measure

Aidan Backus

December 2019

Contents

Chapter 1

Measures

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 1.1. 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* 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 $\mu(X) = 1$.

In all of these example, 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

Definition 1.2. 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.

One easily checks that if $X_1, \dots, X_n \in R$ then $X_1 \cup \cdots \cup X_n \in R$ and $X_1 \cap \cdots \cap X_n \in R$ (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 1.3. 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 1.4. By a *σ-ring* 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.

As before, σ-rings are closed under countable unions. Here σ should be thought of as meaning "countable".

Lemma 1.5. 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

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

□

Lemma 1.6. 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 $X \cup Y, X \setminus Y \in R$. □

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 1.7. 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 1.6, 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 $\mathcal{C}$. □

Definition 1.8. 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 $\sigma(\mathcal{C})$.

We now come across the most important example of a σ-algebra.

Example 1.9. Let X be a topological space (perhaps a metric space, or even $X = \mathbb{R}^d$ at first). Let $\mathcal{T}$ be the topology of X, the set of all open subsets of X (thus if $X = \mathbb{R}^d$, elements of $\mathcal{T}$ are unions of balls $B(x,r) = \{y \in \mathbb{R}^d : |x-y| < r\}$.) We define the *Borel σ-algebra* $\mathcal{B}$ of X by $\mathcal{B} = \sigma(\mathcal{T})$. A *Borel set* in X is an element of $\mathcal{B}$. Thus every open or closed subset of X is Borel.

This definition is deceptively simple. 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. To see this, we will need some powerful set-theoretic machinery; so the reader should skip this example on first reading.

Let $\Sigma_1 = \mathcal{T}$ be the topology of $\mathbb{R}^d$. Given Σ_α, α a countable ordinal (see Definition C.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 C.16). We note that Σ_0 has cardinality $\beth_1$ by Theorem C.24.

Now the mapping $A \mapsto A^c$ is a bijection $\Sigma_\alpha \to \Pi_\alpha$. Moreover if Π_α has cardinality $\beth_1$ (see Definition C.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 C.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 C.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 C.23, since $\aleph_1 \leq \beth_1$ (which follows by Zermelo's well-ordering theorem C.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.

Example 1.10. A σ-algebra may be generated in many different ways. Consider the collection $\mathcal{C}$ of half-open intervals $[a,b)$ in $\mathbb{R}$. We claim that $\sigma(\mathcal{C})$ is the Borel σ-algebra $\mathcal{B}$. To see this, note that any open set can be written as a countable union of open intervals, and any open interval (a,b) can be written as

$$(a,b) = \bigcup_n [a + 1/n, b).$$

Therefore every Borel set can be obtained by applying the operations of countable set and complement to $\mathcal{C}$, so $\mathcal{B} \subseteq \sigma(\mathcal{C})$. Conversely, every half-open interval can be obtained as a countable intersection of open intervals, so $\sigma(\mathcal{C}) \subseteq \mathcal{B}$.

1.2 σ-additive functions

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 $[-\infty, \infty]$, which would include $-\infty$, because the expression $\infty - \infty$ makes no sense.

Definition 1.11. Let B be a Banach space or $(-\infty, \infty]$, and let $\mathcal{C}$ be a collection of sets We say that a function $\mu : \mathcal{C} \to B$ is *σ-additive* if for every disjoint sequence of sets $(X_n)_n$ in Σ such that $\bigcup_n X_n \in \mathcal{C}$,

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

Here the infinite sum is meant in the sense of (A.3) if B is a Banach space.

Definition 1.12. A *measure* is a σ-additive function defined on a σ-algebra Σ on a set X which is not identically ∞.

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*.

Definition 1.13. If the image of a measure μ is contained in $[0, \infty]$, we say that μ is a *positive measure*. If the image of μ is $[0, 1]$, we say that μ is a *probability measure*.

Lemma 1.14. 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 1.15. The examples in Example 1.1 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.

We now introduce an important class of σ-additive functions defined for certain subsets of $\mathbb{R}$, known as Stieltjes premeasures. 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.

Definition 1.16. Let $X \subseteq \mathbb{R}$ be an interval. A function $f : X \to B$ is a *left-continuous function* if for every $x \in X$,

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

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, but we will not need them.

Example 1.17. The *Heaviside step function*

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

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 1.18. Let $f : \mathbb{R} \to \mathbb{R}$ be a nondecreasing, left-continuous function. We introduce the *Stieltjes premeasure* μ_f by declaring that

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

A nondecreasing function can only be discontinuous on a countable set. 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

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

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 1.19. If $f(x) = x$, then the Stieltjes premeasure of an interval $[a,b)$ is just its length $b - a$. More generally, 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, $f' \ge 0$. This is our first clue that there is some connection between integration and measure theory.

Example 1.20. 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, $a \le 0 < b$ 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 1.21. A Stieltjes premeasure μ_f is σ-additive.

Proof. 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

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

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

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

Now $b \geq b_n$ and $a \leq a_n$, 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 $\varepsilon > 0$,

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

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

$$\sum_{n=1}^{\infty} \frac{1}{2^n} = 1 \tag{1.4}$$

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

1.3 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 1.22. 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 $G_1, \ldots, G_m$ disjoint such that
$$E \setminus F = \bigcup_{n=1}^{m} G_n.$$

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

Example 1.23. The set of all half-open intervals $[a, b)$ is a semiring, so a Stieltjes premeasure is a premeasure.

Let us record two properties of semirings and premeasures. Both are straightforward to prove (one by induction, the other by modifying the proof of Lemma 1.14), so we omit the proofs.

Lemma 1.24. Let P be a semiring and μ a premeasure on P. Then:

1. If $E_1, \ldots, E_m \in P$ then there are disjoint $F_1, \ldots, F_n \in P$ such that
$$((((E_1 \setminus E_2) \setminus E_3) \setminus \cdots) \setminus E_m) = \bigcup_{i=1}^{n} F_i.$$

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

Lemma 1.25. 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 $\mu(E) \leq \mu(F)$.

This lemma does not presuppose that $\bigcup E_n \in P$.

Proof. We first prove the first claim. By Lemma 1.24, 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 $n = 1$, $E_1 = F$. $\square$

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 1.26. Let P be a set of sets. A function $\mu : P \to (-\infty, \infty]$ is *σ-subadditive* if for every sequence of $E_i \in P$, if $E \in P$ and $E \subseteq \bigcup_i E_i$, then

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

Lemma 1.27. Every premeasure is σ-subadditive.

Proof. Suppose that $E = \bigcup_i E_i$. By Lemma 1.24, 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 1.25,

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

Thus

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

which was to be shown. $\square$

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

In particular, an outer measure is monotone in the sense that if $E \subseteq F$, $E, F \in P$, then $\mu(E) \leq \mu(F)$.

We now extend any premeasure to an outer measure, but first we should select a domain for our extension.

Definition 1.29. 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 1.30. 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 1.31. 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 sequenes 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 1.32. 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 1.33. 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. $\square$

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 $\mu(A) \leq \mu^*(A)$. $\square$

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 $\bigcup_n E_n \subseteq P$ 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 1.34. 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-subadditity,

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

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.

1.4 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.

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 1.35. Let (X, Σ, μ) be a measured space. We say that μ is a *σ-finite measure* if there are $E_n \in \Sigma$ 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 1.36 (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.

We now begin the proof of the Carathéodory construction theorem. To do this, we need the notion of a clean division.

Definition 1.37. Let μ^* be an outer measure on a hereditary σ-ring $\mathcal{H}$. A *μ^*-measurable set* is a set $A \in \mathcal{H}$ such that for every $E \in \mathcal{H}$,

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \setminus A). \tag{1.5}$$

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.

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

Definition 1.38. Let μ^* be an outer measure on a hereditary σ-ring $\mathcal{H}$. 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.

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.

Theorem 1.39. For any outer measure μ^*, $\mathcal{N}(\mu^*)$ and $\mathcal{M}(\mu^*)$ are σ-rings 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. Finally, μ^* restricts to a measure on $\mathcal{M}(\mu^*)$.

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.

We now show that $\mathcal{M}(\mu^*)$ is a σ-ring. First, if A and B divide E cleanly, then TODO

Now follow Pugh's book to show that $\mathcal{M}(\mu^*)$ is a σ-ring and μ^* is a measure. □

TODO: Show that P has measurable sets

TODO: Completions of measures

TODO: Uniqueness of extensions

1.5 Vector-valued measures

Total variation and stuff

We now prove a version of the triangle and reverse triangle inequalities for total variation.

Lemma 1.40. Let μ, ν be B-valued measures. Then $|\mu+\nu| \leq |\mu|+|\nu|$ and $||\mu|-|\nu|| \leq |\mu-\nu|$.

Proof. Let E be a measurable set. For every disjoint sequence of E_i with $E = \bigcup_i E_i$,

$$\sum_i ||\mu(E_i) + \nu(E_i)|| \leq \sum_i ||\mu(E_i)|| + ||\nu(E_i)|| \leq |\mu|(E_i) + |\nu|(E_i).$$

The reverse triangle inequality now follows from the triangle inequality in the usual way. □

TODO

1.6 Measurable maps

TODO

Chapter 2

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 $f : X \to B$ 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.

2.1 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 2.1. A *simple function* is a function $f : X \to B$ 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 $\mathbf{Simp}(X \to B)$.

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

Definition 2.2. Let $Y \subseteq X$ be a measurable set. The *indicator function* of Y, denoted 1_Y, is the function $X \to \{0, 1\}$, defined by $1_Y(y) = 1$ if $y \in Y$ and $1_Y(y) = 0$ if $y \notin Y$.

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 $f \in \mathbf{Simp}(X \to \mathbb{C})$,

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

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

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

Definition 2.3. A vector space A over $\mathbb{C}$ is called an *algebra* if it is equipped with a multiplication $A \times A \to A$ which is associative, has a unit element, distributes over addition, and satisfies, for every $c, d \in \mathbb{C}$ and $x, y \in A$,

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

In particular, a collection of functions is an algebra iff it is closed under multiplication. The algebraically minded reader will check that given a ring[1] A, we can define an algebra structure on A by choosing a nonzero morphism of rings $\varphi : \mathbb{C} \to A$ (why is $\ker \varphi$ trivial?), where $\varphi(c)a$ defines the action of c as a scalar on a, and conversely any algebra can be expressed in this way.

Lemma 2.4. $\mathbf{Simp}(X \to B)$ is a vector space. In particular, $\mathbf{Simp}(X \to \mathbb{C})$ is a algebra.

Proof. Let $f, g \in \mathbf{Simp}(X \to B)$, 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 $\mathbf{Simp}(X \to B)$ is closed under scaling is similar. Clearly $\mathbf{Simp}(X \to B)$ is nonempty, so this implies that $\mathbf{Simp}(X \to B)$ 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. $\square$

[1] in the sense of abstract algebra, not in the set-theoretic sense that we are using!

Lemma 2.5. Let $f, g \in \mathbf{Simp}(X \to \mathbb{C})$. Then $\max(f, g) \in \mathbf{Simp}(X \to \mathbb{C})$, $\min(f, g) \in \mathbf{Simp}(X \to \mathbb{C})$, and $|f| \in \mathbf{Simp}(X \to \mathbb{C})$.

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

We leave the proof of Lemma 2.5, which similar to the proof of Lemma 2.4, to the reader.

2.2 Measurable functions

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

Definition 2.6. A sequence of functions $f_n : X \to B$ is said to *converge pointwise* to a function f if for every $x \in X$,

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

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.

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 2.7. Let μ be Lebsegue 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 $\mathbb{R} \to \mathbb{C}$, even though it is only defined almost everywhere.

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.

To sate our intuition that two functions that are equal almost everywhere are actually equal, we can (and should, and must!) make a definition.

Definition 2.8. Let μ be a measure on Σ. A sequence of functions $f_n : X \to B$ is said to *converge pointwise almost everywhere* with respect to μ, or simply *almost converge*, to a function $f : X \to B$ if there is a null set Z such that on $X \setminus Z$, $f_n \to f$ pointwise. In this case, we write

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

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

Note that in Definition 2.8, 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 2.9. 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 $f_n \to 0$ everywhere except 1, so $f_n \to 0$ almost everywhere with respect to Lebesgue measure.

Definition 2.10. A *measurable function* $f : X \to B$ is a function such that there is a sequence $f_n \in \mathbf{Simp}(X \to B)$ such that $f_n \to f$ almost everywhere. Let $\mathcal{M}(X \to B)$ denote the set of measurable functions.

The notation $\mathcal{M}(X \to B)$ 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 $\mathcal{M}(X \to B, \mu)$ if μ is not clear from context.

Lemma 2.11. $\mathcal{M}(X \to B)$ is a vector space. In particular, $\mathcal{M}(X \to \mathbb{C})$ is an algebra such that if $f, g \in \mathcal{M}(X \to \mathbb{C})$, then so are $\max(f, g)$, $\min(f, g)$, and $|f|$. Moreover, if $f \in \mathcal{M}(X \to B)$, then $x \mapsto ||f(x)||$ is in $\mathcal{M}(X \to \mathbb{C})$.

Proof. Let $f, g \in \mathcal{M}(X \to B)$, and suppose that $f_n \in \mathbf{Simp}(X \to B)$, $f_n \to f$. Similarly let $g_n \to g$, $g_n \in \mathbf{Simp}(X \to B)$. Then $f_n + g_n \in \mathbf{Simp}(X \to B)$ by Lemma 2.4 and $f_n + g_n \to f + g$.

We leave the other claims as an exercise for the reader. □

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

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. □

Now 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 $\mathcal{M}(X \to B)/N$, two functions f, g are equivalent iff $f - g$ is zero almost everywhere.

Definition 2.13. Let N be the space of all functions that are zero almost everywhere is a vector subspace. We denote its quotient space

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

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

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

2.3 Characterizing measurable functions

The current definition of M is unwieldly. 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.

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

Definition 2.14. A function $f : X \to B$ is *almost separably valued* if there is a null set Z such that $f(X \setminus Z)$ is separable in the topology of B.

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!

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 2.14 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 2.15. The *carrier*[2] of a function $f : X \to B$ is the set $\{x \in X : f(x) \neq 0\}$.

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.

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

Theorem 2.16 (Newberger)**.** Let $f : X \to B$ 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.

We summarize Newberger's theorem in the following corollary:

Corollary 2.17. 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 $f : X \to B$ is measurable iff f is almost separably valued and for every measurable $Y \subseteq B$, $f^{-1}(Y)$ is measurable.

[2]Some books prefer the term "support", but we use "support" to mean the closure of the carrier, whenever (X, μ) has a topology.

The reader should compare this to the result which says that a function f is continuous iff the preimage of an open set is open.

Before we prove Newberger's theorem, we state several lemmata which are useful in their own right.

Lemma 2.18. Let $f_n : X \to B$ 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 ith entry in the above list is clearly equivalent to the $i + 1$th entry. □

Definition 2.19. A function $f : X \to B$ is *separably valued* if the image of f is separable in B.

Lemma 2.20. Let $f : X \to B$ 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 $f_n \to f$ pointwise, $f_n \in \mathbf{Simp}(X \to B)$ 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 2.18 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 2.21. $f_n \to f$ pointwise.

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

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 $||f_n(x) - f(x)|| < \varepsilon$. $\square$

This implies 1. $\square$

Lemma 2.22. Let $f : X \to B$ 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. $\square$

Proof of Newberger's theorem. 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 $f_n \in \mathbf{Simp}(X \to B)$ and a null set Z such that $f_n \to f$ pointwise on $X \setminus Z$ and f is defined on $X \setminus Z$. So $f_n 1_{X \setminus Z} \to f 1_{X \setminus Z}$ (where $f 1_{X \setminus Z} = 0$ if f is undefined), so $f 1_{X \setminus Z}$ meets the criteria of Lemma 2.20; 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 2.20 implies the claim. $\square$

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 2.23. 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 2.18 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. $\square$

By Theorem 2.23, we can lay waste to the possibility that you will *ever* encounter a nonmeasurable function.

Example 2.24. 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 $\mathbb{R} \to B$ 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 Newberger's theorem. 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.

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

To be fair, we should give some examples of nonmeasurable functions.

Example 2.25. An example of a function which is not almost separably valued is given by Example A.16. 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 2.26 (Vitali's set)**.** The circle $\mathbb{T}$ can be identified with $[0, 1)$, by sending $\alpha \in [0, 1)$ to $e^{2\pi i\alpha}$. So it is meaningful to talk about Lebesgue measure μ on $\mathbb{T}$, and 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 C.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.

But should you, dear reader, worry about nonmeasurable functions? Example 2.24 shows that most functions that appear in mainstream analysis, algebra, or number theory are measurable, and I would conjecture that *nonmeasurable functions will never appear in any applications of mathematics to the sciences, nor in any mathematics which can be interpreted as "physically meaningful"*. Example 2.25 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). To paraphrase the analyst Strichartz (TODO cite me), any analyst who concerns themself with the axiom of choice must also worry about monsters under the bed...

2.4 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.

Throughout, we fix a complete measured space (X, Σ, μ) with total variation $|\mu|$ and a Banach space B.

Recall the notion of uniform convergence.

Definition 2.27. 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 $n > N$, $\sup ||f_n - f|| < \varepsilon$.

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 δs, Ns, and so on) may become arbitrarily "bad" as $\varepsilon \to 0$.

Definition 2.28. A sequence of functions $f_n : X \to B$ *converges nearly uniformly* to a function $f : X \to B$ if for every $\varepsilon > 0$ there is a set E_ε such that $|\mu|(X \setminus E_\varepsilon) < \varepsilon$ and $f_n \to f$ uniformly on E_ε.

Thus $f_n \to f$ nearly uniformly iff for every $\varepsilon > 0$ there is a E such that for every $\delta > 0$ there is an a N such that $\mu(E) < \varepsilon$ and for every $n > N$, $\sup_E ||f_n - f|| < \delta$. 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.

The proof of the following lemma is routine.

Lemma 2.29. Suppose that $f_n : X \to B$ and $g_n : X \to B$, $f_n \to f$ and $g_n \to g$ nearly uniformly. Then the following limits also hold nearly uniformly:

1. $f_n + g_n \to f + g$.
2. For every $c \in \mathbb{C}$, $cf_n \to cf$.
3. If $B = \mathbb{C}$, $f_n g_n \to fg$.

4. $(x \mapsto ||f_n(x)||) \to (x \mapsto ||f(x)||)$.

5. $\max(f_n, g_n) \to \max(f, g)$.

6. $\min(f_n, g_n) \to \min(f_n, g_n)$.

Theorem 2.30 (Egorov)**.** Let X be a finite measure space (thus $|\mu|(X) < \infty$) and suppose that $f_n \in M(X \to B)$ almost converge to f, and the f_n are measurable. Then $f_n \to f$ nearly uniformly.

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

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

Since $||f(x) - f_k(x)||$ is measurable, by Newberger's theorem, E_{mn} is measurable. Moreover, if m is fixed, the E_m^n shrink as n increases and $\bigcap_n E_{mn} = \emptyset$, since $f_n \to f$. 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 $n > N$, $\sup_F ||f_n - f|| < \delta$. Indeed, if $1/m < \delta$ and $N = n(m)$, then for every $x \in F$, $x \notin E_N^m$, so if $n > N$ then $||f(x) - f_n(x)|| < \delta$. $\square$

The reader should find an example (say, $\mathbb{R} \to \mathbb{C}$) which shows that the hypothesis that X is a finite measure space in Egorov's theorem not be neglected.

We now define an analogue of Cauchy sequences for nearly uniform convergence.

Definition 2.31. A sequence of functions $f_n : X \to B$ 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 $n_1, n_2 > N$, $\sup ||f_{n_1} - f_{n_2}|| < \delta$.

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 2.32. Let $f_n \in M(X \to B)$ be a nearly uniform Cauchy sequence. Then there is a unique $f \in M(X \to B)$ such that $f_n \to f$ 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 $f_n \to f$ nearly uniformly. In fact, for every $\varepsilon > 0$ we may take $m > 1/\varepsilon$ and let $E_\varepsilon = E_m$. Then $f_n \to f$ uniformly on E_ε.

As for uniqueness, the $f_n \to f$ 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. □

We now introduce another notion of convention.

Definition 2.33. Let $f_n : X \to B$ be a sequence of measurable functions, $f : X \to B$. We say that f_n *converges in measure* to f if for every $\varepsilon > 0$ the set

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

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

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

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

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

That's a pretty hefty definition, but let's consider the intuition for it. 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 $\mu(E)$ the probability that one of the outcomes in E occurs, thus $\mu(X) = 1$. Then

$$\mu(\{x \in X : ||f_n(x) - f(x)|| > \varepsilon\})$$

is the probability that we got an experimental error of size at least ε; as $n \to \infty$, 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 $f_n \to f$ in measure and a priori we can prove no stronger.

We should first check that convergence in measure is well-defined in $M(X \to B)$.

Lemma 2.34. If $f_n \to f$ in measure, then f is unique almost everywhere, hence as an element of $M(X \to B)$.

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

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

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

Again we give a Cauchy criterion.

Definition 2.35. Let $f_n : X \to B$ 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 $n_1, n_2 > N$,

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

We again omit the proof of the usual limit laws.

Lemma 2.36. Suppose that $f_n : X \to B$ and $g_n : X \to B$, $f_n \to f$ and $g_n \to g$ in measure. Then the following limits also hold in measure:

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

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 2.37. Suppose that $f_n \to f$ nearly uniformly; then $f_n \to f$ in measure.

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

Corollary 2.38. If $\mu(X) < \infty$ and $f_n \to f$ almost pointwise, then $f_n \to f$ in measure.

Proof. By Egorov's theorem and Lemma 2.37. □

We now come to a critical result which implies that Cauchyness 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 2.39 (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 $f \in M(X \to B)$ such that:

1. The f_{n_k} are a nearly uniform Cauchy sequence.

2. $f_{n_k} \to f$ nearly uniformly.

3. $f_n \to f$ 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

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

Let $g_k = f_{n_k}$.

Sublemma 2.40. g_k is nearly uniformly Cauchy.

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

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

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 $j > \ell > N$, $x \in F$,

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

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

So by Lemma 2.32, there is an f such that $\lim_k g_k = f$ nearly uniformly. But then Lemma 2.37 implies that $\lim_k g_k = f$ in measure. But

$$\{||f - f_n(x)|| > \varepsilon\} \subseteq \{||f - g_k|| > \varepsilon/2\} \cup \{||f_n - g_k|| > \varepsilon/2\}, \tag{2.1}$$

and the g_k are a subsequence of the Cauchy-in-measure sequence f_n, hence the right hand side of (2.1) is $< \varepsilon$ if n, k are large enough. So $f_n \to f$ in measure. Uniqueness follows by Lemma 2.34. □

Corollary 2.41. If $f_n \to f$ in measure, then there is a subsequence of f_{n_k} such that $f_{n_k} \to f$ 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. □

2.5 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. The reader who is not interested in regularity results and just wants to get to integration can skip this section entirely.

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

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

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

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

We will need Urysohn's lemma:

Lemma 2.42 (Urysohn)**.** Let $K_0, K_1 \subseteq X$ be disjoint closed sets. Then there is a continuous function $f : X \to [0, 1]$ 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.

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 $X = \mathbb{R}^n$, this follows from the Heine-Borel theorem.

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

Proof. Let $f : X \to B$ be continuous; then the f-preimage of an open set is open, hence Borel, hence measurable. We now appeal to Newberger's theorem. □

Definition 2.44. 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 $f : X \to B$ 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.

You should check that your favorite discontinuous function (that isn't the indicator function of the Vitali set) is nearly continuous, and indeed nearly smooth if $X = \mathbb{R}^n$. For example the function $x \mapsto 1/x$ is nearly smooth because if we discard a small neighborhood of 0, then it is smooth.

Definition 2.45. A *Radon measure* on X is a Borel measure μ 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, $\mu(K) < \infty$.

Lemma 2.46. Lebesgue measure is a Radon measure on $\mathbb{R}$.

Proof. Every Stieltjes measure is nonnegative Borel. Local finiteness follows from the Heine-Borel theorem: if K is compact then K is contained in an interval $[\alpha, \beta]$ so $\mu(K) < \beta - \alpha$. 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 $K \subseteq [\alpha + 1/n, \beta - 1/n]$, so $\mu(K) \leq [\alpha + 1/n, \beta - 1/n]$. On the other hand, $\lim_n \mu([\alpha + 1/n, \beta - 1/n]) = \beta - \alpha = \mu((\alpha, \beta))$.

For outer regularity, let W be a Borel set (actually, any Lebesgue measurable set); then
$$\mu(W) = \inf_{(\alpha_n), (\beta_n)} \sum_n \beta_n - \alpha_n$$
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 $E^n_m = (\alpha_n - m^{-1}2^{-n}, \beta_n)$ is an open cover of W for any m, so
$$\mu(W) \leq \sum_n \beta_n - \alpha_n < \frac{1}{m} + \sum_n \beta_n - \alpha_n = \sum_n \beta_n - \alpha_n + m^{-1}2^{-n}.$$
Taking $m \to \infty$ and minimizing $\sum_n \beta_n - \alpha_n$ by varying the α_n and β_n we collapse the above inequalities into infs. $\square$

Throughout the rest of the section, we fix a 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 2.47. If $f : X \to \mathbb{C}^d$ is a measurable function, then there is a sequence of continuous functions $f_n : X \to \mathbb{C}^d$ such that $f_n \to f$ 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 $f_n : X \to [0,1]$, smooth if $X = \mathbb{R}^c$, such that $f_n|K = 1$ and $f_n|(X \setminus U_n) = 0$. Now $f_n \to f$ 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 $X = \mathbb{R}^c$) $f_n^m : X \to [0,1]$ 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 $n \to \infty$ so $f_n^n \to 1_U$ 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 $U_m \supseteq K_{m+1}$, $\mu(U_m \setminus W) < 1/m$. We can find a sequence of continuous (smooth?) functions $f_n^m : X \to [0,1]$ such that $\lim_n f_n^m = 1_{U_m}$ almost pointwise, and $\lim_m 1_{U_m} = 1_W$ almost pointwise, so $f_n^n \to 1_W$ 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 $f \in \mathbf{Simp}(X \to \mathbb{C})$, 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 $X \to \mathbb{C}$, we can approximate f by simple functions.

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

The proof of the above lemma is quite typical. We first check for the simplest case possible, where the claim is quite obvious, and then proceed in stages, each stage allowing the function f to have higher complexity – but only slightly higher, so that no one stage is particularly difficult.

Theorem 2.48 (Luzin)**.** If $\mu(X) < \infty$, then a function $f : X \to \mathbb{C}^d$ 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 2.47, we can find a sequence of continuous (or even smooth) functions f_n with $f_n \to f$ 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. ☐

2.6 Integration of simple functions

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

At first, this seems like it might be 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.

Definition 2.49. An *integrable simple function* is a function $f \in \mathbf{Simp}(X \to B)$ 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 $\mathbf{ISF}(X \to B)$.

We want the integral to be, at first, a linear map $\mathbf{ISF}(X \to B) \to B$. 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 $\int_{-\infty}^{\infty} c1_E = c\mu(E)$.

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,

$$f(x) = \sum_{i=1}^{n} b_i 1_{f^{-1}(b_i)}(x). \tag{2.2}$$

In particular, the indicator functions form a basis of $\mathbf{ISF}(X \to \mathbb{C})$.

Definition 2.50. We call (2.2) the *canonical representation* of the integrable simple function f with image $\{b_1, \dots, b_n\}$.

Definition 2.51. Let $f \in \mathbf{ISF}(X \to B)$ and suppose that (2.2) is the canonical representation of f. Let E be a measurable set. We define the *integral* of f to be

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

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.

We leave it to the reader to verify the following laundry list of properties of the integral on $\mathbf{ISF}(X \to B)$.

Lemma 2.52. Let $f, g \in \mathbf{ISF}(X \to B)$. Then:

1. $\int f + \int g = \int f + g$.
2. For every $c \in \mathbb{C}$, $\int cf = c \int f$.
3. The *triangle inequality*: $|| \int f|| \leq \int ||f(x)|| \; d|\mu|(x)$.
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. If $\int ||f(x)|| \; d|\mu|(x) = 0$ then $f = 0$ almost everywhere.

We want to extend the integral from $\mathbf{ISF}(X \to B)$ to the space $M(X \to B)$ of all measurable functions. But we cannot merely define $\int f = \lim_n f_n$, where the f_n are simple functions that converge to f. Indeed, if $f_n = 1_{[n,n+1]}$ then $f_n \to 0$ 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 2.53. The L^1 *norm* of a function $f \in \mathbf{ISF}(X \to B)$ is defined by

$$||f||_1 = \int_X ||f(x)|| \; d|\mu|(x). \tag{2.3}$$

As the reader should check, the L^1 norm is a seminorm (its kernel, by Lemma 2.52, is the space of functions f with $\int ||f(x)|| \; d|\mu|(x) = 0$). Now when we say that $f_n \to f$ in L^1, we mean that $||f_n - f||_1 \to 0$. Similarly if we say that the f_n are Cauchy in L^1, we mean that $||f_n - f_m||_1 \to 0$.

It is natural to want to extend the L^1 norm to the completion of the space $\mathbf{ISF}(X \to B)$, on which it will actually be a norm. (See Theorem A.10 for more on that.)

Definition 2.54. The completion of $\mathbf{ISF}(X \to B)$ is known as $L^1(X \to B)$.

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

Definition 2.55. Let $f \in L^1(X \to B)$ and suppose that E is a measurable set. Choose a Cauchy sequence of $f_n \in \mathbf{ISF}(X \to B)$ such that $f_n \to f$ in L^1. We define the *integral* of f to be

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

Let's check that the above definition makes sense. When we say that $f_n \to f$ 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 $g_n \to f$ in L^1,

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

the last equality following because the g_n and the f_n are Cauchy equivalent. Therefore the choice of Cauchy sequence in Lemma 2.55 is immaterial, and any choice will return the same integral. As a consequence, the L^1 norm extends to all of $L^1(X \to B)$ by the equation (**??**). Moreover, one can readily verify the following claim.

Lemma 2.56. The conclusion of Lemma 2.52 holds for any $f, g \in L^1(X \to B)$, not just $f, g \in \mathbf{ISF}(X \to B)$.

2.7 The integral in general

The conclusion of the previous section, wherein we treated the integral on $L^1(X \to B)$, 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.

To find the isomorphism, we first show that convergence in L^1 implies several other modes of convergence.

Theorem 2.57 (fundamental theorem of integration, part II)**.** Suppose that $f_n \to f$ in L^1. Then there is a function f' such that $f_n \to f'$ in measure and there is a subsequence of f_{n_k} such that $f_{n_k} \to f'$ 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,

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

Then

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

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

Now 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 2.58 (fundamental theorem of integration, part III)**.** Suppose that $f_n \to f$ in L^1 and $f_n \to f'$ in measure. If $g_n \to f$ in L^1, then $g_n \to f'$ in measure. Conversely, if the g_n are Cauchy in L^1 and $g_n \to f'$ in measure, then $g_n \to f$ in L^1.

Proof. Suppose that $g_n \to f$ 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 $h_k \to 0$ in L^1 then the f_{n_k} and g_{n_k} are Cauchy equivalent.

Lemma 2.59. $h_k \to 0$ in L^1.

Proof. Let $\varepsilon > 0$, and choose N so that if $n_1, n_2 \geq N$ then $||h_n - h_m||_1 < \varepsilon$. 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 $h_N \in \mathbf{ISF}(X \to B)$, 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 $h_n \to 0$ nearly uniformly, there is a measurable set $F \subseteq E$ such that $\mu(E \setminus F) < \varepsilon$ and $h_n \to 0$ uniformly on F. Therefore

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

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

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

So

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

which was to be shown. □

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 $f_n \to f$ in L^1, $g_n \to f$ in L^1. □

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 $f_n \to f$ in L^1, such that $f_n \to f'$ 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 2.60. Let $f \in M$, and suppose the $f_n \in \mathbf{ISF}$ are an L^1 Cauchy sequence. Then the following are equivalent:

1. $f_n \to f$ in measure.

2. $f_n \to f$ nearly uniformly.

3. $f_n \to f$ almost everywhere.

Proof. That 1 implies 2 implies 3 is the content of the fundamental theorem of integration. Now if $f_n \to f$ 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 $f_n \to f$ in measure. □

Elements of M are functions up to the equivalence relation of being equal almost everywhere. Therefore if $f_n \to f$ in L^1 and $f_n \to f'$ 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 2.61. Let $f \in M(X \to B)$, and suppose that there is a L^1 Cauchy sequence $f_n \in \mathbf{ISF}(X \to B)$ such that $f_n \to f$ in measure. Then we say that $f \in L^1(X \to B)$, define for any measurable set E the *integral*

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

and the L^1 *norm*

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

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

We restate the properties of the integral for the reader's convenience.

Theorem 2.62. Let $f, g \in L^1(X \to B)$. Then:

1. $\int f + \int g = \int f + g$.

2. For every $c \in \mathbb{C}$, $\int cf = c \int f$.

3. The *triangle inequality*: $|| \int f|| \leq \int ||f(x)|| \, d|\mu|(x)$.

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. If $\int ||f(x)||\ d|\mu|(x) = 0$ then $f = 0$ almost everywhere.

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

Definition 2.63. Let $f \in M(X \to B)$, and assume that $f \geq 0$. Let E be a measurable set. If for every L^1 Cauchy sequence $f_n \in \mathbf{ISF}(X \to B)$, f_n does not converge to $1_E f$ in measure, then we write

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

Now if $f \in M(X \to \mathbb{C})$, we can write $f = f_a + i f_b$ 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!

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.

Chapter 3

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.

3.1 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 $L^1 = L^1(X \to B)$. 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.

We note that the relationship between L^1 convergence, convergence in measure, nearly uniform convergence and almost everywhere convergence, that we already proved for **ISF**, extends to all of L^1, since L^1 is the completion of **ISF**. That is:

Theorem 3.1. Let f_n be an L^1 Cauchy sequence and $f \in M$. Then $f_n \to f$ nearly uniformly iff $f_n \to f$ in measure iff $f_n \to f$ in L^1.

Since L^1 is the completion of **ISF**, L^1 is a Banach space and **ISF** is dense in L^1. In particular L^1 is complete, so if we are given a L^1 Cauchy sequence, say $f_n \in L^1$, then there is a $f \in L^1$ such that $f_n \to f$ in L^1, and in particular, $\int f_n \to \int f$. Indeed:

Lemma 3.2. If $f_n \to f$ in L^1, then $\int f_n \to \int f$.

Proof. We have

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

which converges to 0. $\square$

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

Proof. Let $f_n \in \mathbf{ISF}$, $f_n \to f$. 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 3.4. Let $f \in L^1$. Then for every $\varepsilon > 0$ there is a measurable set E such that $|\mu|(E) < \infty$ and

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

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

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

□

We now show that L^1 functions are approximately bounded in a certain sense.

Definition 3.5. 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

$$||f||_\infty = \inf_{f'} \sup_x ||f(x)||$$

where the inf is taken over all versions f' of f such that f' is bounded, and the sup is taken over all $x \in X$.

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

Lemma 3.6. 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 $x \in X$,

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

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 $f'_n \to f$ almost uniformly; that is, there is a null set away from which $f'_n \to f$ uniformly. Indeed, given $\varepsilon > 0$ we can find N such that for almost every x and every $n_1, n_2 > N$, $||f'_{n_1}(x) - f'_{n_2}(x)|| < \varepsilon$; this implies that $||f'_n(x) - f(x)|| < \varepsilon$ if $n > N$. Therefore

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

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

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

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 3.7. 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. □

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":

Lemma 3.8. 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! □

Conversely, if the graph of the function cannot be "too skinny", then we have the opposite bound.

Definition 3.9. 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 $\mu(E) \geq \delta$.

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 3.10. 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. □

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 3.11. 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 3.12. 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 $f \in L^1_l$. □

3.2 Indefinite integrals

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

Throughout, let $X = (X, \Sigma, \mu)$ be a complete measured space and B a Banach space.

Definition 3.13. 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*.

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 $f \geq 0$.

We would like ν to be a measure, but first we need to check a certain inequality.

Lemma 3.14. 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,

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

Proof. One checks

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

which proves the claim. □

Theorem 3.15. 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, $||\nu(E) - \rho(E)|| \leq ||f - g||_1 < \varepsilon$.

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 $N \to \infty$, 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 $n > N$,

$$||\rho(F) - \rho(F^N)|| < \varepsilon.$$

In particular,

$$||\nu(F) - \nu(F^N)|| \leq ||\nu(F) - \rho(F)|| + ||\rho(F) - \rho(F^N)|| + ||\rho(F^N) - \nu(F^N)|| < 3\varepsilon.$$

This implies $\nu(F^N) \to \nu(F)$. □

Radon-Nikodym differentiable measures have a particularly easy-to-understand total variation.

Theorem 3.16. Let ν be a Radon-Nikodym differentiable measure. Then for any measurable set E,

$$|\nu|(E) = \int_E \left|\left|\frac{d\nu}{d\mu}(x)\right|\right| \; d|\mu|(x).$$

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

$$\sum_i ||\nu(E_i)|| = \sum_i \left|\left|\int_{E_i} f \; d\mu\right|\right| \leq \int_E ||f(x)|| \; d|\mu|(x).$$

Taking the supremum over all such finite disjoint unions we see that $|\nu|(E) \leq \int_E ||f||$.

We first check the converse when $f \in \mathbf{ISF}$. TODO □

3.3 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 3.17. Let $f_n = 1_{[n,n+1]}$. Then $f_n \to 0$ pointwise but

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f_n(x) \; dx = 1.$$

Here we formulate criteria under which one is allowed to pull a limit inside an integral.

Theorem 3.18 (dominated convergence)**.** Let $f_n \in L^1(X \to B)$, and suppose that $f_n \to f$ almost pointwise. If there is a $g \in L^1(X \to [0, \infty))$ such that for every n and almost every $x \in X$,

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

then $f \in L^1(X \to B)$ and $f_n \to f$ in $L^1(X \to B)$.

Corollary 3.19. Let $f_n \geq 0$ be nonnegative integrable functions and μ a nonnegative measure. Then

$$\int \sum_n f_n = \sum_n \int f_n.$$

Proof. The sequence of partial sums is increasing, so we can apply monotone convergence. □

3.4 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 $x : A \to \bigcup_{\alpha\in A} X_\alpha$ such that for every $a \in A$, $x(\alpha) \in X_\alpha$. We usually write x_α or $\pi_\alpha(x)$ to mean x_α. The maps

$$\pi_\beta : \prod_{\alpha\in A} X_\alpha \to X_\beta$$

are known as *canonical projection*s 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 $X_1 \times \cdots X_n$ to mean the product of sets X_i, $i \in A$. An element of $X_1 \times \cdots \times X_n$ 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 3.20. 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 $X_\alpha \to X'_\alpha$, $x \mapsto (x, \alpha)$, 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 $f : \bigcup_{\alpha\in A} X_\alpha \to A$ 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 C.19, there is an injective map $g : A \to \bigcup_{\alpha\in A} X_\alpha$ such that $f \circ g$ is the identity, and so $g(\alpha) \in X_\alpha$. Define an element x of X_α by letting $x_\alpha = g(\alpha)$. □

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 3.21. 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 3.22. 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. □

Therefore it is reasonable to want to define a premeasure on $\bigoplus_m E_m$, which we do shortly.

Definition 3.23. 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* $\overline{\mu}$ by $\overline{\mu}(E) = \overline{\mu(E)}$. Then we can define the *real part* $\operatorname{Re}\mu = (\mu + \overline{\mu})/2$ and *imaginary part* $\operatorname{Im}\mu = (\mu - \overline{\mu})/2i$. 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 $\mu_\pm$ 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 3.24. Let $(X_1, \Sigma_1, \mu_1), \dots, (X_n, \Sigma_m, \mu_n)$ be measured spaces, where all the μ_m are complex-valued measures. We define a function $\bigoplus_m \mu_m = \mu_1 \oplus \cdots \oplus \mu_n$ 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 3.29.

Lemma 3.25. Let $(X_1, \Sigma_1, \mu_1), \dots, (X_n, \Sigma_m, \mu_n)$ 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 3.19,

$$1_E(x)\nu(F) = \sum_{n=1}^{\infty} 1_{E_n}(x)\int_Y 1_{F_n}(y)\ d\nu(y) = \sum_{n=1}^{\infty} 1_{E_n}(x)\nu(F_n).$$

Applying Corollary 3.19 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 3.26. Let $(X_1, \Sigma_1, \mu_1), \dots, (X_n, \Sigma_m, \mu_n)$ 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 3.25 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 3.27. Let $(X_1, \Sigma_1, \mu_1), \dots, (X_n, \Sigma_m, \mu_n)$ be measured spaces, where all the μ_m are complex-valued measures. Let $\bigotimes_m \mu_m = \mu_1 \otimes \cdots \mu_n$ 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 3.28. Let $A = \{1, \dots, n\}$, equipped with the σ-algebra consisting of *every* subset of A, and consider a function $\beta : A \to [0,1]$ 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 3.29. Let A, β be as in Example 3.28. 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 Example 1.9, 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 nth 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 3.29 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 3.30. 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 αth 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

$$d(x, y) = \max_m d_m(\pi_m(x), \pi_m(y)). \tag{3.1}$$

Lemma 3.31. Let $X_1, \ldots, X_n$ 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 3.1. 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 3.31 to define the Lebesgue integral in general.

3.5 The Lebesgue integral

Let μ be a Stieltjes measure. Then μ is a Borel measure on $\mathbb{R}$, and by Lemma 3.31, 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 3.32. 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 3.33. The Lebesgue measure is a σ-finite Radon measure.

Proof. By induction on d. When $d = 1$ this is the content of Lemma 2.46. 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 disjoint[1] 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^n are disjoint and $L_n = \bigcup_{m \leq n} K_m^m$ is compact. Moreover if $\mu^d(U) = \infty$ then $\mu^d(L_n) \to \infty$; otherwise

$$\mu^d(U) = \sum_n \mu^d(U_n) \leq \sum_n \mu^d(K_n^n) + \frac{1}{2^n 2^n} < \varepsilon + \mu^d(L_n) + \sum_{m>n} \mu^d(K_m^m)$$

and

$$\sum_{m>n} \mu^d(K_m^m) \leq \sum_{m>n} \mu^d(U_m) < \varepsilon$$

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. □

Theorem 3.34. 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. We leave the proof as an exercise. The reader might try to use induction on d. Indeed, when $d = 1$ this is just the fact that $\mu([\alpha, \beta)) = \beta - \alpha$. □

Theorem 3.35. Let $f : [\alpha, \beta] \to \mathbb{R}$ be a bounded Riemann integrable function and let R be its Riemann integral. Then

$$\int_\alpha^\beta f(x)\, dx = R.$$

[1] in the sense that their intersection is Lebesgue null

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 $n \to \infty$. 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. □

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.

Theorem 3.36. Let $f \in L^1(\mathbb{R}^d \to [0, \infty))$. 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. □

3.6 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 3.37 (Fubini)**.** Let (X, Σ, μ) and (Y, Γ, ν) be σ-finite complex-valued measured spaces. Let $f : X \times Y \to \mathbb{C}$ be a nonnegative $(\Sigma \otimes \Gamma)$-measurable function. Then the following are equivalent:

1. $f \in L^1(X \times Y \to \mathbb{C}, \mu \otimes \nu)$.

2. For almost every $x \in X$, $f(x, \cdot) \in L^1(Y \to \mathbb{C}, \nu)$ and the function

   $$x \mapsto \int_Y f(x, y)\ d\nu(y)$$

   is in $L^1(X \to \mathbb{C}, \mu)$.

3. For almost every $y \in Y$, $f(\cdot, y) \in L^1(X \to \mathbb{C}, \mu)$ and the function

$$y \mapsto \int_X f(x,y)\ d\mu(x)$$

is in $L^1(Y \to \mathbb{C}, \nu)$.

Moreover,

$$\int_{X \times Y} f\ d(\mu \otimes \nu) = \int_X \int_Y f(x,y)\ d\mu(x)\ d\nu(y) = \int_Y \int_X f(x,y)\ d\nu(y)\ d\mu(x). \tag{3.2}$$

We will weaken the hypotheses on this theorem somewhat before the end of the section. Because of (3.2) 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 cetra, 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 two of its many applications. First, we note that the following consequence is immediate when we take $X = Y = \mathbb{N}$ and $\mu = \nu$ to be counting measure, thus $\mu(A)$ is the cardinality of A.

Corollary 3.38. 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

$$\sum_{i=1}^\infty \sum_{j=1}^\infty x_{i,j} = \sum_{j=1}^\infty \sum_{i=1}^\infty x_{i,j}.$$

Second, we compute a classical integral as a demonstration.

Example 3.39. Let us show that

$$\int_{-\infty}^\infty e^{-x^2}\ dx = \sqrt{\pi},$$

the *Gaussian integral*. This integral is fundamental in various areas of analysis and its applications, including statistics.

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
$$\int_{-\infty}^{\infty} e^{-x^2}\ dx \int_{-\infty}^{\infty} e^{-y^2}\ dy = \pi. \tag{3.3}$$
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
$$\iint_{\mathbb{R}^2} e^{-(x^2+y^2)}\ dA(x,y) = \int_0^{2\pi} \int_0^{\infty} re^{-r^2}\ dr\ d\theta = 2\pi \int_0^{\infty} re^{-r^2}\ dr.$$
Now this is really just a calculus problem: if $s = r^2$ then $ds = 2r\ dr$, and so one easily checks that
$$\int_0^{\infty} re^{-r^2}\ dr = \frac{1}{2}.$$
Then (3.3) immediately follows.

Monotone Classes

TODO: Prove Fubini

3.7 Differentiation of measures

3.8 Construction of Radon measures

Chapter 4

Hölder duality

4.1 p-norms

4.2 Completeness of L^p

4.3 Convergence theorems

4.4 Proof of Hölder duality

4.5 Compactness of the unit ball

4.6 Riesz-Thorin interpolation

Chapter 5

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 Fubini's theorem, to get motivation for the utility of such results.

5.1 Haar measures

We begin by introducing measure theory on abelian groups. Recall that a group is called abelian if its group operation is commutative. If $(G, +)$ is an abelian group, we let $-g$ denote the inverse of an element g, and let 0 denote the identity of G, so that $g - g = 0$. If the reader is totally unfamiliar with group theory, they could take G to be the real numbers $\mathbb{R}$ under addition, the integers $\mathbb{Z}$ under addition, or the circle $\mathbb{T}$, whose elements are of the form $e^{i\theta}$ and whose group operation is $e^{i\theta}e^{i\varphi} = e^{i(\theta+\varphi)}$.

Definition 5.1. 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 $G \times G \to G$, and $g \mapsto -g$ is continuous.

If the reader is not familiar with the abstract notion of a topology, they can view a locally compact abelian group as a group equipped with a metric $(G, +, d)$, such that $+$ and $g \mapsto -g$ are continuous, and every point in G is contained in an open set U such that the closue $\overline{U}$ is compact.

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 metric.

Given any locally compact abelian group G, we may consider its Borel σ-algebra $\mathcal{B}(G)$. We may define, for any set $A \subseteq G$, its translate

$$A + y = \{x + y \in G : x \in A\}.$$

Definition 5.2. A *Haar measure* on G is a nonnegative measure μ defined on $\mathcal{B}(G)$ with the following properties:

1. *Finiteness on compact sets*: For every compact set K, $\mu(K) < \infty$.

2. *Outer regularity on Borel sets*: For every Borel set B, $\mu(B) = \inf \sum_i \mu(U_i)$, where the inf ranges over all countable open covers $(U_i)_i$ of B.

3. *Inner regularity on open sets*: For every open set U, $\mu(U) = \sup \sum_i \mu(K_i)$, where the sup ranges over all countable sets of disjoint compact sets $(K_i)_i$ such that $K_i \subseteq U$.

4. *Translation invariance*: For every Borel set B and every $x \in G$, $\mu(B + x) = \mu(B)$.

5. *Normalization*: If G is compact, then $\mu(G) = 1$.

Example 5.3. Lebesgue measure is a Haar measure on $\mathbb{R}$, and counting measure is a Haar measure on $\mathbb{Z}$. On finite abelian groups G, we define $\mu(A) = |A|/|G|$ where $|\cdot|$ denotes cardinality.

We define a bijection $f(x) = e^{2\pi i x}$, $f : [0, 1) \to \mathbb{T}$. Then we define a measure μ on the Borel subsets of $\mathbb{T}$ by

$$\mu(B) = \nu(f^{-1}(B))$$

where ν is Lebesgue measure on $[0, 1)$. Then μ is a Haar measure on $\mathbb{T}$, normalized so that $\mu(\mathbb{T}) = 1$ (since $\mathbb{T}$ is compact).

The finite product of Haar measures is a Haar measure, thus we have Haar measures on every locally compact abelian group we have considered.

It is a nontrivial theorem, that we omit, that every locally compact abelian group has a Haar measure, and that the Haar measure is unique up to scalars (or simply unique, for compact abelian groups). Thus the results that we will prove in this chapter will apply to any locally compact abelian group, and we will henceforth consider every locally compact abelian group equipped with a given Haar measure.

5.2 Fourier transform

To motivate what follows, we consider $\mathbb{R}$ with Lebesgue measure and consider a function $f : \mathbb{R} \to \mathbb{C}$. We will assume for simplicity that f is Schwartz:

Definition 5.4. A function $f : \mathbb{R}^d \to \mathbb{C}$ is *Schwartz* if f is smooth and rapidly decreasing in the sense that for every $N > 0$ and $x \in \mathbb{R}$,

$$|f(x)| \lesssim_N |x|^N.$$

Thus we will not have to worry about any issues as to whether certain integrals or derivatives that we consider exist.

We want to view f as a "wave" and decompose f into different waves of the form $e_\xi(x) = e^{2\pi i x\xi}$ for frequencies ξ. Intuitively, one should consider the "inner product"

$$\langle f, e_\xi \rangle = \int_{-\infty}^{\infty} f(x)\overline{e^{2\pi i x\xi}}\, dx = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x\xi}\, dx.$$

This should give the "amplitude" of the piece of f that is oscillating at frequency ξ, which we denote by $\hat{f}(\xi)$. As we will see, we can reconstruct f by the formula

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\xi) e^{2\pi i x \xi}\, dx. \tag{5.1}$$

We prove (5.1), *Fourier's inversion formula*, in the more abstract setting of a locally compact abelian group.

Recall that a *morphism of groups* is a function $f : G \to H$ such that for every $x, y \in G$, $f(x+y) = f(x) + f(y)$. If H is an abelian group, then we can add two morphisms of groups, namely $(f+g)(x) = f(x) + g(x)$. Thus the set of all morphisms of groups into an abelian group is an abelian group.

Definition 5.5. Let G be a locally compact abelian group. The *dual group*, denoted by $\hat{G}$ or $\mathrm{Hom}(G, \mathbb{T})$, is the abelian group of all continuous morphisms of groups $G \to \mathbb{T}$.

One can put a topology on the dual group, turning it into a locally compact abelian group, by declaring that $f_n \to f$ iff for every compact set $K \subseteq G$, $f_n \to f$ uniformly on K. It is useful to know that

$$\widehat{G_1 \times G_2} = \hat{G}_1 \times \hat{G}_2.$$

We let $G \cong H$ to mean that there is a continuous isomorphism $G \to H$, whose inverse is continuous.

Example 5.6. One has $\mathbb{R} \cong \hat{\mathbb{R}}$; namely, $\xi \mapsto e_\xi$, where $e_\xi(x) = e^{2\pi i x \xi}$. One easily checks that $e_\xi \in \hat{\mathbb{R}}$ and that $\xi \mapsto e_\xi$ is injective. To see that is surjective, suppose that $\chi \in \hat{\mathbb{R}}$. We want to show that $\chi(x) = e^{2\pi i x \xi}$ for some ξ. TODO

Example 5.7. Let G be a finite abelian group. Then G can be written as a product of cyclic groups, so we might as well assume that $G \cong \mathbb{Z}/n$ to compute $\hat{G}$. Let $\chi : \mathbb{Z}/n \to \mathbb{T}$ be a morphism of groups; then χ is determined by its value $\chi(1)$ on the generator 1 of $\mathbb{Z}/n$. Now $\chi(1)$ must be an nth root of unity; that is, $\chi(1) = e^{2\pi i k/n}$ for some $k \in \{0, \dots, n-1\}$. This is to ensure that $\chi(1)^n$ is the identity e^0 of $\mathbb{T}$. But the nth roots of unity form a finite abelian group isomorphic to $\mathbb{Z}/n$, and for any root of unity ω we can find a morphism of groups χ such that $\chi(1) = \omega$. Thus $\widehat{\mathbb{Z}/n} \cong \mathbb{Z}/n$.

We leave it as an exercise to check that $\hat{\mathbb{Z}} \cong \mathbb{T}$ and $\hat{\mathbb{T}} \cong \mathbb{Z}$. The isomorphism $\psi : \mathbb{Z} \to \hat{\mathbb{T}}$ is defined by what it does to the generator 1; in fact $\psi(1)$ is the identity on $\hat{\mathbb{T}}$.

Definition 5.8. Assume that G is a locally compact abelian group with Haar measure. Let $f \in L^1(G)$. We define the *Fourier transform* of f by

$$\hat{f}(\chi) = \int_G f(x) \overline{\chi(x)}\, dx.$$

TODO: Convolutions

Theorem 5.9 (Fourier inversion)**.** Suppose that $f \in L^1(G)$ and $\hat{f} \in L^1(\hat{G})$. Then

$$f(x) = \int_{\hat{G}} \hat{f}(\chi) \chi(x)\, d\chi.$$

Proof. One has

$$\int_{\hat{G}} \hat{f}(\chi)\chi(x)\ d\chi = \iint_{\hat{G}\times G} f(x)\chi(x)\overline{\chi(y)}\ dy\ d\chi$$

by Fubini's theorem. Now $\chi(y) = e^{i\theta}$ for some $\theta \in \mathbb{R}$, thus $\overline{\chi(y)} = e^{-\theta} = \chi(y)^{-1}$. Therefore

$$\chi(x)\overline{\chi(y)} = \chi(x-y)$$

so

$$\int_{\hat{G}} \hat{f}(\chi)\chi(x)\ d\chi = \iint_{\hat{G}\times G} f(x)\chi(x-y)\ dy\ d\chi.$$

TODO: Maybe move this to the chapter on Holder duality and use L2 orthogonality. □

TODO: Application to PDE

5.3 Pontryagin duality

5.4 Applications to PDE

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. A *seminormed space* is a vector space V equipped with a function

$$\begin{aligned} V &\to [0, \infty) \\ v &\mapsto ||v||, \end{aligned}$$

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 willl have no use for this.

Example A.2. The most important example of a normed space has as its base space $\mathbb{R}^d$. We define

$$||(x_1, \ldots, x_d)||_2^2 = \sum_{i=1}^{d} |x_i|^2,$$

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.3. We can define several other norms on $\mathbb{R}^d$, closely related to the Euclidean norm $||\cdot||_2$. First we let

$$||(x_1, \ldots, x_d)||_\infty = \max_{i=1}^{d} |x_i|.$$

We then define, for any $p \in [1, \infty)$,

$$||(x_1, \ldots, x_d)||_p^p = \sum_{i=1}^{d} |x_i|^p.$$

Thus when $p = 2$ we recover the Euclidean norm, when $p = 1$ we recover the "sum norm" $||(x_1, \ldots, x_d)||_1 = \sum_i |x_i|$, and in the limit $p \to \infty$ 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

$$\lim_{n\to\infty} v_n = w$$

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.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 $\pi : V \to V'$.

2. For every normed space W and every linear map $T : V \to W$ such that for every v, $||Tv|| = ||v||$, there is a linear map $T' : V/\ker V \to W$ such that $||T'v|| = ||v||$ and the diagram

$$\begin{tikzcd} V \arrow[rr, "T"] \arrow[dr, "\pi"'] & & W \\ & \frac{V}{\ker V} \arrow[ur, "T'"'] & \end{tikzcd} \tag{A.1}$$

   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.10) is unique.

The reader should omit the proof of Theorem A.4 on first reading.

Proof of Theorem A.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

$$0 \longrightarrow \ker V \longrightarrow V \xrightarrow{\pi} \frac{V}{\ker V} \longrightarrow 0 \tag{A.2}$$

(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 $T : V \to W$ 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 $\ker T$. □

We conclude this section with a useful consequence of the triangle inequality.

Lemma A.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 $\max(||v|| - ||w||, ||w|| - ||v||) = |||v|| - ||w|||$. □

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.6. Let $C[0,1]$ denote the vector space of continuous functions $[0,1] \to \mathbb{C}$. 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 $\int_0^1 |g| = 0$, then for every $\varepsilon > 0$ there is a partition of $[0,1]$ into intervals $I_1, \ldots, I_n$ 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 $\varepsilon \to 0$ 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 $g(y) = 0$). 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

$$||f_n - f_m||_2 \leq \frac{4}{\sqrt{n}} \to 0$$

as $n \to \infty$. 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

$$\lim_{n\to\infty} ||f_n - f||_2 = 0;$$

yet f is not continuous, and $|| \cdot ||_2$ is a norm, so f_n cannot converge to any continuous function.

Definition A.7. A *Banach space* is a normed space for which every Cauchy sequence converges.

Example A.8. 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

$$\sum_{n=1}^{\infty} x_n = \lim_{N\to\infty} \sum_{n=1}^{N} x_n \tag{A.3}$$

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

$$\left|\left| y - \sum_{n=1}^{N} x_n \right|\right| < \varepsilon.$$

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

$$\left|\left| \sum_{n=N}^{\infty} x_n \right|\right| < \varepsilon.$$

By the triangle inequality, it in fact suffices to show that for every $\varepsilon > 0$ there is an N such that

$$\sum_{n=N}^{\infty} ||x_n|| < \varepsilon \tag{A.4}$$

to show that the $(x_n)_n$ are summable.

Definition A.9. 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.10. 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 $\iota : V \to W$ 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 $T : V \to X$ such that for every v, $||Tv|| = ||v||$, then there is a linear map $T' : W \to X$ such that the diagram

$$\begin{tikzcd} V \arrow[rr, "T"] \arrow[dr, "\iota"'] & & X \\ & W \arrow[ur, "T'"'] & \end{tikzcd}$$

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

$$||x|| = \lim_{n\to\infty} ||x_n||.$$

(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 $\pi : \mathbf{Cau}(V) \to W$ 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 $\widetilde{x} \in \mathbf{Cau}(\mathbf{Cau}(V))$ such that for every n,

$$\pi(\widetilde{x}_n) = x_n.$$

Now $\widetilde{x}_n$ is a Cauchy sequence, say $(\widetilde{x}_{n,m})_m$. If we replace $\widetilde{x}_n$ with a subsequence (in m) of $\widetilde{x}_n$, $\pi(\widetilde{x}_n)$ will not change. Thus we might as well assume that for any two m, m',

$$||\widetilde{x}_{n,m} - \widetilde{x}_{n,m'}|| < \frac{1}{n}.$$

Now let

$$\widetilde{y}_n = \widetilde{x}_{n,n}.$$

We want $x_n \to \pi(\widetilde{y})$ but it's not obvious that $\widetilde{y}$ 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,

$$\begin{aligned}
||\widetilde{y}_n - \widetilde{y}_{n'}|| &= ||\widetilde{x}_{n,n} - \widetilde{x}_{n',n'}|| \\
&\leq ||\widetilde{x}_{n,n} - \widetilde{x}_{n,j}|| + ||\widetilde{x}_{n,j} - \widetilde{x}_{n',j}|| + ||\widetilde{x}_{n',j} - \widetilde{x}_{n',n'}|| \\
&\leq \frac{1}{n} + ||\widetilde{x}_{n,j} - \widetilde{x}_{n',j}|| + \frac{1}{n'} \\
&< 2\varepsilon + ||\widetilde{x}_{n,j} - \widetilde{x}_{n',j}||.
\end{aligned}$$

But

$$\lim_{j\to\infty} ||\widetilde{x}_{n,j} - \widetilde{x}_{n',j}|| = 0$$

and since j was large enough we can take $||\widetilde{x}_{n,j} - \widetilde{x}_{n',j}|| < \varepsilon$. Then

$$||\widetilde{y}_n - \widetilde{y}_{n'}|| \lesssim \varepsilon.$$

Therefore $\widetilde{y} \in \mathbf{Cau}(V)$, and we can define

$$y = \pi(\widetilde{y}).$$

We similarly choose $N > \varepsilon^{-1}$ such that for every $n, n' \geq N$,

$$||\widetilde{y}_n - \widetilde{y}_{n'}|| < \varepsilon.$$

Then

$$||\widetilde{x}_{n',n} - \widetilde{y}_n|| \leq ||\widetilde{x}_{n',n} - \widetilde{x}_{n',n'}|| + ||\widetilde{y}_{n'} - \widetilde{y}_n|| \lesssim \varepsilon.$$

But

$$\lim_{m\to\infty} ||\widetilde{x}_{n,m} - \widetilde{y}_m|| = ||x_n - y||$$

so $\lim_n x_n = y$.

We can define

$$\iota(v) \mapsto \pi(v, v, v, \dots). \tag{A.5}$$

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 $\widetilde{x} \in \mathbf{Cau}(V)$ with $\pi(\widetilde{x}) = x$. Now $\widetilde{x}$ is a Cauchy sequence in V, so let

$$y_n = \iota(\widetilde{x}_n);$$

then $y \in \mathbf{Cau}(W)$, and the reader can check that $\lim_n y_n = x$.

To check the universal property, let $T : V \to X$ 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.11). 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.13. 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 themself with this material until Chapter TODO.

Definition A.11. Let $T : V \to W$ 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 $B(V \to W)$ the space of bounded linear maps.

Lemma A.12. Let V, W be normed spaces. Then the operator norm is a norm on $B(V \to W)$.

We leave the routine proof to the reader.

Lemma A.13. Let $T : V \to W$ 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 $X \to W$, 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

$$\lim_{n\to\infty} ||Tx_n - Tx'_n|| \leq \lim_{n\to\infty} ||T||_V \cdot ||x_n - x'_n|| = 0$$

since $x_n \to x$ and $x'_n \to x$. Thus Tx does not depend on the choice of Cauchy sequence which approximates x. Similarly,

$$||Tx|| = \lim_{n\to\infty} ||Tx_n|| \leq \lim_{n\to\infty} ||T||_V \cdot ||x_n|| = ||T||_V ||x_n||.$$

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

$$T(x+y) = \lim_{n\to\infty} T(x_n + y_n) = \lim_{n\to\infty} Tx_n + Ty_n = Tx + Ty$$

where $(x_n)_n$ and $(y_n)_n$ are appropriately chosen Cauchy sequences. The proof that $T(cx) = cTx$ is similar. □

A.4 Properties of Banach spaces

Definition A.14. 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.15. $\mathbb{C}^n$ is separable, since $\{(\alpha_1 + i\beta_1, \ldots, \alpha_n + i\beta_n) : \alpha_i, \beta_i \in \mathbb{Q}\}$ 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.16. Let X be the space of all bounded functions $\mathbb{C} \to \mathbb{C}$, 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

Locally compact groups

Appendix C

Foundations

In this appendix we treat the foundations of math: set theory, category thoery, 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.

C.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 C.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 C.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 C.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 C.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 $\exists, \forall, \to, \neg, \in, (,)$, and variables $x_1, x_2, \cdots, y_1, \ldots, y_N$ 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, $\forall x_1((x_1 \in y_1) \to (x_1 \in y_2))$ is a formula which asserts that $y_1 \subseteq y_2$.

First-order formulae allow us to assert the existence of subsets.

Axiom C.5 (restricted comprehension schema)**.** For all sets x, first order-formulae φ with $N+1$ free variables, and sets $w_1, \ldots, w_N$, there is a set y such that $z \in y$ iff $z \in x$ and $\varphi(z, w_1, \ldots, w_N)$ 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 $f : X \to Y$ 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 C.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 $x \to y$.

Theorem C.7 (Cantor-Bernstein)**.** Let x, y be sets. Then $x \cong y$ iff there are injections $x \to y$ and $y \to x$.

In particular, we can ask if a set is finite, countable, etc.

Definition C.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 C.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 C.10. A set x is a *countable set* if there is an injection $x \to \mathbb{N}$. 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 C.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 $x \to 2^x$, given by $y \mapsto \{y\}$.

Theorem C.12 (Cantor's diagonal argument)**.** For every set x, there is no injection $2^x \to x$.

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 C.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 C.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 C.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 C.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 C.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 $\aleph_\omega$ be the least cardinal greater than $\aleph_n$ for every $n \in \mathbb{N}$. It is not possible to define $\aleph_\omega$ yet, but we can do so with the help of a new axiom schema.

Axiom C.18 (replacement schema)**.** For every set X, every first-order formula with $N+2$ free variables φ, and all sets $w_1, \ldots, w_N$, if for every $x \in X$ there is exactly one set y such that $\varphi(x, y, w_1, \ldots, w_N)$ is true, then there is a set z such that $y \in z$ iff there is an $x \in X$ such that $\varphi(x, y, w_1, \ldots, w_N)$ 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 $w_1, \ldots, w_N$, then the image of F is a set. It follows from the replacement schema that $\aleph_\alpha$ 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 C.19 (choice)**.** For every surjection $f : X \to Y$ there is an injection $g : Y \to X$ such that for every $y \in Y$, $f(g(y)) = y$.

In other words if $f : X \to Y$ 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 C.20 (Zermelo's well-ordering theorem)**.** For every set x there is a unique cardinal κ such that $x \cong \kappa$.

Definition C.21. For every set x, the unique cardinal κ such that $x \cong \kappa$ is called the *cardinality* of x.

Definition C.22. We define $\beth_0 = \aleph_0$, and for every α, $\beth_{\alpha+1}$ to be the cardinality of $2^{\beth_\alpha}$. If δ is a limit ordinal we let $\beth_\delta$ be the cardinality of $\bigcup_{\alpha<\delta} \beth_\alpha$.

By Cantor's diagonal argument, $\beth_\alpha < \beth_\beta$ whenever $\alpha < \beta$.

Theorem C.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 C.24. Let $\mathcal{T}$ be the set of all open subsets of $\mathbb{R}$; then $\mathcal{T}$ has cardinality $\beth_1$.

C.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 C.25. 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 C.26 (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 C.27. 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 C.28. A *category* C is a set, whose elements are called *objects*, along with small sets $\mathrm{Hom}(x, y)$ for all objects $x, y \in C$, whose elements are called *morphisms* from x to y, equipped with *composition* operations

$$\begin{aligned} \mathrm{Hom}(x, y) \times \mathrm{Hom}(y, z) &\to \mathrm{Hom}(x, z) \\ (\varphi, \psi) &\mapsto \psi\varphi \end{aligned}$$

such that:

1. For every object $x \in C$, $\mathrm{Hom}(x, x)$ contains a morphism 1_x which is an *identity* in the sense that for all $\varphi \in \mathrm{Hom}(x, y)$, $\varphi = \varphi 1_x$ and for every $\psi \in \mathrm{Hom}(y, x)$, $\psi = 1_x \psi$.

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 $\mathrm{Hom}(x, y)$ be the set of all functions $x \to y$. 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 $\mathbf{Vect}(K)$, 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 $\mathrm{Hom}(x, y)$, 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

$$\begin{tikzcd} x_1 \arrow[r, "\varphi_1"] \arrow[d, "\psi_1"] & x_2 \arrow[d, "\psi_2"] \\ y_1 \arrow[r, "\varphi_2"] & y_2 \end{tikzcd}$$

commutes iff $\psi_1\varphi_2 = \psi_1\varphi_2$.

Definition C.29. Let C be a category, $x, y \in C$ objects, and $\varphi \in \mathrm{Hom}(x, y)$. We say that φ is an *isomorphism*, and x, y are *isomorphic*, if there is a $\psi \in \mathrm{Hom}(y, x)$ which is an *inverse* to φ in the sense that $\varphi\psi = 1_y$ and $\psi\varphi = 1_x$.

Definition C.30. 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 $\mathrm{Hom}(x, y)$. Similarly we say that x is *final* if for every y there is a unique morphism in $\mathrm{Hom}(y, x)$. 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 C.31. 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 $V \to X$ into a Banach space X, and whose morphisms φ are commutative diagrams of linear maps

$$\begin{tikzcd} & X \arrow[dd, "\varphi"] \\ V \arrow[ur] \arrow[dr] & \\ & Y. \end{tikzcd}$$

The initial object of V is the inclusion map $V \to W$, where W is the completion of V. Therefore the completion is well-defined if it exists (which it does, by Theorem A.10). Thus we can take the universal property as the definition of the completion, and take Theorem A.10 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 $\mathbb{Q}$.

C.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 C.32. 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 $U_1, \dots, U_n \mathcal{T}$ then $U_1 \cap \cdots \cap U_n \in \mathcal{T}$.

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 C.33. A set $\mathcal{B}$ of subsets of a set X is called a *basis* if for all $B_1, \dots, B_n \in \mathcal{B}$, $B_1 \cap \cdots \cap B_n \in \mathcal{B}$.

Every basis generates a topology whose elements are arbitrary unions of sets in the basis.

Definition C.34. A *semimetric* d on a set X is a function $d : X \times X \to [0, \infty)$ 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 C.35. 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 want to talk about limits. Unfortunately at the level of abstraction we are working at, this proves quite tricky.

Definition C.36. A *directed set* is a partially ordered set $\mathbb{P}$ such that for all $\alpha_1, \ldots, \alpha_k \in \mathbb{P}$, $\sup(\alpha_1, \ldots, \alpha_k)$ 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 C.37. Let $\mathbb{P}$ be a directed set and X a topological space. A *net* in X, indexed by $\mathbb{P}$, is a function $\mathbb{P} \to X$. If x is a net, we usually write x_α instead of $x(\alpha)$. A *sequence* is a net indexed by $\mathbb{N}$.

Definition C.38. 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 $x_\gamma \to x$ 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, $x_n \to x$ iff for every $\varepsilon > 0$ there is a $N \in \mathbb{N}$ such that for all $n > N$, $d(x_n, x) < \varepsilon$.

Definition C.39. We say that a function $h : S \to \mathbb{P}$ 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 $f : \mathbb{P} \to \mathbb{P}$ between directed sets is *monotone* if for every $\alpha \leq \beta$ in $\mathbb{P}$, $f(\alpha) \leq f(\beta)$.

Definition C.40. 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 $f : \mathbb{P} \to \mathbb{P}'$ such that $x'_\alpha = x_{f(\alpha)}$. A *subsequence* is a sequence which is a subnet of a subsequence.

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 topological space respectively.