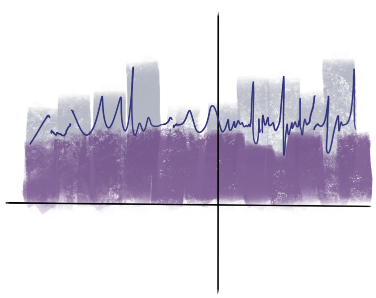
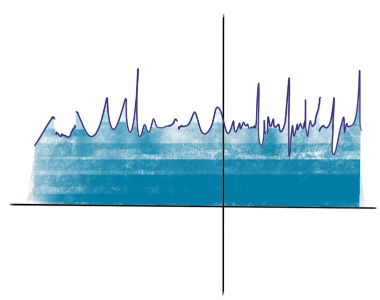
Math 515: Real Analysis I
Lecture Notes

Gabe Khan (based off notes by Tim McNicholl)

July 27, 2023

Contents

Chapter 1

Introduction

The topic of this class is *real analysis*, with a particular focus on measure theory and integration. Since the qualifying exam will primarily focus on functions of a single real variable, this will also be our main focus. We will develop the notion of the Lebesgue integral, which supplants the Riemann integral by being robust enough to handle fairly pathological functions and flexible enough to form the basis for modern functional analysis.

The lecture notes and homework are primarily taken from the textbook by Royden and Fitzpatrick [RF88], which is the textbook for this course. Furthermore, much of the material is taken from previous versions of this course. In particular, this text is mostly modified from the 2015 lecture notes by Timothy McNicholl.

1.1 Lecture 1: A review of Riemann integration

In this introductory lecture, we will review some notions of Riemann integration, which is the historical (and pedagogical) predecessor to the Lebesgue integral that we will study in this course. We start with some basic definitions.

Definition 1. *A partition P of the interval $[a, b]$ is a sequence of real numbers $(x_0, \ldots, x_n)$ with*

$$a = x_0 < x_1 < \ldots x_n = b.$$

Intuitively, the partition cuts up the interval into n subintervals $[x_i, x_{i+1}]$.

Definition 2. *A real valued function* $f : \mathbb{R} \to \mathbb{R}$ *is bounded on the interval* $[a, b]$ *if there is a positive number* M *so that*

$$|f(x)| \leqslant M$$

for all $x \in [a, b]$.

Here, the domain of f can be any set which contains $[a, b]$, and not necessarily the entire real line.

Definition 3. *Suppose that* f *is a real-valued bounded function on* $[a, b]$.

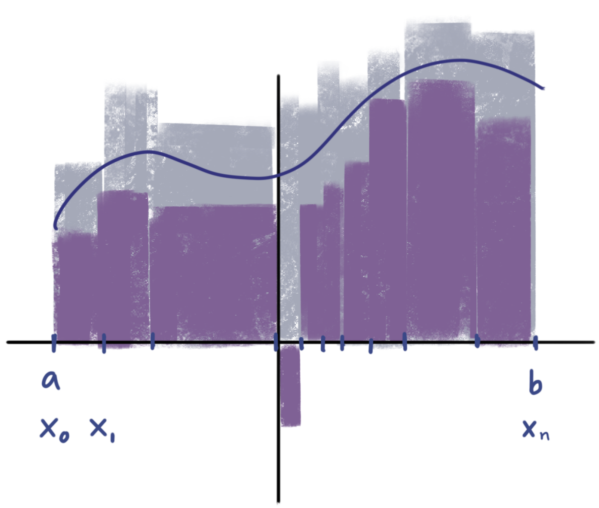
1. *A lower Riemann sum for* f *on* $[a, b]$ *is a sum of the form*

$$\sum_{j=1}^{n} v_j(x_j - x_{j-1})$$

   where $(x_0, \ldots, x_n)$ *is a partition of* $[a, b]$ *and* $v_j \leqslant f(x)$ *whenever* $x \in [x_{j-1}, x_j]$.

2. *An upper Riemann sum for* f *is the same, except now we insist that* $v_j \geqslant f(x)$ *whenever* $x \in [x_{j-1}, x_j]$.

Intuitively, the lower Riemann sum is an *under-estimate* for the area underneath the function $f(x)$ whereas the upper Riemann sum is an *over-estimate.*



By taking finer and finer partitions of the interval $[a,b]$, we can refine our estimates for the area under the curve and (hopefully), compute its Riemann integral.

Definition 4. *The lower Riemann integral* $\underline{\int}$ *is the supremum of all lower Riemann sums with respect to all possible partitions.*

Definition 5. *The upper Riemann integral* $\overline{\int}$ *is the infimum of all upper Riemann sums with respect to all possible partitions.*

We say that a function f is *Riemann integrable* if the lower Riemann integral is equal to the upper Riemann integral.

Example. *If f is continuous on $[a,b]$, then f is integrable.*

Proof. Since f is a continuous function on a closed and bounded interval, it is uniformly continuous. Therefore, for any $\epsilon > 0$, we can find a δ so that

$$|f(x) - f(y)| < \epsilon$$

whenever $|x - y| < \delta$.

We then take a partition P whose mesh[1] is smaller than δ and consider the upper and lower Riemann sums

$$L_P = \sum_{j=1}^{n} f(\underline{x}_j)(x_j - x_{j-1})$$

and

$$U_P = \sum_{j=1}^{n} f(\overline{x}_j)(x_j - x_{j-1}),$$

where $\underline{x}_j = argmin_{x\in[x_{j-1},x_j]} f(x)$ and $\overline{x}_j = argmax_{x\in[x_{j-1},x_j]} f(x)$.

We now compare L_P and U_P using the uniform continuity of f, we have the following estimate

$$\begin{aligned} U_P - L_P &= \sum_{j=1}^{n} f(\overline{x}_j)(x_j - x_{j-1}) - \sum_{j=1}^{n} f(\underline{x}_j)(x_j - x_{j-1}) \\ &= \sum_{j=1}^{n} \left(f(\overline{x}_j) - f(\underline{x}_j)\right)(x_j - x_{j-1}) \\ &< \sum_{j=1}^{n} \epsilon(x_j - x_{j-1}) \\ &= \epsilon \cdot (b-a) \end{aligned}$$

[1] In other words, we consider a partition for which $x_j - x_{j-1} < \delta$ for all j with $1 \leqslant j \leqslant n$.

For any upper and lower sum, we have that

$$L_P \leqslant \underline{\int} \leqslant \overline{\int} \leqslant U_P.$$

By taking ϵ arbitrarily small, this shows that $\mathcal{L} = \mathcal{U}$. □

In fact, we can Riemann integrate functions which have some discontinuities (and we will prove a much stronger version of the following exercise later in the course).

Exercise 1. *Let f be a bounded real-valued function on $[a, b]$ that is continuous except possibly at finitely many points. Show that f is Riemann integrable.*

On the other hand, by considering highly discontinuous functions, we can create functions which are not Riemann integrable.

Non-Example. *The function $f : [0, 1] \to \mathbb{R}$ with*

$$\mathbb{1}_{\mathbb{Q}}(x) = \begin{cases} 1 \textit{ whenever } x \in \mathbb{Q} \\ 0 \textit{ whenever } x \notin \mathbb{Q} \end{cases}$$

is not Riemann integrable.

Proof. In any interval (x_{j-1}, x_j), there are irrational number. Therefore, if $\sum_{j=1}^{n} v_j(x_j - x_{j-1}$ is a lower Riemann sum, then $v_j \leqslant 0$ and thus we have that $\mathcal{L} \leqslant 0$. Similarly, in any interval there are rational numbers, so if $\sum_{j=1}^{n} v_j(x_j - x_{j-1}$ is an upper Riemann sum, then $v_j \geqslant 1$ for all j. As such, $\mathcal{U} \geqslant 1$. □

With slightly more effort, it is possible to show that $\mathcal{L} = 0$ and $\mathcal{U} = 1$, but the key thing is that they are not equal.

In this previous example, the rational numbers are "small" in that they are a countable subset of an uncountable space, and thus we might expect that

$$\int_0^1 \mathbb{1}_{\mathbb{Q}} \, dx = 0.$$

We will later see that this is indeed the case if we use Lebesgue integral.

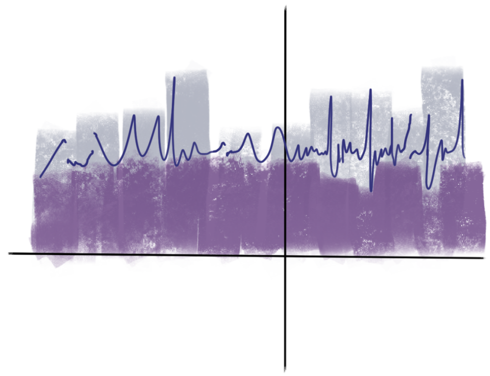


Figure 1.1: An upper and lower sum for a reasonably discontinuous function

1.1.1 Intuition

The idea of the Riemann integral is to slice a function vertically into narrow strips which can be approximated by rectangles. However, what this example shows is that for functions which are extremely discontinuous, it is not possible to determine the correct "height" to make the slices.

The idea of the Lebesgue integral is to instead cut vertically, which transforms the problem from one of determining how high to make the slices to determining the size of sets.

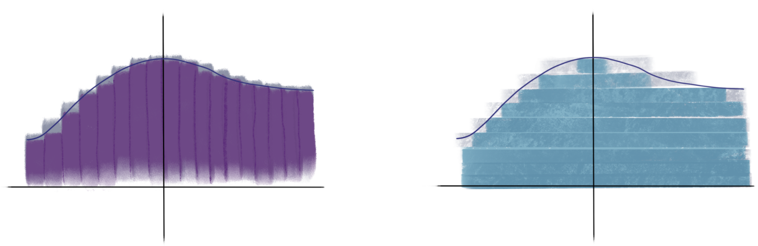


Figure 1.2: A Riemann sum (left) and an approximation by simple functions (right)

1.1.2 Why should we care?

At its core, measure theory studies the "size" of sets. In other words, given a set X, we want a way to determine how large it is. This turns out to be a fairly subtle question, and one with a long history.

There are many ways to discuss the size of a set. To give two basic examples, one could start by simply counting the elements of X (i.e., determining its cardinality)[2]. Another option for topological spaces is to determine the dimension of the set. However, our focus is in measurements which correspond to an integral of one form or another.

For some of you, the idea of extending the Riemann integral to integrate more functions might be interesting and natural. For others, functions which fail to be Riemann integrable are already quite pathological, so the idea of extending the integral might appear to be an esoteric pursuit.

However, the significance of the Lebesgue integral is that it is much more robust than the Riemann integral. Because of this, we will be able to prove strong convergence theorems, which play a central role in modern analysis, PDEs, probability, etc. To give an analogy, the mean and intermediate value theorem play a central role in analysis, but these results do not hold in $\mathbb{Q}$, but require its metric completion $\mathbb{R}$. In this context, the Riemann integral is somewhat akin to the rational numbers, and the Lebesgue integral is $\mathbb{R}$.

1.2 Resources

In this section, I have included some references which might be helpful for your studying.

1. Sheldon Axler wrote a textbook which is freely available online [Axl20]. It's a bit more elementary than our course, so might be a good place to start.

2. I highly recommend the textbook by Gerald Folland [Fol99]. This text considers measures in greater generality from the outset and covers the material in a different order than the other books though, so might be a good reference.

3. Terry Tao wrote a introduction to measure theory which is worth reading [Tao11]. It is also available freely online.

[2]For countable subsets, the cardinality can be phrased as an integral with respect to a so called "counting measure."

4. Gerald Teschl wrote a book on functional analysis which discusses a lot of topics that we cover in the course [Tes98]. It is freely available online.

Chapter 2

Measurable Sets and their Measures

In this chapter, we will build up measure theory with the goal of defining the "length" of a set of real numbers. Our purpose is to motivate the notion of the Lebesgue measure so we will not start with an abstract definition of a measure. Instead, we will start with some natural properties of a "length" and build from there.

There are a few properties of lengths which are "obvious." For instance, any notion of length should satisfy that the length of any finite interval $[a, b]$ is simply $a - b$. Similarly, the length of any unbounded interval should be infinite.

Going further, it is natural to request that the length of a finite (or countable) union of disjoint intervals is the sum of its lengths and that the length of a subset should be less than the length of its parent.

2.1 Outer measures

We now give our first attempt to define the "length" of a set of real numbers, which is the *outer measure,* denoted m^*.

This measure satisfies some of the properties that we desire. For instance, the outer measure of an interval is its length. The outer measure also has the advantage that it is very intuitive and well-defined for any subset of the real numbers. However, this flexibility comes at a cost and we will see that this attempt fails to have some natural properties we might expect for a notion of length.

For instance, given $A, B \subset \mathbb{R}$ with $A \cap B = \varnothing$, it may be that

$$m^*(A \cup B) < m^*(A) + m^*(B).$$

In order to define the outer measure, we start by defining a "length" function $\ell : \mathcal{P}(\mathbb{R}) \to \mathbb{R}$. Here, $\mathcal{P}(\mathbb{R})$ denotes the set of subsets of $\mathbb{R}$

1. If I is a bounded interval, then $\ell(I)$ denotes the usual length of I.
2. If I is an unbounded interval, then $\ell(I) = \infty$.
3. If $\{I_j\}_{j \in F}$ is a countable pairwise disjoint family of intervals[1], then

$$\ell\left(\bigcup_{j \in F} I_j\right) = \sum_{j \in F} \ell(I_j). \tag{2.1}$$

Before moving on, we make some remarks.

1. In Equation 2.1, all the terms in the right hand side of the equation are non-negative. Thus, if the series converges it does so absolutely and the order of summation does not matter.
2. If U is any non-empty open set of reals, there is a countable pairwise disjoint family of open intervals $\{I_j\}_{j \in F}$ so that $U = \bigcup_{j \in F} I_j$. Thus $\ell(U)$ is defined for every non-empty open sets of real numbers.
3. We shall refer to ℓ as the *length* of U

With this in mind, we now define the outer measure.

Definition 6. *Suppose E is a set of real numbers. The outer measure of E, denoted $m^*(E)$, is defined to be the largest real number that is less than or equal to the length of every open set including E. In other words,*

$$m^*(E) = \inf\left\{ \sum_{j \in F} l(I_j) \;\middle|\; E \subset \bigcup_{j \in F} I_j \right\}.$$

Note that this definition is very slightly different from the definition given in the book. However, this difference is not important.

[1]We will often use the notation F to denote a countable set. Such a set may be finite, which is why we are not using $\mathbb{N}$ as our index.

Example. *There are a few outer measures that can readily be computed.*

1. $$m^*(\varnothing) = 0$$

2. $$m^*(\mathbb{R} = \infty$$

3. *Let* $a \in \mathbb{R}$*,* $$m^*(\{a\}) = 0$$

Exercise 2. *Show that the outer measure of a countable set is zero.*

Proposition 1. *The outer measure of an interval is its length.*

Proof. Claim 1: $\ell(I) \leqslant m^*(I)$.

Suppose U is an open set with $I \subset U$. There is a countable pairwise disjoint family of open intervals $\{I_j\}_{j \in F}$ so the $U = \bigcup_{j \in F} I_j$. Since I is connected, there is a $j_0 \in F$ with $I \subset I_{j_0}$. Therefore,

$$\ell(I) \leqslant \ell(I_{j_0}) \leqslant \ell(U).$$

Taking infimums, we find that $\ell(I) \leqslant m^*(I)$.

Claim 2: $m^*(I) \leqslant \ell(I)$.

We break this part of the argument into two cases

1. I is unbounded. Then $\ell(I) = \infty$, and we are done.

2. If I is bounded, let a be the left end-point and b be the right end-point.

   We let $\epsilon > 0$ and take $U = (a - \epsilon, b + \epsilon)$. We see that $I \subset U$. Furthermore, we that that

   $$m^* I \leqslant \ell(U) = (b - a) + 2\epsilon = \ell(I) + 2\epsilon.$$

   Since epsilon was arbitrary, we see that $m^*(I) \leqslant \ell(I)$.

□

Let us use what we've seen to solve a question from the qualifying exams.

Exercise 3 (1997 Qualifying Exam, Problem 6)**.** *For every $\epsilon > 0$ and every subset E of the real numbers, define*

$$\mu_{\varepsilon}(E) = \inf \sum_{n=1}^{\infty} \left[\ell\left(I_n\right)\right]^{\frac{1}{2}}$$

where $\ell(\cdot)$ denotes the length of an interval and the infimum is taken over all countable collections of open intervals $\{I_1, I_2, \ldots\}$ which cover E for which $\ell\left(I_n\right) < \epsilon$ for all n. Prove the following:

1. *Show that μ_ϵ is an outer measure on the real line.*
2. *Show that*

$$\mu(E) := \sup_{\varepsilon > 0} \mu_{\varepsilon}(E) = \lim_{\varepsilon \to 0^+} \mu_{\varepsilon}(E)$$

3. *Show that $\mu(0,1) = +\infty$. (Hint: If one has intervals I_n of length $\leqslant \varepsilon$ which cover $(0,1)$, then $\left[\ell\left(I_n\right)\right]^{\frac{1}{2}} = \frac{\ell(I_n)}{[\ell(I_n)]^{\frac{1}{2}}} \geqslant \epsilon^{-\frac{1}{2}} \ell\left(I_n\right) \cdot \Big)$*

As a challenge (which was not part of the qualifying exam), try to find a set E for which $0 < \mu(E) < \infty$. This is an unfair task without any context, but it's a good thought experiment anyway.

2.1.1 Monotonicity and translation invariance

We now show that the outer measure has two natural and desired properties of length, namely that subsets cannot be bigger than the original set (i.e., the outer measure is monotonic) and that if you translate a set, its outer measure remains the same (i.e., the outer measure is *translation invariant*).

Proposition 2. *If $A \subset B \subset \mathbb{R}$, then $m^*(a) \leqslant m^*(B)$.*

Proof. Suppose $A \subset B \subset \mathbb{R}$. If U is an open set that contains B, then it also contains A. Thus,

$$\inf \left\{\ell(U) B \subset U\right\} \geqslant [\inf \left\{\ell(U) A \subset U\right\},$$

and so

$$m^*(B) \geqslant m^*(A)$$

□

Definition 7. *If $A \subset \mathbb{R}$ and if $a \in \mathbb{R}$, then*

$$a + A = \{a + x \mid x \in A\}.$$

Proposition 3. *Given $A \subset \mathbb{R}$ and $a \in \mathbb{R}$, $m^*(A) = m^*(a + A)$*

Proof. First note that if I is an interval then $\ell(I) = \ell(a + I)$.

We then note that if $\{I_j\}_{j \in F}$ is a pairwise disjoint family of intervals, $\{a + I_j\}_{j \in F}$ is as well and

$$a + \bigcup I_j = \bigcup (a + I_j).$$

If follows that if U is open $\ell(U) = \ell(a + U)$. If U is an open set that includes A, then $a + U$ includes $a + A$ and has the same length as U. Thus,

$$\inf \{\ell(U) | \ A \subset U \text{ and } U \text{ open } \} \geqslant \inf \{\ell(V) | \ a + A \subset U \text{ and } V \text{ open } \}$$

and so,

$$m^*(A) \geqslant m^*(a + A).$$

Note that $A = -a + (a + A)$, so by a symmetric argument we have that

$$m^*(a + A) \geqslant m^*(A).$$

□

Exercise 4. *If $A \subset \mathbb{R}$ and if $\lambda > 0$, then λA is defined to be the set*

$$\lambda A = \{\lambda x \mid x \in A\}.$$

Show that $m^*(\lambda A) = \lambda m^*(A)$.

Here is a more challenging version of the previous exercise (since we haven't defined outer measure in $\mathbb{R}^n$).

Exercise 5. *If $A \subset \mathbb{R}^n$ and if M is an $n \times n$ matrix, then MA is defined to be the set*

$$MA = \{Mx \mid x \in A\}.$$

What is $m^*(MA)$?

Here is some food for thought, which we will revisit later in the course.

Exercise 6. *If $A, B \subset \mathbb{R}$, then*

$$A + B = \{a + b \mid a \in A, b \in B\}.$$

What can we say about $m^*(A + B)$ in terms of $m^*(A)$ and $m^*(B)$? Are there sets with $m^*(A) = m^*(B) = 0$ but $m^*(A + B) \neq 0$?

2.1.2 Countable sub-additivity

We now establish one of the most important properties of length, which is that it is *countably sub-additive.*

Definition 8. *An measure is said to be countably sub-additive if*

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

for any countable collection of sets E_k (disjoint or otherwise).

If one of the E_k 's has infinite outer measure, the right hand side is infinite, and there is nothing to prove. We therefore suppose each of the E_k 's has finite outer measure. Let $\epsilon > 0$. For each natural number k, there is a countable collection $\{I_{k,i}\}_{i=1}^{\infty}$ of open, bounded intervals for which

$$E_k \subseteq \bigcup_{i=1}^{\infty} I_{k,i} \text{ and } \sum_{i=1}^{\infty} \ell\left(I_{k,i}\right) < m^*\left(E_k\right) + \epsilon/2^k.$$

Now, $\{I_{k,i}\}_{1\leqslant i,k<\infty}$ is a countable collection of open, bounded invervals that cover $\bigcup_{k=1}^{\infty} E_k$. Thus we have that

$$\begin{aligned} m^*\left(\bigcup_{k=1}^{\infty} E_k\right) \leqslant \sum_{1\leqslant k,i<\infty} \ell\left(I_{k,i}\right) &= \sum_{k=1}^{\infty}\left[\sum_{i=1}^{\infty} \ell\left(I_{k,i}\right)\right] \\ &< \sum_{k=1}^{\infty}\left[m^*\left(E_k\right) + \epsilon/2^k\right] \\ &= \left[\sum_{k=1}^{\infty} m^*\left(E_k\right)\right] + \epsilon \end{aligned}$$

Note that all of the terms are positive, so we can rearrange terms in the final equality. Since ϵ was arbitrary, we obtain the desired inequality.

Note also that if $\{E_k\}_{k=1}^{n}$ is a finite collection of sets, we can prove finite sub-additivity by taking $E_k = \varnothing$ for $k > n$.

2.2 The need for σ-algebras

The outer measure gives our first definition of length for arbitrary sets, and we have seen that it does a pretty good job. However, it has one major drawback, which is that it fails to be countably additive, or even finitely additive.

In other words, it is possible to find two sets A and B which are *disjoint* and satisfy

$$m^*(A \cup B) < m^*(A) + m^*(B).$$

It is not really possible to construct these sets by hand. In particular, it requires the Axiom of Choice. However, the fact that the counterexamples are so pathological suggests that to fix this problem, all we need to do is restrict our attention to sets which are somehow reasonable, which leads directly into the notion of measurability and σ-algebras.

Definition 9. *Let A be a set of reals. A is* measurable *if $m^*(X) = m^*(A \cap X) + m^*(A^c \cap X)$ for every $X \subseteq \mathbb{R}$.*

Here, we use the notation $X^c = \mathbb{R} - X$ whenever $X \subseteq \mathbb{R}$..

Remark: A is measurable if and only if $m^*(X) \geqslant m^*(A \cap X) + m^*(A^c \cap X)$ for every $X \subseteq \mathbb{R}$.

Example. *We can immediately find some examples of measurable sets.*

1. *If $m^*(A) = 0$, then A is measurable. To see this, suppose $m^*(A) = 0$. Since m^* is monotonic, $m^*(A \cap X) = m^*(A \cap X^c) = 0$ for every $X \subseteq \mathbb{R}$.*

   As a result, $\varnothing$, and every singleton are measurable.

2. *Since $(A^c)^c = A$, for any set A which is measurable, its complement A^c must also be measurable. Thus, $\mathbb{R}$ (the complement of the empty set) is measurable.*

After looking at the definition of measurable sets, we see that there is an algebraic structure to measurable sets, which we can make precise with the notion of a σ-algebra.

Definition 10. *Suppose $\mathcal{S} \subseteq \mathcal{P}(\mathbb{R})$. $\mathcal{S}$ is a* σ-algebra over $\mathbb{R}$ *if it meets the following criteria.*

1. *$\mathbb{R} \in \mathcal{S}$.*

2. *For every $X \in \mathcal{S}$, $X^c \in \mathcal{S}$. (Closure under complements)*

3. *For every sequence of sets $\{A_n\}_{n=0}^{\infty}$ so that $A_n \in \mathcal{S}$ for every $n \in \mathbb{N}$, $\bigcup_{n=0}^{\infty} A_n \in \mathcal{S}$. (Closure under countable unions)*

Suppose $\mathcal{S}$ is a σ-algebra over $\mathbb{R}$.

1. If $X, Y \in \mathcal{S}$, then $X \cup Y \in \mathcal{S}$. Proof: Set $X_0 = X$, $X_1 = Y$, $X_n = \varnothing$ when $n > 1$. Then, $X \cup Y = \bigcup_{n=0}^{\infty} X_n$.

2. If $X_n \in \mathcal{S}$ for each $n \in \mathbb{N}$, then $\bigcap_{n=0}^{\infty} X_n \in \mathcal{S}$.

   Proof: DeMorgan's laws.

Example. *There are a few simple examples of σ-algebras that we can write down immediately.*

1. $\mathcal{P}(\mathbb{R})$ is a σ-algebra over $\mathbb{R}$.

2. $\{\varnothing, \mathbb{R}\}$ is a σ-algebra over $\mathbb{R}$.

Our next goal now is to show that the set of measurable subsets of $\mathbb{R}$ is a σ-algebra. To do so, we will need several lemmas.

Lemma 1. *Suppose $E_1, \ldots, E_m$ are measurable subsets of $\mathbb{R}$. Then, $\bigcup_{j=1}^{m} E_m$ is measurable.*

Proof. We will use induction (on m). The base case, where $m = 1$ is trivial, so the hard work will be the inductive step.

Suppose $\bigcup_{j=1}^{m-1} E_j$ is a measurable set. Let $F_1 = \bigcup_{j=1}^{m-1} E_j$. Let $F_2 = E_m$. We must show that $F_1 \cup F_2$ is measurable. Let $X \subseteq \mathbb{R}$. We need to show

$$m^*(X) = m^*(X \cap (F_1 \cup F_2)) + m^*(X \cap (F_1 \cup F_2)^c).$$

What follows is a sequence of calculations by intersecting various different sets to get the desired equality.

Claim 1: $m^*(X) = m^*(X \cap F_1) + m^*(X \cap F_1^c \cap F_2) + m^*(X \cap (F_1 \cup F_2)^c)$.

Since F_1 is measurable, we have that

$$m^*(X) = m^*(X \cap F_1) + m^*(X \cap F_1^c).$$

Since F_2 is measurable,

$$\begin{aligned} m^*(X \cap F_1^c) &= m^*(X \cap F_1^c \cap F_2) + m^*(X \cap F_1^c \cap F_2^c) \\ &= m^*(X \cap F_1^c \cap F_2) + m^*(X \cap (F_1 \cup F_2)^c). \end{aligned}$$

So:

$$m^*(X) = m^*(X \cap F_1) + m^*(X \cap F_1^c \cap F_2) + m^*(X \cap (F_1 \cup F_c)^c)).$$

Claim 2: $m^*(X \cap F_1) + m^*(X \cap F_1^c \cap F_2) = m^*(X \cap (F_1 \cup F_2))$.

Since F_1 is measurable,

$$\begin{aligned} m^*((X \cap F_1) \cup (X \cap (F_1^c \cap F_2))) &= m^*(((X \cap F_1) \cup X \cap (F_1^c \cap F_2)) \cap F_1) \\ &\quad + m^*(((X \cap F_1) \cup (X \cap (F_1^c \cap F_2))) \cap F_1^c) \\ &= m^*(X \cap F_1) + m^*(X \cap (F_1^c \cap F_2)). \end{aligned}$$

But $m^*(X \cap F_1 \cup X \cap F_1^c \cap F_2) = m^*(X \cap (F_1 \cup F_2))$.

Therefore, $m^*(X) = m^*(X \cap (F_1 \cup F_2)) + m^*(X \cap (F_1 \cup F_2)^c)$. □

This lemma has the following important corollary.

Corollary 1. *If $X, Y \subseteq \mathbb{R}$ are measurable, then so are $X \cap Y$ and $X - Y$.*

The proof of this follows by applying De Morgan's laws, so I'll leave it as a small exercise.

Using the same technique as the previous lemma, it is also possible to prove the following lemma. Since the idea is exactly the same, we will not write down the proof.

Lemma 2. *Suppose $A, E_1, \dots, E_m \subseteq \mathbb{R}$. Suppose that $E_1, \dots, E_m$ are measurable and that $(E_1, \dots, E_m)$ is pairwise disjoint. Then,*

$$m^*(A \cap \bigcup_{j=1}^{m} E_j) = \sum_{j=1}^{m} m^*(A \cap E_j).$$

We will now prove another helpful lemma towards the goal of showing that measurable sets form a σ-algebra.

Lemma 3. *Suppose $\mathcal{S} \subseteq \mathcal{P}(\mathbb{R})$. Then, $\mathcal{S}$ is a σ-algebra over $\mathbb{R}$ if and only if it meets the following three criteria.*

1. *$\mathbb{R} \in \mathcal{S}$.*

2. *$X - Y \in \mathcal{S}$ whenever $X, Y \in \mathcal{S}$.*

3. *$\bigcup_{n=0}^{\infty} A_n \in \mathcal{S}$ whenever $\{A_n\}_{n=0}^{\infty}$ is a pairwise disjoint sequence of sets in $\mathcal{S}$.*

Proof. If $\mathcal{S}$ is a σ-algebra, then by definition and the previous corollary, $\mathcal{S}$ satisfies these three criteria.

Conversely, suppose $\mathcal{S}$ satisfies these three criteria. Then, $\mathbb{R} \in \mathcal{S}$, and (2) implies closure under complementation. It only remains to show that $\mathcal{S}$ is closed under countable unions.

Claim 1: $X \cup Y \in \mathcal{S}$ whenever $X, Y \in \mathcal{S}$.

To show this, suppose $X, Y \in \mathcal{S}$. Set $X_0 = X$, $X_1 = Y - X$, and $X_n = \varnothing$ when $n > 1$. It follows from (2) that $\varnothing \in \mathcal{S}$ and $Y - X \in \mathcal{S}$. By construction, $\{X_n\}_{n=0}^{\infty}$ is a pairwise disjoint sequence. So, by 3, $\bigcup_{n=0}^{\infty} X_n \in \mathcal{S}$. But, $X \cup Y = \bigcup_{n=0}^{\infty} X_n$.

Claim 2: $\mathcal{S}$ is closed under countable unions.

Let $\{A_n\}_{n=0}^{\infty}$ be a sequence of sets in $\mathcal{S}$. For each $n \in \mathbb{N}$, let

$$B_n = A_n - \bigcup_{j<n} A_j.$$

By Claim 1, $\bigcup_{j<n} A_j \in \mathcal{S}$ for each n. By (2), $B_n \in \mathcal{S}$ for each $n \in \mathbb{N}$. By construction, $\{B_n\}_{n=0}^{\infty}$ is pairwise disjoint. Thus, by (3), $\bigcup_{n=0}^{\infty} B_n \in \mathcal{S}$. However, $\bigcup_{n=0}^{\infty} A_n = \bigcup_{n=0}^{\infty} B_n$.

Thus, $\mathcal{S}$ is a σ-algebra. □

With this lemma in hand, we can finally show that the Lebesgue measurable sets form a σ-algebra.

Theorem 4. *The set of all measurable subsets of* $\mathbb{R}$ *is a* σ*-algebra over* $\mathbb{R}$.

Proof. We apply Lemma 3. We have already observed that $\mathbb{R}$ is measurable and that the set of measurable sets is closed under set differences.

What remains to show is that the set of measurable sets is closed under countable disjoint unions. Suppose $\{E_n\}_{n=0}^{\infty}$ is a pairwise disjoint sequence of measurable subsets of $\mathbb{R}$. Set $E = \bigcup_{n=0}^{\infty} E_n$. We need to show that E is measurable, i.e., for $X \subseteq \mathbb{R}$, we have that

$$m^*(X) = m^*(X \cap E) + m^*(X \cap E^c).$$

Claim 1: $m^*(X) \cap \geqslant \sum_{n=0}^{\infty} m^*(X \cap E_n) + m^*(X \cap E^c)$.

Let $m \in \mathbb{N}$. By Lemma 1, $\bigcup_{n=0}^{m} E_n$ is measurable. So,

$$m^*(X) = m^*(X \cap \bigcup_{n=0}^{m} E_n) + m^*(X \cap \left(\bigcup_{n=0}^{m} E_n\right)^c).$$

By Lemma 2,

$$m^*(X \cap \bigcup_{n=0}^{m} E_n) = \sum_{n=0}^{m} m^*(X \cap E_n).$$

At the same time, $(\bigcup_{n=0}^{m} E_n)^c \supseteq E^c$. So,

$$m^*(X \cap \left(\bigcup_{n=0}^{m} E_n\right)^c) \geqslant m^*(X \cap E^c).$$

Thus,

$$m^*(X) \geqslant \sum_{n=0}^{m} m^*(X \cap E_n) + m^*(X \cap E^c)$$

for every $m \in \mathbb{N}$. Thus, Claim 1 holds.

Claim 2: $m^*(X) \geqslant m^*(X \cap E) + m^*(X \cap E^c)$.

By Claim 1,

$$m^*(X) \geqslant \sum_{n=0}^{\infty} m^*(X \cap E_n) + m^*(X \cap E^c).$$

Since outer measure is countably subadditive,

$$\sum_{n=0}^{\infty} m^*(X \cap E_n) \geqslant m^*(\bigcup_{n=0}^{\infty} X \cap E_n).$$

This establishes Claim 2.

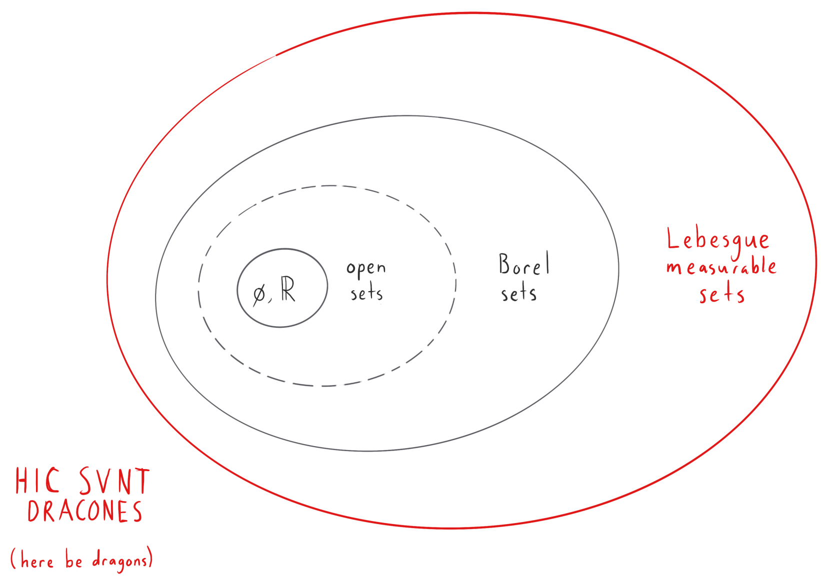
Thus, $m^*(X) = m^*(X \cap E) + m^*(X \cap E^c)$. □

Example. *We are able to use this idea to find many new examples of measurable sets.*

1. *Every countable set of reals is measurable.*
2. *The set of irrational numbers is measurable.*
3. *If A_n is a measurable set of reals for each $n \in \mathbb{N}$, then $\bigcap_{n=0}^{\infty} A_n$ is measurable.*

2.2.1 Borel sets

We now define a class of sets called Borel sets which are "always measurable" (for any reasonable measure). It is worth noting that there are measurable sets[2] which are not Borel. However, Borel sets are those which must be measurable whenever open intervals are measurable.



Definition 11. *Suppose $X \subseteq \mathbb{R}$. X is* Borel *if it belongs to every σ-algebra over $\mathbb{R}$ that contains all the open subsets of $\mathbb{R}$.*

Remarks:

1. Every open set is Borel.
2. If X is a Borel set of reals, then X^c is Borel.
3. If X_n is a Borel set of reals for each each $n \in \mathbb{N}$, then $\bigcup_{n=0}^{\infty} X_n$ and $\bigcap_{n=0}^{\infty} X_n$ are Borel.

Examples:

[2]Here, we mean measurable with respect to the Lebesgue measure.

1. If $x \in \mathbb{R}$, then $\{x\}$ is Borel. Proof: $\{x\} = ((-\infty, x) \cup (x, \infty))^c$.

2. Every countable set of reals is Borel.

3. The set of rational numbers is Borel, and the set of irrational numbers is Borel.

4. $(a, b]$ is Borel if a, b are real numbers such that $a < b$.

Intuitively, the Borel sets are those which you can obtain by taking countable unions, intersections and complements of open sets, all of which preserve measurability. It is possible to explore the structure of these sets in a much deeper way. In particular, it turns out there is a hierarchy of Borel sets depending on how complicated they are (See Chapter 22 of [Kec12] for a background on this). However, we will not need to discuss this topic in this class.

We now show that Borel sets are always measurable. This fact should not be surprising. However, there is still some work to do for this since we haven't yet shown that open intervals are measurable.

Theorem 5. *Every Borel set of reals is measurable.*

To prove this, we only must show that every open set is measurable. In fact, we need only show that if $a < b$, the open interval (a, b) is measurable (since every open set is the countable union of disjoint intervals). Going further, it suffices to show that if $a \in \mathbb{R}$, that the two sets

$$(-\infty, a) \text{ and } (a, \infty)$$

are measurable (Exercise: why?), so that's what we are going to do.

Lemma 4. *If $a \in \mathbb{R}$, the set $(-\infty, a)$ is measurable.*

Proof. Suppose $a \in \mathbb{R}$ and $X \subseteq \mathbb{R}$. We need to show that

$$m^*(X) \geqslant m^*(X \cap (-\infty, a)) + m^*(X \cap [a, \infty)).$$

Note that

$$m^*(X \cap [a, \infty)) \leqslant m^*(X \cap (a, \infty)) + m^*(X \cap \{a\}),$$

by sub-additivity but the latter term is zero, and thus

$$m^*(X \cap [a, \infty)) = m^*(X \cap (a, \infty)).$$

Now we show that

$$m^*(X) = \underbrace{m^*(X \cap (-\infty, a))}_{X_1} + \underbrace{m^*(X \cap (a, \infty))}_{X_2}$$

Since outer measure is monotonic, we can assume that both X_1 and X_2 are non-empty (or else our job is already done). Now we take $r > m^*(X)$. Then there is a non-empty set $U \subseteq \mathbb{R}$ such that $X \subset U$ and $r > \ell(U)$. Let

$$U_1 = U \cap (-\infty, a) \text{ and } U_2 = U \cap (a, \infty).$$

Then U_1 and U_2 are both non-empty and $\ell(U) = \ell(U_1) + \ell(U_2)$. (For the latter claim, see the solution to problem 6 on homework 1).

But $\ell(U_1) \geqslant m^*(X \cap (-\infty, a))$, by definition and $\ell(U_2) \geqslant m^*(X \cap (a, \infty))$, so $r \geqslant m^*(X \cap (-\infty, a)) + m^*(X \cap (a, \infty))$ for all $r > m^*(X)$.

Thus,

$$m^*(X) \geqslant m^*(X \cap (-\infty, a)) + m^*(X \cap (a, \infty)).$$

□

2.3 Approximating Measurable Sets

We now turn our attention to approximating measurable sets. These approximations will be essential when we start using measure theory to compute integrals. To do so, we define the notions of G_δ and F_σ sets, which are used as outer and inner approximations of measurable sets, respectively.

Definition 12. *Let $X \subseteq \mathbb{R}$.*

1. *X is G_δ if it is the intersection of a countable family of open sets.*

2. *X is F_σ if it is the union of a countable family of closed sets.*

Remarks:

1. If $X \subseteq \mathbb{R}$, then X is G_δ if and only if X^c is F_σ.

2. If X is a G_δ set of reals, then there is an increasing sequence of open sets of reals $\{U_n\}_{n=0}^{\infty}$ so that $X = \bigcap_{n=0}^{\infty} U_n$. If X is an F_σ set of reals, then there is a decreasing sequence of closed sets of reals $\{C_n\}_{n=0}^{\infty}$ so that $X = \bigcup_{n=0}^{\infty} C_n$.

Before discussing these sets in detail, let's see a few examples

Example. *1. The set of rational numbers is an F_σ set. Namely,*

$$\mathbb{Q} = \bigcup_{q \in \mathbb{Q}} \{q\}.$$

2. Therefore the set of irrational numbers is a G_δ set.

3. If a, b are real numbers so that $a < b$, then $(a, b]$ is a G_δ set. Proof:

$$(a, b] = \bigcap_{n \in \mathbb{N}} (a, b + 2^{-n}).$$

4. Every open interval is both G_δ and F_σ. Proof: every open set is G_δ. If I is an open interval, then

$$I = \bigcup_{a,b \in I \cap \mathbb{Q}} [a, b].$$

However, not every set will be G_δ. For instance, $\mathbb{Q}$ is not G_δ.

Proof. Suppose that it were. In other words, suppose $\mathbb{Q} = \bigcap_{n=0}^{\infty} U_n$ where U_n is an open set of reals for each $n \in \mathbb{N}$. Let $\{q_n\}_{n=0}^{\infty}$ be an enumeration of the rational numbers. Thus, $U_n - \{q_n\}$ is open and dense for each $n \in \mathbb{N}$. So, by the Baire Category Theorem $\bigcap_{n=0}^{\infty}(U_n - \{q_n\})$ is dense. But, this intersection is empty- a contradiction. □

In order to use G_δ and F_σ sets to approximate measurable sets, we must first establish the following *excision property.*

Lemma 5. *If X is a measurable set of reals that has finite outer measure, and if Y is a set of reals that includes X, then*

$$m^*(Y - X) = m^*(Y) - m^*(X).$$

Proof. Suppose X is a measurable set of reals and $m^*(X) < \infty$. Suppose $\mathbb{R} \supseteq Y \supseteq X$. Since X is measurable,

$$m^*(Y) = m^*(Y \cap X) + m^*(Y \cap X^c) = m^*(X) + m^*(Y - X).$$

Thus, since $m^*(X) < \infty$, $m^*(Y) - m^*(X) = m^*(Y - X)$. □

Using this lemma, we are finally able to prove the simple fact that the outer measure of an open set is its length.

Lemma 6. *If U is a non-empty open set of reals, then $m^*(U) = \ell(U)$.*

Proof. Suppose U is a non-empty open set of reals.
Let

$$S = \{\ell(V) \ : \ \mathbb{R} \supseteq V \supseteq U \mid V \neq \varnothing \text{ is open}\}.$$

Thus, $m^*(U) = \inf(S)$. Since $\ell(U) \in S$, $m^*(U) \leqslant \ell(U)$. If V is a non-empty open set of reals that includes U, then $\ell(V) \geqslant \ell(U)$ (See solution to problem 6 homework 1). Thus, $m^*(U) \geqslant \ell(U)$. □

We can now describe the relationship between measurable sets and their approximations.

Theorem 6. *Let $X \subseteq \mathbb{R}$. Then, the following are equivalent.*

1. *X is measurable.*
2. *For every $\epsilon > 0$, there is an open $U \subseteq \mathbb{R}$ so that $X \subseteq U$ and $m^* * (U - X) < \epsilon$.*
3. *There is a G_δ set of reals V so that $V \supseteq X$ and so that $m^*(V - X) = 0$.*
4. *For every $\epsilon > 0$, there is a closed $C \subseteq \mathbb{R}$ so that $X \supseteq C$ and so that $m^*(X - C) < \epsilon$.*
5. *There is an F_σ set of reals V so that $V \subseteq X$ and $m^*(X - V) = 0$.*

Proving this is somewhat tedious, but it's not too hard.

Proof. Proof:
(1) ⇒ (2): Suppose X is measurable. Let $\epsilon > 0$.

Case 1: $m^*(X) < \infty$.

By the definition of outer measure, there is a non-empty open set of reals U so that $\ell(U) < m^*(X) + \epsilon$ and so that $U \supseteq X$. Since $m^*(X) < \infty$, $\ell(U) - m^*(X) < \epsilon$. By Lemma 6, $\ell(U) = m^*(U)$. Since X is measurable, by Lemma 5, $m^*(U - X) = m^*(U) - m^*(X)$. So, $m^*(U - X) < \epsilon$.

Case 2: $m^*(X) = \infty$.

For each $n \in \mathbb{N}$, let $X_n = X \cap (-n, n)$. Thus, $X = \bigcup_n X_n$. For each $n \in \mathbb{N}$, $m^*(X_n) \leqslant m^*((-n, n)) < \infty$ since outer measure is monotonic. By Case 1, for each $n \in \mathbb{N}$ there is an open set of reals $U_n \supseteq X_n$ so that

$m^*(U_n - X_n) < \epsilon 2^{-(n+1)}$. Set $U = \bigcup_n U_n$. Then, $U - X \subseteq \bigcup_n (U_n - X_n)$. So,

$$m^*(U - X) \leqslant m^*(\bigcup_n U_n - X_n) < \sum_{n=0}^{\infty} \epsilon 2^{-(n+1)} = \epsilon.$$

(2) $\Rightarrow$ (3): Suppose (2). Then, for each $n \in \mathbb{N}$, there is an open set of reals U_n so that $U_n \supseteq X$ and so that $m^*(U_n - X) < 2^{-n}$. Set $V = \bigcap_n U_n$. So, for each $n \in \mathbb{N}$,

$$m^*(V - X) \leqslant m^*(U_n - X) < 2^{-n}.$$

Thus, $m^*(V - X) = 0$.

(3) $\Rightarrow$ (1): Suppose there is a G_δ set V so that $V \supseteq X$ and $m^*(V - X) = 0$. Thus, V is Borel and so V is measurable. Since $m^*(V - X) = 0$, $V - X$ is measurable. Since $(V - X)^c = V^c \cup X$, $X = V \cap (V - X)^c$. Thus, X is measurable.

The remaining implications are now proved by taking complements. $\square$

We can now show that measurable sets are open intervals, modulo a set of small measure.

Theorem 7. *Suppose X is a measurable set of reals whose outer measure is finite. Then, for every $\epsilon > 0$, there are open intervals $I_0, \ldots, I_k$ so that $m^*(X \Delta \bigcup_{n=0}^k I_n) < \epsilon$ and so that $(I_0, \ldots, I_k)$ is pairwise disjoint.*

Proof. Let $\epsilon > 0$. By Theorem 6, there is an open set of reals U so that $U \supseteq X$ and so that $m^*(U - X) < \epsilon/2$. There is a countable and pairwise disjoint family of open intervals $\{I_j\}_{j \in F}$ so that $U = \bigcup_{j \in F} I_j$.

Claim 1: $\ell(U) < \infty$.

Proof Claim 1: Since $m^*(U) \leqslant m^*(U - X) + m^*(X) < \epsilon/2 + m^*(X)$ and $m^*(X) < \infty$, $m^*(U) < \infty$. By Lemma 6, $m^*(U) = \ell(U)$.

Claim 2: There is a finite $F_1 \subseteq F$ so that $\sum_{j \in F - F_1} m^*(I_j) < \epsilon/2$.

Proof of Claim 2: Since $\sum_{j \in F} \ell(I_j) = \ell(U) < \infty$.

Claim 3: $m^*(X \Delta \bigcup_{j \in F_1} I_j) < \epsilon$.

Proof of Claim 3: By monotonicity of outer measure and Claim 2:

$$\begin{aligned} m^*(\bigcup_{j\in F_1} I_j - X) &\leqslant m^*(U - X) < \epsilon/2 \\ m^*(X - \bigcup_{j\in F_1} I_j) &\leqslant m^*(U - \bigcup_{j\in F_1} I_j) < \epsilon/2 \end{aligned}$$

By the countable subadditivity of m^*

$$m^*(X \Delta \bigcup_{j\in F_1} I_j) < \epsilon.$$

□

2.3.0.1 The Vitali covering lemma

Before defining the Lebesgue measure, let us also provide the Vitali covering lemma, which will be quite useful later in the course.

Definition 13. *An interval is* degenerate *if it consists of a single point.*

Definition 14. *Suppose $E \subseteq \mathbb{R}$. Suppose $\mathcal{F}$ is a set of intervals that are closed, bounded, and non-degenerate. $\mathcal{F}$ is a* Vitali covering of E *if for every $x \in E$ and every $\epsilon > 0$ there is an interval $I \in \mathcal{F}$ so that $x \in I$ and $\ell(I) < \epsilon$.*

Here are some examples and non-examples of Vitali coverings.

1. $\{[-1, 2]\}$ is a covering of $[0, 1]$ but not a Vitali covering.
2. $\{[x - 2^{-n}, x + 2^{-n}] \ : \ x \in [0, 1] \text{ and } n \in \mathbb{N}\}$ is a Vitali covering of $[0, 1]$.

Lemma 7 (Vitali Covering Lemma)**.** *Suppose E is a set of reals and $m^*(E) < \infty$. Suppose $\mathcal{F}$ is a Vitali covering of E. Then, for every $\epsilon > 0$, there exist $I_0, \ldots, I_n \in \mathcal{F}$ so that $(I_0, \ldots, I_n)$ is pairwise disjoint and so that*

$$m^*(E - \bigcup_{k=0}^{n} I_j) < \epsilon.$$

Proof. Case 1: There exist $I_0, \ldots, I_n \in \mathcal{F}$ so that $E \subseteq \bigcup_{j=0}^n I_j$ and so that $(I_0, \ldots, I_n)$ is pairwise disjoint.

In this case, there is nothing to prove and we are done.

Case 2: There do not exist $I_0, \ldots, I_n \in \mathcal{F}$ so that $E \subseteq \bigcup_{j=0}^{n} I_j$ and so that $(I_0, \ldots, I_n)$ is pairwise disjoint.

By the definition of outer measure, there is an open set $U \supseteq E$ so that $\ell(U) < \infty$. Let $\mathcal{F}' = \{I \in \mathcal{F} \ : \ I \subseteq U\}$.

Claim 1: $\mathcal{F}'$ is a Vitali covering of E.

Proof: Exercise.

Claim 2: If $I_0, \ldots, I_n \in \mathcal{F}'$, and if $(I_0, \ldots, I_n)$ is pairwise disjoint, then there exists $I \in \mathcal{F}'$ so that $I \cap \bigcup_{j=0}^{n} I_j = \varnothing$.

Proof: Exercise.

We inductively define a sequence of intervals in $\mathcal{F}'$ as follows. Let I_0 be any interval in $\mathcal{F}'$. Assume $I_0, \ldots, I_n$ have been defined. Let

$$\mathcal{F}_n = \{I \in \mathcal{F}' \ : \ I \cap \bigcup_{j=0}^{n} I_j = \varnothing\}.$$

By Claim 2, $\mathcal{F}_n$ is nonempty. Let $s_n = \sup\{\ell(I) \ : \ I \in \mathcal{F}_n\}$. Since $I \subseteq U$ for each $I \in \mathcal{F}'$, $s_n < \infty$. So, there exists $I \in \mathcal{F}_n$ such that $\ell(I) > s_n/2$; let I_{n+1} denote such an interval.

Note that $\{I_n\}_{n=0}^{\infty}$ is pairwise disjoint.

Claim 3: $\sum_{n=0}^{\infty} \ell(I_n) < \infty$.

Proof Claim 3: $\mu(\bigcup_{n=0}^{\infty} I_n) \leqslant \mu(U) < \infty$. By countable additivity, $\mu(\bigcup_{n=0}^{\infty} I_n) = \sum_{n=0}^{\infty} \ell(I_n)$.

Thus, $\lim_{n\to\infty} \ell(I_n) = 0$.

For each $n \in \mathbb{N}$, let c_n = center of I_n and let r_n = radius I_n. i.e. $I_n = [c_n - r_n, c_n + r_n]$. Let $J_n = [c_n - 5r_n, c_n + 5r_n]$. Thus, $\ell(J_n) = 5\ell(I_n)$.

Claim 4: For every $n \in \mathbb{N}$, $E - \bigcup_{k=0}^{n} I_k \subseteq \bigcup_{k=n+1}^{\infty} J_k$.

Proof Claim 4: Let $x \in E - \bigcup_{k=0}^{n} I_k$. There is a $\delta > 0$ so that $(x-\delta, x+\delta) \cap \bigcup_{k=0}^{n} I_k = \varnothing$. There is an $I \in \mathcal{F}'$ so that $x \in I \subseteq (x-\delta, x+\delta)$.

Thus, $I \in \mathcal{F}_n$.

Subclaim 4.1: There exists n_0 so that $I \cap I_{n_0} \neq \varnothing$.

Proof of Subclaim 4.1: Suppose not. Then, $I \in \mathcal{F}_{n_0}$ for all n_0. Thus, $\ell(I_{n+1}) > \ell(I)/2 > 0$ for all n. Thus, $\lim_{n\to\infty} \ell(I_n) \neq 0$- a contradiction.

Choose the least such n_0. Therefore, $n_0 > n$. Also, $I \in \mathcal{F}_{n_0-1}$, so $\ell(I_{n_0}) > \ell(I)/2$.

Subclaim 4.2: $x \in J_{n_0}$.

Proof of Subclaim 4.2: By Subclaim 4.1, there is a point $x_0 \in I \cap I_{n_0}$. Then,

$$|x - x_0| \leqslant \ell(I) < 2\ell(I_{n_0})$$

and thus

$$x_0 - x_{n_0} \leqslant \ell(I_{n_0})/2.$$

So, $|x - x_{n_0}| \leqslant 4r_{n_0} + r_{n_0} = 5r_{n_0}$, which implies $x \in J_{n_0}$.

This proves Claim 4.

Now, let $\epsilon > 0$. There exists n so that

$$\sum_{k=n+1}^{\infty} \ell(I_k) < \epsilon/5.$$

Since $\mu(J_k) = 5\mu(I_k)$, we find that

$$\mu(\bigcup_{n=k+1}^{\infty} J_n) < \epsilon.$$

So, by Claim 4,

$$\mu(E - \bigcup_{k=0}^{n} I_k) < \epsilon.$$

□

2.4 The Lebesgue Measure

With all these preliminaries taken care of, we are finally able to define the notion of the Lebesgue measure.

Definition 15. *If E is a measurable set of reals, then the outer measure of E is also called the* measure *of E and is denoted $\mu(E)$.*

From our previous examples, there are a few sets which we can immediately measure.

Example. *1. The Lebesgue measure of an interval is its length.*

2. $\mu(\mathbb{Q}) = 0$.

Before moving on to proving some properties of the Lebesgue measure, let's solve a problem from a past qualifying exam.

Exercise 7 (2000 Qual Problem 1)**.** *1. Define the measure μ on the Borel-σ-algebra of $\mathbf{R}$ by $\mu(A) = \lambda(A \cap (0,1))$ for Borel sets A. (λ = Lebesgue measure) Let*

$$K = \bigcap \{A : A \text{ is a closed set such that } \mu(A) = 1\}$$

and

$$D = \bigcap \{G : G \text{ is an open set such that } \mu(G) = 1\}$$

Determine precisely which points belong to K and D and prove your claim.

2.4.1 Countable Additivity

The Lebesgue measure has the advantage of being countable additive, which was the entire reason that we needed to discuss measurable sets.

Theorem 8 (Countable Additivity)**.** *: If $\{E_n\}_{n=0}^{\infty}$ is a pairwise disjoint sequence of measurable sets of reals, then $\mu(\bigcup_{n=0}^{\infty} E_n) = \sum_{n=0}^{\infty} \mu(E_n)$.*

Proof. Suppose $\{E_n\}_{n=0}^{\infty}$ is a pairwise disjoint sequence of measurable sets of reals.

Claim 1: $\mu(\bigcup_{n=0}^{\infty} E_n) \leqslant \sum_{n=0}^{\infty} \mu(E_n)$.

This holds since outer measure is countably subadditive.

Claim 2: For each $m \in \mathbb{N}$, $\mu(\bigcup_{n=0}^{\infty} E_n) \geqslant \sum_{n=0}^{m} \mu(E_n)$.

To establish this, we use the finite additivity. Let $m \in \mathbb{N}$. Set $A = \bigcup_{n=0}^{m} E_n$. By Lemma 2,

$$\mu(A \cap \bigcup_{n=0}^{m} E_n) = \sum_{n=0}^{m} \mu(A \cap E_n).$$

But, $A \cap \bigcup_{n=0}^{m} E_n = \bigcup_{n=0}^{m} E_n$ and $A \cap E_n = E_n$. Since outer measure is monotonic,

$$\mu(\bigcup_{n=0}^{m} E_n) \leqslant \mu(\bigcup_{n=0}^{\infty} E_n).$$

Claim 3: $\mu(\bigcup_{n=0}^{\infty} E_n) \geqslant \sum_{n=0}^{\infty} \mu(E_n)$.

To see this, apply Claim 2 and take limits.

□

As a corollary, we immediately obtain finite additivity as well.

Corollary 2. *If $E_0, \ldots, E_m$ are measurable sets of reals, and if $(E_0, \ldots, E_m)$ is pairwise disjoint, then $\mu(\bigcup_{n=0}^{m} E_n) = \sum_{n=0}^{m} \mu(E_n)$.*

We also obtain the measure of the difference of sets is the difference of measures.

Corollary 3. *Suppose X, Y are measurable sets of reals so that $Y \supseteq X$ and so that $\mu(X) < \infty$. Then, $\mu(Y - X) = \mu(Y) - \mu(X)$.*

Proof. By Corollary 2, $\mu(Y) = \mu(Y - X) + \mu(X)$. Since $\mu(X) < \infty$, $\mu(Y) - \mu(X) = \mu(Y - X)$. □

2.4.2 Continuity of measure

The Lebesgue measure has another desirable property, which is that it is continuous with respect to increasing and decreasing sequences of sets.

Theorem 9 (Continuity of measure)**.** *Suppose $\{E_n\}_{n=0}^{\infty}$ is a sequence of measurable sets of reals.*

1. *If $E_n \subseteq E_{n+1}$ for all $n \in \mathbb{N}$, then $\mu(\bigcup_{n=0}^{\infty} E_n) = \lim_{m\to\infty} E_m$.*

2. *If* $E_n \supseteq E_{n+1}$ *for all* $n \in \mathbb{N}$*, and if* $\mu(E_0) < \infty$*, then* $\mu(\bigcap_{n=0}^{\infty} E_n) = \lim_{m\to\infty} \mu(E_m)$*.*

Proof. (1): Suppose $E_n \subseteq E_{n+1}$ for all $n \in \mathbb{N}$.

Case 1: $\mu(E_m) = \infty$ for some m.

By monotonicity of outer measure, $\mu(\bigcup_{n=0}^{\infty} E_n) = \infty$. By monotonicity of outer measure, $\mu(E_n) = \infty$ for all $n > m$. Thus, $\lim_{n\to\infty} \mu(E_n) = \infty$.

Case 2: $\mu(E_m) < \infty$ for all m.

We have $\bigcup_{n=0}^{\infty} E_n = E_0 \cup \bigcup_{n=0}^{\infty} \mu(E_{n+1} - E_n)$. By Theorem 8,

$$\mu(E_0 \cup \bigcup_{n=0}^{\infty} \mu(E_{n+1} - E_n)) = \mu(E_0) + \sum_{n=0}^{\infty} \mu(E_{n+1} - E_n).$$

By Corollary 3, $\mu(E_{n+1} - E_n) = \mu(E_{n+1}) - \mu(E_n)$ for all n. Thus,

$$\sum_{n=0}^{\infty} \mu(E_{n+1} - E_n) = \sum_{n=0}^{\infty} \mu(E_{n+1}) - \mu(E_n) = \lim_{n\to\infty} \mu(E_n) - \mu(E_0).$$

Thus, $\mu(\bigcup_{n=0}^{\infty} E_n) = \lim_{n\to\infty} \mu(E_n)$.

(2): Suppose $E_n \supseteq E_{n+1}$ for all $n \in \mathbb{N}$ and $\mu(E_0) < \infty$. Thus, $\mu(E_n) < \infty$ for all $n \in \mathbb{N}$ and $\mu(\bigcap_{n=0}^{\infty} E_n) < \infty$. Also, $E_0 - E_n \subseteq E_0 - E_{n+1}$ for all $n \in \mathbb{N}$.

Then, using the previous argument (Part 1),

$$\mu(\bigcup_{n=0}^{\infty} E_0 - E_n) = \lim_{n\to\infty} \mu(E_0 - E_n).$$

However, $\bigcup_{n=0}^{\infty} E_0 - E_n = E_0 - \bigcap_{n=0}^{\infty} E_n$. By Corollary 3,

$$\mu(E_0 - \bigcap_{n=0}^{\infty} E_n) = \mu(E_0) - \mu(\bigcap_{n=0}^{\infty} E_n) \text{ and } \mu(E_0 - E_n) = \mu(E_0) - \mu(E_n).$$

As a result, we find that

$$\mu(\bigcap_{n=0}^{\infty} E_n) = \lim_{n\to\infty} \mu(E_n).$$

□

2.4.3 Everywhere and Almost Everywhere

Measure theory is a powerful tool for measuring the size of sets. However, as we have seen it is possible for a non-empty set to have non-zero measure and the Lebesgue measure is unable to distinguish these sets from the empty set. As such, for the rest of the course, we will often talk about properties that hold *almost everywhere*, as opposed to those that hold everywhere.

Definition 16. *We say that a property holds almost everywhere in a set E if the set of all $x \in E$ so that the property* fails *has measure* 0*. The set of all x where the property fails is called an* exceptional set*.*

Example. *Almost all real numbers are irrational.*

The following lemma is extremely important.

Lemma 8 (Borel-Cantelli)**.** *Suppose $\{A_n\}_{n=0}^{\infty}$ is a sequence of measurable sets of reals so that $\sum_{n=0}^{\infty} \mu(A_n) < \infty$. Then, almost every real number belongs to only finitely many A_n's. i.e. the measure of*

$$\{x \in \mathbb{R} \;:\; x \in A_n \textit{ for infinitely many } n\}$$

is zero.

Proof. Let

$$E = \{x \in \mathbb{R} \;:\; x \in A_n \text{ for infinitely many } n\}.$$

We need to show that $\mu(E) = 0$. Note that $x \in E$ if and only if for every $n \in \mathbb{N}$ there exists $m \geqslant n$ so that $x \in A_m$. So:

$$E = \bigcap_{n=0}^{\infty} \bigcup_{m=n}^{\infty} A_m$$

Note that $\bigcup_{m=n} A_m \supseteq \bigcup_{m=n+1} A_m$. Since $\sum_{n=0}^{\infty} \mu(A_n) < \infty$, $\mu(\bigcup_{n=0}^{\infty} A_n) < \infty$. So by Theorem 2,

$$\mu(E) = \lim_{n\to\infty} \mu(\bigcup_{m=n}^{\infty} A_m).$$

However,

$$\mu(\bigcup_{m=n}^{\infty} A_m) \leqslant \sum_{m=n}^{\infty} \mu(A_m).$$

Since $\sum_{m=0}^{\infty} \mu(A_m) < \infty$,

$$\lim_{n\to\infty} \sum_{m=n}^{\infty} \mu(A_m) = 0.$$

Thus, $\mu(E) = 0$. □

2.5 Non-Measurable Sets

Most sets that we encounter "in the wild" are Borel sets, and thus measurable. However, there is certainly some sampling bias going on here, because in some sense, "most" sets are non-measurable[3].

Constructing non-measurable sets requires some work, so we take a brief moment to review equivalence relations. Let us first introduce some notation.

Definition 17. *Suppose X is a set, and suppose $\sim$ is an equivalence relation on X.*

1. *For each $c \in X$, let $[c]_\sim$ = the $\sim$-equivalence class of c.*
2. *$X/\sim$ denotes the set of all $\sim$-equivalence classes.*

Definition 18. *Suppose X is a set, and suppose $\sim$ is an equivalence relation on X. A* system of representatives *for $\sim$ is a set $R \subseteq X$ so that each $\sim$-equivalence class contains exactly one element of R.*

Example. *Let n be a positive integer. Let $\sim$ denote the equivalence modulo n relation on $\mathbb{Z}$. Then $\{0, \ldots, n-1\}$ is a system of representatives for $\sim$. $\{1, \ldots, n\}$ is also a system of representatives for $\sim$.*

The relationship between equivalence classes and system of representatives quickly leads down the rabbit hole of foundational issues.

Question 1. *Does every equivalence relation have a system of representatives?*

Definition 19 (Axiom of Choice)**.** *If X is a nonempty set, then there is a function $F : \mathcal{P}(X) \to X$ so that $F(Y) \in Y$ for each $Y \in \mathcal{P}(X)$.*

With this axiom, we can find a system of representatives for any equivalent relation.

Corollary 4. *Assuming the Axiom of Choice, every equivalence relation has a system of representatives.*

Proof. Suppose X is a set and $\sim$ is an equivalence relation on X. If $X = \emptyset$, then $\emptyset$ is a set of representatives for $\sim$. Let $F : \mathcal{P}(X) \to X$ be a function so that $F(Y) \in Y$ for each $Y \in \mathcal{P}(X)$. Let $R = F[X/\sim]$. We claim that

[3]At least if we assume the axiom of choice.

R is a system of representatives for $\sim$. For, let $\mathbf{a}$ be a $\sim$-equivalence class. Then, $F(\mathbf{a}) \in \mathbf{a}$. Suppose $c \in R$ and $c \neq F(\mathbf{a})$. There exists $\mathbf{b} \in F/\sim$ so that $F(\mathbf{b}) = c$. Thus, $\mathbf{b} \neq \mathbf{a}$. Therefore, $\mathbf{b} \cap \mathbf{a} = \varnothing$. Since $c \in \mathbf{b}$, $c \notin \mathbf{a}$. So, $\mathbf{a}$ contains exactly one element of R. □

The reason to consider equivalence relationships and systems of representatives is that it allows us to construct non-measurable sets.

Definition 20. *When $x, y \in \mathbb{R}$, write $x \sim_{\mathbb{Q}} y$ if $y - x \in \mathbb{Q}$.*

Note that $\sim_{\mathbb{Q}}$ is an equivalence relation on $\mathbb{R}$.

Proposition 10. *Suppose R is a system of representatives for $\sim_{\mathbb{Q}}$. Then, $q + R \cap q' + R = \varnothing$ whenever q, q' are rational numbers so that $q \neq q'$.*

Proof. Suppose $q, q' \in \mathbb{Q}$ and that $q + R \cap q' + R \neq \varnothing$. Then, there exist $c, c' \in R$ so that $q + c = q' + c'$. Thus, $q - q' = c' - c$. Thus, $c \sim_{\mathbb{Q}} c'$. Since R is a system of representatives for $\sim_{\mathbb{Q}}$, $c = c'$. Thus, $q = q'$. □

Theorem 11. *Every system of representatives for $\sim_{\mathbb{Q}}$ is non-measurable.*

Proof. Let R be a set of representatives for $\sim_{\mathbb{Q}}$. By way of contradiction, suppose R is measurable. Let $S_n = R \cap [-n, n]$. Then, S_n is measurable for each $n \in \mathbb{N}$. Also,

$$\mathbb{R} = \bigcup_{\lambda \in \mathbb{Q}} \lambda + R = \bigcup_{\lambda \in \mathbb{Q}} \bigcup_{n=0}^{\infty} \lambda + S_n = \bigcup_{m=0}^{\infty} \bigcup_{\lambda \in \mathbb{Q} \cap [-m,m]} \bigcup_{n=0}^{\infty} \lambda + S_n.$$

We start by showing that if S_n was measurable, then $\mu(S_n) = 0$ for every $n \in \mathbb{N}$.

Proof of Claim: Let $n \in \mathbb{N}$. Then, $\bigcup_{\lambda \in \mathbb{Q} \cap [-1,1]} \lambda + S_n \subseteq [-(n+1), n+1]$. So,

$$\mu(\bigcup_{\lambda \in \mathbb{Q} \cap [-1,1]} \lambda + S_n \subseteq [-(n+1), n+1]) < \infty$$

Thus, by Proposition 10,

$$\sum_{\lambda \in \mathbb{Q} \cap [-1,1]} \mu(\lambda + S_n) < \infty$$

Since outer measure is translation-invariant:

$$\sum_{\lambda \in \mathbb{Q} \cap [-1,1]} \mu(S_n) = 0$$

Therefore, $\mu(S_n) = 0$.

From this, we see that $\mu(\mathbb{R}) = 0$, which is a contradiction. $\square$

As a result, if we assume the axiom of choice, we find that non-measurable sets exist.

Corollary 5. *There is a non-measurable set of reals.*

Proof. By Corollary 4, there is a system of representatives for $\sim_{\mathbb{Q}}$. By Theorem 11, such a system is non-measurable. $\square$

Using this, we can finally go back to show why the outer measure fails finite additivity

Corollary 6. *There exist $A, B \subseteq \mathbb{R}$ so that $A \cap B = \varnothing$ and $m^*(A \cup B) < m^*(A) + m^*(B)$.*

Proof. Let E be a non-measurable set of reals. Then, there is a set of reals X so that $m^*(X) < m^*(E \cap X) + m^*(E^c \cap X)$. $\square$

Proposition 12 (Vitali)**.** *Every set with positive outer measure includes a non-measurable set.*

Proof. Exercise. $\square$

2.5.0.1 Some remarks on non-measurable sets

The existence of measurable sets is an extremely thorny issue which is deeply tied to foundational issues. In fact, Robert Solovay constructed a model (in the sense of mathematical logic) of the real numbers where all subsets are Lebesgue measurable [Sol70]. For this, he used the Zermelo Frenkel axioms with an extra axiom on existence of an inaccessible cardinal. The technical details are outside the scope of this course, but heuristically what this shows is that the existence of non-measurable sets is very nearly *equivalent* to the axiom of choice.

Historically, this phenomena, along with a higher-dimensional version called the Banach-Tarski paradox[4] lead many mathematicians to reject the axiom of choice altogether.

[4]The Banach-Tarski paradox decomposes a solid ball into 5 pieces and then rotates and translates the pieces into two solid balls of the same radius. This construction heavily relies on the axiom of choice and non-measurable sets.

For modern functional analysis, trying to avoid the axiom of choice or the Banach-Tarski paradox is extremely restrictive, because you have to discard the Hahn-Banach theorem [FW91], which is one of the most important results in partial differential equations.

My advice is to not worry too much about the axiom of choice unless you are interested in mathematical logic. In my experience, non-measurable sets and other foundational matters don't show up unless you go looking for them, so it is perfectly possible (and acceptable) for a mathematician to consider such issues only in a superficial way.

2.6 The Cantor Set

We have seen that all countable sets have zero measure and that all Borel sets are measurable. It is natural to ask whether the converse of these results hold. In both cases, the answer is no, and we can see this by studying the Cantor set.

There are several ways to define this set, so we will provide two.

Definition 21 (Cantor set - ternary expansion definition)**.** *The* Cantor middle third set *consists of all numbers in* $[0,1]$ *that have a ternary expansion consisting of* 0*'s and* 2*'s. That is,* x *belongs to the Cantor middle third set if and only if there is a sequence* $\{a_n\}_{n=0}^{\infty}$ *so that* $x = \sum_{n=0}^{\infty} \frac{a_n}{3^{n+1}}$ *and* $a_n \in \{0,2\}$ *for all* $n \in \mathbb{N}$*. We denote this set by* $\mathbf{C}$*.*

Definition 22 (Cantor set - intersection definition)**.** *Consider the closed interval* $I = [0,1]$ *and remove the open set* $(\frac{1}{3}, \frac{2}{3})$*. Set*

$$C_1 = I - \left(\frac{1}{3}, \frac{2}{3}\right) = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right].$$

Repeat the process of removing the middle third from each set to obtain

$$C_2 = C_1 - \left((\frac{1}{9}, \frac{2}{9}) \cup (\frac{7}{9}, \frac{8}{9})\right) = \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right].$$

C_n *is the disjoint union of* 2^n *closed intervals of length* 3^{-n}*. We then define the Cantor set to be*

$$\mathbf{C} = \bigcap_{n=1}^{\infty} C_n.$$

Exercise 8. *Show that these two definitions of the Cantor set agree.*

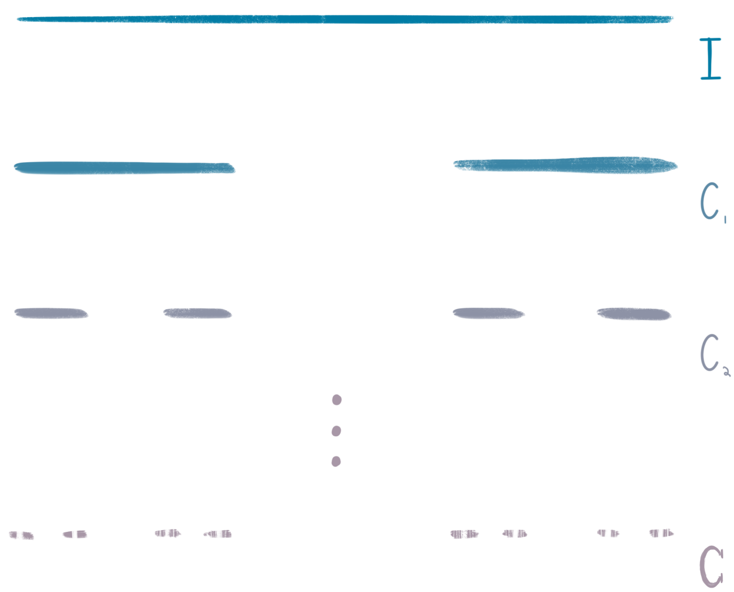


Figure 2.1: The Cantor Set

Exercise 9. *Show that* $\frac{1}{4}$ *is in the Cantor set.*

Proposition 13. $\mathbf{C}$ *is a closed set whose measure is zero.*

Proof. $\mathbf{C}$ is closed since it is the intersection of closed sets. It has measure zero since $\mathbf{C} \subset C_n$ for all n and thus

$$\mu(\mathbf{C}) \leqslant \mu(C_n) = \frac{2^n}{3^n}.$$

□

Exercise 10. *Show that* $\mathbf{C}$ *is uncountable.*

Hint: Use the ternary definition and consider the map sending a number x written in ternary as $.a_1a_2a_3\ldots$ to the number written in binary as $.\frac{a_1}{2}\frac{a_2}{2}\frac{a_3}{2}\cdots$

Exercise 11. *Use the Cantor set to construct a dense uncountable set of real numbers with measure zero.*

2.6.1 Counting subsets

The Cantor set gives an example of an uncountable set with measure zero. However, we can also use it to "construct" a Lebesgue measurable set which is not Borel.

Proposition 14. *Any subset of the Cantor set is measurable.*

Proof. By the monotonicity of the Lebesgue measure, *any* subset of the Cantor set has measure zero. □

To see that this implies that there are Lebesgue measurable sets which are not Borel, we can count the number of subsets of the Cantor sets, of which there are

$$\#(\mathcal{P}(\mathbf{C})) = \#(\mathcal{P}(\mathbb{R})) = 2^{(\#\mathbb{R})}.$$

On the other hand, there are $\#\mathbb{R}$ open subsets of $\mathbb{R}$ (exercise: why?) and every Borel set is constructed from countable intersections and unions of open sets. Using transfinite induction and these two facts, it is possible to show that there are $\#\mathbb{R}$ Borel sets.

We will revisit the Cantor set in the next chapter to provide counterexamples for some intuitive ideas about continuous functions, their derivatives and measures.

Chapter 3

Measurable Functions

We now turn our attention to measurable functions in order to lay the foundations for the Lebesgue integral. The basic idea is to define it in a similar way to continuity, but there are some subtleties.

Definition 23. *A function $f : X \to Y$ is continuous if the pre-image of any open set $f^{-1}(U)$ is open.*

Definition 24. *A function $f : E \to \mathbb{R}$ (or $[-\infty, \infty]$) is measurable if the preimage of any open set $f^{-1}(U)$ is Lebesgue-measurable.*

By considering $f^{-1}(\mathbb{R})$, we can immediately see that the domain of the function E must be measurable. We also see that every continuous real-valued function on $\mathbb{R}$ is measurable.

Proposition 15. *Suppose E is a measurable set of reals and $f : E \to [-\infty, \infty]$. Then, the following are equivalent..*

1. *f is measurable.*
2. *$f^{-1}[(a, \infty]]$ is measurable for every $a \in \mathbb{R}$.*
3. *$f^{-1}[[-\infty, a)]$ is measurable for every $a \in \mathbb{R}$.*
4. *$f^{-1}[[a, \infty]]$ is measurable for every $a \in \mathbb{R}$.*
5. *$f^{-1}[[-\infty, a]]$ is measurable for every $a \in \mathbb{R}$.*

Proof. (Sketch) By definition, 1 implies all the other statements. If any of 2 - 5 hold, then by considering complements, countable unions, and finite intersections, it follows that f is measurable. □

As an immediate consequence, we see the following corollary.

Corollary 7. *Suppose $f : E \to [-\infty, \infty]$ is measurable. Then the preimage $f^{-1}(U)$ of any Borel set is measurable.*

3.0.1 A brief remark on σ-algebras

Much in the same way that we can use function(al)s to induce a topology (by considering the coarsest topology in which a function (or family of functions) is continuous, we can use functions to construct sigma algebras. This idea will not be used heavily in this class, but if you are interested in stochastic calculus or financial math, you will see it there.

Proposition 16. *Suppose $f : E \to \mathbb{R}$ is a measurable function. Let $\mathcal{S} = \{X \subseteq \mathbb{R} \ : \ f^{-1}[X]$ is measurable$\}$. Then, $\mathcal{S}$ is a σ-algebra over $\mathbb{R}$.*

Proof. Since $f^{-1}[\mathbb{R}] = E$, and since E is measurable, $\mathbb{R} \in \mathcal{S}$.

If $f^{-1}[X]$ is measurable, then $f^{-1}[X^c] = E - f^{-1}[X]$ is measurable. So, $\mathcal{S}$ is closed under complementation. If $f^{-1}[A_n]$ is measurable for each $n \in \mathbb{N}$, then $f^{-1}[\bigcup_{n=0}^{\infty} A_n] = \bigcup_{n=0}^{\infty} f^{-1}[A_n]$ is measurable. So, $\mathcal{S}$ is closed under countable unions. □

Now back to the regularly scheduled programming...

3.1 Sums, products, pointwise limits, etc.

We now show that some natural ways to combine measurable functions produce more measurable functions.

Theorem 17. *If $f : E \to [-\infty, \infty]$ is continuous, and if E is measurable, then f is measurable.*

Proof. Let $U \subseteq [-\infty, \infty]$ be an open set of reals. Since f is continuous, there is an open set V so that $f^{-1}[U] = V \cap E$. Since V, E are measurable, $V \cap E$ is measurable. □

Lemma 9. *Suppose $f : E \to [-\infty, \infty]$ and $g : E \to [-\infty, \infty]$ are equal almost everywhere (i.e. $\mu(\{t \in E \ : \ f(t) \neq g(t)\}$. If f is measurable, then g is measurable.*

Proof. Suppose f is measurable. Let $\mathcal{E} = \{x \in E \ : \ f(x) \neq g(x)\}$. Then, for every $U \subseteq [-\infty, \infty]$,

$$g^{-1}[U] = f^{-1}[U] \cap (\mathbb{R} - \mathcal{E}) \cup \ g^{-1}[U] \cap \mathcal{E}.$$

Suppose U is open. Then, $f^{-1}[U]$ is measurable. Since $\mathcal{E} = 0$, $g^{-1}[U] \cap \mathcal{E}$ is measurable. Thus, $g^{-1}[U]$ is measurable. $\square$

Proposition 18. *Suppose $f : E_1 \to [-\infty, \infty]$ is measurable and $g : E_2 \to [-\infty, \infty]$ is continuous. Suppose E_2 is open and $E_2 \supseteq \text{ran}(f)$. Then, $g \circ f$ is measurable.*

Proof. Suppose $U \subseteq [-\infty, \infty]$ is open. Then, $(g \circ f)^{-1}[U] = f^{-1}[g^{-1}[U]]$. Thus, $g^{-1}[U]$ is open. Thus, $(g \circ f)^{-1}[U]$ is measurable. $\square$

Theorem 19. *Suppose $f : E \to [-\infty, \infty]$ and $g : E \to [-\infty, \infty]$ are measurable. Suppose also that f and g are finite almost everywhere. Then, fg is measurable and $\alpha f + \beta g$ is measurable for all $\alpha, \beta \in \mathbb{R}$.*

Proof. By Lemma 9, we can assume that f and g are real-valued.

Claim 1: $f + g$ is measurable.

Proof Claim 1: Set $h = f + g$. Let $a \in \mathbb{R}$.

We have:

$$\begin{aligned} h(x) \in (a, \infty] &\iff f(x) + g(x) > a \\ &\iff f(x) > a - g(x) \\ &\iff \exists q \in \mathbb{Q}\ f(x) > q > a - g(x) \\ &\iff \exists q \in \mathbb{Q}\ f(x) > q \text{ and } g(x) > a - q \end{aligned}$$

So,

$$h^{-1}[(a, \infty]] = \bigcup_{q \in \mathbb{Q}} f^{-1}[(q, \infty]] \cap g^{-1}[(a - q, \infty]].$$

Thus, $h^{-1}[(a, \infty]]$ is measurable. Thus, by Proposition 15, h is measurable.
Claim 2: αf is measurable.

Proof of Claim 2: Set $g(x) = \alpha x$. Then, $\alpha f = g \circ f$. Apply Proposition 18.

Claim 3: fg measurable.
Proof Claim 3: $fg = \frac{1}{2}[(f + g)^2 - (f - g)^2]$. By Proposition 18, $(f + g)^2$ and $(f - g)^2$ are measurable. $\square$

Definition 25. *Suppose $f : E \to [-\infty, \infty]$ and $g : E \to [-\infty, \infty]$. Then, for all $x \in \mathbb{R}$, $\max\{f, g\}(x) = \max\{f(x), g(x)\}$ and $\min\{f, g\}(x) = \min\{f(x), g(x)\}$.*

In other words, $\max\{f,g\}(x)$ is simply the larger of $f(x)$ and $g(x)$ and conversely for $\min\{f,g\}(x)$.

Definition 26. *Suppose $f : E \to [-\infty,\infty]$. The positive part of f, denoted f^+ is defined as $f^+ = \max\{f,0\}$. The negative part of f, denoted f^-, is defined $f^- = -\min\{f,0\}$.*

Proposition 20. *Suppose $f : E \to [-\infty,\infty]$ and $g : E \to [-\infty,\infty]$. Suppose f,g are measurable and finite almost everywhere. Then, $\max\{f,g\}$ and $\min\{f,g\}$ are measurable.*

In fact, it is possible to prove this result without assuming that f and g are finite almost everywhere, but this will not be necessary for our purposes.

Proof. Set $h = \max\{f,g\}$. Suppose U is an open subset of $[-\infty,\infty]$. Then,

$$\begin{aligned} h^{-1}[U] &= \{x \in E : f(x) \in U \text{ and } f(x) > g(x)\} \\ &\quad \bigcup \{x \in E : g(x) \in U \text{ and } g(x) \geqslant f(x)\} \\ &= f^{-1}[U] \cap (f-g)^{-1}[(0,\infty]] \bigcup g^{-1}[U] \cap (g-f)^{-1}[[-\infty,0]]. \end{aligned}$$

Since $f-g$ and $g-f$ are measurable, it follows that $h^{-1}[U]$ is measurable if U is open. Thus, h is measurable.

$\min\{f,g\} = -\max\{-f,-g\}$. Thus, $\min\{f,g\}$ is measurable. □

Corollary 8. *If f is measurable, then f^+ and f^- are measurable.*

It is worth noting that these proofs would be easier were we able to discuss measurability of higher dimensional sets (since max is naturally a two-variable function). This is one of the downsides of sticking with one variable, (the upside is that it is much easier to define the Lebesgue measure on the real line compared to a higher dimensional space).

We now come to one of the most important facts about the measurable functions, which is that the pointwise limit of measurable functions is measurable. Note that the pointwise limit of continuous functions may not be continuous. It turns out that the pointwise limit of Riemann integrable functions need not be Riemann integrable either, so this result will be one of the major advantages of the Lebesgue integral.

Theorem 21. *Suppose E is a measurable set of reals and that $f, f_0, f_1, \ldots$ are extended real-valued functions on E so that $f_0, f_1, \ldots$ are measurable and $\lim_{n\to\infty} f_n(x) = f(x)$ almost everywhere. Then, f is measurable.*

Proof. Let $X = \{x \in E \;:\; \lim_{n\to\infty} f_n(x) \neq f(x)\}$. Suppose $U \subseteq [-\infty, \infty]$ is open. Then,

$$f^{-1}[U] = X \cap f^{-1}[U] \cup X^c \cap f^{-1}[U].$$

Suppose $x \notin X$. Then,

$$\begin{aligned} f(x) \in U \quad &\Leftrightarrow \quad \exists N_0 \in \mathbb{N}\ \forall n \geqslant N_0 f_n(x) \in U \\ &\Leftrightarrow \quad x \in \bigcup_{N_0=0}^{\infty} \bigcap_{n=N_0}^{\infty} f_n^{-1}[U]. \end{aligned}$$

So,

$$X^c \cap f^{-1}[U] = X^c \cap \bigcup_{N_0 \in \mathbb{N}} \bigcap_{n=N_0}^{\infty} f_n^{-1}[U].$$

Thus, $f^{-1}[U]$ is measurable. □

Definition 27. *Notation:* $f_n \to f$ a.e. *is shorthand for* $\{f_n\}_{n=0}^{\infty}$ *converges to* f *almost everywhere.* $f_n \to f$ *is shorthand for* $\{f_n\}_{n=0}^{\infty}$ *converges pointwise to* f*.*

Finally, we note that the restriction of a measurable function to a measurable sub-set is measurable.

Definition 28. $f|_E$ *denotes the restriction of* f *to* E *when* $E \subseteq \mathrm{dom}(f)$*. Namely,*

$$f|_E(x) = \begin{cases} f(x) & \textit{if } x \in E \\ \textit{undefined} & \textit{otherwise} \end{cases}$$

Proposition 22. *If* f *is measurable, and if* E *is a measurable subset of the domain of* f*, then* $f|_E$ *is measurable.*

3.2 The Cantor Function and non-measurable functions

In the previous section, we saw multiple ways to combine measurable functions to obtain new measurable functions. However, there was a conspicuous absence: *composition.*

Observation 1. *The composition of measurable functions is not necessarily measurable!*

To see this, we will revisit the Cantor set and use it define the Cantor-Lebesgue function.

Definition 29. *Suppose $x \in \mathbf{C}$, and let $x = \sum_{n=0}^{\infty} \frac{a_n}{3^{n+1}}$ where $a_n \in \{0,2\}$ for all $n \in \mathbb{N}$. Define*

$$\phi(x) = \sum_{n=0}^{\infty} \frac{a_n/2}{2^{n+1}}.$$

Proposition 23. *Suppose $n_0 \in \mathbb{N}$. If $x_0, x_1 \in \mathbf{C}$, and if $|x_0 - x_1| < 3^{-n_0}$, then $|\phi(x_0) - \phi(x_1)| < 2^{-n_0}$.*

To obtain the full Cantor-Lebesgue function, we extend ϕ to $[0,1]$ as follows. For each $x \in [0,1] - \mathbf{C}$, let

$$\phi(x) = \phi(\sup\{x' \in \mathbf{C} \ : \ x' < x\}).$$

ϕ is called the *Cantor-Lebesgue function.*

It is possible to calculate some values of this function explicitly.

1. If $x \in (\frac{1}{3}, \frac{2}{3})$, then $\phi(x) = \frac{1}{2}$.
2. If $x \in (\frac{7}{9}, \frac{8}{9})$, then $\phi(x) = \frac{3}{4}$.

We will now note some properties of this function

Proposition 24. *If $(a,b) \subseteq [0,1] - \mathbf{C}$, then ϕ is constant on (a,b) and $\phi(a) = \phi(b)$.*

Corollary 9. *ϕ is non-decreasing.*

Proposition 25. *ϕ is continuous.*

Proof. Let $\epsilon > 0$ Choose $n_0 \in \mathbb{N}$ so that $2^{-n_0} < \epsilon$. Suppose $x_0, x_1 \in [0,1]$ and $|x_0 - x_1| < 3^{-n_0}$.

Claim: $|\phi(x_0) - \phi(x_1)| < 2^{-n_0}$.

Proof Claim: WLOG $x_0 < x_1$. Let:

$$\begin{aligned} x_0' &= \inf\{x \in \mathbf{C} \ : \ x_0 \leqslant x\} \\ x_1' &= \sup\{x \in \mathbf{C} \ : \ x \leqslant x_1\} \end{aligned}$$

Thus, $x_0', x_1' \in \mathbf{C}$. By Proposition 24, $\phi(x_0') = \phi(x_0)$ and $\phi(x_1') = \phi(x_1)$.

Case 1: $x_0' > x_1$ or $x_1' < x_0$.

It follows that $(x_0, x_1) \subseteq [0,1] - \mathbf{C}$. Thus, $\phi(x_0) = \phi(x_1)$ by Proposition 24.

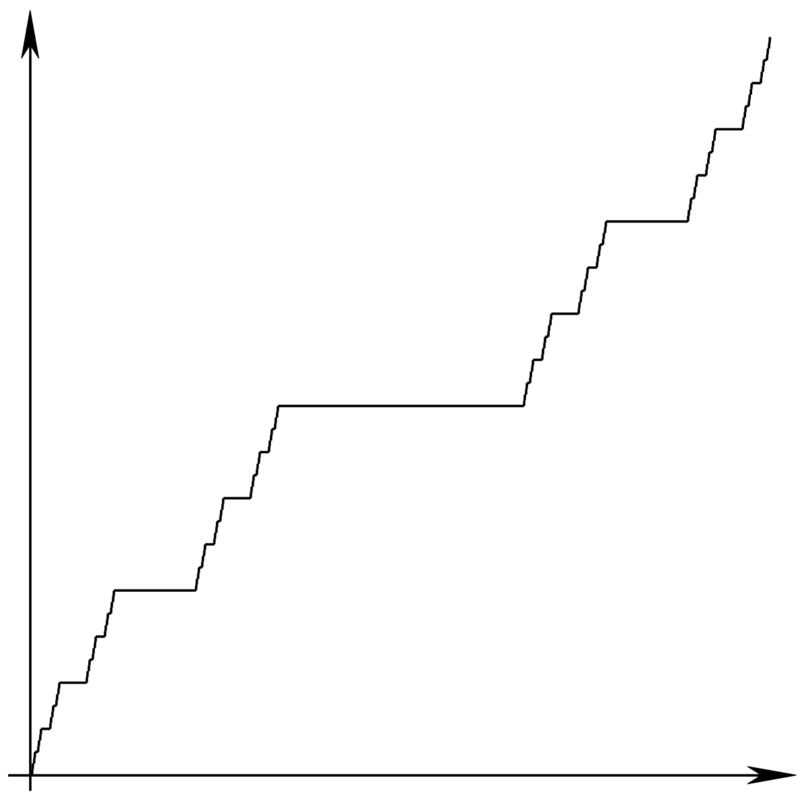


Figure 3.1: An approximation of the Cantor-Lebesgue function

Case 1: $x_0' \leqslant x_1$ and $x_0 \leqslant x_1'$.

Thus, $x_0', x_1' \in [x_0, x_1]$. Therefore, $|x_0' - x_1'| < 3^{-n_0}$. So, by Proposition 23, $|\phi(x_0') - \phi(x_1')| < 2^{-n_0}$. □

This proposition has the (perhaps surprising) consequence that ϕ is surjective on $[0, 1]$

Corollary 10. $\text{ran}(\phi) = [0, 1]$.

Proof. Since $\phi(0) = 0$ and $\phi(1) = 1$ and ϕ continuous. □

The Cantor-Lebesgue function is continuous and increasing, but it will be helpful to modify it so that it is strictly increasing. To do so, we define

$$\psi(x) = x + \phi(x).$$

Observation 2. *The function* ψ *is an increasing and continuous map of* $[0, 1]$ *onto* $[0, 2]$.

Observation 3. *The inverse function* ψ^{-1} *is continuous.*

One somewhat surprising fact is that even though the Cantor set has zero measure, its image under Ψ has positive measure.

Proposition 26. $\mu(\psi[\mathbf{C}]) = 1$.

Proof. Since ψ is one-to-one, $[0, 2] = \psi[\mathbf{C}] \cup \psi[[0, 1] - \mathbf{C}]$. If $(a, b) \subseteq [0, 1] - \mathbf{C}$, then $\mu[\psi[(a, b)]] = \mu((a, b))$. Thus, $\mu(\psi[[0, 1] - \mathbf{C}]) = 1$. Therefore, $\mu(\psi[\mathbf{C}]) = 1$. □

Proposition 27. *There is a measurable* $A \subseteq [0, 1]$ *so that* $\psi[A]$ *is non-measurable.*

Proof. Since $\psi[\mathbf{C}]$ has positive measure, it includes a non-measurable set B. Set $A = \psi^{-1}[B]$. Then, $A \subseteq \mathbf{C}$. So, A is measurable. □

Corollary 11. *There is a non-measurable function.*

Proof. Suppose A is a measurable subset of $[0, 1]$ so that $\psi[A]$ is non-measurable. Let $h = \chi_A \circ \psi^{-1}$. Then, $h^{-1}[(\frac{1}{2}, \frac{3}{2})] = \psi[A]$. Thus, h is non-measurable. □

3.2.1 A remark for those interested in probability

The purpose of this subsection is to make some remarks on the Cantor-Lebesgue function and its relation to probability. The reader is invited to skip this section if they would like.

When one learns probability, one encounters many examples of probability measures (i.e., measures whose total mass is one). These generally fall into two types.

1. The first type are those which are discrete and have probability mass functions, such as the binomial distribution or negative binomial distribution. Such measures are called atomic, since they charge (i.e., give positive measure to) points.

2. The second type are measures such as normal distribution, which has a probability density function

   $$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x)^2}$$

   and where the measure of a subset U can be computed as

   $$\int_U f(x)\,dx$$

   (where this integral is done with respect to the Lebesgue measure).

It is tempting to intuitively think of measures as being some sort of linear combination of a continuous measure and an atomic measure. We will not cover it in this class, but if you continue learning real analysis you will encounter the Radon-Nikodym derivative (of which $f(x)$ is an example) and the Lebesgue decomposition theorem, which support this intuition. Intuitively, these results state that any reasonable[1] measure can be decomposed into a measure which has a probability density function (with respect to the Lebesgue measure) and a measure which is singular (i.e., charges null sets).

However, it is important to realize that singular measures *need not* be atomic. In particular, the Cantor-Lebesgue function can be understood as a cumulative distribution function for a probability measure which is supported on the Cantor set. Although this measure is Lebesgue singular (since the Cantor set has measure zero), the fact that the Cantor-Lebesgue

[1] There is a technical assumption here that the measures be σ-finite, which we will see at the very end of this class.

function is continuous means that the measure is non-atomic. As such, this is an important counter-example to keep in mind if you are studying probability theory.

And now back to our regularly scheduled programming...

3.3 Approximating measurable functions

For Riemann integration, the general idea is to use step functions to approximate the desired function and then compute the area underneath the step function. For the Lebesgue integral, we will instead approximate using *simple functions*. Intuitively, these functions cut horizontally instead of vertically.

Definition 30. *A* simple function *is a real-valued measurable function whose range is finite.*

Example. *1. The Heaviside function*

$$f(x) = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases}$$

2. The Dirichlet function

$$f(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases}$$

Definition 31. *Suppose $E \subseteq \mathbb{R}$. For each $x \in \mathbb{R}$ let*

$$\chi_E(x) = \begin{cases} 1 & x \in E \\ 0 & x \notin E \end{cases}$$

χ_E is called the characteristic function of *E.*

If E is measurable, then χ_E is a simple function.

Proof. Suppose $U \subseteq [-\infty, \infty]$ is open.

$$\chi_E^{-1}[U] = \begin{cases} \mathbb{R} & 0, 1 \in U \\ E^c & 0 \in U \text{ and } 1 \notin U \\ E & 0 \notin U \text{ and } 1 \in U \\ \varnothing & 0, 1 \notin U \end{cases}$$

□

It is not hard to show that every step function is also a simple function

Exercise 12. *Show that every step function is a simple function.*

There is an alternative characterization of simple function and the linear combination of characteristic functions.

Proposition 28. *A function $\phi : \mathbb{R} \to \mathbb{R}$ is simple if and only if there are measurable sets $E_1, \ldots, E_n \subseteq \mathbb{R}$ and $c_1, \ldots, c_n \in \mathbb{R}$ so that*

$$\phi = \sum_{j=1}^{n} c_j \chi_{E_j}$$

Here is a quick sketch of the proof. For the forward direction, suppose that ϕ simple. Let $\{c_1, \ldots, c_n\} = \text{ran}(\phi)$. Take $E_j = \phi^{-1}[\{c_j\}]$. For the reverse direction, note that the range of ϕ is finite. Since χ_{E_j} is measurable, by Theorem 19, we find that ϕ is measurable.

Theorem 29 (Simple Approximation Lemma)**.** *Suppose $f : E \to \mathbb{R}$ is a bounded measurable function. Then, for every $\epsilon > 0$, there exist ϕ, ψ so that ϕ and ψ are simple functions on E, , $\phi \leqslant f \leqslant \psi$ and $\psi - \phi < \epsilon$.*

Proof. Since f is bounded, there is a positive number M so that $|f(x)| < M$ for all $x \in E$. Choose $n_0 \in \mathbb{N}$ so that $2M2^{-n_0} < \epsilon$. Let $c_j = -M + j(2M)2^{-n_0}$, and let $d_j = -M + (j+1)(2M)2^{-n_0}$. Set $V_j = [c_j, d_j)$ when $0 \leqslant j < 2^{n_0}$. Thus, $[-M, M) = \bigcup_{j<2^{n_0}} V_j$. Note $\text{ran}(f) \subseteq [-M, M)$. Set $E_j = f^{-1}[V_j]$. Therefore, E_j is measurable. Set:

$$\begin{aligned} \phi_1 &= \sum_{j<2^{n_0}} c_j \chi_{E_j} \\ \psi_1 &= \sum_{j<2^{n_0}} d_j \chi_{E_j} \end{aligned}$$

By Proposition 28, ϕ_1 and ψ_1 are simple. Let $\phi = \phi_1|_E$, and let $\psi = \psi_1|_E$. Thus, ϕ and ψ are simple and $\phi \leqslant f \leqslant \psi$ by construction. Let $x \in E$. Then, $x \in E_j$ for exactly one j. By construction, $c_j = \phi(x)$ and $d_j = \psi(x)$. But, $d_j - c_j = 2M2^{-n_0} < \epsilon$. □

In fact, we can prove a stronger approximation theorem as well.

Theorem 30. *Suppose $f : E \to [-\infty, \infty]$ and suppose E is measurable. Then, f is measurable. if and only if there is a sequence of simple functions on E $\{\sigma_n\}_{n=0}^{\infty}$ so that $\sigma_n \to f$ a.e. and $|\sigma_n| \leqslant |f|$ for all n. If $f \geqslant 0$, then we can choose $\{\sigma_n\}_{n=0}^{\infty}$ to be non-decreasing.*

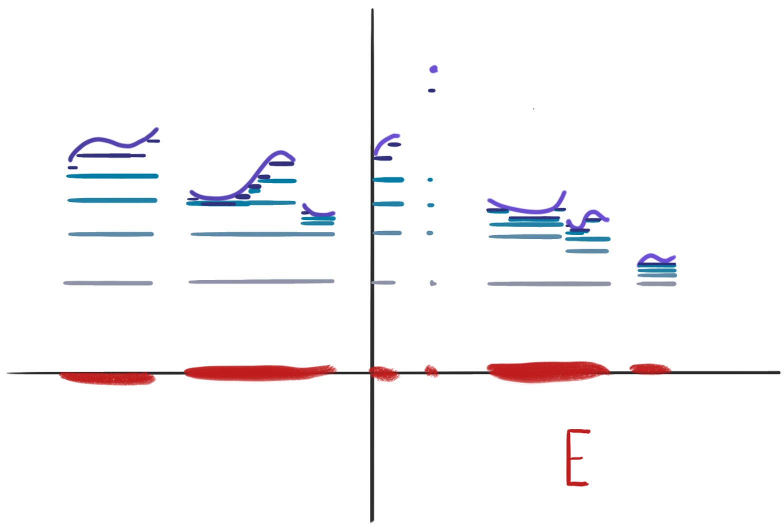


Figure 3.2: The simple functions are converging almost everywhere to the function $f : E \to \mathbb{R}$ (depicted in purple). As a disclaimer, these are not the precise functions that are constructed in the proofs of the simple approximation lemmas.

Proof. The reverse direction follows directly from Theorem 21, so it remains to show the forward direction.

Case 1: $f \geqslant 0$.

Set $E_n = f^{-1}[[0, n+1)]$ for every $n \in \mathbb{N}$. Set $f_n = f \cdot \chi_{E_n}$. Thus, f_n is measurable and bounded. By Theorem 29, there is a simple function ϕ_n so that $\phi_n \leqslant f_n$ and $f_n - \phi_n < 2^{-n}$. Let

$$\tau_n(x) = \begin{cases} \phi_n^+(x) & x \in E_n \\ n+1 & x \notin E_n \end{cases}$$

Thus, τ_n is simple, and $0 \leqslant \tau_n \leqslant f$.

Claim: $\{\tau_n\}_{n=0}^{\infty}$ converges pointwise to f.

Let $x \in E$. Let $\epsilon > 0$. Choose $N_0 \in \mathbb{N}$ so that $2^{-N_0} < \epsilon$ and $f(x) < N_0$. Let $n \geqslant N_0$. Therefore, $x \in E_n$. Thus, $\tau_n(x) = \phi_n^+(x)$.

Subcase a: $\phi_n(x) > 0$.

Then, $\tau_n(x) = \phi_n(x)$ and $|f(x) - \tau_n(x)| = f(x) - \phi_n(x) < \epsilon$.

Subcase b $\phi_n(x) \leqslant 0$:

$|f(x) - \tau_n(x)| = f(x) \leqslant f_n(x) - \phi_n(x) < \epsilon$.

Thus, $\lim_{n\to\infty} \tau_n(x) = f(x)$.

Now, set $\sigma_n = \max_{j\leqslant n} \tau_j$. Thus, σ_n is measurable. $\sigma_n \leqslant \sigma_{n+1}$ by definition. $\tau_j \leqslant f$. Thus, $0 \leqslant \sigma_n \leqslant f$. Since $\{\tau_n\}_{n=0}^{\infty}$ converges pointwise to f, it follows that $\{\sigma_n\}_{n=0}^{\infty}$ converges pointwise to f.

Case 2: $f \ngeqslant 0$.

By Case 1, there exist a sequence of simple functions $\{\sigma_{1,n}\}_{n=0}^{\infty}$ that converges pointwise to f^+ and so that $0 \leqslant \sigma_{1,n} \leqslant f^+$. By Case 1, there is a sequence of simple functions $\{\sigma_{2,n}\}_{n=0}^{\infty}$ that converges pointwise to f^- and so that $0 \leqslant \sigma_{2,n} \leqslant f^-$. Set $\sigma_n = \sigma_{1,n} - \sigma_{2,n}$. Since $f = f^+ - f^-$, $\{\sigma_n\}_{n=0}^{\infty}$ converges to f pointwise. □

3.4 Littlewood's three principles

> The extent of knowledge required is nothing like so great as is sometimes supposed. There are three principles, roughly expressible in the following terms: Every [measurable] set is nearly a finite union of intervals; every [measurable] function is nearly continuous; every pointwise convergent sequence of [measurable] functions is nearly uniformly convergent. Most of the results of [the theory] are fairly intuitive applications of these ideas, and the student armed with them should be equal to most occasions when real variable theory is called for. If one of the principles would be the obvious means to settle the problem if it were 'quite' true, it is natural to ask if the 'nearly' is near enough, and for a problem that is actually solvable it generally is.
>
> J.E. Littlewood

At first, measure theory might seem to be an imposing and very technical subject. However, Littlewood's quote explains that measure theory only improves upon a naive approach to integration by ϵ (albeit a very important ϵ).

We have already seen Littlewood's first principle precisely. The Vitali covering lemma (Theorem 7) shows that given a measurable set E of finite measure, then for each $\epsilon > 0$, there is a finite collection of open intervals whose union U is "nearly equal to" E in the sense that

$$\mu(E - U) + \mu(U - E) < \epsilon.$$

In this section, we will work towards making the other principles precise.

3.4.1 The third principle: Egoroff's theorem

Here, we will show that pointwise convergence of measurable functions is "nearly" uniform convergence. To do so, we start with the following lemma.

Lemma 10. *Suppose $\mu(E) < \infty$, and suppose $\{f_n\}_{n=0}^{\infty}$ is a sequence of measurable functions on E that converges pointwise to a real-valued function f. Then, for all $\epsilon_1, \epsilon_2 > 0$ there exists $N_0 \in \mathbb{N}$ and a measurable $A \subseteq E$ so that $\mu(E - A) < \epsilon_2$ and so that $|f(x) - f_k(x)| < \epsilon_1$ whenever $k \geqslant N_0$ and $x \in A$.*

Proof. Let $\epsilon_1, \epsilon_2 > 0$ and for each $N_0 \in \mathbb{N}$, let

$$E_{N_0} = \bigcap_{n=N_0}^{\infty} \{x \in E \ : \ |f_n(x) - f(x)| < \epsilon_1\}$$

Since $f_n \to f$,

$$E = \bigcup_{N_0 \in \mathbb{N}} E_{N_0}$$

Since $E_{N_0} \subseteq E_{N_0+1}$,

$$\lim_{N_0 \to \infty} \mu(E_{N_0}) = \mu(E).$$

So, since $\mu(E) < \infty$, there is a non-negative integer N_0 so that $\mu(E) - \mu(E_{N_0}) < \epsilon_2$. Set $A = E_{N_0}$. □

With this lemma in hand, we can now prove a somewhat surprising theorem, that pointwise convergence is "nearly" the same as uniform convergence for measurable functions.

Theorem 31. *(Egoroff's Theorem) Suppose $\mu(E) < \infty$ and $\{f_n\}_{n=0}^{\infty}$ is a sequence of measurable functions on E that converges pointwise to a real-valued function f. Then, for every $\epsilon > 0$ there is a closed set $F \subseteq E$ so that $\{f_n\}_{n=0}^{\infty}$ converges uniformly to f on F and $\mu(E - F) < \epsilon$.*

Proof. Let $\epsilon > 0$.

Claim 1: There is a measurable set $A \subseteq E$ so that $\mu(E - A) < \epsilon/2$ and $\{f_n\}_{n=0}^{\infty}$ converges uniformly to f on A.

Proof of Claim 1: By Lemma 10, for each $n \in \mathbb{N}$, there is a non-negative integer M_n and a measurable set $A_n \subseteq E$ so that $\mu(E - A_n) < \epsilon 2^{-(n+2)}$ and $|f(x) - f_k(x)| < 2^{-n}$ whenever $x \in A_n$ and $k \geqslant M_n$. Set $A = \bigcap_{n=0}^{\infty} A_n$.

Subclaim 1.a: $\mu(E - A) < \epsilon/2$.

Proof:

$$\mu(E - A) = \mu(\bigcup_{n=0}^{\infty} E - A_n) < \epsilon \sum_{n=0}^{\infty} 2^{-(n+2)} = \epsilon/2.$$

Subclaim 1.b: $\{f_n\}_{n=0}^{\infty}$ converges uniformly to f on $E - A$.

Proof: Let $\epsilon' > 0$. Choose $n \in \mathbb{N}$ so that $2^{-n} < \epsilon'$. Let $x \in A$, and let $k \geqslant n$. Since $x \in A_n$, $|f(x) - f_k(x)| < 2^{-n} < \epsilon'$.

Claim 2: There is a closed set $F \subseteq E$ so that $\{f_n\}_{n=0}^{\infty}$ converges uniformly on F and $\mu(E - F) < \epsilon$.

Proof: By Theorem 6, there is a closed $F \subseteq A$ so that $\mu(A - F) < \epsilon/2$. So $\mu(E - F) = \mu(E - A) \cup \mu(A - F) < \epsilon$. □

3.4.2 The second principle: Lusin's theorem

We can now make the second principle precise as well, to show that measurable functions are "nearly" continuous functions. To establish this, we start with a simpler version.

Lemma 11. *Let f be a simple function defined on E. Then for each $\epsilon > 0$, there is a continuous function $g : \mathbb{R} \to \mathbb{R}$ and a closed set $F \subseteq E$ for which*

$$f = g \text{ on } F \text{ and } m(E - F) < \epsilon$$

Proof. Let $a_1, a_2, \ldots, a_n$ be the distinct values taken by f, and let them be taken on the sets $E_1, E_2, \ldots, E_n$, respectively. The collection $\{E_k\}_{k=1}^{n}$ is disjoint since the a_k 's are distinct. According to Theorem 4, we may choose closed sets $F_1, F_2, \ldots, F_n$ such that for each index $k, 1 \leqslant k \leqslant n$

$$F_k \subseteq E_k \text{ and } m(E_k - F_k) < \epsilon/n$$

Then we consider the set $F = \bigcup_{k=1}^{n} F_k$, which must be closed (as the finite union of closed sets). Since $\{E_k\}_{k=1}^{n}$ is disjoint,

$$\mu(E - F) = \mu\left(\bigcup_{k=1}^{n} [E_k - F_k]\right) = \sum_{k=1}^{n} m(E_k - F_k) < \epsilon$$

Define g to be a function on F which takes the value a_k on F_k for $1 \leqslant k \leqslant n$. Since the collection $\{F_k\}_{k=1}^{n}$ is disjoint, g is properly defined. Moreover, g is continuous on F since for a point $x \in F_i$, there is an open interval containing x which is disjoint from the closed set $\bigcup_{k \neq i} F_k$ and hence on the intersection of this interval with F the function g is constant. But g can be extended[2] from a continuous function on the closed set F to a continuous function on all of $\mathbb{R}$. The continuous function g on $\mathbb{R}$ has the required approximation properties. □

Using this and Egoroff's theorem, we can prove the full version of Lusin's Theorem.

Theorem 32 (Lusin's Theorem)**.** *Suppose $f : E \to \mathbb{R}$ is measurable. Then, for every $\epsilon > 0$ there is a closed set F and a continuous function $g : \mathbb{R} \to \mathbb{R}$ so that $F \subseteq E$, $\mu(E - F) < \epsilon$, and $g(t) = f(t)$ for all $t \in F$.*

Proof. We consider only provide details for the case that $m(E) < \infty$. According to the Simple Approximation Theorem, there is a sequence $\{f_n\}$ of simple functions defined on E that converges to f pointwise on E. Let n be a natural number. By the preceding proposition, with f replaced by f_n and ϵ replaced by $\epsilon/2^{n+1}$, we may choose a continuous function g_n on $\mathbb{R}$ and a closed set $F_n \subseteq E$ for which

$$f_n = g_n \text{ on } F_n \text{ and } \mu(E - F_n) < \frac{\epsilon}{2^{n+1}}.$$

According to Egoroff's Theorem, there is a closed set F_0 contained in E such that $\{f_n\}$ converges to f uniformly on F_0 and $\mu(E - F_0) < \epsilon/2$. Define $F = \bigcap_{n=0}^{\infty} F_n$. Thus, we have that

[2]This is known as the Tietze Extension Theorem. A proof can be found in your favorite topology book (for instance, see [Mun14] Chapter 4 Section 35).

$$\mu(E - F) = m\left([E - F_0] \cup \bigcup_{n=1}^{\infty} [E - F_n]\right) \leqslant \frac{\epsilon}{2} + \sum_{n=1}^{\infty} \frac{\epsilon}{2^{n+1}} = \epsilon.$$

The set F is closed since it is the intersection of closed sets. Each f_n is continuous on F since $F \subseteq F_n$ and $f_n = g_n$ on F_n. Finally, $\{f_n\}$ converges to f uniformly on F since $F \subseteq F_0$. However, the uniform limit of continuous functions is continuous, so the restriction of f to F is continuous on F. Finally, there is a continuous function g defined on all of $\mathbf{R}$ whose restriction to F equals f (again by the Tietze Extension Theorem). This function g has the required approximation properties. □

Exercise 13. *Prove Lusin's theorem in the case that* $\mu(E) = \infty$.

Chapter 4

Lebesgue Integration

After almost 60 pages of lecture notes (and over 2000 lines of LaTeXcode), we can finally start discussing how to integrate functions.

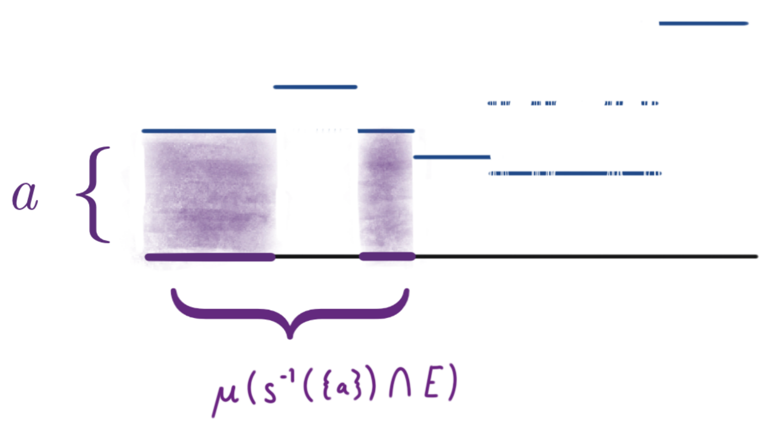
We start by defining the Lebesgue integral for non-negative simple functions, which will be the fundamental building block for the Lebesgue integral.

Definition 32. *If* $s : X \to [0, \infty)$ *is a simple function, and if* $E \subseteq X$ *is measurable, then the* Lebesgue integral of s over E *is defined to be*

$$\int_E s \, d\mu = \sum_{a \in range(s)} a \, \mu\left(s^{-1}(\{a\}) \cap E\right).$$

We denote this by

$$\int_E s \;\; d\mu.$$



Note that in Definition 32, we use the convention $0 \cdot \infty = 0$ (since the measure of the set where s vanishes will often be infinite).

We can also consider integrals over all of X, which is defined as

$$\int_X s\, d\mu = \sum_{a \in \operatorname{ran}(s)} a\mu(s^{-1}[\{a\}])$$

if $s : X \to [0, \infty)$ is simple.

Example. *There are a few examples of integrals that we can compute immediately.*

1. *If E is a measurable set, then $\int_{\mathbb{R}} \chi_E\, d\mu = \mu(E)$. Proof: let $s = \chi_E$. Then, $\operatorname{ran}(s) = \{0, 1\}$. $s^{-1}[\{0\}] = E^c$ and $s^{-1}[\{1\}] = E$. So,*

$$\int_{\mathbb{R}} s\, d\mu = 0\, \mu(E^c) + 1\, \mu(E) = \mu(E).$$

2. *If f is the Dirichlét function, then $\int_{\mathbb{R}} f\, d\mu = 0$.*

3. *If s is a non-negative step function on $[a, b]$, then $\int_{[a,b]} s\, d\mu = \int_a^b s\, dx$. Proof: HW exercise.*

Let us now establish some basic properties of the Lebesgue integral (for simple functions).

Proposition 33. *Suppose $s : X \to [0, \infty)$ is simple and that $E \subseteq X$ is measurable. Then,*

$$\int_E s\, d\mu = \int_X s\chi_E\, d\mu.$$

Proof. Let $t = s\chi_E$. Without loss of generality, we can take $E \neq X$. Then, $\operatorname{ran}(t) = s[E] \cup \{0\}$. Thus, $\operatorname{ran}(t) - \operatorname{ran}(s) \subseteq \{0\}$.

When $a \neq 0$, $t(x) = a$ if and only if $x \in E$ and $s(x) = a$; i.e. $x \in s^{-1}[\{a\}] \cap E$. Thus,

$$\begin{aligned}
\int_X t\, d\mu &= \sum_{a \in \operatorname{ran}(t)} a\mu(t^{-1}[\{a\}] \cap E) \\
&= \sum_{a \in \operatorname{ran}(t) - \{0\}} a\mu(t^{-1}[\{a\}] \cap E) \\
&= \sum_{a \in s[E]} a\mu(s^{-1}[\{a\}] \cap E) \\
&= \sum_{a \in \operatorname{ran}(s)} a\mu(s^{-1}[\{a\}] \cap E)
\end{aligned}$$

□

Lemma 12. *Suppose $s : X \to [0, \infty)$ is a simple function. Suppose E_1, E_2 are measurable sets so that $E_1, E_2 \subseteq X$ and so that $E_1 \cap E_2 = \varnothing$. Then,*

$$\int_{E_1 \cup E_2} s\, \mu = \int_{E_1} s\, d\mu + \int_{E_2} s\, d\mu.$$

Proof.

$$\begin{aligned} \int_{E_1 \cup E_2} s\, d\mu &= \sum_{a \in \operatorname{ran}(s)} a\mu(s^{-1}[\{a\}] \cap (E_1 \cup E_2)) \\ &= \sum_{a \in \operatorname{ran}(s)} a(\mu(s^{-1}[\{a\}] \cap E_1) + \mu(s^{-1}[\{a\}] \cap E_2)) \\ &= \int_{E_1} s\, d\mu + \int_{E_2} s\, d\mu. \end{aligned}$$

□

We now show that the Lebesgue integral is linear, which is one of its most important properties.

Theorem 34. *Suppose $s_1, s_2 : X \to [0, \infty)$ are simple functions.*

1. $\int_X (s_1 + s_2)\, d\mu = \int_X s_1\, d\mu + \int_X s_2\, d\mu$.
2. *If $c \in \mathbb{R}$, then* $\int_X cs_1\, d\mu = c \int_X s_1\, d\mu$. [1]

Proof. Part a:
For each $a \in \operatorname{ran}(s_j)$, let $E_a^{(j)} = s_j^{-1}[\{a\}]$. For each $a \in \operatorname{ran}(s_1)$ and $b \in \operatorname{ran}(s_2)$, let $E_{a,b} = E_a^{(1)} \cap E_b^{(2)}$. Note that:

1. For each $a \in \operatorname{ran}(s_1)$, $E_a^{(1)} = \bigcup_{b \in \operatorname{ran}(s_2)} E_{a,b}$.
2. For each $b \in \operatorname{ran}(s_2)$, $E_b^{(2)} = \bigcup_{a \in \operatorname{ran}(s_1)} E_{a,b}$.
3. $X = \bigcup_{a,b} E_{a,b}$, and
4. $E_{a,b} \cap E_{a',b'} = \varnothing$ if $(a, b) \neq (a', b')$.

[1] If one wants to be pedantic here, they should require that $c > 0$. For the purposes of simplifying the proof of the next theorem, I will leave off this assumption.

By Lemma 12,

$$\int_X (s+t)\,d\mu = \sum_{a,b} \int_{E_{a,b}} (s+t)\,d\mu.$$

On the other hand,

$$\begin{aligned}
\int_{E_{a,b}} (s_1+s_2)\,d\mu &= \sum_{c\in\text{ran}(s+t)} \mu((s_1+s_2)^{-1}[\{c\}]\cap E_{a,b}) \\
&= (a+b)\mu((s_1+s_2)^{-1}[\{a+b\}]\cap E_{a,b}) \\
&= (a+b)\mu(E_{a,b})
\end{aligned}$$

So,

$$\begin{aligned}
\int_E (s_1+s_2)\,d\mu &= \sum_{a,b}(a+b)\mu(E_{a,b}) \\
&= \sum_{a\in\text{ran}(s_1)} \left(\sum_{b\in\text{ran}(s_2)} a\mu(E_{a,b}) \right) + \sum_{b\in\text{ran}(s_2)} \left(\sum_{a\in\text{ran}(s_1)} b\mu(E_{a,b}) \right) \\
&= \sum_{a\in\text{ran}(s_1)} a\mu(E_a^{(1)}) + \sum_{b\in\text{ran}(s_2)} b\mu(E_b^{(2)}) \\
&= \int_E s_1\,d\mu + \int_E s_2\,\mu.
\end{aligned}$$

Part b:
By definition,

$$\begin{aligned}
\int_E cs\,d\mu &= \sum_{ca\in\text{ran}(cs_1)} ca\mu(E_a) \\
&= c\sum_{a\in\text{ran}(s_1)} a\mu(E_a) \\
&= c\int_E s\,d\mu.
\end{aligned}$$

□

We now show that the integral is monotonic, which is another of its essential features.

Theorem 35. *If $s_1 \leqslant s_2$, then $\int_X s_1\,d\mu \leqslant \int_X s_2\,d\mu$.*

Proof. Consider the function $t = s_2 - s_1$. This is positive and simple function, so we have that

$$\begin{aligned}\int_X s_2\, d\mu - \int_X s_2\, d\mu &= \int_X t\, d\mu \\ &= \sum_{a \in \mathrm{ran}(t)} a\mu(E_a) \geqslant 0.\end{aligned}$$

□

Exercise 14. *Rearrange the proof of Theorem 35 to not require the difference of integrals (which we have technically not defined yet).*

4.1 The Lebesgue integral for non-negative functions

We can now define the Lebesgue integral for non-negative functions.

Definition 33. *If $f : X \subseteq [0, \infty]$ is a measurable function, and if E is a measurable subset of X, then the* Lebesgue integral of f over E *is*

$$\sup \left\{ \int_E s\, d\mu \ : 0 \leqslant s \leqslant f \quad \text{and} \quad s \text{ is simple} \right\}.$$

We denote this by

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

Note that if $s : X \to [0, \infty)$ is simple, and if $E \subseteq X$ is measurable, then by Theorem 34, Definitions 32 and 33 yield the same value.

This shows that the definition of the Lebesgue integral is consistent. We can now state two principles for the Lebesgue integral, which are simple consequences of the definition but are also quite useful.

Observation 4. *Suppose $f : X \to [0, \infty]$ is measurable, and suppose $E \subseteq X$ is measurable.*

1. *If $s : X \to \mathbb{R}$ is a simple function so that $0 \leqslant s \leqslant f$, then $\int_E s\, d\mu \leqslant \int_E f\, d\mu$.*
2. *Suppose $\alpha \in [0, \infty]$. $\int_E f\, d\mu \leqslant \alpha$ if and only if $\int_E s\, d\mu \leqslant \alpha$ for every simple $s : X \to \mathbb{R}$ so that $0 \leqslant s \leqslant f$.*

We now establish two propositions about the Lebesgue integral. First we show that the Lebesgue integral of a function on a subset is the same as the integral of the product of the function which the characteristic function

Proposition 36. *Suppose $f : X \to [0, \infty]$ is measurable and that $E \subseteq X$ is measurable. Then, $\int_E f \, d\mu = \int_X f \chi_E \, d\mu$.*

Proof. To do this, we first show that $\int_E f \, d\mu \leqslant \int_X f \cdot \chi_E \, d\mu$.

For this, set $\alpha = \int_X f \cdot \chi_E \, d\mu$ and suppose $s : X \to [0, \infty]$ is a simple function so that $s \leqslant f$.
Then,

$$s \cdot \chi_E \leqslant f \cdot \chi_E.$$

The function $s\chi_E$ is simple so we have that

$$\int_X s \cdot \chi_E \, d\mu \leqslant \int_X f \cdot \chi_E \, d\mu$$

by Definition 33. By Proposition 33, $\int_E s \, d\mu = \int_X s \cdot \chi_E \, d\mu$. So,

$$\int_E s \, d\mu \leqslant \int_X f \chi_E \, d\mu$$

and thus

$$\int_E f \, d\mu \leqslant \int_X f \chi_E \, d\mu.$$

We now show the opposite inequality.

$$\int_X f \cdot \chi_E \, d\mu \leqslant \int_E f \, d\mu.$$

Suppose $s : X \to [0, \infty)$ is a simple function so that $s \leqslant f \cdot \chi_E$. Since f is non-negative, $s \leqslant f$. Thus,

$$\int_E s \, d\mu \leqslant \int_E f \, d\mu.$$

But,

$$\int_E s \, d\mu = \int_X s \, d\mu.$$

So,

$$\int_X s \, d\mu \leqslant \int_E f \, d\mu.$$

□

Proposition 37. *Suppose $f, g : X \to [0, \infty]$ are measurable. If $f \leqslant g$ then $\int_X f \mu \leqslant \int_X g \, d\mu$.*

Proof. Suppose $f \leqslant g$. If s is a simple function so that $0 \leqslant s \leqslant f$, then $s \leqslant g$.

So,

$$\begin{aligned} &\left\{ \int_X s \, d\mu \ : 0 \leqslant s \leqslant f \ \text{ and } \ s \text{ is simple} \right\} \\ \subseteq &\left\{ \int_X s \, d\mu \ : 0 \leqslant s \leqslant g \ \text{ and } \ s \text{ is simple} \right\} \end{aligned}$$

Thus, $\int_E f \, d\mu \leqslant \int_E g \, d\mu$. □

We now establish one of the most important results about the Lebesgue integral, *the monotone convergence theorem*. This is a central and very useful result which shows that we can interchange limits and integration for increasing sequences of functions.

Theorem 38. *(Monotone Convergence Theorem): Suppose $f, f_0, f_1, \ldots : X \to [0, \infty]$ are measurable and that $f_n \leqslant f_{n+1}$ for all $n \in \mathbb{N}$. Suppose $f(t) = \lim_{n\to\infty} f_n(t)$ for all $t \in X$. Then,*

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

Proof. We have that $f_n \leqslant f$ for all $n \in \mathbb{N}$ since $f_n \leqslant f_{n+1}$. Let $\alpha = \lim_{n\to\infty} \int_X f_n \, d\mu$. Then, $\alpha \leqslant \int_X f \, d\mu$.

Claim 1: $\int_X s \, d\mu \leqslant \alpha$ whenever $s : X \to \mathbb{R}$ is a simple function so that $0 \leqslant s < f$.

Proof Claim 1: Suppose $s : X \to \mathbb{R}$ is a simple function so that $0 \leqslant s < f$. For each $n \in \mathbb{N}$, let

$$X_n = \{x \in X \ : \ s(x) \leqslant f_n(x)\}.$$

So, $X = \bigcup_n X_n$ $(s < f)$. Also, $X_n \subseteq X_{n+1}$. On the other hand,

$$\int_{X_n} s \, d\mu \leqslant \int_{X_n} f_n \, d\mu \leqslant \int_X f_n \, d\mu \leqslant \alpha.$$

and

$$\begin{aligned}\lim_{n\to\infty}\int_{X_n} s\,d\mu &= \lim_{n\to\infty}\sum_{a\in\operatorname{ran}(s)} a\mu(s^{-1}[\{a\}]\cap X_n)\\ &= \sum_{a\in\operatorname{ran}(s)} a\mu(s^{-1}[\{a\}]\cap X) = \int_X s\,d\mu.\end{aligned}$$

SO, $\int_X s\,d\mu \leqslant \alpha$.

Claim 2: $\int_X s\,d\mu \leqslant \alpha$ whenever $s : X \to \mathbb{R}$ is a simple function so that $0 \leqslant s \leqslant f$.

Proof Claim 2: Suppose $s : X \to \mathbb{R}$ is a simple function so that $0 \leqslant s \leqslant f$. Let $s_n = (1-2^{-n})s$ for each $n \in \mathbb{N}$. Thus, $0 \leqslant s_n < f$. Hence, $\int_X s_n\,d\mu \leqslant \alpha$. Therefore, $(1-2^{-n})\int_X s_n\,d\mu \leqslant \alpha$ for each $n \in \mathbb{N}$. Thus, $\int_X s\,d\mu \leqslant \alpha$.

Thus, $\int_X f\,d\mu \leqslant \alpha$. □

Using the monotone convergence theorem, we can now show that the Lebesgue integral is linear (we had previously only shown this for simple functions).

Theorem 39. *Suppose $f, g : X \to [0,\infty]$ are measurable.*

1. $\int_X (f+g)\,d\mu = \int_X f\,d\mu + \int_X g\,d\mu$.
2. *If $c \geqslant 0$, then $\int_X cf\,d\mu = c\int_X f\,d\mu$.*

Proof. By the proof of Theorem 30, there exist non-decreasing sequences of non-negative simple functions $\{s_n\}_{n=0}^{\infty}$ and $\{t_n\}_{n=0}^{\infty}$ so that $s_n \to f$ and $t_n \to g$. Apply MCT. □

This has the following corollary, which is essentially a restatement of the Monotone convergence theorem.

Corollary 12. *Suppose $f_0, f_1, \ldots : X \to [0,\infty]$ are measurable and that $f(x) = \sum_{n=0}^{\infty} f_n(x)$ for all $x \in X$. Then,*

$$\int_X f\,d\mu = \sum_{n=0}^{\infty}\int_X f_n\,d\mu.$$

Proof. Set $g_k = \sum_{n=0}^{k} f_n$. Therefore, $0 \leqslant g_k \leqslant g_{k+1}$ and $g_k \to f$. Apply MCT. □

In general, pointwise limits of functions need not preserve the Lebesgue integral. However, there is an inequality which always holds, which is known as Fatou's lemma.

Lemma 13 (Fatou's lemma)**.** *Suppose $f_0, f_1, \ldots : X \to [0,\infty]$ are measurable and that $f(x) = \liminf_n f_n(x)$ for all $x \in X$. Then,*

$$\int_X f\, d\mu \leqslant \liminf_n \int_X f_n\, d\mu$$

Proof. Set $g_k = \inf_{n\geqslant k} f_n$. Thus, $g_k \leqslant f$, $f = \lim_k g_k$, and $g_k \leqslant g_{k+1}$ and $g_k \leqslant f$. Thus,

$$\begin{aligned}\int_X f\, d\mu &= \lim_k \int_X g_k\, d\mu \\ &= \liminf_k \int_X g_k\, d\mu \\ &\leqslant \liminf_k \int_X f_k\, d\mu.\end{aligned}$$

□

To see that Fatou's inequality need not be an equality, consider the following example.

Example 1. *Consider the sequence of functions*

$$f_n = \frac{1}{n}\chi_{[0,n]}.$$

We have that $f_n \to f \equiv 0$ uniformly. However, for all n

$$\int_{\mathbb{R}} f_n\, d\mu = \frac{1}{n}\cdot(n-0) = 1,$$

whereas

$$\int_{\mathbb{R}} f\, d\mu = 0.$$

Before moving on to the Lebesgue integral of functions which can be both positive or negative, let us note one very important inequality in probability, known as Chebychev's inequality.

Theorem 40. *Suppose $f : X \to [0,\infty]$ is measurable. Then, for all $\lambda > 0$,*

$$\mu(\{x \in X \ :\ f(x) \geqslant \lambda\}) \leqslant \frac{1}{\lambda}\int_X f\, d\mu$$

Proof. We can assume that $\int_X f\,d\mu < \infty$, or else the theorem is trivial.

Set $E_\lambda = \{x \in X \ : f(x) \geqslant \lambda\}$.
Set $h = \lambda\chi_{E_\lambda}$. Thus, $0 \leqslant h \leqslant f$. Therefore, $\int_X h\,d\mu \leqslant \int_X f\,d\mu$. But, $\int_X h\,d\mu = \lambda\mu(E_\lambda)$. □

4.2 The Lebesgue integral of real-valued functions

To define the Lebesgue integral of a real-valued function, we restrict ourselves to functions whose absolute value is integrable. This might seem strange initially, but this idea should be somewhat familiar because it is exactly the distinction between convergent series and absolute convergent series.

Definition 34 (Integrable functions)**.** *Suppose $f : X \to [-\infty, \infty]$ is measurable. Suppose E is a measurable subset of X.*

1. *f is* integrable over E *if $\int_E |f|\,d\mu < \infty$.*
2. *f is* integrable *if it is integrable over X.*

For these functions, we can just define the Lebesgue integral of the difference between the integrals of the positive and negative parts.

Definition 35 (Lebesgue integral)**.** *Suppose $f : X \to [-\infty, \infty]$ is integrable and $E \subseteq X$ is measurable. The* Lebesgue integral of f over E *is defined to be*

$$\int_E f^+ d\mu - \int_E f^- \,d\mu.$$

This is denoted by $\int_E f\,d\mu$.

Before moving on, we make two brief remarks.

Observation 5. *1. If $\int_X |f|\,d\mu < \infty$, and if $E \subseteq X$ is measurable, then $\int_E f^+\,d\mu$ and $\int_E f^-\,d\mu$ are finite.*

2. If f is non-negative and integrable, then Definitions 35 and 33 yield same value for $\int_E f\,d\mu$.

Proposition 41. *Suppose $f : X \to [-\infty, \infty]$ is integrable and $E \subseteq X$ is measurable. Then,*

$$\int_E f\,d\mu = \int_X f \cdot \chi_E\,d\mu.$$

Instead of giving a full proof, note that we already know this if $f \geqslant 0$. Otherwise, we set $g = f \cdot \chi_E$. Then we have that $g^+ = f^+ \cdot \chi_E$ and $g^- = f^- \cdot \chi_E$.

We can now show that the integral is monotonic and linear.

Theorem 42. *Suppose $f, g : X \to [-\infty, \infty]$ are integrable and $f \leqslant g$. Then, $\int_X f\, d\mu \leqslant \int_X g\, d\mu$.*

To see this, note that we have already established this if $f \geqslant 0$. Otherwise, note that $f^+ \leqslant g^+$ and $g^- \leqslant f^-$.

Theorem 43. *Suppose $f, g : X \to [-\infty, \infty]$ are integrable and that $\alpha, \beta \in \mathbb{R}$. Then, $\alpha f + \beta g$ is integrable and $\int_X (\alpha f + \beta g)\, d\mu = \alpha \int_X f\, d\mu + \beta \int_X g\, d\mu$.*

Proof. We start by noting that $|\alpha f + \beta g| \leqslant |\alpha||f| + |\beta||g|$. Thus, $\alpha f + \beta g$ is integrable.

Claim 1: $\int_X (f + g)\, d\mu = \int_X f\, d\mu + \int_X g\, d\mu$.

Proof Claim 1: Set $h = f + g$. Then,

$$h^+ - h^- = f^+ - f^- + g^+ - g^-$$

So,

$$h^+ + f^- + g^- = f^+ + g^+ + h^-.$$

Thus,

$$\int_X h^+\, d\mu + \int_X f^+\, d\mu + \int_X g^-\, d\mu = \int_X f^+\, d\mu + \int_X g^+\, d\mu + \int_X h^-\, d\mu.$$

Now rearrange terms.

Claim 2: $\int_X \alpha f\, d\mu = \alpha \int_X f\, d\mu$.

We will skip the proof of this claim.

□

Corollary 13. *Suppose $f : X \to [-\infty, \infty]$ is integrable. Then,*

$$\left| \int_X f\, d\mu \right| \leqslant \int_X |f|\, d\mu.$$

Proof sketch: $f, -f \leqslant |f|$. Apply monotonicity and linearity. This result is known in the book as the Integral Comparison test, where it is stated as follows.

Corollary 14. *Suppose $f : X \to [-\infty, \infty]$ is measurable and*

$$|f| < g$$

for some integrable function g. Then, f is integrable and

$$\left|\int_X f\, d\mu\right| \leqslant \int_X |f|\, d\mu.$$

In any case, we can use this to show the following proposition

Proposition 44. *Suppose $f : X \to [-\infty, \infty]$ is measurable and either integrable or non-negative. If $E \subseteq X$ has measure 0, then $\int_E f\, d\mu = 0$.*

4.2.1 The dominated convergence theorem

We now prove one the most fundamental theorems in measure theory: the dominated convergence theorem.

Theorem 45. *Suppose $f_0, f_1, \ldots : X \to \mathbb{R}$ are measurable and that $f(x) = \lim_{n\to\infty} f_n(x)$ for all $x \in X$. Suppose there is an integrable $g : X \to [-\infty, \infty]$ so that $|f_n(x)| \leqslant g(x)$ for all $x \in X$. Then, f is integrable,*

$$\lim_{n\to\infty} \int_X |f_n - f|\, d\mu = 0, \tag{4.1}$$

and

$$\lim_{n\to\infty} \int_X f_n\, d\mu = \int_X f\, d\mu \tag{4.2}$$

Proof. Since $|f(x)| = \lim_{n\to\infty} |f_n(x)| \leqslant g(x)$. Thus, f is integrable.

To establish Equation 4.1, we note that $|f_n - f| \leqslant 2g$ and that

$$2g = \liminf_n (2g - |f_n - f|).$$

By Fatou's Lemma,

$$\begin{aligned} \int_X 2g\, d\mu &\leqslant \liminf_n \int_X (2g - |f_n - f|)\, d\mu \\ &= \int_X 2g\, d\mu - \limsup_n \int_X |f_n - f|\, d\mu \end{aligned}$$

So,

$$\limsup_n \int_X |f_n - f| \, d\mu \leqslant 0.$$

Thus, $\lim_n \int_X |f_n - f| \, d\mu = 0$.

Equation 4.2 now follows from Equation 4.1. □

As a preliminary application, we can now give a disproportionately sophisticated proof that the harmonic series diverges.

Exercise 15. *Using Example 1, show that the series*

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

diverges. Hint: Consider the function $g(x) = \sup_n f(x)$.

This argument was posted by Tao as an answer on MathOverflow, but I actually find it helpful to remember the dominated convergence theorem and Fatou's inequality.

Theorem 46. *Let* $f : E \to [-\infty, \infty]$ *be integrable. Suppose* $\{E_n\}_{n=0}^{\infty}$ *is a pairwise disjoint sequence of measurable subsets of* E *and set* $E = \bigcup_n E_n$. *Then,*

$$\int_E f \, d\mu = \sum_{n=0}^{\infty} \int_{E_n} f \, d\mu.$$

Proof. Set

$$\begin{aligned} F_n &= \bigcup_{j=0}^{n} E_j \\ f_n &= f \cdot \chi_{F_n} \end{aligned}$$

Thus, $\lim_{n\to\infty} f_n = f$. $|f_n| \leqslant |f|$. So, by DCT

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

But,

$$\int_E f_n \, d\mu = \sum_{k=0}^{n} \int_{E_k} f \, d\mu.$$

□

This implies the following corollary.

Corollary 15. *(Continuity of integration): Suppose $f : X \to [-\infty, \infty]$ is integrable.*

1. *If $E_0 \subseteq E_1 \ldots \subseteq X$ are measurable, and if $E = \bigcup_n E_n$, then*

$$\int_E f \ d\mu = \lim_{n\to\infty} \int_{E_n} f \, d\mu.$$

2. *If $X \supseteq E_0 \supseteq E_1 \supseteq \ldots$, and if $E = \bigcap_n E_n$, then*

$$\int_E f \, d\mu = \lim_{n\to\infty} \int_{E_n} f \, d\mu.$$

4.2.2 Uniform integrability and the Vitali convergence theorem

We conclude our initial discussion of Lebesgue integration by finding a criterion which allows us to interchange limits and integrals. Let us start by making an observation which was actually an earlier homework exercise and says that we can break up sets of finite measure into the disjoint union of sets with small measure.

Lemma 14. *Let E be a set of finite measure and $\delta > 0$. Then E is the disjoint union of a finite collection of sets, each of which has measure less than δ.*

Proof. By the continuity of measure,

$$\lim_{n\to\infty} \mu(E \sim [-n, n]) = \mu(\emptyset) = 0$$

Choose a natural number n_0 for which $\mu\left(E \sim [-n_0, n_0]\right) < \delta$. By choosing a fine enough partition of $[-n_0, n_0]$, express $E \cap [-n_0, n_0]$ as the disjoint union of a finite collection of sets, each of which has measure less than δ. □

Proposition 47. *Let f be a measurable function on E. If f is integrable over E, then for each $\epsilon > 0$, there is a $\delta > 0$ for which if $A \subseteq E$ is measurable and $\mu(A) < \delta$, then*

$$\int_A |f| < \epsilon$$

. Conversely, in the case $\mu(E) < \infty$, if for each $\epsilon > 0$, there is a$\delta > 0$ for which this inequality holds, then f is integrable over E.

Proof. The theorem follows by establishing it separately for the positive and negative parts of f so WLOG we take $f \geqslant 0$ on E. First, we assume f is integrable over E and let $\epsilon > 0$. By the definition of the integral of a nonnegative integrable function, there is a measurable bounded function f_ϵ of finite support which satisfies

$$0 \leqslant f_\epsilon \leqslant f \text{ on } E \text{ and } 0 \leqslant \int_E f - \int_E f_\epsilon < \epsilon/2$$

Since $f - f_\epsilon \geqslant 0$ on E, if $A \subseteq E$ is measurable, then, by the linearity and additivity over domains of the integral,

$$\int_A f - \int_A f_\epsilon = \int_A [f - f_\epsilon] \leqslant \int_E [f - f_\epsilon] = \int_E f - \int_E f_\epsilon < \epsilon/2.$$

But f_ϵ is bounded so we can find $M > 0$ for which $0 \leqslant f_\epsilon < M$ on E_0. Therefore, if $A \subseteq E$ is measurable, then

$$\int_A f < \int_A f_\epsilon + \epsilon/2 \leqslant M \cdot \mu(A) + \epsilon/2 \ .$$

We then take $\delta = \epsilon/2M$. Then we have that if $\mu(A) < \delta$, then

$$\int_A |f| < \epsilon.$$

Conversely, suppose $m(E) < \infty$ and for each $\epsilon > 0$, there is a $\delta > 0$ for which $\mu(A) < \delta$ implies $\int_A |f| < \epsilon$.

Let $\delta_0 > 0$ be such that $\int_A |f| < 1$ whenever $\mu(A) \leqslant \delta_0$. Since $\mu(E) < \infty$, the previous lemma shows that we can consider E as the disjoint union of a finite collection of measurable subsets $\{E_k\}_{k=1}^N$, each of which has measure less than δ_0. Therefore

$$\sum_{k=1}^N \int_{E_k} f < N$$

By the additivity over domains of integration it follows that if h is a nonnegative measurable function of finite support and $0 \leqslant h \leqslant f$ on E, then $\int_E h < N$. Therefore f is integrable. □

With these results in mind, we give the following definition of integrability for a sequence of functions.

Definition 36. *Suppose* $f_0, f_1, \ldots$ *are measurable functions on* X. $\{f_n\}_{n=0}^{\infty}$ *is* uniformly integrable *if for every* $\epsilon > 0$ *there is a* $\delta > 0$ *so that*

$$\int_A |f_n| \, d\mu < \epsilon$$

whenever $n \in \mathbb{N}$ *and* A *is a subset of* X *whose measure is smaller than* δ.

Theorem 48 (Vitali convergence theorem)**.** *Suppose* $\mu(X) < \infty$ *and* $\{f_n\}_{n=0}^{\infty}$ *is a uniformly integrable sequence of measurable functions on* X *that converges pointwise almost everywhere to* f. *Then,* f *is integrable and*

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

We will not cover the proof in class. However, the proof is contained in the book (page 94).

Intuitively, we can think of uniform integrability as being analogous to equicontinuity for a sequence of functions, in that each function in the sequence satisfies a certain estimate uniformly in n. From this perspective, the Vitali convergence theorem is roughly analogous to the Arzela-Ascoli theorem (although for the latter theorem we must pass to a subsequence to find functions which converges uniformly).

4.3 Convergence in measure

This subsection is taken from Section 5.2 of Royden-Fitzpatrick with minimal modifications.

There is one more mode of convergence that we need to know for the qualifying exam, which is *convergence in measure*. This notion is particularly useful in probability theory, because it implies that the measure (i.e., probability) where one function differs greatly from another becomes extremely small.

Definition 37. *Let* $\{f_n\}$ *be a sequence of measurable functions and* f *a measurable function (all defined on* E*), so that* f *and all of the* f_n *are finite a.e. The sequence* $\{f_n\}$ *is said to converge in measure on* E *to* f *provided for each* $\eta > 0$,

$$\lim_{n \to \infty} \mu \{x \in E \mid |f_n(x) - f(x)| > \eta\} = 0$$

It is a good exercise to show that a sequence of (finite a.e.) measurable functions which converge uniformly also converge in measure. However, we can also say something stronger and relate pointwise convergence almost everywhere to convergence in measure.

Proposition 49. *Assume E has finite measure. Let $\{f_n\}$ be a sequence of measurable functions on E that converges pointwise a.e. on E to f and suppose that f is finite a.e. Then $\{f_n\} \to f$ in measure on E.*

Proof. The function f is measurable since it is the pointwise limit almost everywhere of a sequence of measurable functions. To prove convergence in measure we let $\epsilon > 0$ and seek an index N such that

$$\mu\{x \in E | f_n(x) - f(x)| > \eta\} < \epsilon \text{ for all } n \geqslant N,$$

where η is a positive number.

Using Egoroff's Theorem (Theorem 31), we can find a measurable subset F of E with $\mu(E \sim F) < \epsilon$ where $\{f_n\}$ converges to f uniformly. As such, we can find N large enough so that

$$|f_n - f| < \eta \text{ on } F \text{ for all } n \geqslant N.$$

Thus, for $n \geqslant N, \{x \in E \mid \ | f_n(x) - f(x) | > \eta\} \subseteq E \sim F$, which implies that the measure is less than ϵ.

□

This result fails if E has infinite measure (Exercise: find a counterexample). Furthermore, there are sequences which converge in measure which do not converge pointwise, so the converse is false. However, given a sequence of functions which converge in measure, there is a subsequence which converges pointwise almost everywhere.

Theorem 50. *If $\{f_n\} \to f$ in measure on E, then there is a subsequence $\{f_{n_k}\}$ that converges pointwise a.e. on E to f.*

Proof. By the definition of convergence in measure, there is a strictly increasing sequence of natural numbers $\{n_k\}$ for which

$$\mu\{x \in E \mid \ | f_j(x) - f(x) | > 1/k\} < \frac{1}{2^k} \text{ for all } j \geqslant n_k.$$

For each index k, define

$$E_k = \{x \in E \mid \ | f_{n_k} - f(x) | > 1/k\}$$

Since $\mu(E_k) < \frac{1}{2^k}$, we have that $\sum_{k=1}^{\infty} m(E_k) < \infty$. The Borel-Cantelli Lemma then implies that for almost all $x \in E$, there is an index $K(x)$ such that $x \notin E_k$ if $k \geqslant K(x)$, that is,

$$|f_{n_k}(x) - f(x)| \leqslant 1/k \text{ for all } k \geqslant K(x).$$

Therefore

$$\lim_{k\to\infty} f_{n_k}(x) = f(x).$$

In this proof, the index $K(x)$ is allowed to depend on x, as it might not be possible to choose such an index uniformly in x. □

The following proposition shows that many of the theorems in this course can be weakened to require convergence in measure rather than pointwise everywhere convergence.

Proposition 51. *Fatou's Lemma, the Monotone Convergence Theorem, the Lebesgue Dominated Convergence Theorem, and the Vitali Convergence Theorem remain valid if "pointwise convergence a.e." is replaced by "convergence in measure."*

The proof is left as an exercise.

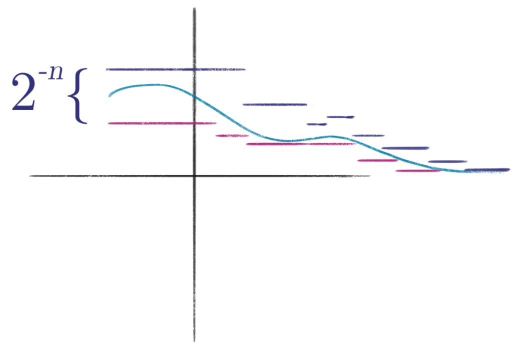
4.4 The Riemann and the Lebesgue integral

At this point, we have defined the Lebesgue integral and discussed some of its fundamental properties. However, it is natural to ask how this integral relates to the Riemann integral, which is the focus of this section. To do so, we start by proving some small lemmas.

Lemma 15. *Suppose $f : [a, b] \to \mathbb{R}$ is Riemann integrable. Then, there is a non-decreasing sequence of step functions $\{s_{1,n}\}_{n=0}^{\infty}$ and a non-increasing sequence of step functions $\{s_{2,n}\}_{n=0}^{\infty}$ so that for all $n \in \mathbb{N}$ $s_{1,n} \leqslant f \leqslant s_{2,n}$ and*

$$\int_a^b (s_{2,n} - s_{1,n}) < 2^{-n}.$$

The proof of this is left as a homework exercise.



We now prove a "sandwich" lemma for these integrals.

Lemma 16. *Suppose $\{\phi_n\}_{n=0}^{\infty}$ is a non-decreasing sequence of integrable function on X, and suppose $\{\psi_n\}_{n=0}^{\infty}$ is a non-increasing sequence of integrable functions on X. Suppose $\phi_n \leqslant f \leqslant \psi_n$ for all $n \in \mathbb{N}$ and that*

$$\lim_{n\to\infty} \int_X (\psi_n - \phi_n) \ d\mu = 0.$$

Then:

1. *$\phi_n \to f$ a.e. and $\psi_n \to f$ a.e.*

2. *f is integrable*

3. $$\lim_{n\to\infty} \int_X \phi_n \, d\mu = \lim_{n\to\infty} \int_X \psi_n \, d\mu = \int_X f \, d\mu.$$

Proof. (a): Let $\phi = \lim_{n\to\infty} \phi_n$ and $\psi = \lim_{n\to\infty} \psi_n$. Thus, $0 \leqslant \psi - \phi \leqslant \psi_n - \phi_n$. Thus, $\int_X (\psi - \phi) \, d\mu = 0$. Therefore, $\psi = \phi$ a.e. Thus, $\phi = f = \psi$ a.e. Thus, f is measurable.

(b): We have $0 \leqslant f - \phi_0 \leqslant \psi_0 - \phi_0$. Thus, $f - \phi_0$ is integrable. Thus, f is integrable.

(c): This follows from monotonicity. □

With these two results in hand, we can now show that the Riemann integral is equal to the Lebesgue integral, whenever the former is well-defined.

Theorem 52. *Suppose $f : [a, b] \to \mathbb{R}$ is Riemann integrable. Then, f is Lebesgue integrable and the two integrals are equal. i.e.*

$$\int_a^b f = \int_{[a,b]} f \, d\mu.$$

Proof. By Lemma 15, there is a non-increasing sequence of step functions $\{s_{2,n}\}_{n=0}^{\infty}$ so that for all $n \in \mathbb{N}$ $s_{1,n} \leqslant f \leqslant s_{2,n}$ and $\int_a^b (s_{n,2} - s_{n,1}) < 2^{-n}$. It follows from Lemma 16 that f is integrable. Since

$$\int_{[a,b]} s_{n,j} \, d\mu = \int_a^b s_{n,j}, \text{ and since } \int_a^b s_{n,1} \leqslant \int_a^b f \leqslant \int_a^b s_{n,2},$$

it follows that

$$\int_a^b f = \lim_{n\to\infty} \int_a^b s_{1,n} = \int_{[a,b]} f \, d\mu.$$

□

As such, all that remains is to determine necessary and sufficient conditions of a function to be Riemann integrable. For this, we have the following result.

Theorem 53. *Suppose $f : [a, b] \to \mathbb{R}$ is bounded. Then, f is Riemann integrable if and only if f is continuous a.e..*

Proof. ($\Rightarrow$): Suppose f is Riemann integrable. By Lemma 15, there is a non-increasing sequence of step functions $\{s_{2,n}\}_{n=0}^{\infty}$ so that for all $n \in \mathbb{N}$ $s_{1,n} \leqslant f \leqslant s_{2,n}$ and $\int_a^b (s_{n,2} - s_{n,1}) < 2^{-n}$.

Let

$$\begin{aligned} E_j &= \{x : \exists n \in \mathbb{N} \ s_{j,n} \text{ is discontinuous at } x\} \ j = 1, 2 \\ E_{j+2} &= \{x : \lim_{n\to\infty} s_{j,n}(x) \neq f(x)\} \ j = 1, 2 \\ E &= E_1 \cup E_2 \cup E_3 \cup E_4 \end{aligned}$$

Thus $E_1 \cup E_2$ is countable. So $\mu(E) = 0$.

Claim: If $x_0 \in [a, b] - E$, then f is continuous at x_0.

Proof Claim: Suppose $x_0 \in [a,b] - E$. Let $\epsilon > 0$. Choose N_0 so that $s_{2,N_0}(x) - s_{1,N_0}(x) < \epsilon$. There exist $\delta > 0$ so that s_{1,N_0} and s_{2,N_0} are constant on $(x_0 - \delta, x_0 + \delta)$. Therefore, if $|x - x_0| < \delta$, then

$$|f(x) - f(x_0)| \leqslant |s_{2,N_0}(x) - s_{1,N_0}(x_0)| < \epsilon.$$

($\Leftarrow$): Suppose f is continuous a.e.. For each n, let P_n denote the uniform partition of $[a,b]$ of width $(b-a)/2^{-n}$. Let $a_j^{(n)}$ denote the j-th point in P_n where $j = 0, \ldots, 2^n$. Set $I_j^{(n)} = (a_j^{(n)}, a_{j+1}^{(n)}]$.

Let

$$\begin{aligned} m_j^{(n)} &= \inf\{f(x') \ : \ x' \in I_j^{(n)}\} \\ M_j^{(n)} &= \sup\{f(x') \ : \ x' \in I_j^{(n)}\} \end{aligned}$$

When $x \in I_j^{(n)}$, set

$$\begin{aligned} \phi_n(x) &= m_j^{(n_0)} \\ \psi_n(x) &= M_j^{(n)}. \end{aligned}$$

We then define the set

$$E = \{x \in [a,b] \ : \ f \text{ is discontinuous at } x\} \bigcup \left\{a_j^{(n)} \ : \ n \in \mathbb{N} \ \text{ and } \ 0 \leqslant j \leqslant 2^n\right\}.$$

Thus, $\mu(E) = 0$.

Claim 1: If $x_0 \in [a,b] - E$, then $\lim_{n\to\infty} \phi_n(x_0) = \lim_{n\to\infty} \psi_n(x_0) = f(x_0)$.

Proof Claim 1: Suppose $x_0 \in [a,b] - E$. Let $\epsilon > 0$. Choose $\delta > 0$ so that

$$|x - x_0| < \delta \ \Rightarrow \ |f(x) - f(x_0)| < \epsilon.$$

Choose $n_0 \in \mathbb{N}$ so that $(b-a)2^{-n_0} < \delta$. Let $n \geqslant n_0$. There is a unique j so that $x \in I_j^{(n)}$. Therefore, $M_j^{(n)} - m_j^{(n)} \leqslant (b-a)2^{-n} \leqslant 2\epsilon$. So, $|s_{2,n}(x) - f(x)|, |s_{1,n}(x) - f(x)| \leqslant 2\epsilon$.

Claim 2: $\lim_{n\to\infty} \int_{[a,b]} (\psi_n - \phi_n)\, d\mu = 0$.

Proof Claim 2: Since f is bounded, there is a positive number M so that $0 \leqslant \psi_n - \phi_n < M$ for all $n \in \mathbb{N}$. Thus, by DCT

$$\lim_{n\to\infty} \int_{[a,b]} (\psi_n - \phi_n)\, d\mu = 0.$$

Claim 3: f is Riemann integrable.

Proof Claim 3:

$$\begin{aligned} \int_{[a,b]} \psi_n \, d\mu &= \int_a^b \psi_n \\ &\geqslant \overline{\int_a^b} f \\ &\geqslant \underline{\int_a^b} f \\ &\geqslant \int_a^b \phi_n = \int_{[a,b]} \phi_n \, d\mu \end{aligned}$$

Therefore,

$$\overline{\int_a^b} f = \underline{\int_a^b} f.$$

□

Although it is genuinely a stronger condition for a function to be Riemann integrable versus Lebesgue integrable, it turns out that the building blocks for each of these integrals are not so different. In fact, we have the following proposition (proof taken from Mathstackexchange [wh])

Proposition 54. *Step functions are dense inside the space of simple functions.*

Proof. It suffices to show that the characteristic of a measurable set of finite measure can be approximated by step functions. Take a measurable set A with $\mu(A) < +\infty$; by regularity, we can find an open set $U \supset A$ such that $\mu(U \backslash A) < \epsilon$.

U can be written as a countable union of disjoint intervals:

$$U = \bigcup_{n=0}^{\infty} I_n$$

so, we can find N such that

$$\mu\left(\bigcup_{n>N} I_n\right) < \epsilon\ .$$

Hence, we define

$$h(x) = \sum_{n=0}^{N} \chi_{I_n}(x)$$

and we have

$$\int |h(x) - \chi_A(x)\| \leqslant 2\epsilon\ .$$

So, the step functions are dense among the simple functions. □

Chapter 5

Differentiation and Integration

The fundamental theorems of calculus provide a foundational relationship between the Riemann integral and the derivative. We now want to establish versions of these theorems for the Lebesgue integral.

5.1 Monotone functions

A function is *monotone* if it is either increasing or decreasing. Such functions play an essential role in the Lebesgue versions of the fundamental theorems of calculus for two reasons.

1. A theorem of Lebesgue asserts that a monotone function on an open interval is differentiable almost everywhere.
2. A theorem of Jordan shows that a very general family of functions (i.e., those of bounded variation) may be expressed as the difference of monotone functions and therefore they also are differentiable almost everywhere on the interior of their domain.

Theorem 55. *Every function that is monotone on an open interval is continuous except at a countable set of points.*

Proof. Suppose I is an open interval, and suppose f is a monotone function on I. Without loss of generality, f is non-decreasing. Set

$$E = \left\{ a \in I \ : \ \lim_{x \to a^-} f(x) \neq \lim_{x \to a^+} f(x) \right\}.$$

For each $a \in E$, let

$$\begin{aligned} \ell_a &= \lim_{x \to a^-} f(x) \\ r_a &= \lim_{x \to a^+} f(x) \end{aligned}$$

Since f is non-decreasing, $\ell_a < r_a$ for all $a \in E$.

Claim 1: if $a, b \in E$, and if $a \neq b$, then $(\ell_a, r_a) \cap (\ell_b, r_b) = \varnothing$.
Proof Claim 1:

Suppose $a, b \in E$ and $a \neq b$. WLOG $a < b$. Since f is non-decreasing, for all $x_0, x_1 \in I$,

$$x_0 < a \text{ and } x_1 > b \Rightarrow f(x_0) \leqslant f(x_1).$$

Thus, $r_a \leqslant \ell_b$.

Claim 2: E is countable.

For each $a \in E$, choose a rational number $q_a \in (r_a, \ell_a)$. By Claim 1, $a \mapsto q_a$ is one-to-one. Thus, E is countable.

□

We want to show that monotone functions are differentiable almost everywhere. For this, we will need to define some concepts related to derivatives.

Definition 38. *Suppose $f : [a, b] \to \mathbb{R}$ is integrable. When $h > 0$ and $x \in [a, b]$, let*

$$\mathrm{Diff}_h(f)(x) = \frac{f(x+h) - f(x)}{h}.$$

We call $\mathrm{Diff}_h(f)$ a divided difference function.

Definition 39. *Suppose $f : A \to \mathbb{R}$ and x_0 belongs to the interior of A.*

1. $\overline{D}(f)(x_0) := \lim_{h \to 0^+} \sup_{0<|t|\leqslant h} \mathrm{Diff}_t(f)(x_0)$. *We call $\overline{D}(f)$ the* upper derivative of f.

2. $\underline{D}(f)(x_0) := \lim_{h \to 0^+} \inf_{0<|t|\leqslant h} \mathrm{Diff}_t(f)(x_0)$. *We call $\underline{D}(f)$ the* lower derivative of f.

If $\overline{D}(f)(x_0) = \underline{D}(f)(x_0)$, *then* f *is* differentiable at x_0 *and we set* $f'(x_0) = \overline{D}(f)(x_0)$.

We can immediately see that $\underline{D}(f)(x_0) \leqslant \overline{D}(f)(x_0)$.

Lemma 17. *Suppose* $f : [a, b] \to \mathbb{R}$ *is non-decreasing. Then:*

1. *For all* $\alpha > 0$,

$$\mu(\{x \in (a, b) \; : \; \overline{D}(f)(x) \geqslant \alpha\}) < \frac{1}{\alpha}(f(b) - f(a)).$$

2. $\mu(\{x \in (a, b) \; : \; \overline{D}(f)(x) = \infty\}) = 0$.

Proof. (1): Suppose $\alpha > 0$. Set

$$E_\alpha = \{x \in (a, b) \; : \; \overline{D}(f)(x) \geqslant \alpha\}.$$

Set

$$\mathcal{F} = \{[c, d] \; : \; [c, d] \subseteq (a, b) \;\text{ and }\; f(d) - f(c) \geqslant \alpha(d - c)\}.$$

Claim 1: $\mathcal{F}$ is a Vitali covering (recall Definition 14) of E_α.

Proof Claim 1: Let $x_0 \in E_\alpha$, and let $\epsilon > 0$. $\overline{D}(f)(x_0) \geqslant \alpha$. So, there is a $\delta > 0$ so that for all $h \in (0, \delta)$, there is a t so that $|t| \leqslant h$ and

$$\frac{f(x_0 + t) - f(x_0)}{t} \geqslant \alpha. \tag{5.1}$$

Choose $h > 0$ so that $h < \delta, \epsilon/2.$ and $[x_0 - h, x_0 + h] \subseteq (a, b)$. Choose t so that $|t| \leqslant h$ and (5.1). Set

$$\begin{aligned} c &= \min\{x_0, x_0 + t\} \\ d &= \max\{x_0, x_0 + t\} \end{aligned}$$

Since f non-decreasing, $f(d) - d(c) \geqslant \alpha(d - c)$. Thus, $[c, d] \in \mathcal{F}$, $x_0 \in [c, d]$, and $d - c < \epsilon$.

Claim 2: $\mu^*(E_\alpha) \leqslant \frac{1}{\alpha}(f(b) - f(a))$.

Proof Claim 2: Let $\epsilon > 0$. By the Vitali Covering Lemma, there exist $[c_0, d_0], \ldots, [c_j, d_j] \in \mathcal{F}$ so that $([c_0, d_0], \ldots, [c_n, d_n])$ is pairwise disjoint and so that

$$\mu^*(E_\alpha - \bigcup_{j=0}^{n} [c_j, d_j]) < \epsilon.$$

Therefore,

$$\begin{aligned} \mu^*(E_\alpha) &\leqslant \epsilon + \sum_{j=0}^{n} (d_j - c_j) \\ &\leqslant \epsilon + \alpha^{-1} \sum_{j=0}^{n} (f(d_j) - f(c_j)) \\ &\leqslant \epsilon + \alpha^{-1}(f(b) - f(a)) \text{ since } f \text{ non-decreasing} \end{aligned}$$

Therefore, $\mu^*(E_\alpha) \leqslant \alpha^{-1}(f(b) - f(a))$.

(2): The second claim follows directly from the first one. □

We can not show that monotone functions are differentiable almost everywhere.

Theorem 56. *Suppose I is an open interval and $f : I \to \mathbb{R}$ is monotone. Then, f is differentiable a.e..*

Proof. Suppose $I = (a, b)$ where $-\infty < a < b < \infty$ and set

$$E = \{x \in (a, b) \ : \ \overline{D}(f)(x) > \underline{D}(f)(x)\}.$$

For all $\alpha, \beta \in \mathbb{Q}$ with $\alpha > \beta$, let

$$E_{\alpha,\beta} = \{x \in (a, b) \ : \ \overline{D}(f)(x) > \alpha > \beta > \underline{D}(f)(x)\}.$$

Therefore, $E = \bigcup_{\alpha,\beta} E_{\alpha,\beta}$.

Claim 1: $\mu^*(E_{\alpha,\beta}) = 0$.

Proof Claim 1: Let $\alpha, \beta \in \mathbb{Q}$ with $\beta > \alpha$ and take $\epsilon > 0$.
Let $\mathcal{O}$ be an open set so that $\mu(\mathcal{O}) \leqslant \mu^*(E_{\alpha,\beta}) + \epsilon$. We then se $\mathcal{F}$ be the collection of closed, bounded intervals $[c, d]$ contained in $\mathcal{O}$ for which $f(d) - f(c) < \beta(d - c)$.

$$\mathcal{F} = \{[c, d] \ : \ [c, d] \subseteq \mathcal{O} \ \text{ and } \ f(d) - f(c) < \beta(d - c)\}.$$

As in the proof of the previous lemma (Lemma 17), $\mathcal{F}$ is a Vitali covering of $E_{\alpha,\beta}$ (since $x \in E_{\alpha,\beta}$ implies that $\underline{D}(f)(x) < \beta$).

By the Vitali Covering Lemma, there exist $[c_0, d_0], \ldots, [c_n, d_n] \in \mathcal{F}$ so that $([c_0, d_0], \ldots, [c_n, d_n])$ is pairwise disjoint and

$$\mu^*(E_{\alpha,\beta} - \bigcup_{j=0}^{n}[c_j, d_j]) < \epsilon.$$

Using Lemma 17, we also have that

$$\mu^*(E_{\alpha,\beta} \cap [c_j, d_j]) \leqslant \alpha^{-1}(f(d_j) - f(c_j)).$$

Combining both of these, we find that

$$\begin{aligned}
\mu^*(E_{\alpha,\beta}) &< \epsilon + \alpha^{-1}\sum_{j=0}^{n}(f(d_j) - f(c_j)) \\
&< \epsilon + \alpha^{-1}\beta\sum_{j=0}^{n}(d_j - c_j) \\
&\leqslant \epsilon + \alpha^{-1}\beta\mu(\mathcal{O}) \\
&\leqslant \epsilon + \alpha^{-1}\beta(\mu^*(E_{\alpha,\beta}) + \epsilon).
\end{aligned}$$

Since ϵ was arbitrary, this shows that

$$\mu^*(E_{\alpha,\beta}) \leqslant \alpha^{-1}\beta\mu^*(E_{\alpha,\beta}).$$

However, since $\alpha^{-1}\beta < 1$ and $\mu^*(E_{\alpha,\beta}) < \infty$,

$$\mu^*(E_{\alpha,\beta}) = 0.$$

□

Our goal now is is prove the fundamental theorem of calculus for monotonic functions. However, before we do so we need to provide two more definitions and prove a small proposition.

Definition 40. *Suppose $f : [a, b] \to \mathbb{R}$ is integrable.*

1. *When $h > 0$ and $x \in [a, b]$, let*

$$\mathrm{Diff}_h(f)(x) = \frac{f(x + h) - f(x)}{h}.$$

We call $\mathrm{Diff}_h(f)$ a divided difference function.

2. When $h > 0$ *and* $x \in [a, b]$*, let*

$$\mathrm{Av}_h(f)(x) = \frac{1}{h} \int_{[x,x+h]} f \ d\mu$$

We call $\mathrm{Av}_h(f)$ *an* average value function.

: In (a) and (b) we take $f(b+h)$ to be $f(b)$.

Proposition 57. *Suppose* $f : [a, b] \to \mathbb{R}$ *is integrable. Then,*

$$\int_u^v \mathrm{Diff}_h(f)\, d\mu = \mathrm{Av}_h(v) - \mathrm{Av}_h(f).$$

whenever $a \leqslant u < v \leqslant b$ *and* $h > 0$*.*

Proof.

$$\int_u^v \mathrm{Diff}_h(f)\, d\mu = \frac{1}{h}\left[\int_u^v f(x+h)\, d\mu(x) - \int_u^v f\, d\mu\right].$$

To see this, note that

$$\int_u^v f(x+h)\, d\mu(x) = \int_{u+h}^{v+h} f\, d\mu.$$

Then note that

$$\int_{u+h}^{v+h} f\, d\mu - \int_u^v f\, d\mu = \int_v^{v+h} f\, d\mu - \int_u^{u+h} f\, d\mu.$$

To see this, note that there are two separate cases, where $u + h < v$ and where $v \leqslant u + h$. Combining these factos together, we find that

$$\int \mathrm{Diff}_h(f)\, d\mu = \mathrm{Av}_h(f)(v) - \mathrm{Av}_h(f)(u).$$

□

Finally, over 80 pages (and over 3000 lines of LaTeX) in, we can (almost) prove the fundamental theorem of calculus (for monotonic functions).

Proposition 58. *Suppose* $f : [a, b] \to \mathbb{R}$ *is non-decreasing. Then,* f' *is integrable and*

$$\int_a^b f' \ d\mu \leqslant f(b) - f(a).$$

Proof. From our previous work, we immediately see the following.

1. f is measurable (applying Theorem 55 since f is monotonic).
2. f is differentiable a.e. (applying Theorem 56).
3. $f' \geqslant 0$. Since f is non-decreasing.

$\int_a^b f' \ d\mu \leqslant \liminf_n [\text{Av}_{2^{-n}}(f)(b) - \text{Av}_{2^{-n}}(f)(a)]$.

Proof Claim 1: We first note that

$$f' = \lim_{n\to\infty} \text{Diff}_{2^{-n}}(f).$$

By Fatou's Lemma and Proposition 57,

$$\begin{aligned} \int_a^b f' \ d\mu &\leqslant \liminf_n \int_a^b \text{Diff}_{2^{-n}}(f) \ d\mu \\ &= \liminf_n [\text{Av}_{2^{-n}}(f)(b) - \text{Av}_{2^{-n}}(f)(a)] \end{aligned}$$

Claim 2: $\text{Av}_{2^{-n}}(f)(b) - \text{Av}_{2^{-n}}(f)(a) \leqslant f(b) - f(a)$.

By definition

$$\text{Av}_{2^{-n}}(f)(b) = 2^n \int_b^{b+2^{-n}} f \ d\mu = 2^n f(b)(b + 2^{-n} - b) = f(b).$$

And, since f is non-decreasing,

$$\text{Av}_{2^{-n}}(f)(a) = 2^n \int_a^{a+2^{-n}} f \ d\mu \geqslant 2^n f(a) 2^{-n} = f(a).$$

□

Note that the inequality in this proposition can be strict. For instance, if ϕ is the Cantor-Lebesgue function, then

$$\int_0^1 \phi' \ d\mu = 0 < \phi(1) - \phi(0).$$

5.2 Functions of bounded variation

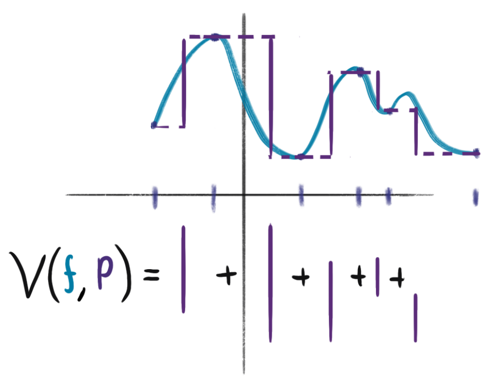
In order to take what we have done for monotonic functions and apply it more generally, we first need to discuss the notion of variation for a function.

Definition 41. *Suppose* $f : [a, b] \to \mathbb{R}$.

1. *If* $P = (x_0, \ldots, x_n)$ *is a partition of* $[a, b]$*, then we define*
$$V(f, P) = \sum_{j<n} |f(x_{j+1}) - f(x_j)|.$$
$V(f, P)$ *is called the* variation of f with respect to P.

2. *The* total variation of f *is*
$$TV(f) := \sup_P V(f, P).$$

3. f *has* bounded variation *if* $TV(f) < \infty$.



We can give some examples and non-examples.

Example. 1. *If* f *is non-decreasing, then* $TV(f) = f(b) - f(a)$.

2. *Lipschitz functions have bounded variation. More precisely, suppose there is a positive real* M *so that* $|f(x) - f(y)| \leqslant M|x - y|$ *whenever* $x, y \in [a, b]$*. Then,* $TV(f) < \infty$.

Non-Example. *When* $0 \leqslant x \leqslant 1$*, let*
$$f(x) = \begin{cases} x\cos(\frac{\pi}{x}) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

For each $n \in \mathbb{N}$, let

$$P_n = (0, \frac{1}{n+1}, \frac{1}{n}, \ldots, \frac{1}{3}, \frac{1}{2}, 1).$$

Then,

$$V(f, P_n) = 1 + 2\sum_{j=2}^{n+1} \frac{1}{j}.$$

So, $TV(f) = \infty$.

The main result about functions with bounded variation that we will use is that they can be written as the difference of two monotonic functions.

Theorem 59. *$f : [a, b] \to \mathbb{R}$ has bounded variation if and only if it can be expressed as a difference of non-decreasing functions.*

We will skip the proof in class, but for completeness I will include the proof in the notes.

Proof. To prove this, we start by defining the total variation function $x \mapsto TV\left(f_{[a,x]}\right)$, which is defined to be the total variation of f restricted to the interval $[a, x]$. We can immediately see several properties of this function.

1. If $a < x < b$, then

   $$TV(f) = TV(f|_{[a,x]}) + TV(f|_{[x,b]}).$$

   (You should convince yourself why this is the case. As a hint, try refining all the partitions with x and see what happens.)

2. Rearranging this equation, we find that

   $$TV\left(f_{[a,v]}\right) - TV\left(f_{[a,u]}\right) = TV\left(f_{[u,v]}\right) \geqslant 0 \text{ for all } a \leqslant u < v \leqslant b,$$

   so the total variation function is increasing.

3. Furthermore, for $a \leqslant u < v \leqslant b$, we have

   $$f(u) - f(v) \leqslant |f(v) - f(u)| = V\left(f_{[u,v]}, P\right) \leqslant TV\left(f_{[u,v]}\right) = TV\left(f_{[a,v]}\right) - TV\left(f_{[a,u]}\right)$$

   which implies that

   $$f(x) + TV\left(f_{[a,x]}\right)$$

   is also an increasing function.

Thus, we can write

$$f(x) = \big(f(x) + TV(f|_{[a,x]})\big) - TV(f|_{[a,x]})$$

as the difference of two increasing functions.

To prove the converse direction, it suffices to note that if $f = g - h$ for non-decreasing functions g and h, then for any partition P of $[a, b]$

$$V(f, P) \leqslant (g(b) - g(a)) + (h(b) - h(a)),$$

which immediately implies that f has bounded total variation. □

Corollary 16. *Every function of bounded variation is differentiable almost everywhere.*

5.3 Absolutely continuous functions

At this point, we have proven the fundamental *inequality* of calculus for a wide class of functions, which are those of bounded variation. However, we have also seen that this inequality need not be an inequality and so it is natural to ask the following question:

"When does the fundamental theorem of calculus hold?"

It turns out that in order to answer this question, we need to define the notion of *absolute continuity.*

Definition 42. *Suppose $f : [a, b] \to \mathbb{R}$. f is* absolutely continuous *if for every $\epsilon > 0$ there is a $\delta > 0$ so that*

$$\sum_{j \leqslant n} |f(y_j) - f(x_j)| < \epsilon$$

whenever (x_0, y_0), …, (x_n, y_n) are disjoint subintervals of $[a, b]$ so that

$$\sum_{j \leqslant n} (y_j - x_j) < \delta.$$

In essence, absolute continuity is a stronger form of uniform continuity where you also require the total change in a function to be small when you take the union of many small intervals (whose total size is less than δ). The main result of this section is the following.

Theorem 60. *Suppose $f : [a, b] \to \mathbb{R}$ is absolutely continuous. Then:*

1. *f is differentiable a.e.*

2. *f' is integrable.*

3. *$\int_{[a,b]} f' \, d\mu = f(b) - f(a)$.*

However, before we prove this, let us first discuss absolute continuity in a bit more detail.

Example. *Every Lipschitz continuous function is absolutely continuous.*

Non-Example. *The Cantor-Lebesgue function is not absolutely continuous.*

Proof. Consider all intervals of the form

$$(0.a_0 a_1 \dots a_n, 0.a_0 \dots a_n 1) \text{ (base 3)}$$

where $a_0, \dots, a_n \in \{0, 2\}$.

The length of each such interval is $1/3^{n+2}$ and there are 2^{n+1} such intervals, so the sum of their lengths is $2^{n+1}/3^{n+2}$ which $\to 0$ as $n \to \infty$. Furthermore, they are non-overlapping.

On the other hand, the change in ϕ across each such interval is $\phi(0.a_0 \dots a_n 1) - \phi(0.a_0 \dots a_n) = 1/2^{n+2}$. So the sum of all these changes in ϕ is $1/2$! Therefore, ϕ is not absolutely continuous. □

Not all functions with bounded variation are absolutely continuous, but all absolutely continuous functions have bounded variation.

Theorem 61. 1. *Every absolutely continuous function has bounded variation.*

2. *In particular, every absolutely continuous function can be expressed as the difference of two non-decreasing absolutely continuous functions.*

Proof. For the first statement, suppose $f : [a, b] \to \mathbb{R}$ is absolutely continuous. Then, there is a $\delta > 0$ so that if (x_0, y_0), …, (x_n, y_n) are subintervals of $[a, b]$ such that

1. $$\sum_{j \leqslant n} (y_j - x_j) < \delta, \text{ and}$$

2. such that $(x_j, y_j) \cap (x_k, y_k) = \varnothing$ whenever $j \neq k$,

then we have that

$$\sum_{j \leqslant n} |f(y_j) - f(x_j)| < 1.$$

Choose $m \in \mathbb{N}$ so that $(b-a)2^{-m} < \delta$.

We now claim that if P is any partition of $[a,b]$, then $V(f,P) < 2^m$. To show this, let P be a partition of $[a,b]$. Let Q denote the uniform partition of $[a,b]$ of width $(b-a)2^{-m}$. Let $I_0, \ldots, I_{2^m-1}$ denote the subintervals of Q. Let $P' = P \cup Q$. Then, $V(f,P) \leqslant V(f,P')$. By the choice of m, δ, $V(f, P' \cap I_j) < 1$ for each j. Thus,

$$V(f,P') = \sum_{j < 2^m} V(f, P' \cap I_j) < 2^m.$$

Thus, f has bounded variation.

To show the second claim, note that f has bounded variation so we set

$$\begin{aligned} g_1(x) &= f(x) + TV(f|_{[a,x]}) \\ g_2(x) &= TV(f|_{[a,x]}) \end{aligned}$$

By our previous work, g_1 and g_2 are non-decreasing.

We now want to show that g_1 and g_2 are absolutely continuous. It suffices to show that g_2 is absolutely continuous. Let $\epsilon > 0$. Choose $\delta > 0$ so that

$$\sum_{j<n} |f(y_j) - f(x_j)| < \epsilon/2$$

whenever (x_0, y_0), ..., (x_n, y_n) are non-overlapping subintervals of $[a,b]$ so that the sum of their lengths is smaller than δ.

Now, suppose (x_0, y_0), ..., (x_n, y_n) are non-overlapping subintervals of $[a,b]$ so that the sum of their lengths is smaller than δ. We want to show that

$$\sum_{j \leqslant n} |g_2(y_j) - g_2(x_j)| < \epsilon.$$

To do so, we use the fact that

$$g_2(y_j) - g_2(x_j) = TV(f|_{[a,y_j]}) - TV(f|_{[a,x_j]}) = TV(f|_{[x_j,y_j]}).$$

If P_j is a partition of $[x_j, y_j]$ for each $j \leqslant n$, then by the choice of δ,

$$\sum_{j \leqslant n} V(f, P_j) < \epsilon/2$$

So,

$$\sum_{j \leqslant n} TV(f|_{[x_j, y_j]}) \leqslant \epsilon/2 < \epsilon.$$

□

With this fact, we can now establish a lemma that gets us most of the way to Theorem 60.

Lemma 18. *Suppose $f : [a, b] \to \mathbb{R}$ is absolutely continuous. Then, $\{\text{Diff}_{2^{-n}}(f)\}_{n=0}^{\infty}$ is uniformly integrable.*

Proof. Suppose f is absolutely continuous. Without loss of generality, we can suppose f is non-decreasing. Let $\epsilon > 0$ and choose $\delta > 0$ as in the definition of absolute continuity for $\epsilon/2$.

Suppose $n \in \mathbb{N}$, and set $h = 2^{-n}$. Now let E is a measurable subset of $[a, b]$ so that $\mu(E) < \delta/2$. Then, there is an open set U so that $U \supseteq E$ and $\mu(U) < \delta$. There is a pairwise disjoint family of bounded open intervals $\{(c_j, d_j)\}_{j \in F}$ so that $U = \bigcup_{j \in F}(c_j, d_j)$. Then:

$$\begin{aligned}
\int_E \text{Diff}_h(f)\, d\mu &= \int_U \text{Diff}_h(f)\, d\mu \\
&= \sum_{j \in F} \int_{(c_j, d_j)} \text{Diff}_h(f)\, d\mu \\
&= \sum_{j \in F} [\text{Av}_h(d_j) - \text{Av}_h(c_j)] \\
&= \frac{1}{h} \sum_{j \in F} \int_0^h [f(d_j + t) - f(c_j + t)]\, d\mu(t)
\end{aligned}$$

If $F' \subseteq F$ is finite, then

$$\begin{aligned}
\frac{1}{h} \sum_{j \in F'} \int_0^h [f(d_j + t) - f(c_j + t)]\ \ d\mu(t) &= \frac{1}{h} \int_0^h \sum_{j \in F'} |f(d_j + t) - f(c_j + t)|\ \ d\mu(t) \\
&< \frac{1}{h} h \frac{\epsilon}{2} = \epsilon/2.
\end{aligned}$$

Thus,

$$\frac{1}{h} \sum_{j \in F} \int_0^h [f(d_j + t) - f(c_j + t)]\, d\mu(t) \leqslant \epsilon/2 < \epsilon.$$

□

Using this lemma, we can now prove the main result from this section.

Proof. The first claim follows immediately from the fact that f has bounded variation.

(2) and (3): By the Lemma, $\{\mathrm{Diff}_{2^{-n}}(f)\}_{n=0}^{\infty}$ is uniformly integrable. Furthermore,

$$\lim_{n\to\infty} \mathrm{Diff}_{2^{-n}}(f) = f' \text{ a.e. by the first claim.}$$

By the Vitali Convergence Theorem, f' is integrable and

$$\int_{[a,b]} f'\,d\mu = \lim_{n\to\infty} \int_{[a,b]} \mathrm{Diff}_{2^{-n}}(f)\,d\mu.$$

However,

$$\lim_{n\to\infty} \int_{[a,b]} \mathrm{Diff}_{2^{-n}}(f)\,d\mu = \lim_{n\to\infty} \mathrm{Av}_{2^{-n}}(f)(b) - \mathrm{Av}_{2^{-n}}(f)(a) = f(b) - f(a).$$

□

5.4 Indefinite Integrals and the Fundamental Theorem of Calculus

We can now use our work to define the notion of the indefinite integral.

Definition 43. *Suppose $f, g : [a,b] \to \mathbb{R}$. We say that f is the* indefinite integral of g *if g is integrable and*

$$f(x) = f(a) + \int_{[a,x]} g\,d\mu$$

for all $x \in [a,b]$.

Theorem 62. *Suppose $f : [a,b] \to \mathbb{R}$. Then, f is absolutely continuous if and only if f is the indefinite integral of a function on $[a,b]$.*

Proof. ($\Rightarrow$): By Theorem 60.

($\Leftarrow$): Suppose $g : [a,b] \to \mathbb{R}$ is integrable and

$$f(x) = f(a) + \int_{[a,x]} g\,d\mu$$

for all $x \in [a, b]$. Let $\epsilon > 0$.

By a homework exercise, there is a $\delta > 0$ so that

$$\int_E |g| \, d\mu < \epsilon$$

whenever E is a measurable subset of $[a, b]$ with $\mu(E) < \delta$. Suppose

$$(x_0, y_0), \ldots, (x_n, y_n)$$

are non-overlapping subintervals of $[a, b]$ so that the sum of their lengths is smaller than δ. Set

$$E = \bigcup_j (x_j, y_j).$$

Then:

$$\begin{aligned} \sum_{j \leqslant n} |f(y_j) - f(x_j)| &= \sum_{j \leqslant n} \left| \int_{[x_j, y_j]} g \, d\mu \right| \\ &\leqslant \sum_{j \leqslant n} \int_{[x_j, y_j]} |g| \, d\mu \\ &= \int_E |g| \, d\mu < \epsilon. \end{aligned}$$

□

For monotone functions, we can state this result a bit differently.

Theorem 63. *Suppose $f : [a, b] \to \mathbb{R}$ is monotone. Then, f is absolutely continuous if and only if*

$$\int_{[a,b]} f' \, d\mu = f(b) - f(a).$$

Proof. Before proving this, it is worth noting that f being monotonic implies that f' is integrable, which implies that the left hand side is well defined.

The forward direction follows immediately from Theorem 60.

For the converse direction, we suppose, without loss of generality, f is non-decreasing.

We need to show f is absolutely continuous. To do so, it suffices to show that f is the indefinite integral of f'. Let $x \in [a, b]$. On the one hand:

$$f(b) - f(a) = \int_{[a,b]} f' \, d\mu = \int_a^x f' \, d\mu + \int_x^b f' \, d\mu$$

But, $f(b) - f(a) = f(b) - f(x) + f(x) - f(a)$. So, we conclude

$$0 = \int_a^x f' \, d\mu - [f(x) - f(a)] + \int_x^b f' \, d\mu - [f(b) - f(x)]$$

However, since f is non-decreasing, by Proposition 58,

$$\int_a^x f' \, d\mu \leqslant f(x) - f(a)$$

$$\int_x^b f' \, d\mu \leqslant f(b) - f(x)$$

Thus, both differences in the above equation are non-positive. So,

$$\int_a^x f' \, d\mu = f(x) - f(a)$$

□

We are almost ready to prove the second form of the fundamental theorem of calculus. Before doing so, we need a small lemma.

Lemma 19. *Suppose $f : [a,b] \to \mathbb{R}$ is integrable Then, $f = 0$ a.e. if and only if*

$$\int_{x_1}^{x_2} f \, d\mu = 0$$

whenever $(x_1, x_2) \subseteq [a,b]$

Instead of giving a full proof of this, note that the forward direction is immediate. For the reverse direction, note that it implies that $\int_U f \, d\mu = 0$ for every open set U. Then, we can approximate measurable sets with open sets to show that

$$\int_E f \, d\mu = 0$$

for any measurable set. Then we can take the positive and negative parts of f and apply Chebychev's inequality (Theorem 40).

Theorem 64. *Suppose $f : [a,b] \to \mathbb{R}$ is integrable. Then,*

$$\frac{d}{dx} \int_a^x f \, d\mu = f(x) \text{ a.e.}$$

Proof. Set

$$F(x) = \int_0^x f \, d\mu$$

So, F is absolutely continuous. Thus, F is differentiable a.e. When $(x_0, x_1) \subseteq [a, b]$, by definition F and Theorem 63

$$\int_{x_0}^{x_1} (F' - f) \, d\mu = 0$$

So by the Lemma, $F' = f$ a.e.

$\square$

5.4.1 The Lebesgue Decomposition

Before concluding this chapter, let me make one further remark about functions of bounded variation. This will probably not be meaningful until you study measure theory in more depth, but it is worth noting now.

Definition 44. *Suppose $f : [a, b] \to \mathbb{R}$. f is* singular *if $f' = 0$ a.e.*

Given any function of bounded variation, it is possible to decompose it as a sum of an absolutely continuous function with a singular function.

Theorem 65. *Suppose $f : [a, b] \to \mathbb{R}$ has bounded variation. Then, f can be written as the sum of an absolutely continuous function and a singular function. This decomposition is known as the* Lebesgue Decomposition.

Proof. Since f has bounded variation, f is differentiable a.e.. Set

$$\begin{aligned} g(x) &= \int_a^x f' \, d\mu \\ h(x) &= f(x) - g(x) \end{aligned}$$

Thus, g is absolutely continuous. $h' = 0$ a.e. and $f = g + h$. $\square$

5.5 A brief remark on Radon-Nikodym derivatives

In this course, we will unfortunately not be able to discuss measures other than the Lebesgue measure. However, the real power of absolute continuity appears when one considers more general measures, so I felt it was necessary to include a brief remark about this.

To begin, we start with a definition of absolute continuity, not for functions, but rather for *measures*.

Definition 45. *A measure μ on Borel subsets of the real line is absolutely continuous with respect to the Lebesgue measure if for every measurable set E, $\mu(E) = 0$ implies $\nu(E) = 0$. This is written as $\nu \ll \mu$.*

There is a straightforward way to create measures which are Lebesgue absolutely continuous. To do so, one takes a non-negative measurable function $f : \mathbb{R} \to \mathbb{R}^+$, and defines the measure to be

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

Example.

1. *The uniform measure on $[0, 1]$ is generated by the function $f = \chi_{[0,1]}$*

2. *The Gaussian measure is generated from the function*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},$$

where μ and σ are parameters.

It turns out that all measures which are absolutely continuous can be expressed in this way.

Theorem 66 (Radon-Nikodym Theorem)**.** *if $\nu \ll \mu$, then there exists a μ-measurable function*[1] *$f : \mathbb{R} \to [0, \infty)$, such that for any measurable set $E \subseteq \mathbb{R}$*

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

In this case, the function f (also written as $\dfrac{d\nu}{d\mu}$ is said to be the Radon-Nikodym derivative of ν with respect to μ.

It is natural to ask what the relationship between absolutely continuous functions and absolutely continuous measures are. In fact, this comes down to the following simple fact.

Proposition 67. *A finite measure ν on Borel subsets of the real line is absolutely continuous with respect to Lebesgue measure if and only if the point function*

$$F(x) = \nu(-\infty, x])$$

is an absolutely continuous real function.

[1]More precisely, a function which is measurable with respect to the associated σ-algebra.

Note that if ν has total mass one, then F is simply the cumulative distribution function. Also note that if you assume Proposition 67, then proving Theorem 66 is essentially just a matter of appealing to Theorem 62, so you can intuitively think about the latter result as being a Radon-Nikodym theorem in disguise.

5.5.1 An even briefer remark on the Lebesgue Decomposition theorem

The motivation for Theorem 65 might seem somewhat mysterious. In fact, this theorem is a special case of a better known theorem for measures, which is the following.

Theorem 68. *Let μ and ν be σ-finite measures on a measurable space $(\Omega, \mathcal{F})$. Then ν can be uniquely decomposed into $\nu = \nu_c + \nu_s$ where $\nu_c \ll \mu$ and $\nu_s \perp \mu$.*

Here, the notation $\nu_s \perp \mu$ means that ν_s is *singular* with respect to μ.

Definition 46. *A measure ν_s is* singular *with respect to μ if it is possible to decompose $\mathcal{F}$ into two disjoint subsets $\mathcal{E}_{\nu_s}$ and $\mathcal{E}_\mu$ so that for any set $A \in \mathcal{E}_{\nu_s}$, $\mu(A) = 0$ and for any $B \in \mathcal{E}_\mu$, $\nu_s(B) = 0$.*

Out of context (and with some of the terms as yet undefined), this result might just be word salad. However, this result, when combined with the Radon-Nikodym theorem, give a good way to understand the relationship between arbitrary measures on reasonable measure spaces.

Chapter 6

Convex functions

We now turn our attention to the notion of convexity.

Definition 47. *Suppose* $f : (a, b) \to \mathbb{R}$ *where* $-\infty \leqslant a < b \leqslant \infty$*.*

1. f *is* convex *if*

$$f(\lambda x_1 + (1-\lambda)x_2) \leqslant \lambda f(x_1) + (1-\lambda)f(x_2)$$

whenever $a < x_1, x_2 < b$ *and* $0 \leqslant \lambda \leqslant 1$*.*

2. f *is* strictly convex *if*

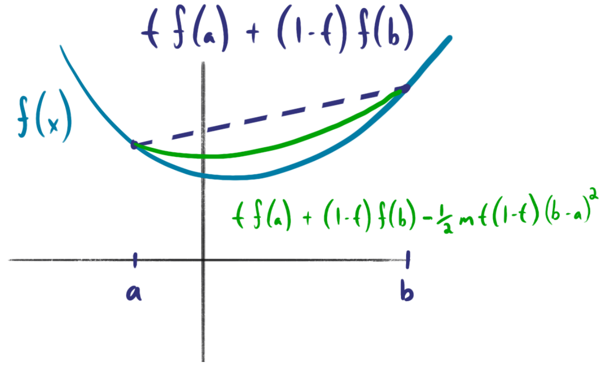
$$f(\lambda x_1 + (1-\lambda)x_2) < \lambda f(x_1) + (1-\lambda)f(x_2)$$

whenever $a < x_1, x_2 < b$ *and* $0 < \lambda < 1$*.*

3. *A function is* strongly convex *if*

$$f(tx + (1-t)y) \leqslant tf(x) + (1-t)f(y) - \frac{1}{2}mt(1-t)(x-y)^2.$$

In other words, convex functions are those which lie below their secant lines, strictly convex functions are those which lie strictly below their secant lines and strongly convex functions lie below a quadratic function



Definition 48. *A function f is* concave *(respectively, strictly, strongly) if $-f$ is convex (strictly convex, strongly convex).*

There are several examples of convex functions which should be familiar.

1. When $p > 0$, $x \mapsto x^p$ strictly convex if $p > 1$, strictly concave if $p < 1$.
2. The exponential function is strictly convex.

We can now show an alternative characterization of convexity in terms of increasing difference quotients.

Proposition 69. *$f : (a, b) \to \mathbb{R}$ is convex if and only if*

$$\frac{f(x) - f(x_1)}{x - x_1} \leqslant \frac{f(x_2) - f(x)}{x_2 - x}$$

whenever $a < x_1 < x < x_2 < b$.

Proof. ($\Rightarrow$): Suppose f is convex. Suppose $a < x_1 < x < x_2 < b$. Then, there exists $0 \leqslant \lambda \leqslant 1$ so that $x = \lambda x_1 + (1 - \lambda)x_2$. Thus,

$$\begin{aligned} \frac{f(x) - f(x_1)}{x - x_1} &\leqslant \frac{\lambda f(x_1) + (1 - \lambda) f(x_2) - f(x_1)}{\lambda x_1 + (1 - \lambda)x_2 - x_1} \\ &= \frac{f(x_2) - f(x_1)}{x_2 - x_1} \end{aligned}$$

Similarly:

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} \leqslant \frac{f(x_2) - f(x)}{x_2 - x}.$$

(⇐): Suppose
$$\frac{f(x)-f(x_1)}{x-x_1} \leqslant \frac{f(x_2)-f(x)}{x_2-x}$$
whenever $a < x_1 < x < x_2 < b$. Suppose $0 \leqslant \lambda \leqslant 1$. We can assume $0 < \lambda < 1$. Set $x = \lambda x_1 + (1-\lambda)x_2$. It follows that
$$\lambda = \frac{x_2-x}{x_2-x_1}$$
and that
$$1-\lambda = \frac{x-x_1}{x_2-x_1}.$$
Since $0 < \lambda < 1$, $x_1 < x < x_2$. Thus,
$$\frac{f(x)-f(x_1)}{x-x_1} \leqslant \frac{f(x_2)-f(x)}{x_2-x}.$$
Therefore
$$\frac{x_2-x}{x_2-x_1}(f(x)-f(x_1)) \leqslant \frac{x-x_1}{x_2-x_1}(f(x_2)-f(x))$$
i.e.
$$\lambda(f(x)-f(x_1)) \leqslant (1-\lambda)(f(x_2)-f(x)).$$
Which rearranges to $f(x) \leqslant \lambda f(x_1) + (1-\lambda)f(x_2)$.

□

6.1 The derivatives and sub-derivatives of convex functions

One nice property of convex functions is that their derivatives satisfy several nice properties. In order to discuss this, we first define the notion of the left and right derivative.

Definition 49. *Suppose* $f : (a,b) \to \mathbb{R}$. *Suppose* $a < x_0 < b$.

1. $\partial_-(f)(x_0) = \lim_{h\to 0^-} \frac{f(x_0+h)-f(x_0)}{h}$.
   $\partial_-(f)$ *is called the* left hand derivative of f.
2. $\partial_+(f)(x_0) = \lim_{h\to 0^+} \frac{f(x_0+h)-f(x_0)}{h}$.
   $\partial_+(f)$ *is called the* right hand derivative of f.

The reason to define these two derivatives is that they are always well defined and increasing for convex functions.

Proposition 70. *Suppose f is convex on (a,b). Then,*

1. *f has left and right hand derivatives at each point in (a,b).*
2. *If $a < u < v < b$, then*

$$\partial_-(f)(u) \leqslant \partial_+(f)(u) \leqslant \frac{f(v)-f(u)}{v-u} \leqslant \partial_-(f)(v) \leqslant \partial_+(f)(v).$$

Proof. (1) : Suppose $x_1 > u$. By Proposition 69,

$$\frac{f(x_1)-f(u)}{x_1-u}$$

decreases as $x_1 \to u$. On the other hand, if $a < c < u$, by Proposition 69 again,

$$\frac{f(u)-f(c)}{u-c} \leqslant \frac{f(x_1)-f(u)}{x_1-u}$$

whenever $x_1 > u$. Thus, $\partial_+(f)(u)$ exists. Similarly, $\partial_-(f)(u)$ exists.

(2): Proof of (1) shows that $\partial_-(f) \leqslant \partial_+(f)$. Suppose $u < x_1 < v$. By Proposition 69,

$$\frac{f(u)-f(x_1)}{u-x_1} \leqslant \frac{f(v)-f(u)}{v-u}.$$

Now, let $x_1 \to u^+$. We obtain

$$\partial_+(f)(u) \leqslant \frac{f(v)-f(u)}{v-u}.$$

We similarly show that

$$\frac{f(v)-f(u)}{v-u} \leqslant \partial_-(f)(v).$$

□

This result has several important corollaries. For instance, it shows that convex functions are Lipschitz.

Corollary 17. *If f is convex on I, and if $[a,b] \subseteq I$, then f is Lipschitz on $[a,b]$.*

Proof. By the Proposition, if $a \leqslant x < y \leqslant b$,

$$\partial_+(f)(a) \leqslant \frac{f(y) - f(x)}{y - x} \leqslant \partial_-(b).$$

Set $M = \max\{|\partial_+(f)(a)|, |\partial_-(f)(b)|\}$. Then, $|f(x) - f(y)| \leqslant M|x - y|$. □

Theorem 71. *Suppose f is convex on I. Then*

1. *f is differentiable except at countably many points of I, and*
2. *f' is non-decreasing.*

Proof. By Proposition 70, $\partial_-(f)$ and $\partial_+(f)$ are non-decreasing. and $\partial_-(f) \leqslant \partial_+(f)$, where the inequality is strict only at the points at which f' does not exist. As such, these points are in one-to-one correspondence with a pairwise disjoint family of open intervals. When $f'(x_0)$ exists, $f'(x_0) = \partial_+(f)(x_0)$. So, f is non-decreasing. □

This has the following corollary.

Corollary 18 (Alexandrov's theorem)**.** *Suppose f is convex on I. Then f is twice differentiable almost everywhere.*

In the proof of Theorem 71, the subdifferential played an important role, so we will investigate it further now.

Definition 50. *The set $\partial f(x) = [\partial_- f(x), \partial_+ f(x)]$ is known as the* sub-differential *of f at x.*

The sub-differential plays a crucial role in the analysis of convex functions. The standard definition of the sub-differential is the following.

$$\partial f(x) = \bigcap_{z \in \operatorname{dom} f} \{v \mid f(z) \geqslant f(x) + v \cdot (z - x)\}$$

Note that by considering v as a vector in Euclidean space, the sub-differential is well defined for functions $f : \mathbb{R}^n \to \mathbb{R}$. It turns out that the sub-differential is always a closed and convex set. This is obvious for convex functions in $\mathbb{R}$, but holds true more generally as well.

6.2 Jensen's inequality

In this course, and on the qualifying exams, one of the best tools that you have involving convex functions is Jensen's inequality. As such, we will spend some time discussing it now.

To begin, we note that convex functions lie above their tangent lines (where tangent is defined to be any point in the sub-differential).

Lemma 20. *Suppose f is convex on (a, b) and that $a < x_0 < b$. If $\partial_-(f)(x_0) \leqslant m \leqslant \partial_+(f)(x_0)$, then $f(x) \geqslant m(x - x_0) + f(x_0)$.*

Using this, we can now state and prove Jensen's inequality.

Theorem 72. *(Jensen's Theorem): If ϕ is a convex function on $\mathbb{R}$, and if f, $\phi \circ f$ integrable on E where $\mu(E) = 1$, then*

$$\phi\left(\int_E f\, d\mu\right) \leqslant \int_E \phi \circ f\, d\mu.$$

Proof. Set $\alpha = \int_E f\, d\mu$. Suppose $\partial_-(\phi)(\alpha) \leqslant m \leqslant \partial_+(f)(\alpha)$. By Lemma 20,

$$\phi(f(x)) \geqslant m(f(x) - \alpha) + \phi(\alpha).$$

Now, take the integral of both sides. □

Example: If $f : [0, 1] \to \mathbb{R}$ is integrable, then

$$\left(\int_0^1 f\, d\mu\right)^2 \leqslant \int_0^1 f^2\, d\mu.$$

Question 2 (2011 Analysis Qualifying Exam, Problem 3)**.** *Let $f : [0, 1] \to \mathbb{R}$ be continuous with $f(x) > 0$ for $x \in [0, 1]$. Show that*

$$\exp\left(\int_0^1 \log f\right) \leqslant \int_0^1 f.$$

6.3 A brief remark on convex duality

Any discussion of convex functions would be incomplete without some mention of convex duality. One of the fundamental properties of convex functions is that they come in pairs.

Definition 51 (Legendre transform)**.** *Let $I \subset \mathbb{R}$ be an interval, and $f : I \to \mathbb{R}$ a function. The* Legendre transform *of f (denoted f^*) is the function $f^* : I^* \to \mathbb{R}$ defined by*

$$f^*(x^*) = \sup_{x \in I} (x^* x - f(x)), \quad x^* \in I^*$$

where the domain I^* is defined to be

$$I^* = \left\{ x^* \in \mathbb{R} : \sup_{x \in I} (x^* x - f(x)) < \infty \right\}$$

Proposition 73. *The Legendre transform of any function is a convex function.*

Proof. Let p and q be two points in the domain I^* where f^* is defined and let t be within $[0, 1]$. Then,

$$\begin{aligned} f^*(tp + (1-t)q) &= \sup_x \{x[tp + (1-t)q] - f(x)\} \\ &= \sup_x \{t[xp - f(x)] + (1-t)[xq - f(x)]\} \\ &\leqslant t \sup_x \{xp_1 - f(x)\} + (1-t) \sup_x \{xq - f(x)\} \\ &= tf^*(p) + (1-t) f^*(q). \end{aligned}$$

□

For convex functions, it turns out that this transformation is involutive. We will not prove this here because I was not able to find a good proof using only what we've covered thus far.

Proposition 74. *A function is convex iff $f^{**} = f$.*

There is a lot more that can be said about the Legendre transform. For instance, there is a deep relationship between the sub-gradients of the Legendre pair.

Furthermore, this duality appears throughout mathematics. To give a short and very incomplete list, the it shows up in

1. Mathematical Physics (Classical mechanics, thermodynamics ...)

2. Optimal transport (see Chapter 5 of [Vil09])

3. Mirror symmetry (in particular, T-duality [Leu05]).

Chapter 7

Basics of functional analysis

In analysis, we will often want to establish quantitative control over our functions. For instance, if we are studying a surface of the ocean, we want to understand how large the waves are and how smooth or rough the chop is. In order to do that, it is necessary to develop a more complete theory of how to bound functions. In this chapter, we cover some preliminary ideas in this direction, which lay the groundwork for modern functional analysis.

7.1 Normed linear spaces and Banach spaces

Definition 52. *Suppose V is a vector space over $\mathbb{R}$. A* norm *on V is a function $\|\cdot\| : V \to [0, \infty)$ so that*

- $\|\alpha v\| = |\alpha|\|v\|$ *for all $\alpha \in \mathbb{R}$ and all $v \in V$. (Homogeneity)*
- *For all $v \in V$, if $\|v\| = 0$, then $v = \mathbf{0}$. (Positivity)*
- *For all $u, v \in V$, $\|u + v\| \leqslant \|u\| + \|v\|$. (The triangle inequality)*

Definition 53. *A* normed linear space *consists of a vector space together with a norm.*

Example. *1. $\mathbb{R}^n$ with its usual norm is a normed linear space.*

2. Let $C[0, 1]$ denote the space of all continuous functions

$$f : [0, 1] \to \mathbb{R}.$$

$C[0, 1]$ is a vector space under the usual pointwise addition and scalar multiplication. For each $f \in C[0, 1]$, we let

$$\|f\|_{\text{sup}} = \max\{|f(t)| \ : \ t \in [0, 1]\}.$$

It follows that $\|f\|_{\sup}$ is a norm on $C[0,1]$.

3. *Integrable functions with the norm*

$$\|f\|_1 = \int |f| \, d\mu.$$

The reason for the "1" in the subscript of the previous norm will soon become clear. The examples that we will be interested in are mostly function spaces such as the latter two spaces, as much of what will say is trivial for finite dimensional spaces.

The main reason to study normed linear spaces is that they induce a notion of convergence.

Definition 54. *Suppose V is a normed linear space, and suppose $\{v_n\}_{n=0}^{\infty}$ is a sequence of vectors.*

1. *We say that $\{v_n\}_{n=0}^{\infty}$* converges in V *(or converges in norm) if there is a vector $v \in V$ so that $\lim_{n\to\infty} \|v_n - v\| = 0$.*

2. *We say that $\{v_n\}_{n=0}^{\infty}$ is* Cauchy *if for every $\epsilon > 0$ there is an $N \in \mathbb{N}$ so that $\|v_n - v_m\| < \epsilon$ whenever $m, n \geqslant N$.*

In this class we will not cover notions of convergence other than convergence in norm, but you may well do so in future classes so we will use the correct terminology from the start.

There is a relationship between convergent sequences and Cauchy sequences, which should hopefully be familiar from a previous course in analysis.

Proposition 75. *Suppose V is a normed linear space, and suppose $\{v_n\}_{n=0}^{\infty}$ is a sequence of vectors in V*

1. *If $\{v_n\}_{n=0}^{\infty}$ converges in V, then there is exactly one vector v in V so that $\lim_{n\to\infty} \|v_n - v\| = 0$; denote this vector by $\lim_{n\to\infty} v_n$.*

2. *If $\{v_n\}_{n=0}^{\infty}$ converges in V, then it is Cauchy.*

3. *Suppose $\{v_n\}_{n=0}^{\infty}$ is Cauchy. If a subsequence of $\{v_n\}_{n=0}^{\infty}$ converges in V, then $\{v_n\}_{n=0}^{\infty}$ converges in V.*

Definition 55. *A normed linear space is* complete *if all of its Cauchy sequences converge; in this case we say that it is a* Banach space.

Example. *$C[0,1]$ with the supremum norm is a Banach space.*

Exercise 16. *Show this fact. Hint: Use uniform convergence.*

Non-Example. $C[0,1]$ *with the norm*

$$\|f\|_{L^2} = \sqrt{\int f^2 \, d\mu}$$

is an incomplete normed linear space.

Exercise 17. *Find a Cauchy sequence in this space which does not converge.*

Before moving on, let us mention two lemmas which will be useful some future considerations.

Lemma 21. *Suppose V is a normed linear space, and suppose $\{v_n\}_{n=0}^{\infty}$ is a Cauchy sequence of vectors in V. Suppose $\{a_j\}_{j=0}^{\infty}$ is a sequence of positive reals that converges to 0. Then, there exist $n_0 < n_1 < \ldots$ so that $\|v_{n_{2j}} - v_{n_{2j+1}}\| < a_j$ for all $j \in \mathbb{N}$.*

The proof is left as a small exercise.

Lemma 22. *Suppose $\{v_n\}_{n=0}^{\infty}$ is a convergent sequence of vectors in a normed linear space V and that v denotes its limit. Then, $\lim_{n\to\infty} \|v_n\| = \|v\|$.*

Proof sketch: Observe that

$$|\|v_n\| - \|v\|| \leqslant \|v_n - v\|.$$

7.2 L^p spaces

Definition 56. *Suppose E is a measurable set of reals. Suppose $1 \leqslant p < \infty$.*

1. *$L^p(E)$ consists of all measurable $f : E \to [-\infty, \infty]$ so that*

$$\int_E |f|^p \, d\mu < \infty$$

2. *When $f \in L^p(E)$, let*

$$\|f\|_p = \left(\int_E |f|^p \, d\mu \right)^{1/p}$$

Before moving on, we make a few remarks. First, note that $L^1(E)$ consists of all integrable functions on E.

Furthermore, when $f, g \in L^p(E)$ and $f = g$ a.e., we identify f and g. As such, L^p spaces are really equivalence classes of functions rather than functions.

There is also an Ł$^\infty$-norm. However, in order to define it, we need to define some more terminology

Definition 57. *Suppose $S \subseteq \mathbb{R}$.*

1. *$M \in (-\infty, \infty]$ is an* essential upper bound on S *if $M > a$ for almost all $a \in S$.*

2. *The* essential supremum of S *is the greatest lower bound of the set of all essential upper bounds of S.*

Exercise 18. *To make sense of this defintion, consider the following space*

$$[0, 1) \cup \{3, 4, 5\}.$$

What is its essential supremum?

With this definition, we can define the L^∞ norm.

Definition 58. *Suppose E is a measurable set of reals.*

1. *$L^\infty(E)$ consists of all measurable $f : E \to [-\infty, \infty]$ so that the essential supremum of* $\operatorname{ran}(|f|)$ *is finite.*

2. *When $f \in L^\infty(E)$, let $\|f\|_\infty$ denote the essential supremum of f.*

The relationship between L^p functions and L^∞ ones can be a bit subtle. Here is an exercise to help think about it.

Exercise 19. *Suppose f is measurable and satisfies*

$$\int_0^1 \exp(|f(x)|)dx < \infty$$

Prove that $f \in L^p([0, 1])$ for all $1 \leqslant p < \infty$. Is it true that such an f must necessarily belong to $L^\infty([0, 1])$?

7.3 Conjugate exponents and some fundamental inequalities

One of the most important features of the L^p-spaces is that they come in conjugate pairs and satisfy several fundamental inequalities.

Definition 59. *Suppose* $1 \leqslant p, q \leqslant \infty$. *We say that* p, q *are* conjugate *if* $\frac{1}{p} + \frac{1}{q} = 1$.

Example. *1.* 1 *and* ∞ *are conjugate.*

2. 2 *and* 2 *are conjugate.*

3. 3 *and* $\frac{3}{2}$ *are conjugate.*

Using some algebra, we can observe two basic, but very important observations

Observation 6. *1. If* $q, p \in \mathbb{R}$ *are conjugate then* $q = p/(p-1)$.

2. If p, q *are conjugate, and if* $p > 2$, *then* $1 \leqslant q < 2$.

7.3.1 Young's inequality and the Peter-Paul version

Theorem 76 (Young's Inequality)**.** *If* p, q *are conjugate then*

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

for all reals $a, b \geqslant 0$.

Proof. Without loss of generality, we can consider $a, b \neq 0$. Set: $s_1 = p\ln(a)$ and $s_2 = q\ln(b)$. Since exp is convex,

$$\begin{aligned} e^{s_1/p + s_2/q} &\leqslant \frac{1}{p}e^{s_1} + \frac{1}{q}e^{s_2} \\ &= \frac{1}{p}a^p + \frac{1}{q}b^q \end{aligned}$$

But, $ab = e^{s_1/p + s_2/q}$. □

Oftentimes in analysis, it will be useful to use a slightly different version of this inequality, which is known as the Peter-Paul inequality. The basic idea is that you "rob Peter in order to pay Paul". In other words, you gain tighter control of the second term at the expense at the cost of losing some control of the first term. For the case $p = q = 2$, this inequality is the following.

$$ab \leqslant \frac{a^2}{2\varepsilon} + \frac{\varepsilon b^2}{2}. \tag{7.1}$$

7.3.2 Holder's Inequality

Theorem 77. *(Hölder's inequality): Suppose E is a measurable set of reals. Suppose p, q are conjugate, $f \in L^p(E)$, and $g \in L^q(E)$. Then, $fg \in L^1(E)$ and*

$$\|fg\|_1 \leqslant \|f\|_p \|g\|_q$$

Proof. We assume that $0 < \|f\|_p, \|f\|_q$, or else the proof is trivial.

Case 1: If one of p, q is infinite, then the inequality holds directly from the monotonicity of the integral.‘

Case 2: $p, q < \infty$.

Claim 1:

$$\frac{|f|}{\|f\|_p}\frac{|g|}{\|g\|_q} \leqslant \frac{1}{p}\frac{|f|^p}{\|f\|_p^p} + \frac{1}{q}\frac{|g|^q}{\|g\|_q^q}.$$

Proof Claim 1: By Young's Inequality.

Claim 2:

$$\int_E \frac{|f||g|}{\|f\|_p \|g\|_q} \leqslant 1.$$

Proof Claim 2: By Claim 1,

$$\begin{aligned} \int_E \frac{|f|}{\|f\|_p}\frac{|g|}{\|g\|_q}\, d\mu &\leqslant \frac{1}{p}\int_E \frac{|f|^p}{\|f\|_p^p}\, d\mu + \int_E \frac{1}{q}\frac{|g|^q}{\|g\|_q^q}\, d\mu \\ &= \frac{1}{p} + \frac{1}{q} = 1. \end{aligned}$$

□Claim 2·

The theorem now follows directly from Claim 2.

□

Holder's inequality is saturated in by the conjugate function, which is the following.

Proposition 78. *If f is not identically* 0, *the function* [2] $f^* = \|f\|_p^{1-p} \cdot \text{sgn}(f) \cdot |f|^{p-1}$ *belongs to* $L^q(X, \mu)$ *and satisfies*

$$\int_E f \cdot f^* = \|f\|_p \text{ and } \|f^*\|_q = 1.$$

Here, $\text{sgn}(f)$ *is the function which is* 1 *whenever* f *is non-negative and* -1 *whenever* f *is negative so that* $\text{sgn}(f) \cdot f = |f|$ *a.e.*

This is a favorite topic of the people who write the qualifying exam.

Exercise 20 (Fall 2021 Qual: Problem 4)**.** *Let* $f \in L^p(\mathbb{R}), \|f\|_p \neq 0$. *Prove that there exists a* $g \in L^q(\mathbb{R})$ *such that* $\|g\|_q = 1$ *and*

$$\int_{\mathbb{R}} f(x)g(x)dx = \frac{1}{2}\|f\|_p.$$

The case of Holder's inequality with $p = q = 2$ is of particular importance, and is better known as the Cauchy-Schwarz inequality.

Corollary 19 (Cauchy-Schwarz inequality)**.** *If* E *is a measurable set of reals, and if* $f, g \in L^2(E)$, *then* $\|fg\|_1 \leqslant \|f\|_2 \|g\|_2$.

7.3.3 Minkowski's inequality

Theorem 79. *Suppose* E *is a measurable set of reals,* $1 \leqslant p \leqslant \infty$, *and* $f, g \in L^p(E)$. *Then,*

$$\|f + g\|_p \leqslant \|f\|_p + \|g\|_p$$

Proof. WLOG $p < \infty$. Let q be conjugate of p.

Claim 1: $(f + g)^p = f(f + g)^{p/q} + g(f + g)^{p/q}$.

To see this, note that since p, q are conjugate, $p - 1 = \frac{p}{q}$. So

$$\begin{aligned} (f + g)^p &= f(f + g)^{p-1} + g(f + g)^{p-1} \\ &= f(f + g)^{p/q} + g(f + g)^{p/q}. \end{aligned}$$

Claim 2:

$$\|f + g\|_p^p \leqslant \||f + g|^{p/q}\|_q \|f\|_p + \|g\|_p).$$

Proof Claim 2: By Claim 1 and the triangle inequality,

$$|f + g|^p \leqslant |f||f + g|^{p/q} + |g||f + g|^{p/q}.$$

Thus,

$$\int_E |f+g|^p \, d\mu \leqslant \int_E |f||f+g|^{p/q} + \, d\mu \int_E |g||f+g|^{p/q} \, d\mu$$

Note that $|f+g|^{p/q} \in L^q(E)$. So, by Hölder's Inequality, $|f||f+g|^{p/q} \in L^1(E)$ and $|g||f+g|^{p/q} \in L^1(E)$ and

$$\begin{aligned}
\int_E |f||f+g|^{p/q} \, d\mu + \int_E |g||f+g|^{p/q} \, d\mu &= \||f||f+g|^{p/q}\|_1 + \||g||f+g|^{p/q}\|_1 \\
&\leqslant \|f\|_p \||f+q|^{p/q}\|_q + \|g\|_p \||f+g|^{p/q}\|_q.
\end{aligned}$$

Finally, we can show the desired inequality.

$$\|f+g\|_p \leqslant \|f\|_p + \|g\|_p.$$

To see this, note that by Claim 2,

$$\|f+g\|_p^p \||f+g|^{p/q}\|_q^{-1} \leqslant \|f\|_p + \|g\|_p$$

But,

$$\|f+g\|_p^p \||f+g|^{p/q}\|_q^{-1} = \left(\int_E (f+g)^p \, d\mu \right)^{1-1/q} = \|f+g\|_p.$$

□

This has the following important corollary, which is that L^p spaces form a normed linear space.

Corollary 20. *If E is a measurable set of reals, and if $1 \leqslant p \leqslant \infty$, then $L^p(E)$ is a normed linear space.*

Before moving on, it is worthwhile to write out some functions which are in various L^p-spaces.

Example. *1. Set $E = (0,1]$. Suppose $1 \leqslant p_1 < p_2 < \infty$. Choose α so that $-1/p_1 < \alpha < -1/p_2$. Set $f(x) = x^\alpha$ when $0 < x \leqslant 1$. It follows that $f \in L^{p_2}(E) - L^{p_1}(E)$.*

2. Set $E = (0,\infty)$. For all $x > 0$, let

$$f(x) = \frac{x^{-1/2}}{1+\ln(x)}$$

It follows that $f \in L^p(E)$ iff $p = 2$. Proof sketch: use change of variables $u = 1+\ln(x)$. When $p > 2$, $\int_1^\infty |f(x)|^p dx = \infty$. When $p < 2$, $\int_0^1 |f(x)|^p dx = \infty$.

7.3.4 Embeddings of L^p-spaces

Theorem 80. *Suppose $\mu(E) < \infty$ and $1 \leqslant p_1 < p_2 < \infty$. Then, $L^{p_2}(E) \subseteq L^{p_1}(E)$.*

Proof. Set $p = p_2/p_1 > 1$. Let $f \in L^{p_2}(E)$. Thus, $|f|^{p_1} \in L^p(E)$. Set $q = \frac{p-1}{p}$ so that p, q are conjugate. Set $g = \chi_E$. Since $\mu(E) < \infty$, $g \in L^q(E)$. So,

$$\begin{aligned} \int_E |f|^{p_1} \, d\mu &= \int_E |f|^{p_1} g \, d\mu \\ &\leqslant \| |f|^{p_1} \|_p \|g\|_q \\ &= \|_{\|} f^{p_1}_{p_2} \mu(E)^q < \infty. \end{aligned}$$

Thus, $f \in L^{p_1}(E)$.

□

7.3.4.1 Interpolation of L^p spaces

On $\mathbb{R}$, none of the L^p spaces are subspaces of others. However, there is a foundational result which states that if $f \in L^{p_0} \cap L^{p_1}$ for $p_0 < p_1$, it is also in L^q for all q with $p_0 < q < p_1$ and we can estimate the L^q norm in terms of the L^{p_0}- and L^{p_1}-norms.

To explain this result, we first introduce some notation.

Definition 60. *Let p_0, p_1 be two numbers such that $0 < p_0 < p_1 \leqslant \infty$. Then for $0 < \theta < 1$ define p_θ by: $\frac{1}{p_\theta} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}$.*

Theorem 81. *Each $f \in L^{p_0} \cap L^{p_1}$ satisfies:*

$$\|f\|_{p_\theta} \leqslant \|f\|_{p_0}^{1-\theta} \|f\|_{p_1}^{\theta}.$$

Proof. Consider $|f| = |f|^\theta |f|^{1-\theta}$ and apply Holder's inequality:

$$\begin{aligned}
\int |f|^{p_\theta} &= \int \left(|f|^\theta |f|^{1-\theta}\right)^{p_\theta} &(7.2)\\
&= \int |f|^{\theta p_\theta} |f|^{(1-\theta)p_\theta} &(7.3)\\
&\leqslant \left(\int \left(|f|^{\theta p_\theta}\right)^{\frac{p_1}{\theta p_\theta}}\right)^{\frac{\theta p_\theta}{p_1}} \left(\int \left(|f|^{(1-\theta)p_\theta}\right)^{\frac{p_0}{(1-\theta)p_\theta}}\right)^{\frac{(1-\theta)p_\theta}{p_0}} &(7.4)\\
&= \left(\int |f|^{p_1}\right)^{\frac{\theta p_\theta}{p_1}} \left(\int |f|^{p_0}\right)^{\frac{(1-\theta)p_\theta}{p_0}} &(7.5)\\
&= \left(\int |f|^{p_1}\right)^{\frac{\theta p_\theta}{p_1}} \left(\int |f|^{p_0}\right)^{\frac{(1-\theta)p_\theta}{p_0}} &(7.6)\\
&= \|f\|_{p_1}^{\theta p_\theta} \|f\|_{p_0}^{(1-\theta)p_\theta} &(7.7)
\end{aligned}$$

Then take the p_θ-th roots of both sides. □

7.4 L^p and pointwise convergence

We can now discuss the relationship between convergence in L^p and usual pointwise convergence of functions.

Theorem 82. *Suppose $1 \leqslant p \leqslant \infty$ and E is a measurable set of reals. Then, every Cauchy sequence of vectors in $L^p(E)$ has a subsequence that converges pointwise almost everywhere.*

Proof. Suppose $\{f_n\}_{n=0}^\infty$ is a Cauchy sequence of vectors in $L^p(E)$. By Lemma 21, there is a sequence $n_0 < n_1 < \ldots$ so that $\|f_{n_j} - f_{n_{j+1}}\|_p^p < 2^{-2j}$ for all j.

For each $j \in \mathbb{N}$, let

$$E_j = \{x \in E \ : \ |f_{n_j}(x) - f_{n_{j+1}}(x)|^p > 2^{-j}\}.$$

Claim 1: $\lim_j f_{n_j}(x)$ exists if $x \in E_j$ for only finitely many j.

Proof Claim 1: Suppose $x \in E_j$ for only finitely many j. Let $\epsilon > 0$. Choose k so that $x \notin E_j$ for all $j \geqslant k$ and $2^{-k+1} < \epsilon$. Suppose $j_0, j_1 \geqslant k$ and $j_1 > j_0$. Then, by the triangle inequality

$$\begin{aligned}
|f_{n_j}(x) - f_{n_{j+1}}(x)| &\leqslant \sum_{j=j_0}^{\infty} 2^{\frac{-j}{p}}\\
&= 2^{-j_0+1} \leqslant 2^{-k+1} < \epsilon.
\end{aligned}$$

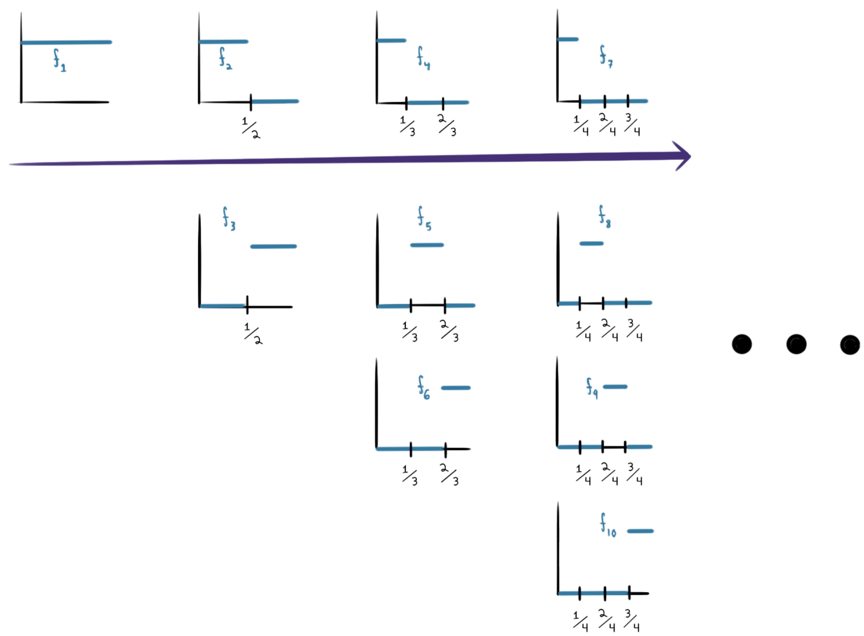


Figure 7.1: A sequence converging in L^p but which does not converge pointwise and a pointwise convergent subsequence.

Claim 2: For almost every $x \in E$, $x \in E_j$ for only finitely many j.

Proof Claim 2: By Chebychev,

$$\mu(E_j) \leqslant 2^j \int_E |f_{n_j} - f_{n_{j+1}}|^p \, d\mu < 2^{-j}.$$

So $\sum_j \mu(E_j) < \infty$. By Borel-Cantelli, Claim 2 follows.

□

Note that the passing to a subsequence might be necessary. It is possible to find sequences which converge to zero in L^p but do not converge pointwise.

Similarly, it is possible to construct families of functions which converge pointwise but do not converge in L^p.

Example 2. *1.* $f_n = \chi_{[n,n+1]}$

2. $f_n = n\chi_{(0,\frac{1}{n}]}$

Lemma 23. *Suppose E is a measurable set of reals. Suppose $1 \leqslant p < \infty$. Suppose $\sum_{n=0}^{\infty} \|f_n\|_p < \infty$. Then:*

1. $\sum_{n=0}^{\infty} f_n$ *converges almost everywhere.*

2. *If $f = \sum_{n=0}^{\infty} f_n$, then $f \in L^p(E)$ and $\lim_{m\to\infty} \|f - \sum_{n=0}^{m} f_n\|_p = 0$ and $\|f\|_p \leqslant \sum_{n=0}^{\infty} \|f_n\|_p$.*

Proof. (2): Suppose $f = \sum_{n=0}^{\infty} f_n$. Set:

$$\begin{aligned} g_m &= \sum_{n=0}^{m} |f_n| \\ g &= \lim_{m\to\infty} g_m \end{aligned}$$

Claim 1: $g \in L^p(E)$ and $\lim_{m\to\infty} \|g_m - g\|_p = 0$.

By definition, $|g_m|^p \leqslant |g_{m+1}|^p$. So, by MCT

$$\lim_{m\to\infty} \int_E |g_m|^p \, d\mu = \int_E |g|^p \, d\mu.$$

i.e.

$$\lim_{m\to\infty} \int_E \|g_m\|_p^p = \|g\|_p^p.$$

But,

$$\|g_m\|_p \leqslant \sum_{n=0}^{m} \|f_n\|_p.$$

Thus, $\|g\|_p < \infty$ since $\sum_{n=0}^{\infty} \|f_n\|_p$.

If $m \in \mathbb{N}$, then

$$\begin{aligned} \|g - g_m\|_p^p &= \int_E \lim_{k\to\infty} \left| \sum_{n=m}^{m+k} |f_n| \right|^p d\mu \\ &= \lim_{k\to\infty} \int_E \left| \sum_{n=m}^{m+k} |f_n| \right|^p d\mu \text{ (MCT)} \\ &= \lim_{k\to\infty} \| \sum_{n=m}^{m+k} f_n\|_p^p \end{aligned}$$

So,

$$\begin{aligned} \|g - g_m\|_p &= \lim_{k\to\infty} \|\sum_{n=m}^{m+k} |f_n|\|_p \\ &\leqslant \lim_{k\to\infty} \sum_{n=m}^{m+k} \|f_n\|_p \\ &= \sum_{n=m}^{\infty} \|f_n\|_p. \end{aligned}$$

Since $\sum_{n=0}^{\infty} \|f_n\|_p < \infty$, it follows that $\lim_{m\to\infty} \|g - g_m\|_p = 0$. $\square_{\text{Claim 1}}$

Claim 2: $\sum_{n=0}^{\infty} f_n$ converges a.e..

Proof Claim 2: Since $g \in L^p(E)$, $g(x) < \infty$ a.e.. Let $x \in E$ so that $g(x) < \infty$. When $k, m \in \mathbb{N}$ and $m > 0$,

$$\left|\sum_{n=m}^{m+k} f_n(x)\right| \leqslant g(x) - g_{m-1}(x)$$

It follows that the partial sums of $\sum_{n=0}^{\infty} f_n(x)$ form a Cauchy sequence. Thus, $\sum_{n=0}^{\infty} f_n$ converges almost everywhere. $\square_{\text{Claim 2}}$.

Set $f = \sum_{n=0}^{\infty}$.

Claim 3: $f \in L^p(E)$.

Proof Claim 3: $|f|^p \leqslant |g|^p$ by definition. Apply Claim 2. $\square_{\text{Claim 3}}$.

Claim 4: $\lim_{m\to\infty} \|f - \sum_{n=0}^{m} f_n\|_p = 0$.

Proof Claim 4: Again, by MCT we have that

$$\|f - \sum_{n=0}^{m} f_n\|_p \leqslant \|g - g_m\|_p.$$

Apply Claim 2. $\square_{\text{Claim 4}}$

Claim 5: $\|f\|_p \leqslant \sum_{n=0}^{\infty} \|f_n\|_p$.

Proof Claim 5: By Claim 4 and Lemma 22. $\square$

Theorem 83. *$L^p(E)$ a Banach space*

Proof. Suppose $\{f_n\}_{n=0}^{\infty}$ is a Cauchy sequence in $L^p(E)$. Choose a sequence $n_0 < n_1 < \ldots$ so that $\|f_{n_{j+1}} - f_{n_j}\|_p \leqslant 2^{-j}$. Set $f = f_{n_0} + \sum_{j=0}^{\infty} f_{n_{j+1}} - f_{n_j}$. By Lemma 23, $f \in L^p(E)$ and $\lim_{m\to\infty} \|f - (f_{n_0} + \sum_{j=0}^{m} f_{n_j})\|_p = 0$. However, $f - (f_{n_0} + \sum_{j=0}^{m} f_{n_j}) = f - f_{n_{j+1}}$. Thus $\lim_{j\to\infty} \|f - f_{n_j}\|_p = 0$. Since $\{f_n\}_{n=0}^{\infty}$ is a Cauchy sequence, $\lim_{n\to\infty} \|f_n - f\|_p = 0$ (see Proposition 75). □

7.5 Approximations and Separability

One major strategy in analysis is to try to approximate functions of one class by those with better properties. This is a version of Littlewood's second principle, but it has many applications (e.g., the continuity method, using test functions to define distributions, etc.). Here, we are able to make this idea precise, using the notion of dense subsets.

Definition 61. *Let X be a normed linear space with norm $\|\cdot\|$. Given two subsets $\mathcal{F}$ and $\mathcal{G}$ of X with $\mathcal{F} \subseteq \mathcal{G}$, we say that $\mathcal{F}$ is dense in $\mathcal{G}$, provided for each function g in $\mathcal{G}$ and $\epsilon > 0$, there is a function f in $\mathcal{F}$ for which $\|f - g\| < \epsilon$.*

Example 3.
1. *The rational numbers $\mathbb{Q}$ are dense in the real numbers $\mathbb{R}$.*
2. *Weierstrauss' theorem shows that the space of polynomials is dense inside $C[a, b]$ for any bounded interval.*

There is an important principle, which is often useful:

Observation 7. *If $\mathcal{F}$ is dense in $\mathcal{G}$ and $\mathcal{G}$ is dense in $\mathcal{H}$, then $\mathcal{F}$ is dense in $\mathcal{H}$.*

Proposition 84. *Let E be a measurable set and $1 \leqslant p \leqslant \infty$. Then the subspace of simple functions in $L^p(E)$ is dense in $L^p(E)$.*

Proof. Suppose that $g \in L^p(E)$. First consider the case $p = \infty$. There is a subset E_0 of E of measure zero for which g is bounded on $E \sim E_0$. From the Simple Approximation Lemma, we find that there is a sequence of simple functions on $E \sim E_0$ that converge uniformly on $E \sim E_0$ to g. Therefore, these functions converge with respect to the $L^\infty(E)$ norm. Thus the simple functions are dense in $L^\infty(E)$.

Now suppose $1 \leqslant p < \infty$. The function g is measurable and therefore, by the Simple Approximation Theorem, there is a sequence $\{\varphi_n\}$ of simple functions on E such that $\{\varphi_n\} \to g$ pointwise on E and

$$|\varphi_n| \leqslant |g| \text{ on } E \text{ for all } n.$$

By comparison, we see that $\varphi_n \in L^p(E)$ for all n. Claim 1: $\{\varphi_n\} \to g$ in $L^p(E)$. Proof of claim 1: For all n

$$|\varphi_n - g|^p \leqslant 2^p \{|\varphi_n|^p + |g|^p\} \leqslant 2^{p+1}|g|^p \text{ on } E.$$

Since $|g|^p$ is integrable over E, we infer from the Lebesgue Dominated Convergence Theorem that $\{\varphi_n\} \to g$ in $L^p(E)$. □

Proposition 85. *Let $[a, b]$ be a closed, bounded interval and $1 \leqslant p < \infty$. Then the subspace of step functions on $[a, b]$ is dense in $L^p[a, b]$.*

The proof of this is essentially the same as the proof of Proposition 54, so we will omit the proof.

Definition 62. *A topological space is* separable *if it admits a dense countable subset.*

Example 4. *The real numbers are separable (which is something you are probably tired of hearing by this point in the course).*

Proposition 86. *Let E be a measurable subset. The space $L^p(E)$ is separable for $1 \leqslant p < \infty$*

Proof. (Sketch) On the interval $[-n, n]$, consider the space $\mathcal{F}_n$ of step functions which take rational values and whose jumps occur at rational numbers and which are defined to be identically zero outside this interval. We consider the union

$$\mathcal{F} = \bigcup_n \mathcal{F}_n.$$

By our previous work, $\mathcal{F}$ is a countable collection of functions that is dense in $L^p(\mathbb{R})$. Finally, let E be a general measurable set. Then the restrictions of functions in $\mathcal{F}$ to E is a countable dense subset of $L^p(E)$. □

Corollary 21. *$C_c(E)$ is dense in $L^p(E)$*

On the other hand $L^\infty(\mathbb{R})$ is *not* separable (see the book for details.)

7.6 The duality of L^p

We have already seen that the conjugate exponents appear in pairs. In this section, we will discuss this duality in a little more depth. We start with some definitions.

7.6.1 Bounded and continuous functionals

Definition 63. *Suppose V is a vector space. A linear functional on V is a map $T : V \to \mathbb{R}$ such that for all $g, h \in V$, $\alpha, \beta \in \mathbb{R}$,*

$$T(\alpha g + \beta h) = \alpha T(g) + \beta T(h).$$

Note that the collection of linear functionals on V is itself a vector space. We can give a few examples.

Example 5.

1. *If $V = \mathbb{R}^n$, the space of linear functionals is simply the space of linear functions.*
2. *Suppose E is measurable, $1 \leqslant p < \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$. For $g \in L^q(E)$, we define the functional $T : L^p(E) \to \mathbb{R}$ by*

$$T(f) = \int_E f g \, d\mu$$

for $f \in L^p(E)$. Using Holder's inequality, we know that $|Tf| \leqslant \|f\|_p \|g\|_q$, so $T(f)$ is well defined.

For normed spaces, we can also discuss the notion of a bounded linear functional.

Definition 64. *For a normed linear space X, a linear functional T is said to be bounded provided there is an $M \geqslant 0$ for which*

$$|T(f)| \leqslant M\|f\| \text{ for all } f \in X.$$

The infimum of the upper-bounds is known as the norm of T and is denoted by $\|T\|_*$.

Exercise 21. *Show that* $\|T\|_* = \sup\{Tf \mid f \in X, \|f\| \leqslant 1\} = \sup\{Tf \mid f \in X, \|f\| = 1\}$

From this, we have the following.

Proposition 87. *The function $\|\cdot\|_*$ is a norm, and the collection of bounded linear functionals on V is a normed linear space, which is denoted V^*.*

Proof. It is immediate that $\|\lambda T\|_* = \lambda \|T\|_*$ and that $\|T\|_* = 0$ if and only if $T \equiv 0$. What remains to show is that

$$\|S + T\|_* \leqslant \|S\|_* + \|T\|_*.$$

To see this, we use the following observation.

$$\begin{aligned} \|S + T\|_* &= \sup \{(S+T)f \mid f \in X, \|f\| = 1\} \\ &\leqslant \sup \{Sf \mid f \in X, \|f\| = 1\} + \sup \{Tf \mid f \in X, \|f\| = 1\} \\ &= [\|S\|_* + [\|T\|_*. \end{aligned}$$

□

One immediate consequence of a linear functional being bounded is the following inequality, which holds for any $f, h \in X$:

$$|T(f) - T(h)| \leqslant \|T\|_* \|f - h\|.$$

From this, it immediately follows that bounded linear functionals are continuous. In other words, whenever a linear functional is bounded, it is also continuous. At first, this might seem like a strange statement, because if you are used to finite dimensional linear algebra, all linear functionals are both continuous and bounded. As such, it's important to keep in mind that things are more complicated for function spaces, which are generally infinite dimensional. Actually, a linear functional being bounded is equivalent to its continuity.

Proposition 88. *For a linear functional T in a normed linear space V, the following conditions are equivalent.*

1. *T is bounded.*

2. *T is continuous.*

3. *T is continuous at one point of V.*

Let us now focus on the case where V is $L^p(E)$, and try to understand the dual space V^*.

Proposition 89. *Let E be measurable and $1 \leqslant p < \infty$. Suppose that q is conjugate to p and that $g \in L^q(E)$.*
Define $T_g : L^p(E) \to \mathbb{R}$ by $T_g(f) = \int_E fgd\ mu$.
Then $T_g \in (L^p(E))^*$ and $\|T\|_* = \|g\|_q$

Proof. Hölder's inequality shows that $T_g \in (L^p(E))^*$ and that $\|T\|_* \leqslant \|g\|_q$.

Therefore, the only thing to prove is the equality. In the case where $1 < p < \infty$, we take

$$f = (sgn(g))\frac{|g|^{q-1}|}{\|g\|_q^{q-1}}$$

to be the conjugate function to g. Doing so, we have that $f \in L^p$ and that $\|f\|_p = 1$. Furthermore, we find that

$$|T_g(f)| = \int \frac{|g|^q|}{\|g\|_q^{q-1}} = \|g\|_q,$$

which implies that

$$\|T_g\|_* \geqslant \|g\|_q.$$

We then turn to the case where $p = 1$. For the sake of contradiction, we assume that $\|T_g\| < \|g\|_\infty$.

Then there is a set A with $\mu(A) > 0$ for which $|g| > \|T_g\|$ on A. Then we take $f = sgn(g)\frac{\chi_A}{\mu(A)}$. Then $f \in L^1$ and $\|f\|_1 = 1$. However,

$$T_g(f) = \int fg = \int |g|\frac{\chi_A}{\mu(A)} > \|T_g\|_*,$$

which is a contradiction. □

Before we come to the main result of this section, let us state a small lemma which will be helpful.

Lemma 24. *Let E be a measurable set, $1 \leqslant p < \infty$ and we have a function g which is integrable on E and satisfies the following property.*

There exists and $M \geqslant 0$ so that

$$\left|\int_E gf\right| \leqslant M\|f\|_p$$

for every simple function $f \in L^p(E)$.

Then $g \in L^q(E)$ and $\|g\|_q \leqslant M$.

Proof. (Sketch)

1. g is finite almost everywhere since it is integrable.

2. As such, we can use the simple approximation lemma to find a sequence of simple functions ϕ_n so that $0 \leqslant \phi_n \leqslant |g|$ and $\phi_n \to |g|$ as $n \to \infty$.

3. Then we consider the function

$$f_n = sgn(g)\phi_n^{q-1}$$

   and compute its integral against g.

4. Finally, we apply Fatou's lemma.

□

We can now state and prove the main result of this section, which characterizes the dual space of L^p (except when p is infinite). We start with the case where the domain is an interval.

Theorem 90 (Riesz representation theorem)**.** *Let $[a, b]$ be a closed, bounded interval and $1 \leqslant p < \infty$. Suppose T is a bounded linear functional on $L^p[a, b]$. Then there is a function $g \in L^q[a, b]$, for which*

$$T(f) = \int_I g \cdot f \text{ for all } f \text{ in } L^p[a, b]$$

Proof. We will only consider the case $p > 1$ (the proof of the case $p = 1$ is similar). For x in $[a, b]$, we define

$$\Phi(x) = T\left(\chi_{[a,x)}\right).$$

Intuitively, Φ is analogous to a "cumulative distribution function" for T, so we can expect that Φ' is the corresponding density (i.e., the function g we are trying to find). However, the work in this proof will be to make this precise.

We first show that this real-valued function Φ is absolutely continuous on $[a, b]$, so that it may be differentiated almost everywhere. By the linearity of T, for each $[c, d] \subseteq [a, b]$, since $\chi_{[c,d)} = \chi_{[a,d)} - \chi_{[a,c)}$,

$$\Phi(d) - \Phi(c) = T\left(\chi_{[a,d)}\right) - T\left(\chi_{[a,c)}\right) = T\left(\chi_{[c,d)}\right)$$

Thus if $\{(a_k, b_k)\}_{k=1}^n$ is a finite disjoint collection of intervals in (a, b), we have that

$$\begin{aligned}\sum_{k=1}^n \left|\Phi(b_k) - \Phi(a_k)\right| &= \sum_{k=1}^n \operatorname{sgn}\left[\Phi(b_k) - \Phi(a_k)\right] \cdot T\left(\chi_{[a_k,b_k)}\right) \\ &= T\left(\sum_{k=1}^n \operatorname{sgn}\left[\Phi(b_k) - \Phi(a_k)\right] \cdot \chi_{[a_k,b_k)}\right)\end{aligned}$$

Then, we consider the simple function $f = \sum_{k=1}^{n} \operatorname{sgn}\left[\Phi\left(b_k\right) - \Phi\left(a_k\right)\right] \cdot \chi_{[a_k, b_k)}$. Evaluating $T(f)$ we find the following

$$|T(f)| \leqslant \|T\|_* \cdot \|f\|_p \text{ and } \|f\|_p = \left[\sum_{k=1}^{n}\left(b_k - a_k\right)\right]^{1/p}$$

Thus,

$$\sum_{k=1}^{n}\left|\Phi\left(b_k\right) - \Phi\left(a_k\right)\right| \leqslant \|T\|_* \cdot \left[\sum_{k=1}^{n}\left(b_k - a_k\right)\right]^{1/p} \tag{7.8}$$

For $\epsilon > 0$, we take $\delta = \left(\epsilon / \|T\|_*\right)^p$, which implies that the left hand side of Equation **??** is less than ϵ, and thus that Φ is absolutely continuous on $[a, b]$.

Now that we have shown that Φ is absolutely continuous, we know that it is differentiable almost everywhere and that the function $g = \Phi'$ is integrable over $[a, b]$. Furthermore, by the fundamental theorem of calculus, we have that

$$\Phi(x) = \int_0^x g \text{ for all } x \in [a, b]$$

Therefore, for each $[c, d] \subseteq (a, b)$

$$T\left(\chi_{[c,d)}\right) = \Phi(d) - \Phi(c) = \int_a^b g \cdot \chi_{[c,d)}.$$

Now all that remains to show is that the functionals T and $f \mapsto \int_a^b g \cdot f$ are the same. To do so, we use the density of step functions in L^p.

More precisely, since the functional T and the functional $f \mapsto \int_a^b g \cdot f$ are linear on the linear space of step functions, it follows that

$$T(f) = \int_a^b g \cdot f \text{ for all step functions } f \text{ on } [a, b]$$

If f is a simple function on $[a, b]$, there is a sequence of step functions $\{\varphi_n\}$ which converges to f in $L^p[a, b]$ and also is uniformly pointwise bounded on $[a, b]$. Since the linear functional T is bounded on $L^p[a, b]$, it follows from the continuity of the functional that

$$\lim_{n \to \infty} T\left(\varphi_n\right) = T(f)$$

On the other hand, the Lebesgue Dominated Convergence Theorem implies that

$$\lim_{n \to \infty} \int_a^b g \cdot \varphi_n = \int_a^b g \cdot f$$

Therefore

$$T(f) = \int_a^b g \cdot f \text{ for all simple functions } f \text{ on } [a, b]$$

Since T is bounded,

$$\left| \int_a^b g \cdot f \right| = |T(f)| \leqslant \|T\|_* \cdot \|f\|_p \text{ for all simple functions } f \text{ on } [a, b].$$

According to Lemma 24, g belongs to $L^q[a, b]$ so the linear functional $f \mapsto \int_a^b g \cdot f$ is bounded on $L^p[a, b]$. This functional agrees with the bounded functional T on the simple functions, which is a dense subspace of $L^p[a, b]$, so these two functionals agree on all of $L^p[a, b]$.

□

In fact, this result holds for general measurable sets, not just intervals.

Theorem 91 (The Riesz Representation Theorem for the Dual of L $^p(E)$)**.** *Let E be a measurable set, $1 \leqslant p < \infty$, and q the conjugate of p. For each $g \in L^q(E)$, define the bounded linear functional $\mathcal{R}_g$ on $L^p(E)$ by*

$$\mathcal{R}_g(f) = \int_E g \cdot f \text{ for all } f \text{ in } L^p(E)$$

Then for each bounded linear functional T on $L^p(E)$, there is a unique function $g \in L^q(E)$ for which

$$\mathcal{R}_g = T, \text{ and } \|T\|_* = \|g\|_q$$

Observation 8. *Remark Let $[a, b]$ be a nondegenerate closed, bounded interval. Hölder's Inequality shows that if f belongs to $L^1[a, b]$, then the functional $g \mapsto \int_a^b f \cdot g$ is a bounded linear functional on $L^\infty[a, b]$. However, there are bounded linear functionals on $L^\infty[a, b]$ that are not of this form.*

In Chapter 19 Section 3 of the book, Royden-Fitzpatrick explain a result of Kantorovich which details the dual of L^∞. Feel free to look at this section for details.

7.7 Problems

L^p spaces appear on nearly every qualifying exam, so we will spend some time working on problems involving them.

1. August 2020 # 2 Let $1 < p, q < \infty$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. For any $f \in L^p([0,1])$, set
$$g(x) = \int_0^x f(t)dt.$$
Prove that $g \in L^q([0,1])$ and $\|g\|_q \leqslant 2^{-1/q}\|f\|_p$.

2. January 2020 # 4 Let $f, g \in L^1(\mathbb{R})$. Given $n \in N$, define $(T_n f) = f(x-n)$. Prove that
$$\lim_{n\to\infty} \|T_n f + g\|_1 = \|f\|_1 + \|g\|_1$$

3. Fall 2011 # 2 Suppose f is a measurable function on $[0,1]$ such that for every $1 \leqslant p < \infty, f \in L^p[0,1]$, and suppose there exists a B such that $\|f\|_p \leqslant B$. Prove that $f \in L^\infty[0,1]$.

4. Fall 2014 # 3 Let $p \geqslant 1$ and (f_n) be a sequence of measurable functions in $L^p[0,1]$ such that $\lim_{n\to\infty} f_n(x) = f(x)$ almost everywhere. Show that $\lim_{n\to\infty} \|f_n - f\|_p = 0$ if and only if $\lim_{n\to\infty} \|f_n\|_p = \|f\|_p$

5. Fall 2016 # 1 Let $f : [0,1] \to \mathbb{R}$ be a Lebesgue integrable function and $I_n = \int_0^1 |f(x)|^n dx$.

   (a) if $\|f\|_\infty > 1$, show that $\lim_{n\to\infty} I_n$ is equal to ∞.

   (b) if $\|f\|_\infty \leqslant 1$, show that $\lim_{n\to\infty} I_n$ exists and evaluate the limit.

Chapter 8

Fubini and Tonelli Theorems in $\mathbb{R}^n$

In this chapter, we discuss multivariate integration and determine when it is possible to compute a multiple integral as an iterated integral, as you saw in multivariate calculus. This material is covered in Chapter 20 of Royden-Fitzpatrick. Unfortunately, it uses quite a few terms that we have not yet defined, so it might be worthwhile to read [Bal08] for more details.

8.1 A crash course in more general measures and integration

In order to discuss this, we need to introduce measurable spaces in a bit more generality.

Definition 65. *A measure space (X, M, μ) is a set X with a σ-algebra M of subsets of X and a measure μ, which is a function $\mu : M \to [0, \infty]$ which satisfies*

1. $\mu(\emptyset) = 0$
2. *μ is countably additive.*

Definition 66. *A measure space is*

1. *finite if $\mu(X) < \infty$*
2. *σ-finite if X is the union of countably many sets, each of which has finite measure.*

3. *complete if $\mu(E) = 0$ implies that every subset of E is measurable.*

Example 6. *1. The Lebesgue measure on $[0, 1]$ is finite, σ-finite and complete.*

2. The Lebesgue measure on $\mathbb{R}$ is not finite, but is σ-finite and complete.

3. The Borel measure on $\mathbb{R}$ is σ-finite, but neither finite nor complete.

4. The counting measure on $\mathbb{R}$ is neither finite nor σ-finite, but is complete.

8.1.1 Product measures

The goal now is define the Lebesgue measure on $\mathbb{R}^2$ (or $\mathbb{R}^n$). To this end, we will define the product measure from the Lebesgue measure on $\mathbb{R}$. More generally, given two measure spaces $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$, we can build a product measure on the Cartesian product $X \times Y$.

Definition 67 (Measurable rectangles)**.** *If $A \subseteq X$ and $B \subseteq Y$, we call $A \times B$ a rectangle. If $A \in \mathcal{A}$ and $B \in \mathcal{B}$, we call $A \times B$ a measurable rectangle.*

We can define the measure of a measurable rectangle to be the product of the measures of A and B.

Definition 68. *If $R = A \times B$ is a measurable rectangle, we define the measure of R to be $\lambda(R) = \mu(A) \times \nu(B)$.*

Example 7. *Let $[a, b] \subset \mathbb{R}$ and $[c, d] \subset \mathbb{R}$, both induced with the Lebesgue measure. What is $\lambda([a, b] \times [c, d])$?*

To make sure that this definition is consistent, we'll verify the following lemma.

Lemma 25. *Let $\{A_k \times B_k\}_{k=1}^{\infty}$ be a countable disjoint collection of measurable rectangles whose union also is a measurable rectangle $A \times B$. Then*

$$\mu(A) \times \nu(B) = \sum_{k=1}^{\infty} \mu(A_k) \times \nu(B_k)$$

Proof. Fix a point $x \in A$. For each $y \in B$, the point (x, y) belongs to exactly one $A_k \times B_k$. Therefore we have the following disjoint union:

$$B = \bigcup_{\{k | x \in A_k\}} B_k$$

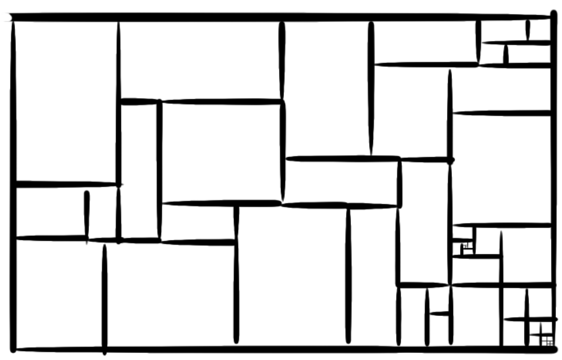


Figure 8.1: A measurable rectangle as the union of countable measurable rectangles

By the countable additivity of the measure ν,

$$\nu(B) = \sum_{\{k | x \in A_k\}} \nu\left(B_k\right)$$

Rewriting this equality in terms of characteristic functions, we find the following

$$\nu(B) \cdot \chi_A(x) = \sum_{k=1}^{\infty} \nu\left(B_k\right) \cdot \chi_{A_k}(x) \text{ for all } x \in A$$

Since each A_k is contained in A, this equality also clearly holds for $x \in X \backslash A$. Therefore

$$\nu(B) \cdot \chi_A = \sum_{k=1}^{\infty} \nu\left(B_k\right) \cdot \chi_{A_k} \text{ on } X$$

By the Monotone Convergence Theorem,

$$\mu(A) \times \nu(B) = \int_X \nu(B) \cdot \chi_A \, d\mu = \sum_{k=1}^{\infty} \int_X \nu\left(B_k\right) \cdot \chi_{A_k} \, d\mu = \sum_{k=1}^{\infty} \mu\left(A_k\right) \times \nu\left(B_k\right)$$

□

Technically speaking, λ is actually a pre-measure, and we need to consider its Caratheodory extension to define the measure. Getting into the details of this would take several weeks, so we will not do so. However, the Lebesgue measure is a complete outer measure, so there is a short-cut.

Definition 69. *Given two measure spaces* $(X, \mathcal{A}, \mu)$ *and* $(Y, \mathcal{B}, \nu)$ *and a set* $E \subset X \times Y$*, we define the outer measure*

$$(\mu \times \nu)^*(E) = \inf \sum_{k=1}^{\infty} \lambda(R_k),$$

where the infimum is taken over all countable collections of measurable rectangles R_k *which cover* E*.*

Definition 70. *A set* $E \subset X \times Y$ *is measurable if*

$$(\mu \times \nu)^*(S) = (\mu \times \nu)^*(E \cap S) + (\mu \times \nu)^*(E^c \cap S)$$

for all sets S*.*

Combining these, we can define the product measure.

Definition 71. *The space* $(X \times Y, \text{measurable sets}, \mu \times \nu)$ *is a measure space, and the measure is known as the product measure.*

8.1.2 Some measurable sets

It is helpful to know a good collection of measurable sets in $\mathbb{R}^n$, since our definition of measurable makes it hard to construct them by hand. For this, we have the following proposition.

Proposition 92. *Any open subset* U *of* $\mathbb{R}^2$ *is measurable with respect to the product measure.*

In fact, open sets are Borel (i.e., the intersection/union/complements of countably many measurable rectangles. However, it will not be necessary to prove this.

As with the case in $\mathbb{R}$, there are many Lebesgue measurable sets which are not Borel. As a note of caution, measurable sets in higher dimensions can be very complicated, so it is worthwhile to verify that you are sets are measurable.

8.2 Integration with respect to a general measure

Before we can discuss the Fubini and Tonelli theorems, we need to say what it means to integrate with respect to a more general measure. The way that this is done is essentially the same as in the one-dimensional case, by using simple functions.

Definition 72 (Simple functions)**.** *Given a measure space* $(X, \mathcal{A}, \mu)$*, a simple function* $s : X \to \mathbb{R}$ *is a function whose range is finite and for which the inverse image*

$$s^{-1}(a) = \{x \in X \mid f(x) = a\}$$

is a measurable set (i.e., in $\mathcal{A}$*) for all* $a \in \mathbb{R}$*.*

Definition 73. *For non-negative simple functions, we define the integral*

$$\int_X s\, d\mu = \sum_{a \in \ range(s)} a\mu(s^{-1}(a)).$$

Then, given a function $f : X \to \mathbb{R}^+$ which is non-negative and measurable (i.e., the pre-image of every open set is in $\mathcal{A}$), we define the integral of f as

$$\int_X f\, d\mu = \sup \left\{ \int_X s\, d\mu \mid s \text{ simple and } s \leqslant f \right\}.$$

Finally, we say that a function is integrable if both its positive and negative parts are integrable.

8.3 Product measures and iterated integrals

We are now able to discuss the Fubini and Tonelli theorems. Before stating these theorems, let us give one further definition.

If $f : x \times y \to \mathbb{R}$ is a measurable function, for $x \in X$, we define the function $f(x,) : Y \to \mathbb{R}$ to be the x-slice of f, For $y \in Y$, we say that the function $f(\cdot, y) : X \to \mathbb{R}$ is the y-slice of f.

Theorem 93 (Fubini's Theorem)**.** *Let* $(X, \mathcal{A}, \mu)$ *and* $(Y, \mathcal{B}, \nu)$ *be two measure spaces and* ν *be complete. Let* f *be integrable*[1] *over* $X \times Y$ *with respect to the product measure* $\mu \times \nu$*. Then for almost all* $x \in X$*, the* x*-slice of* $f, f(x, \cdot)$*, is integrable over* Y *with respect to* ν *and*

$$\int f d(\mu \times \nu) = \int \left[\int f(x, y) d\nu(y) \right] d\mu(x).$$

[1]This means that f is measurable and that $\int_{X \times Y} |f|\, d(\mu \times \nu) < \infty$.

Theorem 94 (Tonelli's theorem)**.** *let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be σ-finite measure spaces. Let f be a measurable function on $X \times Y$ with respect to $(\mu \times \nu)$. If $0 \leqslant f \leqslant \infty$, then*

1. *for almost all $x \in X$, the x-slice is ν-measurable,*
2. *for almost all $y \in X$, the y-slice is μ-measurable,*
3. *the function $\int_Y f(\cdot, y)\, d\nu(y)$ is μ-measurable,*
4. *the function $\int_X f(x, \cdot)\, d\mu(x)$ is ν-measurable.*
5. $$\int_{X\times Y} f d(\mu\times\nu) = \int_X \left[\int_Y f(x,y) d\nu(y)\right] d\mu(x) = \int_Y \left[\int_X f(x,y)\, d\mu(x)\right] d\nu(y)$$

The proof of Fubini and Tonelli's theorem in Royden requires a fair amount of background, so I recommend reading the proof in Axler [Axl20], where it is Theorem 5.28. We will not cover the proofs in this class, but you should read them to understand the basic idea.

Before we finish up, let us provide two examples which show why the assumptions in these theorems are necessary.

Non-Example.

1. *Prove* $\int_0^1 \left[\int_0^1 \frac{x^2-y^2}{(x^2+y^2)^2} dy\right] dx = +\frac{\pi}{4}$
2. *Prove* $\int_0^1 \left[\int_0^1 \frac{x^2-y^2}{(x^2+y^2)^2} dx\right] dy = -\frac{\pi}{4}$
3. *Explain why the answer to the above parts do not violate Fubini's theorem.*

Proof. The function $f = \frac{x^2-y^2}{(x^2+y^2)^2}$ is continuous on $[0,1] \times [0,1]$ except at the origin, hence is measurable.

Doing the first integral, we find that

$$\int_0^1 \frac{x^2-y^2}{(x^2+y^2)^2} dy = \left[\frac{y}{x^2+y^2}\right]_{y=0}^1 = \frac{1}{1+x^2}$$

and

$$\int_0^1 \left(\int_0^1 \frac{x^2-y^2}{(x^2+y^2)^2} dy\right) dx = \int_0^1 \frac{1}{1+x^2} dx = \frac{\pi}{4}$$

Proceeding to the integral computation in the opposite order we get

$$\int_0^1 \left(\int_0^1 \frac{x^2 - y^2}{(x^2 + y^2)^2} dx \right) dy = -\frac{\pi}{4}.$$

To understand why this doesn't contradict Fubini-Tonelli, we estimate $\int_{I \times I} |f(x, y)| d(x, y)$. We have

$$\begin{aligned}
\int_{I \times I} \left| \frac{x^2 - y^2}{(x^2 + y^2)^2} \right| d(x, y) &\geqslant \int_0^1 \left(\int_0^{\pi/2} \frac{|r^2 \cos^2(\theta) - r^2 \sin^2(\theta)|}{r^4} r d\theta \right) dr \\
&= \int_0^1 \int_0^{\pi/2} \frac{|\cos(2\theta)|}{r^2} r d\theta dr \\
&= \int_0^1 2 \int_0^{\pi/4} \frac{\cos(2\theta)}{r} d\theta dr \\
&= 2\frac{1}{2} \int_0^1 \frac{1}{r} d\theta dr \\
&= \lim r \to 0 - ln(r) = \infty.
\end{aligned}$$

As such, $f \notin L^1(I \times I)$, so this does not contradict Tonelli's theorem. □

It's also good to have an example in mind of why it is necessary to assume that the measures are σ-finite. For this, we consider the following example, which is taken from Axler [Axl20] 5.30.

Example 8. *Suppose $\mathcal{B}$ is the σ-algebra of Borel subsets of $[0, 1], \lambda$ is Lebesgue measure on $([0, 1], \mathcal{B})$, and μ is counting measure on $([0, 1], \mathcal{B})$. Let D denote the diagonal of $[0, 1] \times [0, 1]$; in other words,*

$$D = \{(x, x) : x \in [0, 1]\}.$$

Then

$$\int_{[0,1]} \int_{[0,1]} \chi_D(x, y) \, d\mu(y) \, d\lambda(x) = \int_{[0,1]} 1 \, d\lambda = 1,$$

but

$$\int_{[0,1]} \int_{[0,1]} \chi_D(x, y) \, d\lambda(x) \, d\mu(y) = \int_{[0,1]} 0 \, d\mu = 0.$$

Bibliography

[Axl20] Sheldon Axler. *Measure, integration & real analysis*. Springer Nature, 2020.

[Bal08] Prakash Balachandran. 241 lecture: Product measures - Fubini and Tonelli, 2008.

[Fol99] Gerald B Folland. *Real analysis: modern techniques and their applications*, volume 40. John Wiley & Sons, 1999.

[FW91] Matthew Foreman and Friedrich Wehrung. The hahn-banach theorem implies the existence of a non-lebesgue measurable set. *Fundamenta Mathematicae*, 138:13–19, 1991.

[Kec12] Alexander Kechris. *Classical descriptive set theory*, volume 156. Springer Science & Business Media, 2012.

[Leu05] Naichung Conan Leung. Mirror symmetry without corrections. *Communications in Analysis and Geometry*, 13(2):287–331, 2005.

[Mun14] James Munkres. *Topology*. Pearson Education, 2014.

[RF88] Halsey Lawrence Royden and Patrick Fitzpatrick. *Real analysis*, volume 32. Macmillan New York, 1988.

[Sol70] Robert M Solovay. A model of set-theory in which every set of reals is lebesgue measurable. *Annals of Mathematics*, pages 1–56, 1970.

[Tao11] Terence Tao. *An introduction to measure theory*, volume 126. American Mathematical Society Providence, RI, 2011.

[Tes98] Gerald Teschl. Topics in real analysis. *Amer. Math. Soc., Providence, to appear*, 1998.

[Vil09] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.

[wh] wisefool (https://math.stackexchange.com/users/51807/wisefool). Simple function approximation of a function in l^p. Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/256589 (version: 2012-12-11).