \|w\|$. \square

A.3 Linear maps

Though we will use this material before then, the reader need not familiarize themselves with this material until Chapter TODO.

Definition A.3.1. Let $T : V \rightarrow W$ be a linear map between normed spaces. We say that T is a *bounded linear map* if there is a $C > 0$ such that for every $v \in V$,

$$\|Tv\| \leq C\|v\|.$$

The infima of all choices of C is called the *operator norm* of T , denoted $\|T\|$. We denote by $B(V \rightarrow W)$ the space of bounded linear maps.

Lemma A.3.2. Let V, W be normed spaces. Then the operator norm is a norm on $B(V \rightarrow W)$.

We leave the routine proof to the reader.

Lemma A.3.3. Let $T : V \rightarrow W$ be a bounded linear map between normed spaces, and suppose that V is a dense subspace of a normed space X . Then there is a unique extension of T to a bounded linear map $X \rightarrow W$, which has the same operator norm.

Proof. Let $x \in X$, and let $(x_n)_n$ be a sequence in V with $\lim_n x_n = x$. Define $Tx = \lim_n Tx_n$. We let $\|T\|_V$ denote the operator norm of T with domain V and $\|T\|_X$ with domain X .

To see that this is well-defined, suppose that $\lim_n x'_n = x$ as well. Then

$$\lim_{n \rightarrow \infty} \|Tx_n - Tx'_n\| \leq \lim_{n \rightarrow \infty} \|T\|_V \cdot \|x_n - x'_n\| = 0$$

since $x_n \rightarrow x$ and $x'_n \rightarrow x$. Thus Tx does not depend on the choice of Cauchy sequence which approximates x . Similarly,

$$\|Tx\| = \lim_{n \rightarrow \infty} \|Tx_n\| \leq \lim_{n \rightarrow \infty} \|T\|_V \cdot \|x_n\| = \|T\|_V \|x\|.$$

Thus $\|T\|_X \leq \|T\|_V$, but V is a subspace of X so clearly $\|T\|_V \leq \|T\|_X$.

To check that T is linear on X , note that T is continuous and hence can be commuted with limits. This means that

$$T(x + y) = \lim_{n \rightarrow \infty} T(x_n + y_n) = \lim_{n \rightarrow \infty} Tx_n + Ty_n = Tx + Ty$$

where $(x_n)_n$ and $(y_n)_n$ are appropriately chosen Cauchy sequences. The proof that $T(cx) = cTx$ is similar. \square

A.4 Properties of Banach spaces

Definition A.4.1. Let B be a Banach space and let $X \subseteq B$. We say that X is *separable* if there is a countable dense subset of X .

If X is separable and $Y \subseteq X$, then Y is separable; in fact, if C is countable and dense in X , then $C \cap Y$ is countable and dense in Y .

Example A.4.2. \mathbb{C}^n is separable, since $\{(\alpha_1 + i\beta_1, \dots, \alpha_n + i\beta_n) : \alpha_i, \beta_i \in \mathbb{Q}\}$ is countable and dense in \mathbb{C}^n . In particular any finite-dimensional space is separable. Moreover, most spaces that we consider will turn out to be separable.

Example A.4.3. Let X be the space of all bounded functions $\mathbb{C} \rightarrow \mathbb{C}$, where

$$\|f - g\| = \sup_{z \in \mathbb{C}} |f(z) - g(z)|.$$

Then X is not separable. In fact, if $z \in \mathbb{C}$, let $1_z(z) = 1$ and $1_z(w) = 0$ if $w \neq z$. Then $Y = \{1_z : z \in \mathbb{C}\}$ is an uncountable subset of X such that for every $z_1 \neq z_2$, $\|1_{z_1} - 1_{z_2}\| = 1$. So Y is discrete and uncountable, and hence cannot be separable.

A.5 Convexity

Appendix B

Foundations

In this appendix we treat the foundations of math: set theory, category theory, and point-set topology. We omit the proofs, not because they are uninteresting or unworthy of being learned, but because these topics are not analysis; this is an analysis book, and it is reasonable to use material that is not analysis as a black box.

B.1 Axiomatic set theory

We assume familiarity with naive set theory; this will suffice for everything in this text except for a few examples, which require more sophisticated set-theoretic techniques that we record here. We refer the reader to CITE:Kunen for more details.

Let us record the first few axioms of set theory. It will be convenient to assume that *every* mathematical object is a set. This is no loss of generality, because one can define the natural numbers by declaring that $0 = \emptyset$ and for every natural number n , $n = \{0, \dots, n-1\}$. The definition of real numbers in terms of Dedekind cuts, where each real number x is by definition the set $\{y \in \mathbb{Q} : y < x\}$, is similar. More generally, any mathematical object can be encoded as a set in some appropriate way.

The first three axioms are presumably uncontroversial to anyone who has studied naive set theory.

Axiom B.1.1 (extensionality). For all sets x, y, z , if $z \in x$ implies $z \in y$, and $z \in y$ implies $z \in x$, then $x = y$.

Axiom B.1.2 (pairing). For all sets x, y , there exists a set z , usually denoted $\{x, y\}$, such that for every set w , $w \in z$ iff $w = x$ or $w = y$.

Axiom B.1.3 (union). For all sets x, y , there exists a set z , usually denoted $x \cup y$, such that for every set w , $w \in z$ iff $w \in x$ or $w \in y$.

Now to avoid Russell's paradox, and avoid other logic technicalities we want to forbid that a set include itself.

Axiom B.1.4 (foundation). For all sets x such that there is a set $w \in x$, there is a set $y \in x$ such that for all sets $z \in y$, $z \notin x$.

If $x \in x$, then by pairing, $\{x\}$ is a set, and this contradicts foundation.

A *first-order formula* with N free variables in the language of set theory is a string consisting only of the symbols $\exists, \forall, \rightarrow, \neg, \in, (,),$ and variables $x_1, x_2, \dots, y_1, \dots, y_N$ that refer to sets, which is meaningful (so the formula $\forall () \implies ()$ is not a first-order formula), such that the x_i always appear after a quantifier \forall or \exists , and the y_i never do. Plugging in sets for the y_i gives a statement which is either true or false. For example, $\forall x_1((x_1 \in y_1) \rightarrow (x_1 \in y_2))$ is a formula which asserts that $y_1 \subseteq y_2$.

First-order formulae allow us to assert the existence of subsets.

Axiom B.1.5 (restricted comprehension schema). For all sets x , first order-formulae φ with $N + 1$ free variables, and sets w_1, \dots, w_N , there is a set y such that $z \in y$ iff $z \in x$ and $\varphi(z, w_1, \dots, w_N)$ is true.

Henceforth we adopt the usual notation $\subseteq, \subset, \{y \in x : \varphi(y)\}$, etc. We can also now define the empty set \emptyset and various other interesting sets.

Already we have developed enough to study finite sets. For example, we can define intersection \cap and ordered pairs (x, y) in terms of the notions we introduced in the above axioms. We can also define functions; a function $f : X \rightarrow Y$ is just a set f of ordered pairs (x, y) such that for every $x \in X$ there is exactly one $y \in Y$ such that $(x, y) \in f$ (though we usually write $f(x) = y$ to mean $(x, y) \in f$). So we can talk about injections and bijections, and ask if two sets have the same cardinality.

Definition B.1.6. Let x, y be sets. We say that x, y have the same *cardinality*, and write $x \cong y$, if there is a bijection $x \rightarrow y$.

Theorem B.1.7 (Cantor-Bernstein). Let x, y be sets. Then $x \cong y$ iff there are injections $x \rightarrow y$ and $y \rightarrow x$.

In particular, we can ask if a set is finite, countable, etc.

Definition B.1.8. A *finite set* is a set x such that for every $y \subseteq x$, if $x \cong y$, then $x = y$. Otherwise, the set x is an *infinite set*.

However, it is not obvious (and in fact, may even be false) that there is an infinite set at this stage.

Axiom B.1.9 (infinity). There exists a set x , usually denoted \mathbb{N} , such that $\emptyset \in x$ and for all sets $y \in x$, $\{y \cup \{y\}\} \in \mathbb{N}$.

From \mathbb{N} we may construct \mathbb{Z} , \mathbb{Q} , and other familiar countable objects.

Definition B.1.10. A set x is a *countable set* if there is an injection $x \rightarrow \mathbb{N}$. Otherwise, the set x is an *uncountable set*.

Again we get stuck; it is not obvious (and may be false) that there can be uncountable sets.

The axioms that follow are highly dangerous. They assert the existence of deeply infinitary sets, whose elements cannot be easily described. This is the essence of the paradoxical examples that we will use axiomatic set theory to prove the existence of. Famously, Lebesgue rejected the axiom of choice (while implicitly using weak forms of it); modern constructivists, finitists, and intuitionists reject some or all of the following axioms (or even the axiom of infinity).

Axiom B.1.11 (power set). For every set x , there exists a set y , usually denoted 2^x , such that for all sets z , $z \in y$ iff $z \subseteq x$.

The axioms that we have developed up to this point define what is known as *Zermelo set theory*.

There is clearly an injection $x \rightarrow 2^x$, given by $y \mapsto \{y\}$.

Theorem B.1.12 (Cantor's diagonal argument). For every set x , there is no injection $2^x \rightarrow x$.

In particular, $2^{\mathbb{N}}$ is uncountable. We may now prove the existence of \mathbb{R} , $2^{\mathbb{R}}$, the topology of \mathbb{R} , and so on. For example, we have:

Theorem B.1.13 (Dedekind). There exists a set \mathbb{R} , whose elements are nonempty proper subsets x of \mathbb{Q} such that for each $q \in x$, if $r \in \mathbb{Q}$ and $r < q$ then $r \in x$, and there is an $s \in x$ such that $s > x$. There exists a ring structure on \mathbb{R} which turns \mathbb{R} into an ordered field of characteristic 0 such that for every set $X \in 2^{\mathbb{R}}$, $\sup X$ is well-defined. Moreover, \mathbb{R} is unique up to unique isomorphism of ordered fields.

We now pause to introduce the notion of *transfinite induction*, which we will use a few times.

Definition B.1.14. A *transitive set* is a set x such that $x \subset 2^x$.

Thus every element of a transitive set is also a subset of x .

Definition B.1.15. An *ordinal* is a transitive set x such that for every $y \in x$, y is an ordinal.

This definition may seem circular – but it is not. \emptyset is an ordinal, usually denoted 0 when we think of it as an ordinal, and every natural number is an ordinal (why?), but so is \mathbb{N} , which we usually denote ω when we think of it as an ordinal. From there we keep going: given an ordinal α we define its *successor* $\alpha + 1 = \alpha \cup \{\alpha\}$. Not every ordinal is a successor or 0; for example ω was not. Such ordinals are known as *limits*. They are limit points in the order topology.

Theorem B.1.16 (transfinite induction). Fix a limit ordinal κ . Let X be a set of ordinals such that:

1. $0 \in X$.
2. For every ordinal $\alpha \in X$, $\alpha + 1 \in X$.
3. If $\delta < \kappa$ is a limit ordinal and for every $\alpha < \delta$, $\alpha \in X$, then $\delta \in X$.

Then $X = \kappa$.

The trivial example of transfinite induction is when $\kappa = \omega$; then there are no limit ordinals to consider and the theorem collapses down to induction on \mathbb{N} . From this we can introduce *transfinite recursion*. Suppose that κ is a limit ordinal, and we want to define sets x_α for every $\alpha < \kappa$. Then we just have to:

1. Define x_0 .
2. Show that if we can define x_α , then we can define $x_{\alpha+1}$.
3. Show that for every limit ordinal $\delta < \kappa$, if we can for every $\alpha < \delta$ define x_α , then we can define x_δ .

Then the theorem of transfinite induction will imply that, for every $\alpha < \kappa$, x_α is defined.

The set of all countable ordinals is an uncountable ordinal (in fact, the smallest uncountable ordinal), which we denote ω_1 .

If X is a set of ordinals, then X has a least element; this is similar to the fact that every set of natural numbers has a least element.

Definition B.1.17. An ordinal κ is a *cardinal* if κ is the least element of $\{\alpha : \alpha \cong \kappa\}$.

The idea is that cardinals should be canonical representatives of the equivalence class of all sets with a given cardinality.

We let \aleph_0 be the least infinite cardinal (thus $\aleph_0 = \omega$), and for every cardinal \aleph_n , we define \aleph_{n+1} to be the least cardinal greater than \aleph_n . We would like to continue this recursion and let \aleph_ω be the least cardinal greater than \aleph_n for every $n \in \mathbb{N}$. It is not possible to define \aleph_ω yet, but we can do so with the help of a new axiom schema.

Axiom B.1.18 (replacement schema). For every set X , every first-order formula with $N+2$ free variables φ , and all sets w_1, \dots, w_N , if for every $x \in X$ there is exactly one set y such that $\varphi(x, y, w_1, \dots, w_N)$ is true, then there is a set z such that $y \in z$ iff there is an $x \in X$ such that $\varphi(x, y, w_1, \dots, w_N)$ is true.

In other words, if F is a function which can be explicitly defined by a first-order formula in terms of N parameters w_1, \dots, w_N , then the image of F is a set. It follows from the replacement schema that \aleph_α is defined for every ordinal α .

Zermelo set theory along with the replacement schema is known as *Zermelo-Fraenkel set theory*.

Zermelo-Fraenkel set theory cannot prove that every set is in bijection with a cardinal, or that every vector space has a basis. To prove that every set is in bijection with a cardinal, Zermelo introduced the so-called axiom of choice.

Axiom B.1.19 (choice). For every surjection $f : X \rightarrow Y$ there is an injection $g : Y \rightarrow X$ such that for every $y \in Y$, $f(g(y)) = y$.

In other words if $f : X \rightarrow Y$ is a surjection then for every $y \in Y$ we may choose $x \in X$ so that $f(x) = y$. This follows from Zermelo-Fraenkel set theory if we only have to choose finitely many x , but in general, X may be uncountable, so that we may have to make uncountably many choices when we define g .

The above axioms comprise *Zermelo-Fraenkel set theory with choice*. Throughout the text, we assume Zermelo-Fraenkel set theory with choice.

Theorem B.1.20 (Zermelo's well-ordering theorem). For every set x there is a unique cardinal κ such that $x \cong \kappa$.

Definition B.1.21. For every set x , the unique cardinal κ such that $x \cong \kappa$ is called the *cardinality* of x . We denote the cardinality of x by $\text{card } x$.

Proof of Zermelo's well-ordering theorem. Uniqueness is obvious since \cong is an equivalence relation. So we need a cardinal κ and a bijection $f : \kappa \rightarrow x$.

Let λ be a cardinal. We first build a function $f_\lambda : \lambda \rightarrow x$ by transfinite recursion, using the axiom of choice. Choose $y \in x$ and set $f_\lambda(0) = y$. Suppose $\alpha < \lambda$ and we have defined $f_\lambda(\beta)$ for $\beta < \alpha$. Choose $z \in x$ such that for every such β , $f_\lambda(\beta) \neq z$, if such a z exists; otherwise set $f_\lambda(\alpha) = z$. By transfinite recursion, this defines a function f_λ . If f_λ is a bijection we can set $\kappa = \lambda$ and we're done.

Otherwise suppose f_λ is not injective. Then there is an ordinal $\beta < \lambda$ (so in particular, $\text{card } \beta < \lambda$) such that $f_\lambda|_\beta$ is injective and β is maximal possible (so in particular, $f_\lambda|_\beta$ is surjective), by construction of f_λ . Let $\kappa = \text{card } \beta$ and choose $g : \beta \rightarrow \kappa$. Then set $f = (f_\lambda|_\beta) \circ g^{-1}$.

Finally suppose f_λ is not surjective. Let $\delta = \text{card } 2^\lambda$ and consider f_δ instead. If the process of passing from λ to δ does not halt, then there are arbitrarily large cardinals δ which admit injections $\delta \rightarrow x$, so the universe of all sets embeds in x . This breaks Russell's paradox. \square

Note that while we technically don't need Zermelo's well-ordering theorem to prove the Cantor-Bernstein theorem, it is an immediate consequence.

Definition B.1.22. We define $\beth_0 = \aleph_0$, and for every α , $\beth_{\alpha+1}$ to be the cardinality of 2^{\beth_α} . If δ is a limit ordinal we let \beth_δ be the cardinality of $\bigcup_{\alpha < \delta} \beth_\alpha$.

By Cantor's diagonal argument, $\beth_\alpha < \beth_\beta$ whenever $\alpha < \beta$.

Theorem B.1.23. Let $\lambda \leq \kappa$ be infinite cardinals. Suppose that X is a set of cardinality λ , whose elements are sets of cardinality κ . Then $\bigcup_{x \in X} x$ has cardinality κ . Moreover, if $\lambda < \kappa$ and $\kappa = \beth_\alpha$ for some α , then the Cartesian product of λ many copies of κ has cardinality κ .

In particular, if λ is countable or $\lambda = \beth_1$, then a union of λ many sets of cardinality \beth_1 has cardinality \beth_1 .

Theorem B.1.24. Let \mathcal{T} be the set of all open subsets of \mathbb{R} ; then \mathcal{T} has cardinality \beth_1 .

Another consequence of Zermelo's well-ordering theorem is Zorn's lemma, which is frequently useful. The point is to allow us to build up "deeply infinitary" objects (that cannot be completed in ω many steps!) without explicitly using transfinite recursion. Of course the proof will be by transfinite recursion, but mathematicians who are not logicians tend to find the transfinite terrifying, and use Zorn's lemma as something of a black box.

Definition B.1.25. Let \mathbb{P} be a partially ordered set. A *chain* \mathcal{C} in \mathbb{P} is a set such that for every $x, y \in \mathcal{C}$, $x \leq y$ or $y \leq x$.

Theorem B.1.26 (Zorn's lemma). Let \mathbb{P} be a partially ordered set such that every chain in \mathbb{P} has an upper bound. Then \mathbb{P} has a maximal element.

Proof. Let $\kappa = \text{card } \mathbb{P}$ and choose a bijection $f : \kappa \rightarrow \mathbb{P}$. We will define an ordinal δ and an injection $g : \delta + 1 \rightarrow \mathbb{P}$ as follows. Let $g(0) = f(0)$. If $g(\alpha)$ has been defined and is not maximal, there is a $\beta < \kappa$ such that $f(\beta) > g(\alpha)$; choose the least such β , and set $g(\alpha + 1) = f(\beta)$. If $g(\alpha)$ has been defined for all $\alpha < \gamma$, then let $\mathcal{C}_\gamma = \{g(\alpha) : \alpha < \gamma\}$; then \mathcal{C}_γ is a chain, so it has an upper bound, which we define to be $g(\gamma)$.

If the above process stops at some δ , then $g(\delta)$ is maximal in \mathbb{P} . Otherwise, there are arbitrarily large ordinals δ such that there are injections $\delta \rightarrow \mathbb{P}$, hence $\delta \rightarrow \kappa$. In particular we could take $\delta = \text{card } 2^\kappa$ and get an injection $2^\kappa \rightarrow \kappa$, a contradiction. \square

B.2 Universal properties

This section has nothing to do with measure theory. It solely exists to justify certain algebraic handwaves in the appendices. The reader who is interested in category theory should read TODO:Cite Aluffi and Riehl, and most other readers, except those who are completely ignorant of algebra, should ignore this section entirely.

To avoid technicalities we adjoin a new axiom to the Zermelo-Fraenkel set theory with the axiom of choice to obtain a weak form of *Tarski-Grothendieck set theory*.

Definition B.2.1. A transitive set U is a *universe* if $\mathbb{N} \in U$, U is closed under pairing and power set, and for every subset $V \subset U$ of strictly less cardinality than U , $\bigcup_{x \in V} x \in U$.

Axiom B.2.2 (universe). There exists a universe.

Henceforth we fix a universe U . One can show that universes are closed under every relevant operation that mathematicians might care about, so any object that will ever appear in this book is contained in U . For example $\mathbb{R} \in U$, any measurable subset of \mathbb{R} is in U , and any Banach space we ever consider is in U .

Definition B.2.3. A *small set* is an element of U .

So anything that we will ever have reason to care about, except in this section, is small.

Definition B.2.4. A *category* C is a set, whose elements are called *objects*, along with small sets $\text{Hom}(x, y)$ for all objects $x, y \in C$, whose elements are called *morphisms* from x to y , equipped with *composition* operations

$$\begin{aligned} \text{Hom}(x, y) \times \text{Hom}(y, z) &\rightarrow \text{Hom}(x, z) \\ (\varphi, \psi) &\mapsto \psi\varphi \end{aligned}$$

such that:

1. For every object $x \in C$, $\text{Hom}(x, x)$ contains a morphism 1_x which is an *identity* in the sense that for all $\varphi \in \text{Hom}(x, y)$, $\varphi = \varphi 1_x$ and for every $\psi \in \text{Hom}(y, x)$, $\psi = 1_x \psi$.
2. Composition is *associative* in the sense that $\psi(\varphi\rho) = (\psi\varphi)\rho$ whenever $\psi\varphi$ and $\varphi\rho$ are defined.

For example U forms a category, whose objects are small sets, and whose morphisms are defined by letting $\text{Hom}(x, y)$ be the set of all functions $x \rightarrow y$. We dare not define a category whose objects consist of all sets, due to Russell's paradox; but we will abuse terminology and call U the category of sets all the same. We denote it by **Set**. Similarly we define **Grp**, the category of (small) groups where the morphisms are group homomorphisms, and **Vect**(K), the category of (small) vector spaces over a (small) field K where the morphisms are linear maps.

It can be convenient to draw diagrams of morphisms, which are graphs where the nodes are objects, an edge from x to y is a morphism in $\text{Hom}(x, y)$, and the diagram *commutes* if for any two objects x, y in the diagram and any two paths $\varphi_1 \cdots \varphi_n$ and $\psi_1 \cdots \psi_m$ from x, y , $\varphi_1 \cdots \varphi_n = \psi_1 \cdots \psi_m$. For example, the diagram

$$\begin{array}{ccc} x_1 & \xrightarrow{\varphi_1} & x_2 \\ \downarrow \psi_1 & & \downarrow \psi_2 \\ y_1 & \xrightarrow{\varphi_2} & y_2 \end{array}$$

commutes iff $\psi_1 \varphi_2 = \psi_2 \varphi_1$.

Definition B.2.5. Let C be a category, $x, y \in C$ objects, and $\varphi \in \text{Hom}(x, y)$. We say that φ is an *isomorphism*, and x, y are *isomorphic*, if there is a $\psi \in \text{Hom}(y, x)$ which is an *inverse* to φ in the sense that $\varphi\psi = 1_y$ and $\psi\varphi = 1_x$.

Definition B.2.6. Let C be a category and $x \in C$ an object. We say that x is *initial* in C if for every $y \in C$ there is a unique morphism in $\text{Hom}(x, y)$. Similarly we say that x is *final* if for every y there is a unique morphism in $\text{Hom}(y, x)$. Either way, we say that x is *terminal*.

If x and y are terminal, then there is a *unique* isomorphism between x and y . In category theory one cannot distinguish between two objects between which there is a unique isomorphism, so we abuse terminology and say that $x = y$, even if x, y are not equal in the sense of the axiom of extensionality. Thus a terminal object, if it exists, is unique. For example, the trivial vector space is the unique terminal object in **Vect**.

Definition B.2.7. Let P be a property held by objects x in a certain category C . We say that P is a *universal property* if, for every x that holds property P , x is terminal in C .

Therefore an object with a universal property is unique.

We can now make rigorous our handwaving about the universal property of the completion. Let V be a normed space. We define a category $C(V)$, whose objects are norm-preserving linear maps $V \rightarrow X$ into a Banach space X , and whose morphisms φ are commutative diagrams of linear maps

$$\begin{array}{ccc} & & X \\ & \nearrow & \downarrow \varphi \\ V & & \\ & \searrow & \\ & & Y \end{array}$$

The initial object of V is the inclusion map $V \rightarrow W$, where W is the completion of V . Therefore the completion is well-defined if it exists (which it does, by Theorem A.2.5). Thus we can take the universal property as the definition of the completion, and take Theorem A.2.5 as just an indication that this definition makes sense; we can then forget about such hideous objects as $\mathbf{Cau}(\mathbf{Cau}(V))$.

We paraphrase the above universal property by saying that the completion of V is the initial Banach space which contains a copy of V . A similar argument shows that \mathbb{R} is uniquely defined as the initial complete metric space which contains a copy of \mathbb{Q} .

B.3 Point-set topology

We briefly sketch ideas from point-set topology that we will need. We refer the reader to TODO:Cite Munkres or Bradley for reference.

Definition B.3.1. A *topology* \mathcal{T} in a set X is a set of subsets of X such that:

1. $\emptyset, X \in \mathcal{T}$.
2. If $\mathcal{U} \subseteq \mathcal{T}$, then the union $\bigcup_{U \in \mathcal{U}} U$ of all elements of \mathcal{U} is also in \mathcal{T} .
3. If $U_1, \dots, U_n \in \mathcal{T}$ then $U_1 \cap \dots \cap U_n \in \mathcal{T}$.

A pair (X, \mathcal{T}) is called a *topological space* and usually just denoted X . Elements of \mathcal{T} are called *open sets*. The complement of an open set is called a *closed set*. If $x \in X$, a *neighborhood* of x is an open set containing x .

It follows that the arbitrary union and finite intersection of open sets is open, and the arbitrary intersection and finite union of closed sets. The point of a topology is that if K is a closed set, K is closed under taking limits, as we will see.

The obvious examples of topologies on a set X are the discrete topology (wherein every set is open) and the indiscrete topology (wherein the only open sets are \emptyset, X).

Definition B.3.2. A set \mathcal{B} of subsets of a set X is called a *basis* if for all $B_1, \dots, B_n \in \mathcal{B}$, $B_1 \cap \dots \cap B_n \in \mathcal{B}$.

Every basis generates a topology whose elements are arbitrary unions of sets in the basis.

Definition B.3.3. A *semimetric* d on a set X is a function $d : X \times X \rightarrow [0, \infty)$ such that:

1. For all x , $d(x, x) = 0$.
2. For all x, y , $d(x, y) = d(y, x)$.
3. The *triangle inequality*: For all x, y, z , $d(x, y) \leq d(x, z) + d(z, x)$.

If the only pairs (x, y) such that $d(x, y) = 0$ are those with $x = y$, we say that d is a *metric* and call $X = (X, d)$ a *metric space*. We call sets of the form $B(x, \varepsilon) = \{y \in X : d(x, y) < \varepsilon\}$, where $\varepsilon > 0$ and $x \in X$, *open balls*.

The open balls of a semimetric on X form a basis of subsets of X , so every semimetric induces a topology, whose open sets are unions of open balls.

Example B.3.4. Every seminorm induces a semimetric by $d(x, y) = \|x - y\|$. In particular, \mathbb{R}^d has a norm (namely its absolute value), so \mathbb{R}^d is a topological space.

We now describe two different constructions of spaces: product spaces and subspaces.

Definition B.3.5. Let \mathcal{X} be a set of topological spaces. The *product space* $\prod \mathcal{X}$ is the space X which has $\prod \mathcal{X}$ as a set, and has the smallest topology \mathcal{T} such that for every open set $U \subseteq Y$, $Y \in \mathcal{X}$, $\pi : X \rightarrow Y$ the canonical projection, $\pi^{-1}(U) \in \mathcal{T}$.

For example $X \times Y$ is a product space. The open sets are generated by those of the form $U \times V$ where $U \subseteq X$ and $V \subseteq Y$ are open.

Definition B.3.6. Let X be a topological space and $Y \subseteq X$ a subset. We say that $U \subseteq Y$ is *open* in Y (or *relatively open*) if there is an open set $V \subseteq X$ such that $U = V \cap Y$. With respect to that topology, we call Y a *subspace* of X .

So a subspace is a topological space in its own right.

Topological spaces allow us to define continuity in a high level of abstraction. This agrees with the usual definition of a continuous function $[0, 1] \rightarrow \mathbb{R}$.

Definition B.3.7. Let X, Y be topological spaces and $f : X \rightarrow Y$. We say that f is *continuous* if for every $U \subseteq Y$ open, $f^{-1}(U)$ is open.

Lemma B.3.8. Let X, Y be metric spaces and $f : X \rightarrow Y$. Then f is continuous iff for every $\varepsilon > 0$ and $x \in X$, there is a $\delta > 0$ such that for every $x' \in X$, if $d(x, x') < \delta$, then $d(f(x), f(x')) < \varepsilon$.

Lemma B.3.9. Let \mathcal{X} be a set of topological spaces. Then $X = \prod \mathcal{X}$ satisfies the *universal property of products* in the category of topological spaces: for every $Y \in \mathcal{X}$, $\pi_Y : X \rightarrow Y$ the canonical projection, π_Y is continuous, and for every topological space Z , if we are given maps $f_Y : Z \rightarrow Y$ for every Y , there is a unique map $f : Z \rightarrow X$ such that $\pi_Y \circ f = f_Y$.

We now want to talk about limits. Unfortunately at the level of abstraction we are working at, this proves quite tricky.

Definition B.3.10. A *directed set* is a partially ordered set \mathbb{P} such that for all $\alpha_1, \dots, \alpha_k \in \mathbb{P}$, $\sup(\alpha_1, \dots, \alpha_k)$ exists.

For example \mathbb{N} is directed, and in fact so is any ordinal. A topology, equipped with the relation \subseteq , is also directed.

Definition B.3.11. Let \mathbb{P} be a directed set and X a topological space. A *net* in X , indexed by \mathbb{P} , is a function $\mathbb{P} \rightarrow X$. If x is a net, we usually write x_α instead of $x(\alpha)$. A *sequence* is a net indexed by \mathbb{N} .

Definition B.3.12. Let \mathbb{P} be a directed set and X a topological space. Let x be a net in X indexed by \mathbb{P} . Let $x^* \in X$. We say that x^* is the *limit* of x if for every neighborhood U of x , there is an $\alpha \in \mathbb{P}$ such that for every $\beta \geq \alpha$, $x_\beta \in U$. In this case we write $\lim_\gamma x_\gamma = x^*$ or $x_\gamma \rightarrow x^*$ if it is clear from context that γ is a dummy variable. If x has a limit we say that X *converges*.

For example, in a metric space, $x_n \rightarrow x$ iff for every $\varepsilon > 0$ there is a $N \in \mathbb{N}$ such that for all $n > N$, $d(x_n, x) < \varepsilon$.

Lemma B.3.13. Let $f : X \rightarrow Y$. Then f is continuous iff for every net $x_\gamma \rightarrow x^*$ in X , $f(x_\gamma) \rightarrow f(x^*)$ in Y .

Definition B.3.14. We say that a function $h : S \rightarrow \mathbb{P}$ is *cofinal* if for every $\alpha \in \mathbb{P}$ there is a β in the image of h such that $\beta \geq \alpha$. We say that a function $f : \mathbb{P} \rightarrow \mathbb{P}$ between directed sets is *monotone* if for every $\alpha \leq \beta$ in \mathbb{P} , $f(\alpha) \leq f(\beta)$.

Definition B.3.15. Let X be a topological space. Let x be a net indexed by \mathbb{P} and x' a net indexed by \mathbb{P}' . We say that x' is a *subnet* of x if there is a monotone cofinal function $f : \mathbb{P} \rightarrow \mathbb{P}'$ such that $x'_\alpha = x_{f(\alpha)}$. A *subsequence* is a sequence which is a subnet of a subsequence.

Lemma B.3.16. Let X be a topological space. A set $K \subseteq X$ is closed iff for every net x in K which converges in X , x converges in K . Moreover, given a set $A \subseteq X$, the smallest closed set K containing A is the set of all limits of nets in A .

Subsequences are significantly easier to work with than subnets, as one can eliminate the need to worry about cofinality; y is a subsequence of x iff there is an increasing sequence n of natural numbers such that $x_{n_k} = y_k$ for all $k \in \mathbb{N}$.

At this point our definition of limit has three problems:

1. We want to eliminate the need to choose \mathbb{P} whenever possible, and just work with $\mathbb{P} = \mathbb{N}$.
2. We want to be able to guarantee that if a net has a limit x^* , then x^* is unique.
3. We want to be able to guarantee that every net has a subnet which converges.

These are the essences of the definitions of a first-countable, Hausdorff, and compact space respectively.

Definition B.3.17. Let X be a topological space and $x \in X$. A *fundamental system of neighborhoods* at x is a set \mathcal{U} of open sets $U \ni x$ such that for every open set $V \ni x$ there is a $U \in \mathcal{U}$ with $U \subseteq V$. If, for every $x \in X$, x has a countable fundamental system of neighborhoods, then we say that X is a *first-countable space*.

Lemma B.3.18. Let X be a first-countable space. A set $K \subseteq X$ is closed iff for every sequence x in K which converges in X , x converges in K . Moreover, given a set $A \subseteq X$, the smallest closed set K containing A is the set of all limits of sequences in A .

Lemma B.3.19. Every metric space is first-countable, and a product of countably many first-countable spaces is first-countable.

Now we discuss Hausdorffness.

Lemma B.3.20. Let X be a topological space. The following are equivalent:

1. For every two points $x_1, x_2 \in X$, there are open sets $U_1 \ni x_1, U_2 \ni x_2$ such that $U_1 \cap U_2$ is empty.
2. The *diagonal* $\Delta(X) = \{(x, x) : x \in X\}$ is a closed subset of $X \times X$.
3. For every sequence $x \in X$, x has at most one limit.

Here $X \times X$ is viewed as a product space.

This lemma motivates a definition.

Definition B.3.21. A *Hausdorff space* is a topological space such that for every two points $x_1, x_2 \in X$, either $x_1 = x_2$ or there are open sets $U_1 \ni x_1, U_2 \ni x_2$ such that $U_1 \cap U_2$ is empty. If X is a Hausdorff space, we also say that X satisfies *Axiom T^2* .

There are other properties called *Axiom T^s* , where if $s' \leq s$ then T^s implies $T^{s'}$. For example:

Definition B.3.22. Let X be a topological space. We say that X satisfies:

1. *Axiom T^0* , if for every $x_1, x_2 \in X$, either $x_1 = x_2$ or there is an open set U such that $x_1 \in U$ and $x_2 \notin U$.
2. *Axiom T^1* , if for every $x \in X$, $\{x\}$ is a closed set.
3. *Axiom T^4* , if for every two closed sets $K_1, K_2 \subseteq X$, either $K_1 = K_2$ or there are open sets U_1, U_2 such that $K_1 \subseteq U_1, K_2 \subseteq U_2$, and $U_1 \cap U_2$ is empty.

Axioms of the form *Axiom T^s* are called *separation axioms*.

Lemma B.3.23. One has a string of implications: *Axiom T^4* implies *Axiom T^2* , which implies *Axiom T^1* , which implies *Axiom T^0* .

Lemma B.3.24. Let X be a semimetric space. The following are equivalent:

1. X is a metric space.
2. X satisfies *Axiom T^4* .
3. X satisfies *Axiom T^0* .

Moreover, if R is the equivalence relation defined by $x_1 R x_2$ iff $d(x_1, x_2) = 0$, then X/R can be given the structure of a metric space.

In particular, every metric space is Hausdorff and every semimetric space can be turned into a metric space by identifying points that are indistinguishable by its topology.

Finally let us discuss compactness.

Definition B.3.25. An *open cover* \mathcal{U} of a topological space X is a set of open sets such that $\bigcup \mathcal{U} = X$. A *subcover* of \mathcal{U} is an open cover $\subseteq \mathcal{U}$.

Definition B.3.26. A *compact space* is a topological space such that every open cover has a finite subcover. A *locally compact space* is a topological space X such that for every $x \in X$ there is an open set $U \ni x$ such that the smallest closed set containing U is compact.

Lemma B.3.27. Let X be a topological space. Then X is compact iff for every net x in X , there is a subnet x' such that x' is convergent. Moreover, if X is first-countable, then we can replace the word “net” by “sequence” in the previous sentence.

Lemma B.3.28. Let X be a topological space. Then:

1. If X is Hausdorff and $K \subseteq X$ is compact, then K is closed.
2. If $f : X \rightarrow Y$ is continuous and X is compact, then $f(X)$ is compact.

In particular, if X is compact, Y is Hausdorff, and $f : X \rightarrow Y$ is continuous, then f sends closed sets to closed sets.

Theorem B.3.29 (Heine-Borel). \mathbb{R}^d is locally compact, and $K \subseteq \mathbb{R}^d$ is compact iff K is closed and bounded.

Lemma B.3.30. If K is a compact space, then K is Hausdorff iff K satisfies Axiom T^4 .

We now arrive at the two nontrivial theorems of point-set topology which will be used in the main text as black boxes. Since they are so important we sketch their proofs, but since these results rely on the above lemmata which we have not proved, these arguments are incomplete; see TODO:Cite for a complete exposition.

Theorem B.3.31 (Urysohn lemma and Tietze extension). Let X be a topological space, $A, K \subseteq X$ are disjoint, and A is closed. Suppose that one of the following is true.

1. X satisfies Axiom T^4 and K is closed.
2. X is a locally compact Hausdorff space and A is compact.

Then:

1. There is a continuous function $f : X \rightarrow [0, 1]$ such that $f|_A = 0$ and $f|_K = 1$.
2. For every interval $L \subseteq \mathbb{R}$ and every continuous function $g : K \rightarrow L$, there is a continuous function $G : X \rightarrow L$ such that $G|_K = g$.

Proof. We prove this assuming Axiom T^4 . The proof assuming that X is a locally compact Hausdorff space follows from the case of Axiom T^4 and the previous lemma after a little work, which we omit.

We now prove the existence of f . Let \bar{U} denote the smallest closed set containing an open set U . Set $U(1/2)$ to be an open set containing K which is disjoint from a neighborhood of A , which exists by Axiom T^4 . Then $U(1/2)^c$ is closed and so by Axiom T^4 we can find $U(1/4)$ open such that $\overline{U(1/4)} \subseteq U(1/2)$. Similarly we can find $U(3/4)$ open such that $\overline{U(3/4)} \subseteq U(1/2)$. Repeating this process we define $U(p/2^n)$ for every $0 < p < 2^n$, $n \in \mathbb{N}$, such that $\overline{U(s)} \subseteq U(r)$. Now set

$$f(x) = \inf_{x \in U(r)} r$$

if such an r exists, otherwise $f(x) = 1$. Since $\{p/2^n : 0 < p < 2^n, n \in \mathbb{N}\}$ is dense in $[0, 1]$, one can check that f is continuous, so f is as desired.

We may assume that $L = [0, 1]$; we omit the details. We may then set $G|A = 0$, and use the function f to interpolate between A, K . We omit the details. \square

Note carefully that the above result does not imply that every locally compact Hausdorff space satisfies Axiom T^4 – in fact this is false.

Theorem B.3.32 (Tychonoff). Let \mathcal{X} be a nonempty set of nonempty compact spaces. Let $X = \prod \mathcal{X}$. Then X is compact and nonempty.

Proof. By Zermelo's well-ordering theorem we may write $\mathcal{X} = (X_\alpha)_{\alpha < \kappa}$ where κ is a cardinal and the α are ordinals. We proceed by transfinite induction, since κ can be viewed as an ordinal. The base case is $\kappa = 1$, in which case $X = X_0$ is obviously compact.

The inductive case is that $\kappa = \beta + 1$, where $Y = \prod_{\alpha < \beta} X_\alpha$ is compact and X_β is compact. So we must show that $X = Y \times X_\beta$ is compact. Let $x = (y, z)$ be a net in X , where y is a net in Y and z is a net in X_β . By compactness there is a subnet $x' = (y', z')$ such that y' is a convergent net in Y . Then by compactness again there is a subnet $x'' = (y'', z'')$ of x' such that z'' is a convergent net in X_β . But y' is convergent in Y and y'' is a subnet of y' so y'' is convergent in Y . Therefore x'' is convergent in X .

The interesting case is the limit case, thus if $\delta < \kappa$ then $Y_\delta = \prod_{\alpha < \delta} X_\alpha$ is compact, and κ is not the successor of any ordinal. In that case $X = \prod_{\alpha < \kappa} X_\alpha$, and if x is a net in X then we can let y^δ be the projection of x onto Y_δ . So we can find subnets z^δ of y^δ which converge.

Let \mathbb{P} be the domain of x . We may assume that $\mathbb{P} = \gamma$ for some cardinal γ , by the lemma:

Lemma B.3.33. Let Q be a partially ordered set. Then there is a cardinal γ and an injection $f : Q \rightarrow \gamma$ such that if $x \leq y$ then $f(x) \leq f(y)$.

This is a straightforward consequence of Zorn's lemma, once we use Zermelo's well-ordering theorem to find $\gamma > \text{card } Q$ large enough. We omit the details.

By replacing κ with a larger cardinal λ , and setting X_β to be a point for all $\kappa \leq \beta < \lambda$, we may assume that $\gamma \leq \kappa$. For every α let $x'_\alpha = z^\alpha$. Then x' converges since z^α does for every α . So x' has a limit. Therefore X is compact. \square

Note that the above use of the axiom of choice (as Zermelo's well-ordering theorem and Zorn's lemma) cannot be avoided. There is a theorem of Kelley which says that if we are in Zermelo-Fraenkel set theory possibly without the axiom of choice, but Tychonoff's theorem holds, then so does the axiom of choice.